Springer Proceedings in Mathematics & Statistics

Aref Jeribi
Mohamed Ali Hammami
Afif Masmoudi *Editors*

# Applied Mathematics in Tunisia

## International Conference on Advances in Applied Mathematics (ICAAM), Hammamet, Tunisia, December 2013

Springer

**Springer Proceedings in Mathematics & Statistics**

Volume 131

More information about this series at http://www.springer.com/series/10533

# Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of select contributions from workshops and conferences in all areas of current research in mathematics and statistics, including OR and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

Aref Jeribi • Mohamed Ali Hammami
Afif Masmoudi
Editors

# Applied Mathematics in Tunisia

## International Conference on Advances in Applied Mathematics (ICAAM), Hammamet, Tunisia, December 2013

Springer

*Editors*
Aref Jeribi
Department of Mathematics
University of Sfax
Sfax, Tunisia

Mohamed Ali Hammami
Department of Mathematics
University of Sfax
Sfax, Tunisia

Afif Masmoudi
Department of Mathematics
University of Sfax
Sfax, Tunisia

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-319-18040-3 ISBN 978-3-319-18041-0 (eBook)
DOI 10.1007/978-3-319-18041-0

Library of Congress Control Number: 2015947258

Mathematics Subject Classification (2010): 01A10; 34K05; 34K20; 34K25; 34K28; 34K35; 34K50; 34L05; 34L10; 34L30; 35A01; 35P05; 41A15; 47-00; 60E07; 60G07; 62E20; 62F86; 62H12; 47H10; 54C15; 54C20; 54C55; 55M15; 47A10; 47A53; 47A11; 47A55

Springer Cham Heidelberg New York Dordrecht London


Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

# Preface

The ICAAM (International Conference on Advances in Applied Mathematics) provides an international forum for scientists and researchers and is aimed at highlighting some of the major theoretical advances in mathematics and their applications that are promising issues in widely different areas of science and technology. Themes of the ICAAM 2013 focused on: spectral theory, operator theory, optimization, numerical analysis, partial differential equations, ordinary differential equations, control theory, dynamical systems, nonlinear systems and matrices, probability, and statistics.

This volume contains a collection of scientific papers presented at ICAAM, which was held in Hammamet, Tunisia, 16–19 December 2013.

The conference is supported by several institutions, including the following:

- Tunisian Association of Applied Mathematics and Industry (ATMAI)
- Mediterranean Institute for the Mathematical Sciences (MIMS)
- University of Sfax
- Faculty of Sciences of Sfax (FSS)
- Tunisia Polytechnic School
- National Engineering School of Tunis

**Program Committee**
Azgal Abichou, Tunisia
Pietro Aiena, Italy
Amel Ben Abda, Tunisia
Mohamed Ben Ayed, Tunisia
Yacine Chitour, France
Jean-Michel Coron, France
Mohamed Ali Hammami, Tunisia
Mohamed Ali Jendoubi, Tunisia
Aref Jeribi, Tunisia
Fuad Kittaneh, Jordan
Afif Masmoudi, Tunisia

Maher Mnif, Tunisia
Zuhair Nashed, USA
Ali Wehbe, Libanon

**Organizing Committee**
Azgal Abichou, Univ of Carthage
Hammadi Baklouti, Univ of Sfax
Amel Ben Abda, Univ of El Manar
Mohamed Ben Ayed, Univ of Sfax
Mongi Ben Hamadou, Univ of Sfax
Hanen Ben Hassen, Univ of Sfax
Mohamed Amine Ghorbel, Univ of Sfax
Mohamed Ali Hammami, Univ of Sfax
Mohamed Ali Jendoubi, Univ of Carthage
Aref Jeribi, Univ of Sfax
Emna Kallel, Univ of Sfax
Mourid Marrakchi, Univ of Sfax
Afif Masmoudi, Univ of Sfax
Maher Mnif, Univ of Sfax
Nedra Moalla, Univ of Sfax
Ali Wehbe, Univ of Beirut

# Acknowledgements

The ICAAM is an international applied mathematics conference held in Tunisia. Several academic institutions contributed to the organization of the ICAAM 2013 and they are acknowledged in the preamble of this book. Further, we are grateful to all members of the Program and Organizing committees for their active help, and we address special thanks to Dr. Hanen Ben Hassen for her valuable participation in the elaboration of this book. We also thank all participants for their contributions to the success of the conference.

Sfax, Tunisia

Aref Jeribi
Mohamed Ali Hammami
Afif Masmoudi

# Contents

# List of Contributors

**Abderrahmane Senoussaoui** Faculté des Sciences Département de Mathématiques, Université d'Oran, Oran, Algeria

**Mohamed Abid** Laboratory CES-Lab, National School of Engineering of Sfax, Sfax, Tunisia

**Yassine Adjabi** Faculty of Sciences, Department of Mathematics, University of M'hamed Bougra, Boumerdes, Algeria

**J.J. Anza Aguirrezabala** Department of Applied Mathematics, ETS de Ingeniería de Bilbao, Universidad del País Vasco UPV/EHU, Vizcaya, Spain

**Pietro Aiena** Dipartimento di Metodi e Modelli Matematici, Facoltà di Ingegneria, Università di Palermo, Palermo, Italy

**F. Aliouane** Jjiel University, Jijel, Algeria

**Taher Ben Arab** Faculté des Sciences de Sfax, Laboratoire de Probabilités et Statistique, Sfax, Tunisie

Faculté des Sciences de Sfax, Laboratoire de Probabilités et Statistique, Université de Sfax, Sfax, Tunisie

**D. Azzam-Laouir** Jijel University, Jijel, Algeria

**Giorgia Bellomonte** Dipartimento di Matematica e Informatica, Università degli Studi di Palermo, Palermo, Italy

**Salvatore Di Bella** Dipartimento di Matematica e Informatica, Università di Palermo, Palermo, Italy

**Zehour Nedjraoui Benbouziane** Département de Mathématiques, Université de Tlemcen, Tlemcen, Algérie

**Nawel Benouhiba** Laboratory of Applied Mathematics, Badji Mokhtar-Annaba University, Annaba, Algeria

**Nadira Boukli-Hacene** Département de Mathématiques, Université de Tlemcen, Tlemcen, Algérie

**Amina Bounouala** Laboratory of Applied Mathematics, Badji Mokhtar-Annaba University, Annaba, Algeria

**E. Alberdi Celaya** Department of Applied Mathematics, EUIT de Minas y Obras Públicas, Universidad del País Vasco UPV/EHU, Vizcaya, Spain

**Wajdi Chaker** Higher institute of informatics and multimedia of Gabes., Gabès, Tunisie

**Salma Charfi** Faculté des Sciences de Sfax, Département de Mathématiques, Université de Sfax, Sfax, Tunisie

**Hanen Damak** University of Monastir Tunisia, Monastir, Tunisia

**Hajer Dbebria** Laboratory of Engineering Mathematics, Polytechnic School of Tunisia, University of Carthage, Carthage, Tunisia

**Smail Djebali** Laboratoire "Théorie du point Fixe et Applications", Algiers, Algeria

**Asrar Elleuch** Faculté des Sciences de Sfax, Département de Mathématiques, Université de Sfax, Sfax, Tunisie

**Rahmet Savaş Eren** Department of Mathematics, Istanbul Medeniyet University, Istanbul, Turkey

**Khalil Ezzinbi** Faculté des Sciences Semlalia, Département de Mathématiques, Université Cadi Ayyad, Marrakech, Morocco

**Abdellah Gherbi** EPST d'Oran, Oran, Algeria

**Abdelaziz Ghribi** Higher Institute of Business Administration of Sfax, Sfax, Tunisia

**Joachim Gwinner** Fakultät für Luft- und Raumfahrttechnik, Institut für Mathamatik und Rechneranwendung, Universität der Bundeswehr München, Neubiberg/München, Germany

**Atizez Hadrich** Faculté des Sciences de Sfax, Laboratoire de Probabilités et Statistique, Universitá de Sfax, Sfax, Tunisie

Laboratoire d'Informatique, Signal et Image de la Côte d'Opale (LISIC-EA 4491), Calais Cedex, France

**Mohamed Ali Hammami** University of Sfax Tunisia, Sfax, Tunisia

**Bernard Helffer** Département de Mathématiques, Université Paris-Sud 11, Orsay, France

Laboratoire Jean Leray, Université de Nantes, Nantes, French

**Aref Jeribi** Faculty of Science of Sfax, University of Sfax, Soukra Road Km 3.5, B.P. 1171, 3000, Sfax, Tunisia

**Arezki Kessi** Faculty of Mathematics, USTHB, Algiers, Algeria

**Bilel Krichen** Preparatory Engineering Institute, Sfax, Tunisia

**Hela Louati** University of Sfax Tunisia, Sfax, Tunisia

**Mahdi Louati** Faculty of Science, Department of Mathematics, University of Sfax, Sfax, Tunisia

**Fatma Mallouli** Laboratory CES-Lab, National School of Engineering of Sfax, Sfax, Tunisia

**Afif Masmoudi** Faculté des Sciences de Sfax, Laboratoire de Probabilités et Statistique, Université de Sfax, Sfax, Tunisie

**Atef Masmoudi** Laboratory of Probability and Statistics, Faculty of Sciences of Sfax, Universitá de Sfax, Sfax, Tunisie

**Bekkai Messirdi** University of Oran, Oran, Algeria

**Farouk Mselmi** Faculty of Science, Department of Mathematics, University of Sfax, Sfax, Tunisia

**Hatem Najar** Faculté des Sciences de Moanstir, Département de Mathématiques, Université de Monastir, Monastir, Tunisia

Laboratoire de recherche: Algèbre Géométrie et Théorie Spectrale : LR11ES53

**Ali Salem** College of Arts and Sciences in Noairyya, University of Dammam, King of Saudi Arabia, Dammam, Saudi Arabia

Laboratory of Engineering Mathematics, Polytechnic School of Tunisia, University of Carthage, Carthage, Tunisia

**Ekrem Savaş** Department of Mathematics, Istanbul Commerce University, Uskudar, Istanbul, Turkey

**Mourad Zribi** Laboratoire d'Informatique, Signal et Image de la Côte d'Opale (LISIC-EA 4491), ULCO, 50 Rue Ferdinand, Buisson BP 719, 62228, Calais Cedex, France

# Polaroid operators and Weyl type theorems

**Pietro Aiena**

**Abstract** Weyl type theorems have been proved for a considerably large number of classes of operators. In this work, after introducing the class of polaroid operators and some notions from local spectral theory, we determine a theoretical and general framework from which Weyl type theorems may be promptly established for many of these classes of operators. The theory is exemplified by given several examples of hereditarily polaroid operators.

**Keywords** Localized SVEP • polaroid type operators • Weyl type theorems

*Mathematics Reviews* Primary 47A10, 47A11. Secondary 47A53, 47A55

## 1 Introduction

This note is a free-style paraphrase of a presentation at the *International Conference on Advances in Applied Mathematics* held in Hammamet (Tunisia), 16–19 December 2013. I would like to thank the organizers for their kind invitation and, overall, for the generous hospitality.

A bounded linear operator $T \in L(X)$, $X$ a complex infinite dimensional Banach space, which satisfies Weyl's theorem has a very special structure of its spectrum $\sigma(T)$, precisely, $T$ is said to verify Weyl's theorem, if the complement in the spectrum of the Weyl spectrum coincides with the isolated points of the spectrum which are eigenvalues of finite multiplicity. In his pioneering work H. Weyl [53] discovered that every hermitian operator on a Hilbert space has a such structure of the spectrum,and many years later it was proved by Coburn [26] that also Toeplitz operators and hyponormal operators satisfy this property. Later, Berberian [17, 23], showed that several other classes of operators, including seminormal operators,

P. Aiena (✉)
Dipartimento di Metodi e Modelli Matematici, Facoltà di Ingegneria, Università di Palermo, Viale delle Scienze, I-90128 Palermo, Italy
e-mail: paiena@unipa.it



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_1

satisfy Weyl'theorem. In [13] it has been observed that Weyl's theorem holds for every convolution operator acting on the group algebra $L^1(G)$, $G$ a locally compact abelian group.

There are also several variants of Weyl's theorem that have been studied in the last two decades by several authors, for instance $a$-Weyl's theorem and property $(w)$ and their generalized versions. Most of these Weyl type theorems were essentially proved for special classes of operators. Many times the arguments used for each one of these classes of operators, are rather similar. In this paper we outline an useful and general theoretical framework, which entails, as a particular case, Weyl type theorems for almost all these classes of operators. Our framework combines classical arguments of Fredholm theory and a localized version of the single-valued extension property.

## 2 Definitions and preliminary results

Let $T \in L(X)$ be a bounded linear operator on an infinite-dimensional complex Banach space $X$, and denote by $\alpha(T)$ and $\beta(T)$, the dimension of the kernel $\ker T$ and the codimension of the range $R(T) := T(X)$, respectively. Let

$$\Phi_+(X) := \{T \in L(X) : \alpha(T) < \infty \text{ and } T(X) \text{ is closed}\}$$

denote the class of all *upper semi-Fredholm* operators, and let

$$\Phi_-(X) := \{T \in L(X) : \beta(T) < \infty\}$$

denote the class of all *lower semi-Fredholm* operators. If $T \in \Phi_\pm(X) := \Phi_+(X) \cup \Phi_-(X)$, the *index* of $T$ is defined by $\text{ind}(T) := \alpha(T) - \beta(T)$. If $\Phi(X) := \Phi_+(X) \cap \Phi_-(X)$ denotes the set of all *Fredholm* operators, the set of *Weyl operators* is defined by

$$W(X) := \{T \in \Phi(X) : \text{ind}\, T = 0\},$$

the class of *upper semi-Weyl operators* is defined by

$$W_+(X) := \{T \in \Phi_+(X) : \text{ind}\, T \leq 0\},$$

and class of *lower semi-Weyl operators* is defined by

$$W_-(X) := \{T \in \Phi_-(X) : \text{ind}\, T \geq 0\}.$$

Clearly, $W(X) = W_+(X) \cap W_(X)$. The classes of operators above defined generate the following spectra: the *Weyl spectrum*, defined by

$$\sigma_{\mathrm{w}}(T) := \{\lambda \in \mathbb{C} : \lambda I - T \notin W(X)\};$$

and the *upper semi-Weyl spectrum*, defined by

$$\sigma_{\mathrm{uw}}(T) := \{\lambda \in \mathbb{C} : \lambda I - T \notin W_+(X)\}.$$

Let $p := p(T)$ be the *ascent* of an operator $T$; i.e. the smallest nonnegative integer $p$ such that $\ker T^p = \ker T^{p+1}(X)$. If such integer does not exist we put $p(T) = \infty$. Analogously, let $q := q(T)$ be the *descent* of $T$; i.e the smallest nonnegative integer $q$ such that $T^q(X) = T^{q+1}(X)$, and if such integer does not exist we put $q(T) = \infty$. It is well known that if $p(T)$ and $q(T)$ are both finite then $p(T) = q(T)$, see [1, Theorem 3.3].

The class of all *Browder operators* is defined

$$B(X) := \{T \in \Phi(X) : p(T), q(T) < \infty\};$$

while the class of all *upper semi-Browder operators* is defined

$$B_+(X) := \{T \in \Phi_+(X) : p(T) < \infty\}.$$

Obviously, $B(X) \subseteq W(X)$ and $B_+(X) \subseteq W_+(X)$, see [1, Theorem 3.4].

Semi-Fredholm operators have been generalized by Berkani [18, 20] and [19] in the following way: for every $T \in L(X)$ and a nonnegative integer $n$ let us denote by $T_{[n]}$ the restriction of $T$ to $T^n(X)$, viewed as a map from the space $T^n(X)$ into itself (we set $T_{[0]} = T$). $T \in L(X)$ is said to be *semi B-Fredholm*, (resp. *B-Fredholm, upper semi B-Fredholm*, *lower semi B-Fredholm,*) if for some integer $n \geq 0$ the range $T^n(X)$ is closed and $T_{[n]}$ is a semi-Fredholm operator (resp. Fredholm, upper semi-Fredholm, lower semi-Fredholm). In this case $T_{[m]}$ is a semi-Fredholm operator for all $m \geq n$ [20] with the same index of $T_{[n]}$. This enables one to define the index of a semi B-Fredholm as $\operatorname{ind} T = \operatorname{ind} T_{[n]}$.

A bounded operator $T \in L(X)$ is said to be *B-Weyl* (respectively, *upper semi B-Weyl*, *lower semi B-Weyl*) if for some integer $n \geq 0$ the range $T^n(X)$ is closed and $T_{[n]}$ is Weyl (respectively, upper semi-Weyl. lower semi-Weyl). The *B-Weyl spectrum* is defined by

$$\sigma_{\mathrm{bw}}(T) := \{\lambda \in \mathbb{C} : \lambda I - T \text{ is not B-Weyl}\},$$

and, analogously, the *upper semi B-Weyl spectrum* of $T$ is defined by

$$\sigma_{\mathrm{ubw}}(T) := \{\lambda \in \mathbb{C} : \lambda I - T \text{ is not upper semi B-Weyl}\}.$$

The concept of Drazin invertibility has been introduced in a more abstract setting than operator theory. In the case of the Banach algebra $L(X)$, $T \in L(X)$ is said to be *Drazin invertible* (with a finite index) if $p(T) = q(T) < \infty$, and this is equivalent to saying that $T = T_0 \oplus T_1$, where $T_0$ is invertible and $T_1$ is nilpotent, see [42, Corollary 2.2] and [41, Prop. A]. Every B-Fredholm operator $T$ admits the representation $T = T_0 \oplus T_1$, where $T_0$ is Fredholm and $T_1$ is nilpotent [19], so every Drazin invertible operator is B-Fredholm. Drazin invertibility for bounded operators suggests the following definition:

**Definition 2.1.** $T \in L(X)$ is said to be *left Drazin invertible* if $p := p(T) < \infty$ and $T^{p+1}(X)$ is closed, while $T$ is said to be *right Drazin invertible* if $q := q(T) < \infty$ and $T^q(X)$ is closed.

Note that the concept of Drazin invertibility may be given in terms of B-Fredholm theory: indeed, $T$ is Drazin invertible (respectively, left Drazin invertible, right Drazin invertible) if and only if $T$ is B-Browder (respectively, upper semi B-Browder, lower semi B-Browder), see [7].

The *Drazin spectrum* is then defined as

$$\sigma_{\mathrm{d}}(T) := \{\lambda \in \mathbb{C} : \lambda I - T \text{ is not Drazin invertible}\},$$

while the *left Drazin spectrum* is defined as

$$\sigma_{\mathrm{ld}}(T) := \{\lambda \in \mathbb{C} : \lambda I - T \text{ is not left Drazin invertible}\}.$$

In the sequel we denote by $\sigma_{\mathrm{a}}(T)$ the *approximate point spectrum*, defined by

$$\sigma_{\mathrm{a}}(T) := \{\lambda \in \mathbb{C} : \lambda I - T \text{ is not bounded below}\},$$

where an operator is said to be *bounded below* if it is injective and has closed range. The classical *surjective spectrum* of $T$ is denoted by $\sigma_{\mathrm{s}}(T)$.

Denote by $T'$ the dual of $T \in L(X)$, and if $T$ is defined on a Hilbert space denote by $T^*$ the Hilbert adjoint of $T$. The concepts of left or right Drazin invertibility lead to the concepts of left or right pole. Let us denote by $\sigma_{\mathrm{a}}(T)$ the classical *approximate point spectrum* and by $\sigma_{\mathrm{s}}(T)$ the *surjectivity spectrum*. It is well known that $\sigma_{\mathrm{a}}(T') = \sigma_{\mathrm{s}}(T)$, where $T'$ denotes the dual of $T$, and $\sigma_{\mathrm{s}}(T') = \sigma_{\mathrm{a}}(T)$. Evidently, $\sigma_{\mathrm{uw}}(T) \subseteq \sigma_{\mathrm{a}}(T)$.

**Definition 2.2.** Let $T \in L(X)$, $X$ a Banach space. If $\lambda I - T$ is left Drazin invertible and $\lambda \in \sigma_{\mathrm{a}}(T)$, then $\lambda$ is said to be a *left pole* of the resolvent of $T$. A left pole $\lambda$ is said to have *finite rank* if $\alpha(\lambda I - T) < \infty$. If $\lambda I - T$ is right Drazin invertible and $\lambda \in \sigma_{\mathrm{s}}(T)$, then $\lambda$ is said to be a *right pole* of the resolvent of $T$.A right pole $\lambda$ is said to have *finite rank* if $\beta(\lambda I - T) < \infty$.

Evidently, $\lambda$ is a pole of $T$ if and only if $\lambda$ is both a left and a right pole of $T$. Moreover, $\lambda$ is a pole of $T$ if and only if $\lambda$ is a pole of $T'$. In the case of Hilbert space operators, $\lambda$ is a pole of $T'$ if and only if $\overline{\lambda}$ is a pole of $T^*$.

**Definition 2.3.** Let $T \in L(X)$. Then

(i) $T$ is said to be *left polaroid* if every isolated point of $\sigma_{\mathrm{a}}(T)$ is a left pole of the resolvent of $T$.

(ii) $T \in L(X)$ is said to be *right polaroid* if every isolated point of $\sigma_s(T)$ is a right pole of the resolvent of $T$.
(iii) $T$ is said to be *a-polaroid* if every $\lambda \in \text{iso}\,\sigma_a(T)$ is a pole of the resolvent of $T$.

The concept of left and right polaroid is dual each other:

**Theorem 2.4 ([5]).** *If $T \in L(X)$, $X$ a Banach space, then the following equivalences hold:*

*(i) $T$ is left polaroid if and only if $T'$ is right polaroid.*
*(ii) $T$ is right polaroid if and only if $T'$ is left polaroid.*
*(iii) $T$ is polaroid if and only if $T'$ is polaroid.*

*If $T$ is a Hilbert space operator, then in the equivalences* (i), (ii)*, and* (iii) *$T'$ may be replaced by $T^*$. Moreover, $T'$ is a-polaroid if and only if $T^*$ is a-polaroid.*

Polaroid operators on infinite dimensional complex Banach spaces have been recently investigated, together with the related conditions for an operator of being left, right polaroid or *a*-polaroid [6, 30, 31, 34].

The *quasi-nilpotent part* of $T \in L(X)$ is defined as the set

$$H_0(T) := \{x \in X : \lim_{n\to\infty} \|T^n x\|^{\frac{1}{n}} = 0\}.$$

or, alternatively, $H_0(T)$ in terms of local spectral theory may be defined as the *glocal subspace* associated with the set $\{0\}$, see [43] or [1]. Clearly, $\ker T^n \subseteq H_0(T)$ for every $n \in \mathbb{N}$. The *analytic core* of $T$ is defined $K(T) := \{x \in X :$ there exist $c > 0$ and a sequence $(x_n)_{n\geq 1} \subseteq X$ such that $Tx_1 = x, Tx_{n+1} = x_n$ for all $n \in \mathbb{N}$, and $||x_n|| \leq c^n||x||$for all $n \in \mathbb{N}\}$.

Note that $K(T)$ may be also defined as the *local spectral space* associated with $\mathbb{C} \setminus \{0\}$, see again [43] or [1]. Note that $T(K(T)) = K(T)$, see [1, Theorem 1.21].

**Theorem 2.5 ([9, Theorem 2.2]).** *If $T \in L(X)$, the following statements hold:*

*(i) $T$ is polaroid if and only if there exists $p := p(\lambda I - T) \in \mathbb{N}$ such that*

$$H_0(\lambda I - T) = \ker(\lambda I - T)^p \quad \text{for all } \lambda \in \text{iso}\,\sigma(T). \tag{1}$$

*(ii) If $T$ is left polaroid, then there exists $p := p(\lambda I - T) \in \mathbb{N}$ such that*

$$H_0(\lambda I - T) = \ker(\lambda I - T)^p \quad \text{for all } \lambda \in \text{iso}\,\sigma_a(T). \tag{2}$$

It is easily seen that the following implications hold:

$$T\ a\text{-polaroid} \Rightarrow T \text{ left polaroid} \Rightarrow T \text{ polaroid}$$

Furthermore, if $T$ is right polaroid, then $T$ is polaroid.

The study of Weyl type theorems needs some typical tools originating from local spectral theory. In particular, we consider the following basic property that has relevant role in the theory of decomposable operators, as well as in Fredholm theory, see Laursen and Neumann [43] and [1].

**Definition 2.6.** Let $X$ be a complex Banach space and $T \in L(X)$. The operator $T$ is said to have *the single valued extension property* at $\lambda_0 \in \mathbb{C}$ (abbreviated SVEP at $\lambda_0$), if for every open disc $\mathbb{D}$ of $\lambda_0$, the only analytic function $f : U \to X$ which satisfies the equation $(\lambda I - T)f(\lambda) = 0$ for all $\lambda \in \mathbb{D}$ is the function $f \equiv 0$.

An operator $T \in L(X)$ is said to have SVEP if $T$ has SVEP at every point $\lambda \in \mathbb{C}$. Evidently, $T \in L(X)$ has SVEP at every isolated point of the spectrum.

We also have

$$p(\lambda I - T) < \infty \Rightarrow T \text{ has SVEP at } \lambda, \tag{3}$$

and dually,

$$q(\lambda I - T) < \infty \Rightarrow T' \text{ has SVEP at } \lambda, \tag{4}$$

see [1, Theorem 3.8]. Furthermore, from definition of localized SVEP it is easily seen that

$$\sigma_{\mathrm{a}}(T) \text{ does not cluster at } \lambda \Rightarrow T \text{ has SVEP at } \lambda, \tag{5}$$

and dually,

$$\sigma_{\mathrm{s}}(T) \text{ does not cluster at } \lambda \Rightarrow T' \text{ has SVEP at } \lambda. \tag{6}$$

Note that $H_0(T)$ generally is not closed and [1, Theorem 2.31 ]

$$H_0(\lambda I - T) \text{ closed} \Rightarrow T \text{ has SVEP at } \lambda. \tag{7}$$

*Remark 2.7.* The converse of the implications (1)–(5) holds if $\lambda I - T$ is semi-Fredholm, see [1, Chapter 3], or if $\lambda I - T$ is semi B-Fredholm [3].

Recall that a bounded operator $K \in L(X)$ is said to be *algebraic* if there exists a non-constant polynomial $h$ such that $h(K) = 0$. Trivially, every nilpotent operator is algebraic and it is well known that if $K^n(X)$ has finite dimension for some $n \in \mathbb{N}$ then $K$ is algebraic.

**Theorem 2.8 ([11]).** *If $T \in L(X)$ has SVEP and $K \in L(X)$ is an algebraic operator which commutes with $T$, then $T + K$ has SVEP.*

The polaroid type conditions for $T$ (respectively, for $T'$) are equivalent, assuming the SVEP at the points of certain sets:

**Theorem 2.9.** *Let $T \in L(X)$. Then we have*

(i) *If $T'$ has SVEP at every $\lambda \notin \sigma_{\mathrm{uw}}(T)$ then the properties of being polaroid, a-polaroid and left polaroid for $T$ are all equivalent.*
(ii) *If $T$ has SVEP at every $\lambda \notin \sigma_{\mathrm{lw}}(T)$, then the properties of being polaroid, a-polaroid and left polaroid for $T'$ are all equivalent.*

The polaroid condition is preserved by the functional calculus:

**Theorem 2.10 ([4]).** *For an operator $T \in L(X)$ the following statements are equivalent.*

*(i) $T$ is polaroid;*
*(ii) $f(T)$ is polaroid for every $f \in \mathcal{H}_{nc}(\sigma(T))$;*
*(iii) there exists a non-trivial polynomial $p$ such that $p(T)$ is polaroid;*
*(iv) there exists $f \in \mathcal{H}_{nc}(\sigma(T))$ such that $f(T)$ is polaroid.*

## 3 Examples of polaroid operators

The following class of operators has been introduced by Oudghiri [51].

**Definition 3.1.** A bounded operator $T \in L(X)$ is said to belong to the class $H(p)$ if there exists a natural $p := p(\lambda)$ such that:

$$H_0(\lambda I - T) = \ker(\lambda I - T)^p \quad \text{for all } \lambda \in \mathbb{C}. \tag{8}$$

In the case that $p = p(\lambda) = 1$ for every $\lambda \in \mathbb{C}$ we shall say that $T$ belongs to the class $H(1)$. Every convolution operator of the group algebra $L^1(G)$ is $H(1)$ [13]. In the sequel we show that some other important classes of operators are $H(1)$.

(a) **Totally paranormal operators.** Recall that $T \in L(X)$ is said *paranormal* if $\|Tx\| \leq \|T^2x\|\|x\|$ for all $x \in X$. The property of being paranormal is not translation-invariant. $T \in L(X)$ is called *totally paranormal* if $\lambda I - T$ is paranormal for all $\lambda \in \mathbb{C}$. Every totally paranormal operator has property $H(1)$ [44]. In fact, if $x \in H_0(\lambda I - T)$, then $\|(\lambda I - T)^n x\|^{1/n} \to 0$ and since $T$ is totally paranormal we then have $(\lambda I - T)^n x\|^{1/n} \geq \|(\lambda I - T)x\|$. Therefore, $H_0(\lambda I - T) \subseteq \ker(\lambda I - T)$, and since the reverse inclusion holds for every operator then $H_0(\lambda I - T) = \ker(\lambda I - T)$.
(b) **Hyponormal operators.** A bounded operator $T \in L(H)$ on a Hilbert space is said to be *hyponormal* if $\|T^*x\| \leq \|Tx\|$ for all $x \in H$, or equivalently $T^*T \geq TT^*$. It is easily seen that every hyponormal operator is totally paranormal, hence $H(1)$. The class of totally paranormal operators includes also subnormal operators and quasi-normal operators, since these operators are hyponormal, see [27] or [37].

Two operators $T \in L(X)$, $S \in L(Y)$, $X$ and $Y$ Banach spaces, are said to be *intertwined* by $A \in L(X, Y)$ if $SA = AT$; and $A$ is said to be a *quasi-affinity* if it has a trivial kernel and dense range. If $T$ and $S$ are intertwined by a quasi-affinity, then $T$ is called a *quasi-affine transform* of $S$, and we write $T \prec S$. If both $T \prec S$ and $S \prec T$ hold, then $T$ and $S$ are said to be *quasi-similar*.

The next result shows that property $H(1)$ is preserved by quasi-affine transforms.

**Theorem 3.2.** *Suppose that $S \in L(Y)$ has property $H(1)$ and $T \prec S$. Then $T$ has property $H(1)$. Analogously, if $S \in L(Y)$ has property $H(p)$ and $T \prec S$, then $T$ has property $H(p)$*

*Proof.* Suppose $S$ has property $H(1)$, $SA = AT$, with $A$ injective. If $\lambda \in \mathbb{C}$ and $x \in H_0(\lambda I - T)$, then

$$\|(\lambda I - S)^n Ax\|^{1/n} = \|A(\lambda I - T)^n x\|^{1/n} \le \|A\|^{1/n}\|(\lambda I - T)^n x\|^{1/n},$$

from which it follows that $Ax \in H_0(\lambda I - S) = \ker(\lambda I - S)$. Hence $A(\lambda I - T)x = (\lambda I - S)Ax = 0$ and, since $A$ is injective, this implies that $(\lambda I - T)x = 0$, i.e. $x \in \ker(\lambda I - T)$. Therefore $H_0(\lambda I - T) = \ker(\lambda I - T)$ for all $\lambda \in \mathbb{C}$.

The more general case of $H(p)$-operators is proved by a similar argument.

■

For $T \in L(H)$ let $T = W|T|$ be the polar decomposition of $T$. Then $R := |T|^{1/2}W|T|^{1/2}$ is said the *Aluthge transform* of $T$. If $R = V|R|$ is the polar decomposition of $R$, define $\tilde{T} := |R|^{1/2}V|R|^{1/2}$.

(c) **Log-hyponormal operators.** An operator $T \in L(H)$ is said to be *log-hyponormal* if $T$ is invertible and satisfies $\log(T^*T) \ge \log(TT^*)$. If $T$ is log-hyponormal, then $\tilde{T}$ is hyponormal and $T = K\tilde{T}K^{-1}$, where $K := |R|^{1/2}|T|^{1/2}$, see [25, 52]. Hence $T$ is similar to a hyponormal operator and therefore, by Theorem 3.2, has property $H(1)$.

(d) **$p$-hyponormal operators.** An operator $T \in L(H)$ is said to be *$p$-hyponormal*, with $0 < p \le 1$, if $(T^*T)^p \ge (TT^*)^p$. Every $p$-hyponormal operator is paranormal, see [16]. Every invertible $p$-hyponormal $T$ is quasi-similar to a log-hyponormal operator and consequently, by Theorem 3.2, it has property $H(1)$ [14, 29]. This is also true for $p$-hyponormal operators which are not invertible, see [33].

**Theorem 3.3 ([51]).** *For a bounded operator $T \in L(X)$ the following assertions are equivalent:*

(i) *$T$ has the property $H(p)$;*
(ii) *$f(T)$ has the property $H(p)$ for every $f \in \mathscr{H}(\sigma(T))$;*
(iii) *There exists an analytic function $h$ defined in an open neighborhood $\mathscr{U}$ of $\sigma(T)$, non-identically constant in any component of $\mathscr{U}$, such that $h(T)$ has the property $H(p)$.*

*Moreover, if $T$ has the property $H(p)$, then any restriction $T|M$ on a closed $T$-invariant subspace has the property $H(p)$.*

An obvious consequence of Theorem 3.3 is that $T \in L(H)$ is *algebraically hyponormal* (i.e., there exists a non-trivial polynomial $h$ for which $h(T)$ is hyponormal) then $T$ is $H(p)$. In [37, §2.72] is given an example of a hyponormal operator $T$ for which $T^2$ is not hyponormal. However, an important consequence of Theorem 3.3 is that $T^2$ inherits from $T$ the property of being $H(1)$.

**Theorem 3.4.** *If $T \in L(X)$ has the property $H(p)$, then $T$ is hereditarily polaroid. Moreover, $T$ has SVEP.*

*Proof.* Evidently, Theorem 2.5 entails that $T$ is polaroid, while, since $H_0(\lambda I - T)$ is closed for all $\lambda \in \mathbb{C}$, then $T$ has SVEP by the inclusion (7). Furthermore, $T$ is hereditarily polaroid by Theorem 3.3. ∎

An operator similar to a restriction of a generalized scalar operator to one of its closed invariant subspaces is called *subscalar*. The interested reader can find a well-organized study of these operators in the Laursen and Neumann book [43].

**Theorem 3.5 ([51]).** *Every subscalar operator $T \in L(X)$ is $H(p)$.*

Therefore, we have

$$\text{subscalarity} \ \Rightarrow \text{property } H(p) \Rightarrow \text{SVEP}.$$

Classical example of subscalar operators are hyponormal operators. Theorem 3.5 implies that some other important classes of operators are $H(p)$.

(e) ***M*-hyponormal operators.** Recall that $T \in L(H)$ is said to be *$M$-hyponormal* if there exists $M > 0$ such that $TT^* \leq MT^*T$. Every $M$-hyponormal operator is subscalar [43, Proposition 2.4.9] and hence $H(p)$.

(f) ***w*-hyponormal operators.** If $T \in L(H)$ and $T = U|T|$ is the polar decomposition, define $\hat{T} := |T|^{\frac{1}{2}} U |T|^{\frac{1}{2}}$. $T \in L(H)$ is said to be *$w$-hyponormal* if $|\hat{T}| \geq |T| \geq |\hat{T}^*|$. Examples of $w$-hyponormal operators are $p$-hyponormal operators and log-hyponormal operators. All $w$-hyponormal operators are subscalar (together with its Aluthge transformation, see [47]), and hence $H(p)$ (precisely, $H(1)$, see [39, Theorem 2.5].

(g) ***p*-quasihyponormal operators.** A Hilbert space operator $T \in L(H)$ is said to be *$p$-quasihyponormal* for some $0 < p \leq 1$ if

$$T^*|T^*|^{2p}T \leq T * |T|^{2p}T.$$

Every $p$-quasi-hyponormal is paranormal [46].

Let us denote by $p_* - QH$ the class of all $p$-quasihyponormal operators $T$ for which $\ker T \subseteq \ker T^*$. The following result is due to Duggal and Jeon [35, Theorem 2.2 and Theorem 2.12].

**Theorem 3.6.** *Every $p_* - QH$ operator is $H(1)$.*

A bounded linear operator $T \in L(X)$, defined on a complex infinite dimensional Banach space $X$, is said to be *normaloid* if $\|T\| = r(T)$, $r(T)$ the spectral radius of $T$. An operator $T \in L(X)$ is said to be *hereditarily normaloid*, $T \in \mathcal{H}N$, if the restriction $T|M$ of $T$, to any closed $T$-invariant subspace $M$, is normaloid. Finally, $T \in L(X)$ is said to be *totally hereditarily normaloid*, $T \in \mathcal{T}\mathcal{H}\mathcal{N}$, if $T \in \mathcal{H}N$ and every invertible restriction $T|M$ has a normaloid inverse. Totally hereditarily normaloid operators were introduced in [32], and have since investigated in [30], and [34], for establishing Weyl type theorems.

**Theorem 3.7 ([4]).** *If $T \in L(X)$ is $\mathcal{T}\mathcal{H}\mathcal{N}$, then $T$ is polaroid. If $X$ is a separable Banach space, or a Hilbert space, then $T$ has SVEP.*

Now, let $\mathcal{C}$ be any class of operators. We say that $T$ is an *analytically $\mathcal{C}$*-operator if there exists some analytic function $f \in \mathcal{H}_{nc}(\sigma(T))$ such that $f(T) \in \mathcal{C}$. It should be noted that the property of being analytically $\mathcal{C}$ is translation invariant.

In the sequel we list examples of $\mathcal{T}\mathcal{H}\mathcal{N}$-operators:

(h) **Paranormal operators.** Paranormal operators on Banach spaces are $\mathcal{T}\mathcal{H}\mathcal{N}$-operators. Note that a paranormal operator need not to be $H(p)$. A subclass of paranormal operators is given by the class of all $p$-quasi-hyponormal operators on Hilbert spaces, see [35], where an operator $T \in L(H)$, $H$ a separable infinite dimensional Hilbert space, is said to be *p-quasi-hyponormal*, for some $0 < p \leq 1$, if

$$T^*(|T|^{2p} - |T^*|^{2p}T \geq 0,$$

where $|T| := (T^*T)^{1/2}$. Another subclass of paranormal operators on Hilbert spaces is given by the $A$ class of operators introduced by Furuta *et al.* [37], where $T \in L(H)$ is said to be *a class A* operator if $|T|^2 \leq |T^2|$.

(g) **quasi *-paranormal operators.** An operator $T \in L(H)$, $H$ a Hilbert space, is called *quasi *-paranormal* if

$$\|T^*Tx\|^2 \leq \|T^3x\|\|Tx\| \quad \text{for all unit vectors } x \in H$$

Every quasi *-paranormal operator is totally hereditarily normaloid, see [50]. The class of quasi *-paranormal contains the class of all *-paranormal operators, i.e. the class of $T \in L(H)$ for which

$$\|T^*x\|^2 \leq \|T^2x\| \quad \text{for all unit vectors } x \in H,$$

see [49] for details. Every *quasi-hyponormal operator* is quasi *-paranormal, see [49].

(f) **k-quasi-*-class A operators.** A bounded operator $T \in L(H)$, $H$ a separable Hilbert space, is said to be *k-quasi-*-class A operator* if

$$T^{*k}|T^2|T^k \geq T^{*k}|T^{*2}||T^*$$

Every k-quasi-*-class A operator is totally hereditarily normaloid, see [48]. For $k = 1$ we obtain the class of all quasi-*-class A operators, which is included in the class of all quasi *-paranormal operators.

In order to give some other examples of polaroid operator, we introduce a new class of operators. In the sequel by $\overline{Y}$ we denote the closure of $Y \subseteq X$.

**Definition 3.8.** An operator $T \in L(X)$, $X$ a Banach space, is said to be *k-quasi totally hereditarily normaloid*, $k$ a nonnegative integer, if the restriction $T|\overline{T^k(X)}$ is $\mathcal{THN}$.

Evidently, every $\mathcal{THN}$-operator is quasi-$\mathcal{THN}$, and if $T^k(X)$ is dense in $X$ then a quasi-$\mathcal{THN}$ operator $T$ is $\mathcal{THN}$.

We recall now some elementary algebraic facts. Suppose that $T \in L(X)$ and $X = M \oplus N$, with $M$ and $N$ closed subspace of $X$, $M$ invariant under $T$. With respect to this decomposition of $X$ it is known that $T$ may be represented by an upper triangular operator matrix $\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$, where $A \in L(M)$, $C \in L(N)$ and $B \in L(N, M)$. It is easily seen that for every $x = \begin{pmatrix} x \\ 0 \end{pmatrix} \in M$ we have $Tx = Ax$, so $A = T|M$. Let us consider now the case of operators $T$ acting on a Hilbert space $H$, and suppose that $T^k(H)$ is not dense in $H$. In this case we can consider the nontrivial orthogonal decomposition

$$H = \overline{T^k(H)} \oplus \overline{T^k(H)}^{\perp}, \tag{9}$$

where $\overline{T^k(H)}^{\perp} = \ker(T^*)^k$, $T^*$ the adjoint of $T$. Note that the subspace $\overline{T^k(H)}$ is $T$-invariant, so we can represent, with respect the decomposition (9), $T$ as an upper triangular operator matrix

$$\begin{pmatrix} T_1 & T_2 \\ 0 & T_3 \end{pmatrix}, \tag{10}$$

where $T_1 = T|\overline{T^k(H)}$. Moreover, $T_3$ is nilpotent. Indeed, if $x \in \overline{T^k(X)}^{\perp}$, an easy computation yields $T^k x = T \begin{pmatrix} 0 \\ x \end{pmatrix} = {T_3}^k x$. Hence ${T_3}^k x = 0$, since $T^k x \in \overline{T^k(H)} \cup \overline{T^k(H)}^{\perp} = \{0\}$. Therefore we have:

**Theorem 3.9.** *Suppose that $T \in L(H)$ and $T^k(H)$ non-dense in $H$. Then, according to the decomposition (9), $T = \begin{pmatrix} T_1 & T_2 \\ 0 & T_3 \end{pmatrix}$ is quasi-$\mathscr{T}HN$ if and only if $T_1$ is $\mathscr{T}\mathscr{H}\mathscr{N}$. Furthermore,*

$$\sigma(T) = \sigma(T_1) \cup \sigma(T_3) = \sigma(T_1) \cup \{0\}.$$

*Proof.* The first assertion is clear, since $T_1 = T|\overline{T^k(H)}$. The second assertion follows from the following general result: if $T := \begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$ is an upper triangular operator matrix acting on some direct sum of Banach spaces and $\sigma(A) \cap \sigma(B)$ has no interior points, then $\sigma(T) = \sigma(A) \cup \sigma(B)$; see [45]. ∎

Upper triangular operator matrices have been studied by many authors, see, for instance, [24, 28, 38, 56]. The next result improves Theorem 3.7.

**Theorem 3.10 ([10]).** *If $T \in L(H)$ is an analytically quasi-$\mathscr{T}\mathscr{H}\mathscr{N}$ operator, then $T$ is hereditarily polaroid. Moreover, $T$ has SVEP.*

In the sequel we give some examples of operators which are quasi-totally hereditarily normaloid.

(i) $(n, k)$**-quasiparanormal operators.** The class of quasi-paranormal operators may be extended as follows: $T \in L(H)$ is said to be $(n, k)$-*quasiparanormal* if

$$\|T^{k+1}x\| \leq \|T^{1+n}(T^k x)\|^{\frac{1}{1+n}} \|T^k x\|^{\frac{n}{1+n}} \quad \text{for all } x \in H.$$

The class of $(1, k)$- quasiparanormal operators has been studied in [55]. The $(1, 1)$-quasiparanormal operators has been studied in [54]. If $T^k(H)$ is not dense then, in the triangulation $T = \begin{pmatrix} T_1 & T_2 \\ 0 & T_3 \end{pmatrix}$, $T_1 = T|\overline{T^k(H)}$ is $n$-quasiparanormal, and hence $\mathscr{T}\mathscr{H}\mathscr{N}$, see [55].

(l) $k$**-quasiclass** $A$ **operators.** An extension of class $A$ operators is given by the class of all $k$-*quasiclass $A$ operators*, where $T \in L(H)$, $H$ a separable infinite dimensional Hilbert space, is said to be a $k$-quasiclass $A$ operator if

$$T^{*k}(|T|^2 - |T|^2)T^k \geq 0.$$

Every $k$-quasiclass $A$ operator is quasi-$\mathscr{T}\mathscr{H}\mathscr{N}$. Indeed, if $T$ has dense range, then $T$ is a class $A$ operator and hence paranormal. If $T$ does not have dense range, then $T$ with respect to the decomposition $H = \overline{T^k(H)} \oplus \ker T^{*k}$ may be represented as a matrix $T = \begin{pmatrix} T_1 & T_2 \\ 0 & T_3 \end{pmatrix}$, where $T_1 := T|\overline{T^k(H)}$ is a class $A$ operator, and hence $\mathscr{T}HN$, see [52].

As it has been observed in [36, Example 0.2], a quasi-class $A$ operator (i.e., $k = 1$) need not to be normaloid. This shows that, in general, a quasi-$\mathcal{THN}$ operator is not normaloid, so the class of quasi-$\mathcal{THN}$ operators properly contains the class of $\mathcal{THN}$ operators.

(m) ***k*-quasi *-paranormal operators.** An operator $T \in L(H)$, $H$ a separable infinite dimensional Hilbert space, is said to be *k-quasi *-paranormal*, $k \in \mathbb{N}$, if

$$\|T^*T^kx\|^2 \leq \|T^{k+2}x\|\|T^kx\| \quad \text{for all unit vectors } x \in H.$$

This class of operators contains the class of all quasi- $*$-paranormal operators (which corresponds to the value $k = 1$). Every $k$-quasi *-paranormal operator is quasi-$\mathcal{THN}$. Indeed, if $T^k$ has dense range then, $T$ is $*$-paranormal and hence $\mathcal{THN}$. If $T^k$ does not dense range, then $T$ may be decomposed, on $H = \overline{T^k(H)} \oplus \ker T^{*k}$, as $T = \begin{pmatrix} T_1 & T_2 \\ 0 & T_3 \end{pmatrix}$, where $T_1 = T|\overline{T^k(H)}$ is $*$-paranormal, hence $\mathcal{THN}$, see [49, Lemma 2.1].

(n) **$(p,k)$-quasihyponormal operators.** An extension of $p$-quasi-hyponormal operators is defined as follows: an operator $T \in L(H)$ is said to be $(p,k)$-*quasihyponormal* for some $0 < p \leq 1$ and $k \in \mathbb{N}$, if

$$T^{*k}|T^*|^{2p}T^k \leq T^{*k}|T|^{2p}T^k.$$

Every $(p,k)$-quasihyponormal operator $T$ with respect to the decomposition $H = \overline{T^k(H)} \oplus \ker T^{*k}$ may be represented as a matrix $T = \begin{pmatrix} T_1 & T_2 \\ 0 & 0 \end{pmatrix}$, where $T_1 := T|\overline{T^k(H)}$ is $k$-hyponormal (hence paranormal) and consequently $\mathcal{THN}$, see [40].

## 4 Weyl type theorems

If $T \in L(X)$, define

$$E(T) := \{\lambda \in \operatorname{iso}\sigma(T) : 0 < \alpha(\lambda I - T)\},$$

and

$$E^a(T) := \{\lambda \in \operatorname{iso}\sigma_a(T) : 0 < \alpha(\lambda I - T)\}.$$

Evidently, $E^0(T) \subseteq E(T) \subseteq E^a(T)$ for every $T \in L(X)$. Define

$$\pi_{00}(T) := \{\lambda \in \operatorname{iso}\sigma(T) : 0 < \alpha(\lambda I - T) < \infty\},$$

and

$$\pi^a_{00}(T) := \{\lambda \in \operatorname{iso}\sigma_{\rm a}(T) : 0 < \alpha(\lambda I - T) < \infty\}.$$

Let $p_{00}(T) := \sigma(T) \setminus \sigma_{\rm b}(T)$, i.e. $p_{00}(T)$ is the set of all poles of the resolvent of $T$.

**Definition 4.1.** A bounded operator $T \in L(X)$ is said to satisfy *Weyl's theorem*, in symbol *(W)*, if $\sigma(T) \setminus \sigma_{\rm w}(T) = \pi_{00}(T)$. $T$ is said to satisfy *a-Weyl's theorem*, in symbol *(aW)*, if $\sigma_{\rm a}(T) \setminus \sigma_{\rm uw}(T) = \pi^a_{00}(T)$. $T$ is said to satisfy property $(w)$, if $\sigma_{\rm a}(T) \setminus \sigma_{\rm uw}(T) = \pi_{00}(T)$.

Recall that $T \in L(X)$ is said to satisfy *Browder's theorem* if $\sigma_{\rm w}(T) = \sigma_{\rm b}(T)$, while $T \in L(X)$ is said to satisfy *a-Browder's theorem* if $\sigma_{\rm uw}(T) = \sigma_{\rm ub}(T)$. Weyl's theorem for $T$ entails Browder's theorem for $T$, while $a$-Weyl's theorem entails $a$-Browder's theorem. Either $a$-Weyl's theorem or property (w) entails Weyl's theorem. Property $(w)$ and $a$-Weyl's theorem are independent, see [12].

The generalized versions of Weyl type theorems are defined as follows:

**Definition 4.2.** A bounded operator $T \in L(X)$ is said to satisfy *generalized Weyl's theorem*, in symbol, *(gW)*, if $\sigma(T) \setminus \sigma_{\rm bw}(T) = E(T)$. $T \in L(X)$ is said to satisfies *generalized a-Weyl's theorem*, in symbol, *(gaW)*, if $\sigma_{\rm a}(T) \setminus \sigma_{\rm ubw}$ $(T) = E^a(T)$. $T \in L(X)$ is said to satisfy *generalized property* $(w)$, in symbol, $(gw)$, if $\sigma_{\rm a}(T) \setminus \sigma_{\rm ubw}(T) = E(T)$.

Recall that $T \in L(X)$ is said to satisfy *generalized Browder's theorem* if $\sigma_{\rm bb}(T) = \sigma_{\rm bw}(T)$, while $T \in L(X)$ is said to satisfy *generalized a-Browder's theorem* if $\sigma_{\rm ubb}(T) = \sigma_{\rm ubw}(T)$. Browder's theorem and generalized Browder's theorem are equivalent, and analogously $a$-Browder's theorem and generalized $a$-Browder's theorem are equivalent, see [15]. $a$-Browder's theorems entails Browder's theorems and if $T$, or $T'$, has SVEP then $a$-Browder's theorem holds for $T$. Generalized $a$-Weyl's theorem, as well as generalized property $(w)$, entails generalized $a$-Browder's theorem.

In the following diagrams we resume the relationships between all Weyl type theorems:

$$(gw) \quad \Rightarrow (w) \Rightarrow (\mathrm{W})$$

$$(\mathrm{gaW}) \Rightarrow (\mathrm{aW}) \Rightarrow (\mathrm{W}),$$

see [21, Theorem 2.3], [12] and [22]. Generalized property $(w)$ and generalized $a$-Weyl's theorem are also independent, see [21]. Furthermore,

$$(gw) \;\Rightarrow (\mathrm{gW}) \Rightarrow (\mathrm{W})$$

$$(\mathrm{gaW}) \Rightarrow (\mathrm{gW}) \Rightarrow (\mathrm{W})$$

see [21] and [22]. The converse of all these implications in general does not hold. Furthermore, by [2, Theorem 3.1],

$$(W) \text{ holds for } T \;\Leftrightarrow \text{ Browder's theorem holds for } T \text{ and } p_{00}(T) = \pi_{00}(T).$$

Under the polaroid conditions we have a very clear situation:

**Theorem 4.3 ([6]).** *Let $T \in L(X)$. Then we have:*

(i) *If $T$ is polaroid, then $(W)$ and $(gW)$ for $T$ are equivalent.*
(ii) *If $T$ is left-polaroid, then $(aW)$ and $(gaW)$ are equivalent for $T$, while $(W)$ and $(gW)$ are equivalent for $T$.*
(iii) *If $T$ is a-polaroid, then $(aW)$, $(gaW)$, $(w)$ and $(gw)$ are equivalent for $T$, while $(W)$ and $(gW)$ are equivalent for $T$.*

Although the polaroid conditions are neither necessary nor sufficient for an operator to satisfying Weyl type theorems, almost all of the commonly considered classes of operators satisfy Weyl type theorems since they are polaroid type and have the single valued extension property (SVEP). Indeed, we have:

**Theorem 4.4.** *Let $T \in L(X)$ be polaroid and suppose that either $T'$ has SVEP at every $\lambda \notin \sigma_{\mathrm{uw}}(T)$ or $T$ has SVEP at every $\lambda \notin \sigma_{\mathrm{lw}}(T)$. Then both $T$ and $T'$ satisfy Weyl's theorem.*

*Proof.* Each one of the assumptions on the SVEP ensures that $T$, or equivalently $T'$, satisfies Browder's theorem. In fact, if $T'$ has SVEP at every $\lambda \notin \sigma_{\mathrm{uw}}(T)$, then $a$-Browder's theorem (and hence Browder's theorem) holds for $T$, while if $T$ has SVEP at every $\lambda \notin \sigma_{\mathrm{lw}}(T)$ then $a$-Browder's theorem (and hence Browder's theorem) holds for $T'$, see [8, Theorem 2.3] The polaroid condition for $T$ entails that $p_{00}(T) = \pi_{00}(T)$, so Weyl's theorem holds for $T$. If $T$ is polaroid, then $T'$ is polaroid and hence $p_{00}(T') = \pi_{00}(T')$, so Weyl's theorem holds also for $T'$. ∎

For a bounded operator $T \in L(X)$, define $\Pi^a(T) := \sigma_{\mathrm{a}}(T) \setminus \sigma_{\mathrm{ld}}(T)$. It is clear that $\Pi^a_{00}(T)$ is the set of all left poles of the resolvent.

**Theorem 4.5.** *Let $T \in L(X)$ be left polaroid and suppose that either $T$ or $T'$ has SVEP. Then $T$ satisfies generalized a-Weyl's theorem.*

*Proof.* $T$ satisfies $a$-Browder's theorem and the left polaroid condition entails that $\Pi^a(T) = E^a(T)$. By [14, Theorem 2.18] then $(gaW)$ holds for $T$. ∎

**Theorem 4.6.** *Let $T \in L(X)$ be polaroid. Then we have:*

(i) *if $T'$ has SVEP at every $\lambda \notin \sigma_{\mathrm{uw}}(T)$, then $(gaW)$ and $(gw)$ hold for $T$.*
(ii) *If $T$ has SVEP at every $\lambda \notin \sigma_{\mathrm{lw}}(T)$, then $(gaW)$ and $(gw)$ hold for $T'$.*

*Proof.* (i) We show that $\sigma_{\mathrm{uw}}(T) = \sigma_{\mathrm{w}}(T)$. Let $\lambda \notin \sigma_{\mathrm{uw}}(T)$. Then $\lambda I - T \in W_+(X)$, so $\mathrm{ind}(\lambda I - T) \le 0$. Since $T'$ has SVEP at $\lambda$ by Remark 2.7 we have $q(\lambda I - T) < \infty$. Consequently, $\mathrm{ind}(\lambda I - T) \ge 0$, see [1, Theorem 3.4], and hence $\lambda \notin \sigma_{\rm w}(T)$. This shows that $\sigma_{\rm w}(T) \subseteq \sigma_{\rm uw}(T)$ and since the opposite inclusion holds for every operator then $\sigma_{\rm uw}(T) = \sigma_{\rm w}(T)$. Now, as we have seen in the proof of Theorem 2.9, we have $\sigma_{\rm a}(T) = \sigma(T)$, thus $\sigma(T) \setminus \sigma_{\rm w}(T) = \sigma_{\rm a}(T) \setminus \sigma_{\rm uw}(T)$ and $\pi_{00}(T) = \pi_{00}^a(T)$. Therefore, $(aW)$ holds for $T$ and this by Theorem 4.3 is equivalent, since by Theorem 2.9 $T$ is $a$-polaroid, to saying that $(gaW)$ and $(gw)$ hold for $T$.

(ii) We show that $\sigma_{\rm uw}(T') = \sigma_{\rm w}(T')$, or equivalently $\sigma_{\rm lw}(T) = \sigma_{\rm w}(T)$ . Let $\lambda \notin \sigma_{\rm lw}(T)$. Then $\lambda I - T \in W_-(X)$, so $\text{ind}(\lambda I - T) \geq 0$. Since $T$ has SVEP at $\lambda$ by Remark 2.7 we then have $p(\lambda I - T) < \infty$. Consequently, $\text{ind}(\lambda I - T) \leq 0$, see [1, Theorem 3.4], and hence $\lambda \notin \sigma_{\rm w}(T)$. It is easily seen that $\sigma_{\rm a}(T') = \sigma_{\rm s}(T) = \sigma(T) = \sigma(T')$ from which we obtain $\pi_{00}(T') = \pi_{00}^a(T')$. Therefore, $(aW)$ holds for $T'$ and this by Theorem 4.3 is equivalent, since by Theorem 2.9 $T'$ is $a$-polaroid, to saying that $(gaW)$ and $(gw)$ hold for $T'$. ■

Let $\mathscr{H}_{nc}(\sigma(T))$ denote the set of all analytic functions, defined on an open neighborhood of $\sigma(T)$, such that $f$ is non-constant on each of the components of its domain. Define, by the classical functional calculus, $f(T)$ for every $f \in \mathscr{H}_{nc}(\sigma(T))$.

**Theorem 4.7.** *Suppose that $T \in L(X)$ has SVEP and let $f \in \mathscr{H}_{nc}(\sigma(T))$.*

(i) *If $T$ is polaroid, then $f(T)$ satisfies $(gW)$.*
(ii) *If $T$ is left polaroid, then $f(T)$ satisfies $(gaW)$.*
(iii) *If $T$ is a-polaroid, then $f(T)$ satisfies both $(gaW)$ and $(gw)$.*

*Proof.* (i) $f(T)$ is polaroid and by [1, Theorem 2.40] has SVEP. Combining Theorem 4.4 and Theorem 4.3 we then conclude that $f(T)$ satisfies $(gW)$ .

(ii) $f(T)$ is left polaroid by [6, Lemma 3.11] and has SVEP. Combining Theorem 4.5 and Theorem 4.3 it then follows that $f(T)$ satisfies $(gaW)$.

(iii) By part (ii) $f(T)$ satisfies $(gaW)$, since it is also left polaroid. $f(T)$ is $a$-polaroid by [6, Lemma 3.11] and has SVEP. By Theorem 4.3 then $f(T)$ satisfies also $(gw)$. ■

The next two examples show that the assumption of being polaroid in part (i) of Theorem 4.7 is not sufficient to ensure property $(gaW)$, or $(gw)$.

*Example 4.8.* Let $R \in L(\ell^2(\mathbb{N}))$ be the right shift and let $Q$ denote the quasi-nilpotent operator defined as

$$Q(x_1, x_2, \dots) := (0, \frac{x_2}{2}, \frac{x_3}{3}, \dots) \quad \text{for all } x = (x_1, x_2, \dots) \in \ell^2(\mathbb{N}).$$

Let $T := R \oplus Q$. Then $T$ has SVEP, since both $R$ and $Q$ have SVEP, and is polaroid, since $\sigma(T) = D(0, 1)$, where $D(0, 1)$ is the closed unit disc of $\mathbb{C}$ centered at 0 and radius 1, has no isolated points. We also have $\sigma_{\rm a}(T) = \Gamma \cup \{0\}$, where $\Gamma$ denotes the unit circle of $\mathbb{C}$. Therefore, $\sigma_{\rm uw}(T) \subseteq \sigma_{\rm a}(T) = \Gamma \cup \{0\}$. Now, for every $\lambda \notin \sigma_{\rm uw}(T)$ the SVEP of $T$ at $\lambda$ implies that $\lambda \notin \text{acc}\, \sigma_{\rm a}(T) = \Gamma$, thus $\Gamma \subseteq \sigma_{\rm uw}(T)$. Clearly,

$p(T) = p(R) + p(Q) = \infty$, so $0 \in \sigma_{\mathrm{ub}}(T) = \sigma_{\mathrm{uw}}(T)$, where the last equality holds since $T$ satisfies $a$-Browder's theorem. Therefore, $\sigma_{\mathrm{uw}}(T) = \Gamma \cup \{0\}$, hence $\sigma_{\mathrm{a}}(T) \setminus \sigma_{\mathrm{uw}}(T) = \emptyset$. But $\pi^a_{00}(T) = \{0\}$, so $a$-Weyl's theorem does not hold for $T$.

Note that property $(gw)$ holds for $T$. Indeed, $\sigma_{\mathrm{ubw}}(T) \subseteq \sigma_{\mathrm{uw}}(T) = \Gamma \cup \{0\}$, and repeating the same argument used above, and generalized $a$-Browder's theorem for $T$) we easily obtain $\sigma_{\mathrm{ubw}}(T) = \Gamma \cup \{0\}$. Clearly, $E(T) = \emptyset$ and hence $E(T) = \sigma_{\mathrm{a}}(T) \setminus \sigma_{\mathrm{ubw}}(T)$.

Also the assumption of being left polaroid in part (ii) of Theorem 4.7 is not sufficient to ensure property $(gw)$:

*Example 4.9.* Let $T$ be the hyponormal operator given by the direct sum of the 1-dimensional zero operator $U$ and the unilateral right shift $R$ on $\ell^2(\mathbb{N})$. Evidently, $T$ has SVEP and iso $\sigma_{\mathrm{a}}(T) = \{0\}$ since $\sigma_{\mathrm{a}}(T) = \Gamma \cup \{0\}$. Clearly, $T \in \Phi_+(X)$, and hence $T^2 \in \Phi_+(X)$, so $T^2(X)$ is closed, and since $p(T) = p(U) = 1$it then follows that 0 is a left pole of $T$, i.e. $T$ is left polaroid. We show that $T$ does not satisfy $(w)$ (and hence $(gw)$). We know that $\sigma_{\mathrm{uw}}(T) \subseteq \sigma_{\mathrm{a}}(T) = \Gamma \cup \{0\}$ and repeating the same argument of Example 4.8 we have $\Gamma \subseteq \sigma_{\mathrm{uw}}(T) \subseteq \Gamma \cup \{0\}$. Since $T \in B_+(X) \subseteq W_+(X)$ it then follows that $0 \notin \sigma_{\mathrm{uw}}(T)$, so $\sigma_{\mathrm{uw}}(T) = \Gamma$, and hence

$$\sigma_{\mathrm{a}}(T) \setminus \sigma_{\mathrm{uw}}(T) = \{0\} \neq \pi_{00}(T) = \emptyset,$$

thus $T$ does not satisfy $(w)$ (and hence $(gw)$.

The following perturbation result has been proved in [5, Theorem 3.12].

**Theorem 4.10.** *Suppose that $T \in L(X)$ and $K \in L(X)$ an algebraic operator commuting with $T \in L(X)$. If $T \in L(X)$, or $T^*$, has SVEP and $T$, or $T^*$, is hereditarily polaroid, then $f(T + K)$ and $f(T^* + K^*)$ ) satisfies $(gW)$ for every $f \in \mathcal{H}_{nc}(\sigma(T + K))$.*

By using Theorem 4.10 and Theorem 4.4 we obtain the following result.

**Theorem 4.11 ([10]).** *Let $T \in L(H)$ be an analytically quasi-$\mathcal{THN}$ operator on a Hilbert space $H$, and let $K \in L(H)$ be an algebraic operator commuting with $T$. Then both $f(T + K)$ and $f(T' + K')$ satisfy $(gW)$ for every $f \in \mathcal{H}_{nc}(\sigma(T + K))$.*

Since $H(p)$ operators are hereditarily polaroid, we also have:

**Theorem 4.12.** *Suppose that $T \in L(X)$ has property $H(p)$, and let $K \in L(H)$ be an algebraic operator commuting with $T$. Then both $f(T + K)$ and $f(T' + K')$ satisfies $(gW)$ for every $f \in \mathcal{H}_{nc}(\sigma(T + K))$.*

For the dual $f(T' + K')$ we can say much more.

**Corollary 4.13.** *Suppose that $T \in L(X)$ has property $H(p)$, or $T \in L(H)$ be an analytically quasi-$\mathcal{THN}$ operator on a Hilbert space $H$. If $K \in L(H)$ is an algebraic operator commuting with $T$, then $f(T' + K')$ all Weyl type theorems for every $f \in \mathcal{H}_{nc}(\sigma(T + K))$.*

*Proof.* Suppose that $T$ has property $H(p)$. Since $T$ is hereditarily polaroid then $T+K$, is polaroid and hence also $T'+K'$ is polaroid. By Theorem 2.10 then $T'+K'$ is polaroid. Moreover, $T'+K'$ has SVEP and hence $f(T'+K')$ has SVEP for every $f \in \mathscr{H}_{nc}(\sigma(T+K))$, by [1, Theorem 2.40]. By Theorem 2.9 we then have that $f(T'+K')$ is $a$-polaroid, so Theorem 4.6 entails that $f(T'+K')$ satisfies all Weyl type theorems.

The proof for the case that $T$ is analytically quasi-$\mathscr{T}\mathscr{H}\mathscr{N}$ on a Hilbert space $H$ is the same. ∎

## References


1. P. Aiena: *Fredholm and local spectral theory, with application to multipliers.* Kluwer Acad. Publishers (2004).
2. P. Aiena: *Classes of Operators Satisfying a-Weyl's theorem.* Studia Math. **169** (2005), 105–122.
3. P. Aiena: *Quasi Fredholm operators and localized SVEP*, (2006). Acta Sci. Math. (Szeged), **73** (2007), 251–263.
4. Aiena, Pietro: *Algebraically paranormal operators on Banach spaces*. Banach J. Math. Anal. 7 (2013), no. 2, 136–145.
5. P. Aiena, E. Aponte: *Polaroid type operators under perturbations*. Studia Math. **214**, (2013), no. 2, 121–136.
6. P. Aiena, E. Aponte, E. Bazan: *Weyl type theorems for left and right polaroid operators.* Inte. Equa. Oper. Theory 66 (2010), no. 1, 1–20.
7. P. Aiena, M. T. Biondi, C. Carpintero: *On Drazin invertibility*, Proc. Amer. Math. Soc. **136**, (2008), 2839–2848.
8. P. Aiena, C. Carpintero, E. Rosas: *Some characterization of operators satisfying a-Browder theorem*. J. Math. Anal. Appl. **311**, (2005), 530–544.
9. P. Aiena, M. Chō, M. González: *Polaroid type operator under quasi-affinities.* J. Math. Anal. Appl. **371**, (2010), 485–495.
10. P. Aiena, J. Guillen, P. Peña : *A unifying approach to Weyl type theorems for Banach space operators*. Integ. Equ. Oper. Theory **77** (2013), 371–384.
11. P. Aiena, M. M. Neumann: *On the stability of the localized single-valued extension property under commuting perturbations*. Proc. Amer. Math. Soc. **141** (2013), no. 6, 2039–2050.
12. P. Aiena, P. Peña: *A variation on Weyl's theorem*. J. Math. Anal. Appl. **324** (2006), 566–579.
13. P. Aiena, F. Villafãne: *Weyl's theorem for some classes of operators.* Int. Equa. Oper. Theory **53**, (2005), 453–466.
14. A. Aluthge.: *On p-hyponormal operators for* $1 < p < 1$, Integral Equations Operator Theory **13**, (1990), 307–315.
15. M. Amouch, H. Zguitti: *On the equivalence of Browder's and generalized Browder's theorem*. Glasgow Math. Jour. **48**, (2006), 179–185.
16. T. Ando: *Operators with a norm condition*, Acta Sci. Math. (Szeged) **33**, (1972), 169–178.
17. S. K. Berberian: *An extension of Weyl's theorem to a class of not necessarily normal operators*. Michigan Math. J. **16** (1969),273–279.
18. M. Berkani: *On a class of quasi-Fredholm operators*. Integral Equations and Operator Theory **34** (1), (1999), 244–249.
19. M. Berkani: *Index of B-Fredholm operators and generalization of a Weyl's theorem*, Proceding American Mathematical Society, vol. 130, **6**, (2001), 1717–1723.
20. M. Berkani, M. Sarih: *On semi B-Fredholm operators.* Glasgow Mathematics Journal **43**, No. 4, (2001), 457–465.

21. M. Berkani, M. Amouch: *On the property (gw)*. Mediterr. J. Math. 5 (2008), no. 3, 371–378.
22. M. Berkani, J. J. Koliha: *Weyl type theorems for bounded linear operators*, Acta Sci. Math. (Szeged) **69** (2003), no. 1–2, 359–376.
23. M. Berkani, M. Sarih: *On semi B-Fredholm operators*, Glasgow Math. J. **43** (2001), 457–465.
24. X. Cao, M. Guo, B. Meng: *Weyl's theorem for upper triangular operator matrices.* Linear Alg. and Appl. **402** (2005), 61–73.
25. M. Chō, I. H. Jeon, J. I. Lee: *Spectral and structural properties of log-hyponormal operators* Glasgow Math. J. **42** (2000), 345–350.
26. L.A. Coburn: *Weyl's theorem for nonnormal operators.* Research Notes in Mathematics **51**, (1981).
27. J. B. Conway: *Subnormal operators* Michigan Math. J. **20** (1970), 529–544.
28. S.V. Djordjevíc , H. Zguitti: *Essential point spectra of operator matrices trough local spectral theory*, J. Math. Anal. and Appl. **338** (2008), 285–291.
29. B. P. Duggal: *Quasi-similar p-hyponormal operators.* Integ. Equ. Oper. theory **26**, (1996), 338–345.
30. B.P. Duggal: *Polaroid operators satisfying Weyl's theorem.* Linear Algebra Appl. **414** (2006), 271–277.
31. B.P. Duggal: *Hereditarily polaroid operators, SVEP and Weyl's theorem.* J. Math. Anal. Appl. **340** (2008), 366–373.
32. B.P. Duggal, S.V. Djordjevíc: *Generalized Weyl's theorem for a class of operators satisfying a norm condition.* Math. Proc. Royal Irish Acad. **104**A, (2004), 75–81.
33. B. P. Duggal, H. Jeon: *Remarks on spectral properties of p-hyponormal and log-hyponormal operators.* Bull. Kor. Math. Soc. **42**, (2005), 541–552.
34. B.P. Duggal, R. E. Harte, I. H. Jeon: *Polaroid operators and Weyl's theorem.* Proc. Amer. Math. Soc. **132** (2004), 1345–1349.
35. B. P. Duggal, I. H. Jeon<. *On p-quasi-hyponormal operators.* Linear Alg. and Appl. **422**, (2007), 331–340.
36. B. P. Duggal, I. H. Jeon, I H. Kim: *On Weyl's theorem for quasi-class A operators.* J. Korean Math. Soc. **43**, (2006), N. 4, 899–909.
37. T. Furuta: *Invitation to linear operators.*, Taylor and Francis, London-N. York 2001.
38. Y. K. Han, H. Y. Lee, W. Y. Lee: *Invertible completions of* $2 \times 2$ *upper triangular operator matrices.* Proc. Amer. Math. Soc. **129** (2001), 119–123.
39. Y. M. Han, J. I. Lee, D. Wang: *Riesz idempotent and Weyl's theorem for w-hyponormal operator.* Integ. Equ. Oper. theory **53**, (2005), 51–60.
40. I. H. Kim: *On* $(p,k)$*-quasihyponormal operators.* Math. Inequal. and Appl.7, **4**, (2004), 629–638.
41. J. J. Koliha: *Isolated spectral points* Proc. Amer. Math. Soc. **124** (1996), 3417–3424.
42. D. C. Lay: *Spectral analysis using ascent, descent, nullity and defect.* Math. Ann. **184** (1970), 197–214.
43. K. B. Laursen, M. M. Neumann: *Introduction to local spectral theory.*, Clarendon Press, Oxford 2000.
44. K.B. Laursen: *Operators with finite ascent.* Pacific J. Math. **152**, (1992), 323–36.
45. W. Y. Lee: *Weyl'spectra of operator matrices.* Proc. Amer. Math. Soc. **129** (2001), 131–138.
46. W. Y. Lee, S. H. Lee: *Some generalized theorems on p-quasihyponormal operators for* $0 < p < 1$. Nihonkai Math. J. **8**, (1997), 109–115.
47. C. Lin, Y. Ruan, Z. Yan: *w-hyponormal operators are subscalar.* Integr. Equ. oper. theory. **50**, (2004), 165–168.
48. S. Mecheri: *Isolated points of spectrum of k-quasi-*-class A operators*. Studia Math. 208 (2012), no. 1, 87–96.
49. S. Mecheri: *On a new class of operators and Weyl type theorems*. Filomat **27**(4), (2013), 629—636.
50. S. Mecheri, L. Braha: *Polaroid and p−paranormal operators.* Mathematical Inequalities and Appl. **16** (2013), 557–568.

51. M. Oudghiri: *Weyl's and Browder's theorem for operators satisfying the SVEP* Studia Math. **163**, 1, (2004), 85–101.
52. K. Tanahashi: *On log-hyponormal operators.* Integral Equations Operator Theory **34**, (1999), 364–372.
53. H. Weyl: *Uber beschrankte quadratiche Formen, deren Differenz vollsteig ist.* Rend. Circ. Mat. Palermo **27**, (1909), 373–92.
54. J. Yuan, Z. Gao: *Weyl spectrum of class A(n) and n-paranormal operators.* Integral Equations Operator Theory **60** (2008), 289–298.
55. J. T. Yuan, G. X. Ji: *On* $(n,k)$*-quasi paranormal operators.* Studia Math. **209**, (2012), 289–301.
56. E. H. Zerouali, H. Zguitti: *Perturbation of spectra of operator matrices and local spectral theory*. J. Math. Anal. and Appl. **324** (2006), 292–1005.

# On non self-adjoint spectral problems occurring in superconductivity

**Bernard Helffer**

**Abstract** In this survey we would like to discuss spectral properties of non self-adjoint operators appearing in the analysis of the long time behavior of the solutions of the time-dependent Ginzburg Landau system (due to Eliashberg-Gorkov) and to consider in particular the global stability of the stationary normal solutions. We will first recall some standard results on the time independent model including the Giorgi-Phillips Theorem and will then focus on the role of the electric current in comparison with the role of the exterior magnetic field for the time independent problem. The recent theorems have been obtained in collaboration with Y. Almog, X. Pan, R. Henry, K. Beauchard, and L. Robbiano or by R. Henry alone. This survey is a short version of a course given in July 2013 in Berder, supported by the programme ANR 2011 BS01019 01 NOSEVOL.

## 1 The Ginzburg-Landau model for superconductivity

### *1.1 The Ginzburg-Landau functional*

Let us describe the mathematical problem. It is naturally posed for domains in $\mathbb{R}^3$, but for cylindrical domains in $\mathbb{R}^3$, it is natural to consider a functional defined in a domain $\Omega \subset \mathbb{R}^2$, where $\Omega$ is the cross-section of the cylinder. This explains why we also consider models in $\mathbb{R}^2$ and we will here only consider this case. We assume to simplify that $\Omega$ is connected and simply connected. The Ginzburg-Landau functional is defined by

B. Helffer (✉)
Département de Mathématiques, Université Paris-Sud 11, UMR 8628 du CNRS, Bat. 425, Orsay, France

Laboratoire Jean Leray, Université de Nantes, Nantes, France



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_2

$$\mathscr{G}(\psi,\mathbf{A}) = \int_\Omega \left( |\nabla_{\kappa\sigma\mathbf{A}}\psi|^2 - \kappa^2|\psi|^2 + \frac{\kappa^2}{2}|\psi|^4 \right) d\mathbf{x} \\ + (\kappa\sigma)^2 \int_\Omega |\text{curl}\,\mathbf{A} - \beta|^2\, d\mathbf{x}\,, \qquad (1)$$

where, with $\mathbf{x} = (x, y)$, $d\mathbf{x}$ denotes the Lebesgue measure $dxdy$. Here the function $\psi$ is called the order parameter (or sometimes the wave function) and $\mathbf{A}$ is a magnetic potential. For $\mathbf{A} = (A_1, A_2)$, curl $\mathbf{A} = \partial_x A_2 - \partial_y A_1$ and $\nabla_{\kappa\sigma\mathbf{A}}$ denotes the magnetic gradient: $\nabla + i\kappa\sigma\mathbf{A}$. The symbol $\beta$ denotes a reference magnetic field and is called the external magnetic field or the applied magnetic field (in the constant magnetic field case, we take $\beta \equiv 1$), which is assumed to be in $L^2(\Omega)$. The parameter $\kappa > 0$ (the Ginzburg-Landau parameter) depends on the material, and $\sigma > 0$ (or rather the product $\kappa\sigma$) is a measure of the strength of the external magnetic field.

We are concerned with the analysis of the asymptotic regime $\kappa \to +\infty$, which corresponds to strong type II samples.

We will sometime write $\mathscr{G} = \mathscr{G}_{\kappa,\sigma}$, if we want to mention the parameters involved in the definition of the functional.

The natural domain of the functional is $H^1(\Omega, \mathbb{C}) \times H^1(\Omega, \mathbb{R}^2)$. However, due to the gauge invariance of $\mathscr{G}$, it is better to restrict the functional to the smaller set $H^1(\Omega, \mathbb{C}) \times H^1_{\text{div}}(\Omega)$, where

$$H^1_{\text{div}}(\Omega) = \left\{ \mathbf{V} = (V_1, V_2) \in H^1(\Omega, \mathbb{R})^2 \;\middle|\; \text{div}\,\mathbf{V} = 0 \text{ in } \Omega\,,\ \mathbf{V}\cdot\nu = 0 \text{ on } \partial\Omega \right\}. \qquad (2)$$

We define the Ginzburg-Landau ground state energy to be the infimum of the functional, i.e.

$$E(\kappa,\sigma) := \inf_{(\psi,\mathbf{A})\in H^1(\Omega)\times H^1_{\text{div}}(\Omega)} \mathscr{G}_{\kappa,\sigma}(\psi,\mathbf{A})\,. \qquad (3)$$

If $\mathbf{F}$ is the unique magnetic potential in $H^1_{\text{div}}(\Omega)$ such that:

$$\text{curl}\mathbf{F} = \beta\,,$$

we observe that:

$\mathscr{G}_{\kappa,\sigma}(\psi \equiv 0, \mathbf{F}) = 0$, which implies the inequality:

$$E(\kappa,\sigma) \le 0\,. \qquad (4)$$

The pair $(0, \mathbf{F})$ is called *normal state* in Physics and we will, in particular, study when we have equality or strict inequality in (4).

## 1.2 Minimizers and Ginzburg-Landau equations

As $\Omega$ is bounded, the existence of a minimizer is rather standard. A minimizer should satisfy the Euler-Lagrange equation, which is called in this context the Ginzburg-Landau system and reads:

$$\left.\begin{aligned} -\Delta_{\kappa\sigma\mathbf{A}}\psi &= \kappa^2(1-|\psi|^2)\psi\,, \\ \text{curl}\,\big(\text{curl}\,\mathbf{A}-\beta\big) &= -\tfrac{1}{\kappa\sigma}\Re\left(\overline{\psi}\,\nabla_{\kappa\sigma\mathbf{A}}\psi\right), \end{aligned}\right\} \quad \text{in} \quad \Omega\,, \tag{5a}$$

$$\left.\begin{aligned} \nu\cdot\nabla_{\kappa\sigma\mathbf{A}}\psi &= 0\,, \\ \text{curl}\,\mathbf{A} - \beta &= 0\,, \end{aligned}\right\} \quad \text{on} \quad \partial\Omega\,. \tag{5b}$$

Here, $-\Delta_{\kappa\sigma\mathbf{A}}$ is the magnetic Laplacian:

$$-\Delta_{\kappa\sigma\mathbf{A}} := (D_x+\kappa\sigma A_1)^2 + (D_y+\kappa\sigma A_2)^2\,, \text{ with } D_x = -i\partial_x\,,\, D_y = -i\partial_y\,,$$

and

$$\text{curl}\,^2\mathbf{A} = (\partial_y(\text{curl}\,\mathbf{A}), -\partial_x(\text{curl}\,\mathbf{A}))\,.$$

The analysis of the system (5) can be performed by PDE techniques which are recalled in [18]. We note that this system is nonlinear, that $H^1(\Omega)$ is, when $\Omega$ is bounded and regular in $\mathbb{R}^2$, compactly embedded in $L^p(\Omega)$ for all $p\in[1,+\infty)$.

Actually, the nonlinearity is weak in the sense that the principal part is a linear elliptic system. One can show in particular that the solution in $H^1(\Omega,\mathbb{C})\times H^1_{\text{div}}(\Omega)$ of the elliptic system (5) is actually, when $\Omega$ is regular, in $C^\infty\big(\overline{\Omega}\big)$.

## 1.3 Basic properties for solutions of the Ginzburg-Landau equations

The first important property which is a consequence of the maximum principle is

**Proposition 1.** *If $(\psi,\mathbf{A})\in H^1(\Omega)\times H^1(\Omega,\mathbb{R}^2)$ is a (weak) solution to (5), then*

$$\|\psi\|_{L^\infty(\Omega)} \leq 1\,. \tag{6}$$

Using Proposition 1, we can get (see [18] for details) various a priori estimates on solutions to the Ginzburg-Landau equations (5), which play an important role in the whole theory.

**Proposition 2.** *Let $\Omega\subset\mathbb{R}^2$ be bounded and smooth, and let $\beta\in L^2(\Omega)$ be given. Then for all $p\geq 2$, there exists a constant $C=C(p)>0$ such that for all solutions $(\psi,\mathbf{A})\in H^1(\Omega)\times H^1_{\text{div}}(\Omega)$ to (5), we have*

$$\|\nabla^2_{\kappa\sigma\mathbf{A}}\psi\|_p \leq \kappa^2\|\psi\|_p\,, \tag{7}$$

$$\|\nabla_{\kappa\sigma\mathbf{A}}\psi\|_2 \leq \kappa\|\psi\|_2\,, \tag{8}$$

$$\|curl\,\mathbf{A}-\beta\|_{W^{1,p}(\Omega)} \leq \frac{C}{\kappa\sigma}\|\psi\|_\infty\,\|\nabla_{\kappa\sigma\mathbf{A}}\psi\|_p\,. \tag{9}$$

## 1.4 The Giorgi-Phillips Theorem for minimizers

We observe that $(0,\mathbf{F})$ is a trivial critical point of the functional $\mathscr{G}$, i.e., a trivial solution of the Ginzburg-Landau system (5). The pair $(0,\mathbf{F})$ is often called the *normal state* or *normal solution*. It is natural to discuss—as a function of $\sigma$—whether this pair is a local or global minimizer. When $\sigma$ is large, one will show that this solution is effectively the unique global minimizer. One says that in this case the superconductivity is destroyed. In other words, the order parameter is identically zero in $\Omega$ . Let us give a rather simple proof of this result that roughly says that $(0,\mathbf{F})$ is the unique minimizer of the functional when the strength of the exterior magnetic field is sufficiently large. We will actually show a stronger result for all the solutions of the associated Ginzburg-Landau system.

So we assume that we have a **nonnormal** stationary point $(\psi,\mathbf{A})$ for $\mathscr{G}$, that is a solution $(\psi,\mathbf{A})\in H^1(\Omega)\times H^1_{\mathrm{div}}(\Omega)$ of (5) satisfying

$$\int_\Omega |\psi(\mathbf{x})|^2\,d\mathbf{x} > 0\,. \tag{10}$$

By (8), (9), and (6), and using a standard inequality on the curl $-$ div system for controlling $\|\mathbf{A}-\mathbf{F}\|^2$ in $\Omega$ by $\|\mathrm{curl}\,\mathbf{A}-\beta\|^2$, we get

$$\|\nabla_{\kappa\sigma\mathbf{A}}\psi\|_2^2 + (\kappa\sigma)^2\|\mathbf{A}-\mathbf{F}\|_2^2 \leq C_\Omega\kappa^2\|\psi\|_2^2\,. \tag{11}$$

Writing $\mathbf{A}=\mathbf{A}-\mathbf{F}+\mathbf{F}$ and implementing (6) and (11) give

$$\int_\Omega |(\nabla + i\kappa\sigma\mathbf{F})\psi|^2\,d\mathbf{x} \leq 2C_\Omega\kappa^2\int_\Omega |\psi(\mathbf{x})|^2\,d\mathbf{x}\,. \tag{12}$$

Since $\psi$ satisfies (10), we obtain

$$\lambda_1^N(\sigma\kappa\mathbf{F}) \leq 2C_\Omega\kappa^2\,, \tag{13}$$

where $\lambda_1^N(\sigma\kappa\mathbf{F})$ denotes the ground state energy of the Neumann realization of $-\Delta_{\sigma\kappa\mathbf{F}}$ in $\Omega$.

We observe that $\lambda_1^N(\sigma\kappa\mathbf{F}) > 0$. So by combining an analysis in the small $B$ regime (perturbation theory) and for large $B$ (see below Theorem 2), and the continuity of $B\mapsto\lambda_1^N(B\mathbf{F})$, we get the existence of a constant $C_0>0$ such that

$$\lambda_1^N(\sigma\kappa\mathbf{F}) \geq \frac{1}{C_0}\min(\sigma\kappa, (\sigma\kappa)^2)\,. \tag{14}$$

Thus, we find that if a nontrivial stationary point $(\psi, \mathbf{A})$ exists, then

$$\sigma \leq C\,(1+\kappa)\,.$$

What we have obtained can be reformulated as the following theorem.

**Theorem 1 (Giorgi-Phillips).** *Let $\Omega \subset \mathbb{R}^2$ be smooth, bounded, and simply connected, and let $\beta$ in (5) be continuous and satisfy*

$$\beta(x) \geq c_{min} > 0\,, \qquad \forall x \in \Omega\,.$$

*Then there exists a constant $C$ such that if*

$$\sigma \geq C\max\{\kappa, 1\}\,,$$

*then the pair $(0, \mathbf{F})$ is the unique solution to (5) in $H^1(\Omega)\times H^1_{\rm div}(\Omega)$.*

We have used in the proof of Theorem 1:

**Theorem 2 (Lu-Pan).**

$$\lambda_1^N(BF) = B\min(b, \Theta_0 b') + o(B)\,,$$

*where $\Theta_0 \in (0,1)$, $b = \inf_{x\in\Omega}\beta(x)$ and $b' = \inf_{x\in\partial\Omega}\beta(x)$.*

Two models are indeed involved in the proof by localization: the model with constant magnetic fields

$$(D_x - \frac{B}{2}\beta(x_j,y_j)y)^2 + (D_y + \frac{B}{2}\beta(x_j,y_j)x)^2\,,$$

in $\mathbb{R}^2$ and the Neumann realization of the same operator in $\mathbb{R}^2_+$.

The bottom of the spectrum of the first one is $B|\beta(x_j,y_j)|$ and the bottom of the spectrum of the second one is $\Theta_0 B|\beta(x_j,y_j)|$.

*Remark 1.* In this form this theorem is due to Lu-Pan [32]. Many improvements concerning the control the remainder term $o(B)$ have been obtained (see [18] and the references therein and for more recent references [34] or [24]).

*Remark 2.* The Giorgi-Phillips statement is the starting point of the analysis of the third critical field corresponding to the transition between normal minimizers and non-normal minimizers. We refer to the books of Fournais-Helffer [18] and Sandier-Serfaty [37] for a detailed analysis of the behavior of these critical fields and the references therein.

To conclude this section, we treat the case when $\beta$ vanishes in $\Omega$ but

$$|\beta(\mathbf{x})| + |\nabla\beta(\mathbf{x})| \geq \hat{c}_{min} > 0. \tag{15}$$

More precisely, introducing $\mathscr{Z}(\beta) = \beta^{-1}(0)$, one has the theorem:

**Theorem 3 (Pan-Kwek).**

$$\lim_{B\to+\infty} \frac{\lambda_1^N(B\mathbf{A})}{B^{\frac{2}{3}}} = [\alpha_1(\beta)]^{\frac{2}{3}}, \tag{16}$$

*where*

$$\alpha_1(\beta) = \min\left\{\frac{1}{2}\hat{\nu}_0^{\frac{3}{2}} \inf_{\mathbf{x}\in\Omega\cap\mathscr{Z}(\beta)} |\nabla\beta(\mathbf{x})| \,,\, \inf_{\mathbf{x}\in\partial\Omega\cap\mathscr{Z}(\beta)} \zeta(\vartheta(\mathbf{x}))^{\frac{3}{2}}|\nabla\beta(\mathbf{x})|\right\}, \tag{17}$$

*where $\hat{\nu}_0$ is defined in (29) and $\vartheta(\mathbf{x})$ denotes the angle between curl $\beta$ and the tangent vector of $\partial\Omega$ at $\mathbf{x}$ and $\zeta(\vartheta)$ denotes the lowest eigenvalue of the Neumann realization of $-\Delta_{\mathbf{A}_\vartheta}$ in $\mathbb{R}^2_+$ with $\mathbf{A}_\vartheta = -\frac{|\mathbf{x}|^2}{2}(\cos\vartheta, \sin\vartheta)$ .*

As a consequence, we can extend the Giorgi-Phillips theorem to this case.

**Theorem 4.** *Let $\Omega \subset \mathbb{R}^2$ be smooth, bounded, and simply connected, and let $\beta$ be in $C^\infty(\overline{\Omega})$ and satisfying (15). Then there exists a constant $C$ such that if*

$$\sigma \geq C \max\{\kappa^2, 1\},$$

*then the pair $(0, \mathbf{F})$ is the unique solution to (5) in $H^1(\Omega) \times H^1_{\rm div}(\Omega)$ .*

Notice that a more precise statement is given in [33] in the limit $\kappa$ large.

## 2 Time-Dependent Ginzburg Landau I: models

### *2.1 The model in superconductivity*

The physical problem is posed in a domain $\Omega$ with specific boundary conditions which will be discussed later. We will first analyze here limiting situations where the domain, possibly after a blowing argument, becomes the whole space (or the half-space). We work in dimension 2. We assume that a magnetic field of magnitude $\mathscr{H}_e$ is applied perpendicularly to the sample and identified (via its intensity) with a function. We denote the normal conductivity of the sample by $\varsigma$. Then the time-dependent Ginzburg-Landau system (also known as the Gorkov-Eliashberg equations) is in $(0, T) \times \Omega$ :

$$\begin{cases} \partial_t\psi + i\kappa\phi\psi = \Delta_{\kappa\mathbf{A}}\psi + \kappa^2(1-|\psi|^2)\psi \,, \\ \kappa^2 \text{curl}\,^2\mathbf{A} + \varsigma(\partial_t\mathbf{A} + \nabla\phi) = \kappa\,\text{Im}\,(\bar{\psi}\cdot\nabla_{\kappa\mathbf{A}}\psi) + \kappa^2\text{curl}\,\mathscr{H}_e \,, \end{cases} \tag{18}$$

where the new object is the electric potential $\phi$.

In addition, we assume that $(\psi, \mathbf{A}, \phi)$ satisfies an initial condition at $t = 0$. Note that many physicists are assuming that curl $\mathscr{H}_e = 0$ and this is what we will do in the next section.

In order to solve this equation, one should also define a gauge (Coulomb, Lorentz,...). The orbit of $(\psi, \mathbf{A}, \phi)$ by the gauge group is

$$\{(\exp(i\kappa q)\,\psi, \mathbf{A} + \nabla q, \phi - \partial_t q)\,|\, q \in \mathscr{Q}\}\,,$$

where $\mathscr{Q}$ is a suitable space of regular functions of $(t, x, y)$. We refer to Bauman-Jadallah-Phillips [6] (Paragraph B in the introduction). We will choose the Coulomb gauge which reads $\operatorname{div} \mathbf{A} = 0$ for any $t$. Another possibility could be to take $\operatorname{div} \mathbf{A} = \omega\phi$ (Lorentz gauge). As in the analysis of the time independent case, the "normal" solutions will play an important role.

## 2.2 From Ginzburg-Landau to TDGL

Let us make the parallel between the standard GL case and TDGL at the level of the models. The Schrödinger operators with constant magnetic field in $\mathbb{R}^2$ and in $\mathbb{R}^2_+$ are the basic models for analyzing the general Schrödinger operator in $\Omega$. For TDGL, the models are $D_x^2 + D_y^2 + i\,c\,y$, in $\mathbb{R}^2$, $D_x^2 + D_y^2 + i\,c\,(x\cos\theta + y\sin\theta)$ in $\mathbb{R}^2_+$ (affine case), $D_x^2 + (D_y - \alpha x^2)^2 + i\,c\,y$ in $\mathbb{R}^2$ analyzed in [3] and in $\mathbb{R}^2_+$ in [4, 5]

$$D_x^2 + (D_y - \alpha(x\sin\theta - y\cos\theta)^2)^2 + ic(x\cos\theta + y\sin\theta)$$

(only in the case $\theta = \frac{\pi}{2}$). Here we have used the notation $D_x = -i\partial_x$, $D_y = -i\partial_y$.

The results obtained in [3–5] correspond in some sense to the results which have been obtained for the Schrödinger operator with constant magnetic field for the analysis of the time independent problem. In the TDGL case, we are facing many new difficulties:

- Treat the spectral analysis of non self-adjoint problems. Already in the linear case, the decay of the associated semi-group does not depend uniquely on the knowledge of the spectrum, but also on resolvent estimates in the complex planes.
- The notion of stationary solutions has to be defined.
- The global existence of solutions has to be proved.
- The notion of stability has to be defined. Roughly speaking we hope to find conditions on the initial data and on the current implying the convergence of the solution to the stationary one and to measure the decay.
- The technical problems relative to the existence of corners have to be controlled.

### 2.2.1 Stationary normal solutions: first analysis

We now determine the stationary (i. e. time independent) normal solutions of the system. From (64), we see that if $(0, \mathbf{A}, \phi)$ is such a solution, then $(\mathbf{A}, \phi)$ satisfies the system

$$\kappa^2 \text{curl}\,(\text{curl}\,\mathbf{A}) + \varsigma \nabla\phi = \kappa^2 \text{curl}\,\mathscr{H}_e\,,\ \text{div}\,\mathbf{A} = 0 \quad \text{in } \Omega\,. \tag{19}$$

Interpreting these two equations as the Cauchy-Riemann equations, this can be rewritten (in addition to the divergence free condition) as the property that

$$\kappa^2(\text{curl}\,\mathbf{A} - \mathscr{H}_e) + \, i\varsigma\phi\,,$$

is an holomorphic function in $\Omega$.

## 2.3 *Special situation: $\phi$ affine*

Here we follow the exposition of [22] and the reader can also look for a more elementary presentation at the last chapters of [23]. As simplest non trivial example, we observe that, if $\Omega = \mathbb{R}^2$, (18) has the following stationary normal state solution

$$\mathbf{A} = \frac{1}{2J}(Jx+h)^2(0,1)\,, \quad \phi = \frac{\kappa^2 J}{\varsigma}y\,. \tag{20}$$

Note that curl $\mathbf{A} = (Jx+h)$, that is, the induced magnetic field equals the sum of the applied magnetic field $h$ and the magnetic field produced by the electric current $Jx$.

For this normal state solution, the linearization of (18) with respect to the order parameter is

$$\partial_t\psi + \frac{i\kappa^3 Jy}{\varsigma}\psi = \Delta\psi + \frac{i\kappa}{J}(Jx+h)^2\partial_y\psi - (\frac{\kappa}{2J})^2(Jx+h)^4\psi + \kappa^2\psi\,. \tag{21}$$

Applying the transformation $x \to x - h/J$ and taking for simplification $\kappa = 1$, the time-dependent linearized Ginzburg-Landau equation takes the form

$$\frac{\partial\psi}{\partial t} + i\frac{J}{\varsigma}y\psi = \Delta\psi + iJx^2\frac{\partial\psi}{\partial y} - \Big(\frac{1}{4}J^2x^4 - 1\Big)\psi\,. \tag{22}$$

Rescaling $x$ and $t$ by applying: $t \to J^{2/3}t$ ; $(x,y) \to J^{1/3}(x,y)$ , yields

$$\partial_t u = -(\mathscr{A}_{0,c} - \lambda)u\,, \tag{23}$$

where

$$\mathscr{A}_{0,c} := D_x^2 + (D_y + \frac{1}{2}x^2)^2 + i\,c\,y\,, \tag{24}$$

and

$$c = 1/\varsigma\ ;\ \lambda = \frac{1}{J^{2/3}}\ ;\ u(x,y,t) = \psi(J^{-1/3}x, J^{-1/3}y, J^{-2/3}t)\,.$$

Our main problem will be to analyze the long time property of the attached semi-group. We now apply the transformation

$$u \to u\, e^{icyt}$$

to obtain

$$\partial_t u = -\left(D_x^2 u + (D_y + \frac{1}{2}x^2 - ct)^2 u - \lambda u\right)\,. \tag{25}$$

Note that considering the partial Fourier transform with respect to the $y$ variable, we obtain for the partial Fourier transform $\hat{u}$ of $u$:

$$\partial_t \hat{u} = -D_x^2 \hat{u} - \left[\left(\frac{1}{2}x^2 + (-ct+\omega)\right)^2 - \lambda\right]\hat{u}\,. \tag{26}$$

This can be rewritten as the analysis of a family (depending on $\omega \in \mathbb{R}$) of time-dependent problems on the line

$$\partial_t \hat{u} = -\mathscr{M}_{\beta(t,\omega)}\hat{u} + \lambda \hat{u}\,, \tag{27}$$

with $\mathscr{M}_\beta$ being the well-known anharmonic oscillator (also called the Montgomery operator in other contexts):

$$\mathscr{M}_\beta = D_x^2 + (\frac{1}{2}x^2 + \beta)^2\,, \tag{28}$$

and

$$\beta(t,\omega) = -ct + \omega\,.$$

An important quantity appearing also in Theorem 3 is

$$\hat{\nu}_0 = \inf_\beta \nu(\beta)\,, \tag{29}$$

where $\nu(\beta)$ is the ground state energy of $\mathscr{M}_\beta$.

## 2.4 The results by Almog-Helffer-Pan [3]

The main point concerning the previously defined operator $\overline{\mathscr{A}_{0,c}}$ is to obtain an optimal control of the decay of the associated semi-group as $t \to +\infty$.

**Theorem 5.** *If $c \neq 0$, $\mathscr{A} = \overline{\mathscr{A}_{0,c}}$ has compact resolvent, empty spectrum, and there exists $C > 0$ such that*

$$\| \exp(-t\mathscr{A}) \| \leq \exp\Big(-\frac{2\sqrt{2c}}{3} t^{3/2} + C t^{3/4}\Big)\,, \tag{30}$$

*for any $t \geq 1$ and*

$$\|(\mathscr{A} - \lambda)^{-1}\| \leq \exp\Big(\frac{1}{6c} (\Re\,\lambda)^3 + C\,(\Re\,\lambda)^{3/2}\Big)\,, \tag{31}$$

*for all $\lambda$ such that $\Re\,\lambda \geq 1$.*

Here a semi-classical analysis of the operator $\mathscr{M}_\beta$ as $|\beta| \to \pm\infty$ plays an important role. We refer to [3] for details and to [21] for the involved semi-classical analysis.

If we consider instead the Dirichlet realization $\mathscr{A}_c^D$ of $\mathscr{A}_{0,c}$ in $\{y > 0\}$, it is easily proven that $\mathscr{A}_c^D$ has compact resolvent if $c \neq 0$. We prove in [4] that if the spectrum of $\mathscr{A}_c^D$ is not empty then the decay of the semi-group $\exp -t\mathscr{A}_c^D$ is exponential with a rate corresponding to $\inf_{z \in \varsigma(\mathscr{A}_c^D)} \Re\, z$. We will explain the argument in the case of a simpler model: the complex Airy operator. We also conjecture in [4] that $\varsigma(\mathscr{A}_c^D)$ is not empty and give a proof of the statement for $|c|$ large enough and in [5] for $|c|$ small enough.

## 2.5 A simplified model : no magnetic field

We assume, following Almog [1], that a current of constant magnitude $J$ is being flown through the sample in the $x$ axis direction, and that there is no applied magnetic field: $h = 0$. Then (18) has (in some asymptotic regime) the following stationary normal state solution

$$\mathbf{A} = 0\,, \quad \phi = Jx\,. \tag{32}$$

For this normal state solution, the linearization of (18) gives

$$\partial_t \psi + iJx\psi = \Delta_{x,y}\psi + \psi\,, \tag{33}$$

whose analysis is (see ahead) strongly related to the Airy equation.

The complex Airy operator in $\mathbb{R}$

This operator can be defined as the closed extension $\mathscr{A}$ of the differential operator on $C_0^\infty(\mathbb{R})$ $\mathscr{A}_0^+ := D_x^2 + ix$. We observe that $\mathscr{A} = (\mathscr{A}_0^-)^*$ with $\mathscr{A}_0^- := D_x^2 - ix$ and that its domain is

$$D(\mathscr{A}) = \{u \in H^2(\mathbb{R})\,,\, x\,u \in L^2(\mathbb{R})\}\,.$$

In particular $\mathscr{A}$ has compact resolvent.

It is also easy to see that

$$\Re\, \langle \mathscr{A} u \,|\, u\rangle \geq 0\,. \tag{34}$$

Hence $-\mathscr{A}$ is the generator of a semi-group $S_t$ of contraction,

$$S_t = \exp -t\mathscr{A}\,. \tag{35}$$

Hence all the results of this theory can be applied.

In particular, we have, for $\Re\,\lambda < 0$

$$||(\mathscr{A} - \lambda)^{-1}|| \leq \frac{1}{|\Re\,\lambda|}\,. \tag{36}$$

A very special property of this operator is that, for any $a \in \mathbb{R}$,

$$T_a \mathscr{A} = (\mathscr{A} - ia) T_a\,, \tag{37}$$

where $T_a$ is the translation operator $(T_a u)(x) = u(x-a)$.

As immediate consequence, we obtain that the spectrum is empty and that the resolvent of $\mathscr{A}$, which is defined for any $\lambda \in \mathbb{C}$ satisfies

$$||(\mathscr{A} - \lambda)^{-1}|| = ||(\mathscr{A} - \Re\,\lambda)^{-1}||\,. \tag{38}$$

One can also look at the semi-classical question, i.e. consider the operator

$$\mathscr{A}_h = h^2 D_x^2 + ix\,, \tag{39}$$

and observe that it is the toy model for some results of Dencker-Sjöstrand-Zworski [13]. We refer for more details to the lectures by J. Sjöstrand [38].

The most interesting property is the control of the resolvent for $\Re\,\lambda \geq 0$.

**Proposition 3 (W. Bordeaux-Montrieux [8]).** *As $\Re\,\lambda \to +\infty$, we have*

$$||(\mathscr{A} - \lambda)^{-1}|| \sim \sqrt{\frac{\pi}{2}}(\Re\,\lambda)^{-\frac{1}{4}} \exp \frac{4}{3}(\Re\,\lambda)^{\frac{3}{2}}\,, \tag{40}$$

This improves a previous result by J. Martinet (see in [23]). The proof of the (rather standard) upper bound is based on the direct analysis of the semi-group in the Fourier representation. We note indeed that

$$\mathscr{F}(D_x^2 + i\,x)\mathscr{F}^{-1} = \xi^2 - \frac{d}{d\xi}\,. \tag{41}$$

Then we have

$$\mathscr{F} S_t \mathscr{F}^{-1} v = \exp(-\xi^2 t - \xi t^2 - \frac{t^3}{3}) v(\xi + t)\,, \tag{42}$$

and this implies immediately

$$||S_t|| = \exp \max_{\xi}(-\xi^2 t - \xi t^2 - \frac{t^3}{3}) = \exp(-\frac{t^3}{12})\,. \tag{43}$$

Then one can get an estimate of the resolvent by using, for $\lambda \in \mathbb{C}$, the formula

$$(\mathscr{A} - \lambda)^{-1} = \int_0^{+\infty} \exp -t(\mathscr{A} - \lambda)\, dt\,. \tag{44}$$

For a closed accretive operator, (44) is standard when $\Re\,\lambda < 0$, but estimate (43) on $S_t$ gives immediately an holomorphic extension of the right-hand side to the whole space, showing independently that the spectrum is empty (see Davies [12]) and giving for $\lambda > 0$ the estimate

$$||(\mathscr{A} - \lambda)^{-1}|| \le \int_0^{+\infty} \exp(\lambda t - \frac{t^3}{12})\, dt. \tag{45}$$

The asymptotic behavior as $\lambda \to +\infty$ of this integral is immediately obtained by using the Laplace method.

## *2.6 Pseudo-spectra and semi-groups*

We now analyze the properties of a contraction semi-group $\exp -t\mathscr{A}$, with $\mathscr{A}$ maximally accretive. As before, we have, for $\Re\,\lambda < 0$,

$$||(\mathscr{A} - \lambda)^{-1}|| \le \frac{1}{|\Re\,\lambda|}\,. \tag{46}$$

If we add the assumption that $\mathrm{Im} < \mathscr{A}u, u > \ge 0$ for all $u$ in the domain of $\mathscr{A}$ and if $\mathrm{Im}\,\lambda < 0$ one gets also a similar inequality, so the main remaining question is the analysis of the resolvent in the set $\Re\,\lambda \ge 0\,,\ \mathrm{Im}\,\lambda \ge 0$, which corresponds to the numerical range of the operator.

We recall that for any $\epsilon > 0$, we define the $\epsilon$-pseudospectra by

$$\Sigma_\epsilon(\mathscr{A}) = \{\lambda \in \mathbb{C} \,|\, ||(\mathscr{A} - \lambda)^{-1}|| > \frac{1}{\epsilon}\}\,, \tag{47}$$

with the convention that $||(\mathscr{A} - \lambda)^{-1}|| = +\infty$ if $\lambda \in \sigma(\mathscr{A})$.

We have

$$\cap_{\epsilon>0}\, \Sigma_\epsilon(\mathscr{A}) = \sigma(\mathscr{A})\,. \tag{48}$$

We define, for any $\epsilon > 0$, the $\epsilon$-pseudospectral abscissa by

$$\hat{\alpha}_\epsilon(\mathscr{A}) = \inf_{z\in\Sigma_\epsilon(\mathscr{A})} \Re z\,, \tag{49}$$

and the growth bound of $\mathscr{A}$ by

$$\hat{\omega}_0(\mathscr{A}) = \lim_{t\to+\infty} \frac{1}{t} \log || \exp -t\mathscr{A} ||\,. \tag{50}$$

Of course, we have

$$\lim_{\epsilon\to+\infty} \hat{\alpha}_\epsilon(\mathscr{A}) \leq \inf_{z\in\sigma(\mathscr{A})} \Re z\,, \tag{51}$$

but the equality is wrong in general. The right behavior of the semi-group as $t \to +\infty$ is given by:

**Theorem 6 (Gearhart-Prüss).** *Let $\mathscr{A}$ be a densely defined closed operator in a Hilbert space $X$ such that $-\mathscr{A}$ generates a contraction semi-group, then*

$$\lim_{\epsilon\to 0} \hat{\alpha}_\epsilon(\mathscr{A}) = -\hat{\omega}_0(\mathscr{A})\,. \tag{52}$$

We refer to [15] for a proof and to [25] for a more quantitative version of this theorem which is particularly useful when parameters are involved.

## 2.7 *The complex Airy operator in* $\mathbb{R}^+$

Spectral analysis

Here we mainly describe some results presented in [1], who refers to [30]. We consider the Dirichlet realization $\mathscr{A}^D$ of the complex Airy operator $D_x^2 + ix$ on the half-line. We have

$$\Re\, \langle \mathscr{A}^D u \,|\, u\rangle \geq 0\,,\ \forall u \in D(\mathscr{A}^D)\,, \tag{53}$$

and $\mathscr{A}^D$ is the generator of a semi-group of contraction, whose adjoint is described by replacing in the previous description $(D_x^2 + ix)$ by $(D_x^2 - ix)$. The operator is injective and its spectrum is contained in $\Re\lambda > 0$. Moreover, it has a compact inverse, hence the spectrum (if any) is discrete.

Using what is known on the usual Airy operator, Sibuya's theory and a complex rotation, we obtain [1] that the spectrum of $\mathscr{A}^D$ is given by

$$\sigma(\mathscr{A}^D) = \cup_{j=1}^{+\infty}\{\lambda_j\}, \tag{54}$$

with

$$\lambda_j = -(\exp i\frac{\pi}{3})\,\mu_j, \tag{55}$$

the $\mu_j$'s being real zeroes of the Airy function satisfying

$$0 > \mu_1 > \cdots > \mu_j > \mu_{j+1} > \cdots . \tag{56}$$

It is shown in [1] that the vector space generated by the corresponding eigenfunctions is dense in $L^2(\mathbb{R}^+)$. But there is no way to normalize these eigenfunctions for getting a good basis of $L^2(\mathbb{R}^+)$. We refer to Y. Almog [1], E.B. Davies [11] and to R. Henry [27, 28] who shows that the norm of the spectral projector $\pi_n$ associated with the $n$-th eigenvalue increases exponentially like $\exp \alpha n$ for some $\alpha > 0$.

Decay of the semi-group

We now apply Gearhardt-Pruss theorem to $\mathscr{A}^D$ and our main theorem is

**Theorem 7.**

$$\hat{\omega}_0(\mathscr{A}^D) = -\Re\lambda_1 . \tag{57}$$

This statement was established by Almog [1] in a much weaker form. Using the first eigenfunction it is easy to see that

$$|| \exp -t\mathscr{A}^D|| \geq \exp -\Re\lambda_1\, t . \tag{58}$$

Hence we have immediately

$$0 \geq \hat{\omega}_0(\mathscr{A}^D) \geq -\Re\lambda_1 . \tag{59}$$

To prove the reverse inequality, it is enough to show the following lemma.

**Lemma 1.** *For any $\alpha < \Re\lambda_1$, there exists a constant $C$ such that, for all $\lambda$ s.t. $\Re\lambda \leq \alpha$*

$$||(\mathscr{A}^D - \lambda)^{-1}|| \leq C . \tag{60}$$

*Proof.* By assumption, $\lambda$ is not in the spectrum. Hence the problem is just a control of the resolvent as $|\operatorname{Im}\lambda| \to +\infty$. The case, when $\operatorname{Im}\lambda < 0$ has already been considered, so it remains to control the norm of the resolvent as $\operatorname{Im}\lambda \to +\infty$ and $\Re\lambda \in [-\alpha, +\alpha]$. The main idea is that when $\operatorname{Im}\lambda \to +\infty$, we have to inverse the operator

$$D_x^2 + i(x - \operatorname{Im}\lambda) - \Re\lambda\,.$$

If we consider the Dirichlet realization in the interval $]0, \frac{\operatorname{Im}\lambda}{2}[$ of $D_x^2 + i(x - \operatorname{Im}\lambda) - \Re\lambda$, it is easy to see that the operator is invertible by considering the imaginary part of this operator and that this inverse $R_1(\lambda)$ satisfies

$$||R_1(\lambda)|| \leq \frac{2}{\operatorname{Im}\lambda}.$$

Far from the boundary, we can use the resolvent of the problem on the line for which we have a uniform control of the norm for $\Re\lambda \in [-\alpha, +\alpha]$.

Physical interpretation

Coming back to the application in superconductivity (with $\kappa = 1$), one is looking at the semi-group associated with $\mathscr{A}_J := D_x^2 + iJx - 1$ (where $J \geq 0$ is a parameter). The stability analysis leads to a critical value

$$J_c = (\Re\lambda_1)^{-\frac{3}{2}}\,, \tag{61}$$

such that :

- For $J \in [0, J_c[$, $||\exp -t\mathscr{A}_J|| \to +\infty$ as $t \to +\infty$.
- For $J > J_c$, $||\exp -t\mathscr{A}_J|| \to 0$ as $t \to +\infty$.

This was obtained in [22] improving Lemma 2.4 in Almog [1], who gets only this decay for $||\exp -t\mathscr{A}_J\psi||$, with $\psi$ in a specific dense space.

## 2.8 Higher dimension problems relative to Airy

Here we refer to [1] and [26].

The model in $\mathbb{R}^2$

We consider the operator $\mathscr{A}_2 := -\Delta_{x,y} + i\,x$, and first show:

**Proposition 4.**

$$\sigma(\mathscr{A}_2) = \emptyset\,. \tag{62}$$

*Proof.* After a Fourier transform in the $y$ variable, it is enough to show that $(\widehat{\mathscr{A}_2} - \lambda)$ is invertible with $\widehat{\mathscr{A}_2} = D_x^2 + i\,x + \eta^2$ . We have just to control for a given $\lambda \in \mathbb{C}$, the resolvent $(D_x^2 + i\,x + \eta^2 - \lambda)^{-1}$ (whose existence is given by the one-dimensional result) uniformly in $\mathscr{L}(L^2(\mathbb{R}))$ with respect to $\eta$.

The model in $\mathbb{R}^2_+$ : perpendicular current

Here it is useful to reintroduce the parameter $J$, which is assumed to be positive. Hence we consider the Dirichlet realization $\mathscr{A}_2^{D,\perp} : = -\Delta_{x,y} + i\,Jx$ , in $\mathbb{R}^2_+ = \{x > 0\}$ .

**Proposition 5.**

$$\sigma(\mathscr{A}_2^{D,\perp}) = \cup_{r\geq 0, j\in\mathbb{N}^*}(\lambda_j + r)\,. \tag{63}$$

*Proof.* For the inclusion $\cup_{r\geq 0, j\in\mathbb{N}^*}(\lambda_j + r) \subset \sigma(\mathscr{A}_2^{D,\perp})$, we can use $L^\infty$ eigenfunctions in the form $(x, y) \mapsto u_j(x) \exp iy\eta$, where $u_j$ is the eigenfunction associated with $\lambda_j$. for the reverse inclusion, we observe that we can control uniformly the resolvent $(\mathscr{A}^D - \lambda + \eta^2)^{-1}$ with respect to $\eta$ under the condition that

$$\lambda \notin \cup_{r\geq 0, j\in\mathbb{N}^*}(\lambda_j + r)\,.$$

It is enough to observe the uniform control as $\eta^2 \to +\infty$ which results of (46).

*Remark 3.* The case when the current is not perpendicular has been treated by B. Helffer [22], R. Henry [26, 29]. The spectrum is actually empty.

*Remark 4.* The analysis of the previous models permits actually the semi-classical analysis of the spectrum and of the resolvent for the Dirichlet realization of $-h^2\Delta + iV(x)$ in $L^2(\Omega)$. Here $V$ is a $C^\infty$ potential such that $\nabla V \neq 0$ in $\bar{\Omega}$. Then using the results for the models, one can (see [1, 29]) get a lower bound for

$$\liminf_{h\to 0} h^{-\frac{2}{3}}\,(\inf \Re\, \sigma(\mathscr{A}_h))\ .$$

## 3 Time-Dependent Ginzburg-Landau equation II: general case

The starting point on the mathematical side is a paper of Yaniv Almog [1]. This work was continued in collaboration with Y. Almog and X. Pan [3–5] by the analysis of specific toy models. In [2] (in collaboration with Y. Almog) a rather general situation is considered showing how the toy models are involved in the question.

## 3.1 Boundary conditions

We consider a superconductor placed at a temperature lower than the critical one. It is well understood from numerous experimental observations that a sufficiently strong current, applied through the sample, will force the superconductor to arrive at the normal state. To explain this phenomenon mathematically, we use the time-dependent Ginzburg-Landau model which was already defined in (18) without to make explicit the boundary conditions and in a different scaling. Hence we consider more precisely the following system of equations, referred to as (TDGL1) (Time-Dependent Ginzburg-Landau equation),

$$\frac{\partial \psi}{\partial t} + i\phi\psi = \Delta_{\mathbf{A}}\psi + \psi\left(1 - |\psi|^2\right) , \qquad \text{in } \mathbb{R}_+ \times \Omega , \tag{64a}$$

$$\kappa^2 \text{curl}\,^2\mathbf{A} + \varsigma\left(\frac{\partial \mathbf{A}}{\partial t} + \nabla\phi\right) = \operatorname{Im}(\bar{\psi} \cdot (\nabla + i\mathbf{A})\psi) , \qquad \text{in } \mathbb{R}_+ \times \Omega , \tag{64b}$$

$$\psi = 0 , \qquad \text{on } \mathbb{R}_+ \times \partial\Omega_c , \tag{64c}$$

$$(\nabla + i\mathbf{A})\psi \cdot \nu = 0 , \qquad \text{on } \mathbb{R}_+ \times \partial\Omega_i , \tag{64d}$$

$$\varsigma\left(\frac{\partial \mathbf{A}}{\partial t} + \nabla\phi\right) \cdot \nu = J , \qquad \text{on } \mathbb{R}_+ \times \partial\Omega_c , \tag{64e}$$

$$\varsigma\left(\frac{\partial \mathbf{A}}{\partial t} + \nabla\phi\right) \cdot \nu = 0 , \qquad \text{on } \mathbb{R}_+ \times \partial\Omega_i , \tag{64f}$$

$$\frac{1}{|\partial\Omega|}\int_{\partial\Omega} \text{curl}\,\mathbf{A}(t,x)\,ds = h_{ex} , \qquad \text{on } \mathbb{R}_+ , \tag{64g}$$

$$\psi(0,x) = \psi_0(x) , \qquad \text{in } \Omega , \tag{64h}$$

$$\mathbf{A}(0,x) = \mathbf{A}_0(x) , \qquad \text{in } \Omega . \tag{64i}$$

In the above, $ds$ denotes the induced measure on $\partial\Omega$. The domain $\Omega \subset\subset \mathbb{R}^2$, occupied by the superconducting sample, has a smooth interface, denoted by $\partial\Omega_c$, with a conducting metal which is at the normal state. The rest of the boundary, denoted by $\partial\Omega_i$, is adjacent to an insulator. To simplify some of the arguments (or simply have a proof) we introduce the following geometrical assumption on $\partial\Omega$:

$$(R1) \begin{cases} (a)\ \partial\Omega_i \text{ and } \partial\Omega_c \text{ are of class } C^3\,; \\ (b)\quad \text{Near each edge, } \partial\Omega_i \text{ and } \partial\Omega_c \text{ are flat} \\ \quad \text{and meet with an angle of } \frac{\pi}{2}\,. \end{cases} \tag{65}$$

We also require:

$$(R2) \qquad \text{Both } \partial\Omega_c \text{ and } \partial\Omega_i \text{ have two components.} \tag{66}$$

Figure 1 presents a typical sample with properties (R1) and (R2).

We require that $J$ is a smooth current

$$J = hJ_r \tag{67}$$

satisfying

$$(J1) \quad J_r \in C^2(\overline{\partial\Omega_c}), \tag{68}$$

$$(J2) \quad \int_{\partial\Omega_c} J_r\, ds = 0\,, \tag{69}$$

and

$$(J3) \quad \text{the sign of } J_r \text{ is constant on each connected component of } \partial\Omega_c\,. \tag{70}$$

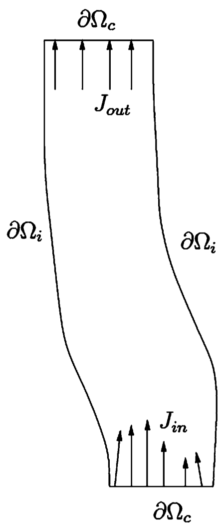


**Fig. 1** Typical superconducting sample. The arrows denote the direction of the current flow ($J_{in}$ for the inlet, and $J_{out}$ for the outlet).

We assume, for the initial conditions (64h,i), that

$$\psi_0 \in H^1(\Omega, \mathbb{C}) \text{ and } \mathbf{A}_0 \in H^1(\Omega, \mathbb{R}^2)\,, \tag{71}$$

and:

$$\|\psi_0\|_\infty \leq 1\,. \tag{72}$$

We consider Coulomb gauge solutions of (64):

$$\operatorname{div} \mathbf{A} = 0 \text{ in } \Omega,\ \mathbf{A}\cdot\nu = 0 \text{ on } \partial\Omega\,. \tag{73}$$

Note, however, that for the proof of existence of solutions it is better to consider first solutions in the Lorentz gauge: $\operatorname{div} \mathbf{A} = \omega\,\phi$, keeping the condition $\mathbf{A}\cdot\nu = 0$ on $\partial\Omega$.

Equivalent boundary conditions: From (TDGL1) to (TDGL2)

Instead of considering the boundary conditions (64e,f,g), it is possible to use an equivalent boundary condition where we prescribe instead the magnetic field (see (80) below. By (64b,e,f), on each point on $\partial\Omega$, except for the corners, we have

$$\frac{\partial}{\partial\tau}\text{curl } \mathbf{A}(t,\cdot) = \frac{1}{\kappa^2}J(\cdot)\,, \tag{74}$$

where $\partial/\partial\tau$ denotes the tangential derivative along $\partial\Omega$ in the positive direction. For convenience we set

$$J_r(x) \equiv 0 \text{ on } \partial\Omega_i\,. \tag{75}$$

Thus, if we introduce on the boundary the function $B$ by

$$\text{curl } \mathbf{A}(t,x) = h\,B_r(t,x) \quad \text{on } \partial\Omega\,, \tag{76}$$

where $h$ denotes a parameter measuring the intensity of the magnetic field.

One can recover the magnetic field $B(t,\cdot)$

$$B_r(t,x) = h_r - \frac{1}{\kappa^2\,|\partial\Omega|}\int_{\partial\Omega} |\Gamma(\tilde{x},x)|\,J_r(\tilde{x})ds(\tilde{x}) \text{ for } x \in \partial\Omega\,. \tag{77}$$

where

$$h_r = h_{ex}/h \tag{78}$$

and $|\Gamma(\tilde{x},x)|$ is the length inside the boundary between $x$ and $\tilde{x}$.

This shows that $B_r(t,x) = B_r(x)$ on the boundary, hence is time independent.

Note also that the condition (74) gives:

$$\textit{The magnetic field } B \textit{ is constant along each component of } \partial\Omega_i\,. \tag{79}$$

Hence the system (TGDL1) is equivalent to the system (TGDL2), consisting in the same equations except (1e-1g) replaced by:

$$\text{curl}\,\mathbf{A}(t,x) = h\,B_r(x)\,,\ \text{on}\ \mathbb{R}_+ \times \partial\Omega\,, \tag{80}$$

where $B_r$ is given by (77).

Of course functional spaces should be introduced to give a precise mathematical sense to this statement of equivalence.

Conversely, a solution of (TGDL2) must satisfy (TGDL1) with

$$J_r = \kappa^2 \frac{\partial B_r}{\partial \tau}\ \text{on}\ \partial\Omega\,,$$

and

$$h_r = \frac{1}{|\partial\Omega|} \int_{\partial\Omega} B_r(x)ds\,.$$

## 3.2 Stationary normal solutions

If we assume a time independent solution of (TDGL1) in the form $(0, \mathbf{A}_n, \phi_n)$, we get for the magnetic and electric normal potentials $\mathbf{A}_n$ and $\phi_n$ the following equations:

$$\begin{array}{ll} -c\,\text{curl}\,^2\mathbf{A}_n + \nabla\phi_n = 0 & \text{in}\ \Omega\,, \\ -\varsigma\,\frac{\partial\phi_n}{\partial\nu} = J_r & \text{on}\ \partial\Omega\,, \\ \frac{1}{|\partial\Omega|}\int_{\partial\Omega} \text{curl}\,\mathbf{A}_n\,ds = h_r\,, & \end{array} \tag{81}$$

in which

$$c = \kappa^2/\varsigma\,. \tag{82}$$

If we fix the Coulomb gauge for $\mathbf{A}_n$, we can prove the existence, uniqueness, and regularity of solutions to the above problem.

Note that $\phi_n$ is a solution of

$$\Delta\phi_n = 0\ \text{in}\ \Omega\,,\ \int_\Omega \phi_n\,d\mathbf{x} = 0\,, \tag{83}$$

and

$$-\,\varsigma\frac{\partial\phi_n}{\partial\nu} = J_r\ \text{on}\ \partial\Omega\,. \tag{84}$$

This is the problem with Neumann boundary condition but in a domain with corners. One can prove the $H^2$-regularity when the angle is $\frac{\pi}{2}$. (See Kondratev [31], Grisvard [20], Dauge [10] for these questions of regularity).

The next assumption (which can be expressed in term of $J$ and $h_{ex}$) is

$$(B) \quad B_n := \operatorname{curl} \mathbf{A}_n \neq 0 \ \text{ at the corners}\,. \tag{85}$$

For some of the results, we assume for technical reasons

$$(C) \quad \nabla\phi_n \perp \partial\Omega \ \text{ on } B_n^{-1}(0) \cap \partial\Omega\,. \tag{86}$$

To recover $\mathbf{A}_n$ we first determine $B_n$ modulo a constant. The constant is fixed by the mean value. We recover $\mathbf{A}_n$ uniquely by choosing the Coulomb gauge.

We can now state:

**Theorem 8.** *Suppose that $\Omega$ satisfies condition (R1) and that $B$ is in $H^{\frac{1}{2}}(\partial\Omega)$ (on each regular component of $\partial\Omega$). Suppose further that $(\psi_0, \mathbf{A}_0)$ satisfies (71) and (72). Then, there exists a unique weak solution $(\psi_c, \mathbf{A}_c, \phi_c)$ of (TGDL2) in the Coulomb gauge. Moreover, this solution is strong and*

$$\|\psi_c(t,\cdot)\|_\infty \leq 1\,,\ \forall t > 0\,. \tag{87}$$

*Finally, let $\mathbf{A}_1 = \mathbf{A}_c - h\mathbf{A}_n$ where $\mathbf{A}_n$ is the previously constructed normal solution. Then*

$$\mathbf{A}_1 \in L^2_{loc}([0,+\infty); H^2(\Omega,\mathbb{R}^2))\,. \tag{88}$$

We can now return to the solution of (TGDL1).

**Theorem 9.** *Under the assumptions of Theorem 8, assuming that $\mathbf{j}$ is given by (68)–(69), and $B_r$ by (77), the solution of (TDGL2) has the additional property that $\phi_c \in C([0,+\infty); W^{1,p}(\Omega))$ for all finite $p$, and is a solution of (TDGL1).*

## 3.3 *The question of stability*

Here we continue to discuss the results of [2]. One possible mechanism which contributes to the breakdown of superconductivity by a strong current is the magnetic field induced by the current. In the absence of electric current, it was proved (see our first section) by Giorgi-Phillips in [19] that, when a sufficiently strong magnetic field is applied on the sample's boundary (or when $h_{ex}$ is sufficiently large), the normal state, for which $\psi \equiv 0$, becomes the unique solution for the steady-state version of (64) (cf. also Fournais-Helffer [18] and the references therein).

For the time-dependent Ginzburg-Landau equations it was proved in Feireisl-Takac [16] that every solution must reach an equilibrium in the long-time limit. When combined with the results in [19] it follows that when the applied magnetic field is sufficiently large the normal state becomes globally stable. This will be discussed in Subsection 3.6.

No such result was available in the presence of electric currents before [2]. The results in [16] are based on the fact that, in the absence of currents, the Ginzburg-Landau energy functional serves as a Lyapunov functional. In the presence of a current one has to take account of the work it produces, which does not necessarily decrease the energy (cf. [35] for instance).

Moreover, the magnetic field is not the only mechanism which forces the sample into the normal state when the electric current is sufficiently large.

Consider the reduced model where one neglects the induced magnetic field and set $\mathbf{A} \equiv 0$ in (18). It has been proved in [1, 30, 36] that the normal state is at least locally stable when the current is sufficiently strong. In a recent contribution [4], it has been shown that the critical current where the normal state loses its local stability tends to the critical value for the reduced model [30] in the small conductivity limit, or when $c \to \infty$. This result suggests that stability is being forced not only by the magnetic field that the current induces, but also by the potential term in (64a).

In [2] we proved global stability of the normal state, as a solution of (64), for sufficiently large currents (see Section 2). We begin by proving global existence and uniqueness of solutions for (64) and obtain their regularity. While these questions have previously addressed (cf. [9, 17], and [14] to name just a few references) the fact that the boundary is not smooth at the corners requires in [2] some additional attention.

### 3.4 A non self-adjoint operator

Let

$$\mathscr{L}_h = -\Delta_{h\mathbf{A}_n} + i\, h\phi_n \,,$$

be defined (Dirichlet-Neumann problem) over the domain

$$D(\mathscr{L}_h) = \{u \in H^2(\Omega) \mid u|_{\partial\Omega_c} = 0 \;;\; \nabla u \cdot \nu|_{\partial\Omega_i} = 0\}\,.$$

We prove in [2] that a proper bound on the resolvent of $\mathscr{L}_h$, which is the elliptic operator in (64a) linearized near $(0, h\mathbf{A}_n, h\phi_n)$ gives the stability.

**Theorem 10.** *Let $\nu \geq 0$. There exists $\kappa_0 > 0$ and $C_1 > 0$ such that, if for some $\kappa > \kappa_0$ we have*

$$\sup_{\gamma \in \mathbb{R}} \|(\mathscr{L}_h - i\gamma - \nu)^{-1}\| < 1 - \frac{C_1}{\kappa^2}\,, \tag{89}$$

*then, any solution of* (64) *must satisfy*

$$\int_0^\infty e^{2\nu t}\, \|\psi(t,\cdot)\|_2^2\, dt < \infty\,. \tag{90}$$

Assumption (89) does not guarantee that the semigroup necessarily becomes a contraction in the long-time limit. The above stability is proved in the large $\kappa$ limit.

As the resolvent of $\mathscr{L}_h$ in an arbitrary domain is difficult to control, we provide an estimate of its norm for large values of $h$, which can be applied for either large domains, or large $\kappa$ values.

## 3.5 Large domains $\Omega_R$

Having in mind the assumptions in Theorem 10, our aim is to show that the norm of the resolvent can be controlled from two approximated problems, with constant current defined either in $\mathbb{R}^2$ or in $\mathbb{R}^2_+$ with Dirichlet boundary conditions.

From resolvent estimates, together with the results of Almog-Helffer-Pan in [3–5], we deduce that the critical current, for which the normal state loses its local stability, can be approximated by the same critical current obtained for the above $\mathbb{R}^2_+$ problem. Before to give a precise statement let us describe the toy models.

### Two toy models

We now give the definitions of these model operators in $\mathbb{R}^2$ and $\mathbb{R}^2_+ = \{y > 0\}$.

These models depend on two real parameters $c \neq 0$ and $\mathrm{j}$.

The first one is

$$\mathscr{A}(\mathrm{j}, c) = D_x^2 + (D_y - \mathrm{j}x^2)^2 + ic\mathrm{j}y\,, \tag{91}$$

defined on

$$D(\mathscr{A}) = \{u \in L^2(\mathbb{R}^2) \mid \mathscr{A}u \in L^2(\mathbb{R}^2)\}\,. \tag{92}$$

It has empty spectrum and we have a good control of the resolvent depending only on the real part of the spectral parameter.

The second one is $\mathscr{A}_+(\mathrm{j}, c)$, which is defined (via the Lax-Milgram theorem) by the same differential formula of $\mathscr{A}$ but on the domain

$$D(\mathscr{A}_+) = \{u \in \tilde{V} \,:\, \mathscr{A}_+ u \in L^2(\mathbb{R}^2_+, \mathbb{C})\}, \tag{93}$$

where

$$\tilde{V} = H_0^{1,\mathrm{mag}}(\mathbb{R}^2_+, \mathbb{C}) \cap L^2(\mathbb{R}^2_+, \mathbb{C}; y\, dxdy)\,. \tag{94}$$

Here the analysis of the spectrum is more difficult. The guess is that it is non-empty. This is only proven for $|c|$ large enough or small enough [4, 5].

Towards the next theorem

We set, for $z \in \bar{\Omega}$,

$$\jmath(z) := h|\nabla B_n(z)| = \frac{h}{c}|\nabla \phi_n(z)|\,, \tag{95}$$

and then define

$$\mathscr{A}(z) = \mathscr{A}(\jmath(z), c) \quad ; \quad \mathscr{A}_+(z) = \mathscr{A}_+(\jmath(z), c) \tag{96}$$

Under all of the above assumptions $B_n^{-1}(0)$ is either empty, or else consists of a single curve $\Gamma$ connecting between the two connected components of $\partial\Omega_c$.

We treat the second case. We denote the two points of intersection by $z_1$ and $z_2$ and then set

$$\nu_m(z_1, z_2, c) = \min_{i=1,2} \inf_{\lambda \in \sigma(\mathscr{A}_+(z_i))} \Re\,\lambda\,. \tag{97}$$

Large domain limit

Let then $R > 0$. We denote by $\Omega_R$ the image of $\Omega$ under the dilation $x \to R\,x$ We assume that the domain $\Omega$ has the property (R1)-(R2) and that assumptions (J1)–(J3), (B) and (C) are met.

Denote the transformed electric field by $\phi_R$. It satisfies the problem

$$\begin{cases} \Delta\phi_R = 0 & \text{in } \Omega_R\,, \\ \frac{\partial\phi_R}{\partial\nu} = -\frac{J_R(x)}{\varsigma} & \text{on } \partial\Omega_R\,, \end{cases}$$

where

$$J_R(x) = J_r(x/R)\,.$$

Note that

$$\phi_R(x) = R\,\phi_n(x/R)\,.$$

The transformed magnetic potential, which we denote by $A_R$ then satisfies

$$A_R(x) = R^2\, \mathbf{A}_n(x/R)\,.$$

Let then

$$\mathscr{L}_h^R = -\nabla_{hA_R}^2 + i\,h\,\phi_R\,, \tag{98}$$

and let

$$\mu(R) = \inf_{\lambda\in\sigma(\mathscr{L}_h^R)} \Re\,\lambda \quad\text{and}\quad \mu_\infty = \liminf_{R\to\infty} \mu(R)\,. \tag{99}$$

The following theorem is proved in [2]:

**Theorem 11.** *Under the previous assumptions, $\mu(R)$ has a limit as $R \to +\infty$, which is given by*

$$\mu_\infty = \nu_m\,.$$

*Furthermore, let us assume that $\nu < \mu_\infty$. Then there exist $R_0$, $C$, such that, for $R \geq R_0$,*

$$\sup_{\gamma\in\mathbb{R}} \|(\mathscr{L}_h^R - \nu - i\gamma)^{-1}\| \leq \max\Big(\sup_{z_0\in\Gamma} \|(\mathscr{A}(z_0) - \nu)^{-1}\|,\ \sup_{\substack{\gamma\in\mathbb{R}\\ i=1,2}} \|(\mathscr{A}_+(z_i) - \nu - i\gamma)^{-1}\|\Big)\Big(1 + \frac{C}{R^{1/4}}\Big) + \frac{C}{R^{1/4}}. \tag{100}$$

One can deduce from (100) an upper bound for the critical current where the normal state $(0, h\mathbf{A}_n, h\phi_n)$ becomes globally stable. Let

$$\mathrm{j}_m = \inf_{z\in\Gamma} \mathrm{j}(z)\,, \tag{101a}$$

and

$$\mathrm{j}_+ = \inf_{i=1,2} \mathrm{j}(z_i)\,. \tag{101b}$$

When the domain size is multiplied by $R$, the resolvent norm of $\mathscr{L}_h$ is given by the left-hand side of (100). By (89) it then follows that if $R$ and $\kappa$ are sufficiently large, and if

$$\mathrm{j}_m > \|\mathscr{A}^{-1}(1,c)\|^{3/2} \tag{102a}$$

and

$$\mathrm{j}_+ > \sup_{\gamma\in\mathbb{R}} \|(\mathscr{A}_+(1,c) - i\gamma)^{-1}\|^{3/2}\,, \tag{102b}$$

then the normal state must be globally stable. The above conditions serve as an upper bound for the critical current where the normal state becomes globally stable.

On the semiclassical side

The above regime corresponds to the spectral analysis of

$$\sum_j (\hbar D_{x_j} + A_j)^2 + i\hbar\phi(x),$$

in the limit $\hbar \to 0$. With $\phi = 0$, this analysis plays an important role in the analysis of the superconductivity. In the above questions, we have $\nabla\phi \cdot \nabla \mathrm{curl}\, \mathbf{A} = 0$ (see the first line in (81)) and the zero set of curl $\mathbf{A}$ consists in a curve $\Gamma$ joining two points of the boundary where the Dirichlet condition is assumed.

When $\mathbf{A} = \mathbf{0}$, a connected problem is to determine the bottom of the (real part of the) spectrum under the assumption that $\phi$ is a Morse function and has no critical point at the boundary. The answer depends on the presence or not of critical sets inside $\Omega$. When there are no critical points, the case is treated in [1] (see also [29]). One should look at all the points where $\nabla\phi$ is orthogonal to the boundary. Assuming that these points are isolated, we will get the result by looking at the transversal Airy operators computed at these points. That is looking at

$$\hbar^2 D_t^2 + i\hbar|\nabla\phi(x_\ell)|t + i\hbar\phi(x_\ell)$$

in $\mathbb{R}^+$, with Dirichlet condition at 0. With $j(x_\ell) = |\nabla\phi(x_\ell)|$, the smallest real part is $j(x_\ell)^{\frac{2}{3}}\hbar^{\frac{4}{3}} \cos\frac{\pi}{3}\,\alpha$, where $\alpha$ is the lowest eigenvalue of the standard Airy operator on $\mathbb{R}^+$. Actually, depending on the angle of $\nabla\phi$ with the normal, we get a model in $\mathbb{R}^2_+$:

$$\hbar^2(D_t^2 + D_s^2) + iJ(\cos\theta t + \sin\theta s)\,,$$

with boundary condition at $t = 0$.

As we have seen in the study of models, the only case when spectrum is present is the case when $\theta = 0$.

In the case where there are critical points in $\Omega$, we consider the complex harmonic oscillator in $\mathbb{R}^2$ obtained by considering the quadratic approximation of $\phi$ at the various critical points $\mathbf{x}_\ell$:

$$\hbar^2(D_x^2 + D_y^2) + i\hbar\langle \mathrm{Hess}\,\phi(\mathbf{x}_\ell)(x,y)\,,\,(x,y)\rangle + i\hbar\phi(\mathbf{x}_\ell).$$

In this case the bottom is of order $\mathscr{O}(h^{\frac{3}{2}})$ and this explains why these points will have the dominant role.

A similar question coming from control theory

In the $(1D)$-case, this question appears also in control theory (Beauchard, Helffer, Henry, and Robbiano [7]) for two models

$$\mathscr{A}_{(-R,R)} = -\frac{d^2}{dy^2} + iy \quad and \quad \mathscr{H}_{[-R,R]} = -\frac{d^2}{dy^2} + iy^2$$

defined on the segment $(-R, R)$, $R > 0$, with Dirichlet boundary conditions at the ends $y = \pm R$, with domains

$$\mathscr{D}(\mathscr{A}_{(-R,R)}) = \mathscr{D}(\mathscr{H}_{(-R,R)}) = H_0^1((-R,R);\mathbb{C}) \cap H^2((-R,R);\mathbb{C}).$$

More precisely, we study the asymptotic behavior, as $R \to +\infty$, of the bottom of the spectrum of $\mathscr{A}_{(-R,R)}$ and $\mathscr{H}_{(-R,R)}$ and we use the following two theorems.

**Theorem 12.** *Let $\mu_1 < 0$ be the first zero of the Airy function. Then,*

$$\lim_{R\to\infty} \left(\inf \Re\, \sigma(\mathscr{A}_{(-R,R)})\right) = \frac{|\mu_1|}{2}\,, \tag{103}$$

*where $\sigma(\mathscr{A}_{(-R,R)})$ denotes the spectrum of $\mathscr{A}_{(-R,R)}$.*

Now, let us consider the case of Davies operator (or 'complex hamonic oscillator')

**Theorem 13.**

$$\lim_{R\to\infty} \left(\inf \Re\, \sigma(\mathscr{H}_{(-R,R)})\right) = \frac{\sqrt{2}}{2}, \tag{104}$$

*where $\sigma(\mathscr{H}_{(-R,R)})$ denotes the spectrum of $\mathscr{H}_{(-R,R)}$.*

Analogous questions have been considered in [1, 3–5] and [2]. These two operators are analyzed thanks to technics developed in these references. The study of more general cases (dimension 2) complementary to those studied in [1] and [2] is done by R. Henry in [29].

## 3.6 The Giorgi-Phillips type theorem for stationary solutions

We finally come back to the discussion started in the second paragraph of Subsection 3.3. The aim is to establish the equivalent of Giorgi-Phillips theorem for the stationary solutions of time-dependent equations. Of course, one can consider different regimes according to the parameters. We only present one possible choice.

We assume that $\kappa^2/\varsigma = c > 0$, where $c$ is fixed. In [2], it appears useful in order to get a good scaling to take

$$J_r = \kappa^2 \tilde{J}_r\,. \tag{105}$$

With this scaling, we assume that $\tilde{J}_r$ and $h_r$ are independent of $h$ and $\kappa$, so $B_r$ is also independent of $h$ and $\kappa$ (see (80) and (77)). Looking at (83) and (84), we get a $\phi_n$ which is independent of $h$ and $\kappa$, and through (81) the same property for $\mathbf{A}_n$.

We now assume that we have a **nonnormal** stationary point $(\psi, \mathbf{A}, \phi)$ of (64), with condition (80) and that

$$\int_\Omega |\psi(\mathbf{x})|^2\, d\mathbf{x} > 0\,. \tag{106}$$

Then we get:

$$\begin{aligned}
&i\phi\psi = \Delta_{\mathbf{A}}\psi + \psi(1-|\psi|^2)\,, && \text{in } \Omega\,, && (107a)\\
&\kappa^2 \text{curl}\,^2\mathbf{A} + \varsigma\nabla\phi = \text{Im}\,(\bar{\psi}\,\cdot\,(\nabla + i\mathbf{A})\psi)\,, && \text{in } \Omega\,, && (107b)\\
&\psi = 0\,, && \text{on } \partial\Omega_c\,, && (107c)\\
&(\nabla + i\mathbf{A})\psi\cdot\nu = 0\,, && \text{on } \partial\Omega_i\,, && (107d)\\
&\text{curl}\,\mathbf{A} = hB_r\,, && \text{on } \partial\Omega_i\,. && (107e)
\end{aligned}$$

Taking the scalar product with $\psi$ in the first line, we get (using also the boundary condition)

$$i\int_\Omega \phi(\mathbf{x})|\psi(\mathbf{x})|^2\, d\mathbf{x} + ||(\nabla + i\mathbf{A})\psi||^2 + \int_\Omega |\psi(\mathbf{x})|^4\, d\mathbf{x} = ||\psi||^2\,. \tag{108}$$

Now for the second equation, we take the scalar product with $\mathbf{A} - h\mathbf{A}_n$, where $(0, \mathbf{A}_n, \phi_n)$ is the normal stationary solution, and observing that $\text{div}\, A = 0$, we obtain:

$$\kappa^2 ||\text{curl}\,(\mathbf{A} - h\mathbf{A}_n)||^2 = \int_\Omega \big((\mathbf{A} - h\mathbf{A}_n)\cdot\,\text{Im}\,(\bar{\psi}\,\cdot\,(\nabla + i\mathbf{A})\psi)\big)\, d\mathbf{x}\,. \tag{109}$$

Now (108) implies

$$||(\nabla + i\mathbf{A})\psi||^2 \le ||\psi||^2\,. \tag{110}$$

Playing with (109) leads first to

$$\kappa^2 ||\mathbf{A} - h\mathbf{A}_n||^2 \le C_\Omega \kappa^2 ||\text{curl}\,(\mathbf{A} - h\mathbf{A}_n)||^2 \le \hat{C}_\Omega ||(\mathbf{A} - h\mathbf{A}_n)||\, ||(\nabla + i\mathbf{A})\psi||\,.$$

Hence

$$\kappa^2||\mathbf{A} - h\mathbf{A}_n|| \le \hat{C}_\Omega\, ||(\nabla + i\mathbf{A})\psi|| \ \le ||\psi||$$

and we get

$$\kappa^4||\mathbf{A} - h\mathbf{A}_n||^2 + ||(\nabla + i\mathbf{A})\psi||^2 \le \tilde{C}_\Omega||\psi||^2 .$$

Comparing $\int_\Omega |(\nabla + ih\mathbf{A}_n)\psi|^2\, d\mathbf{x}$ and $\int_\Omega |(\nabla + i\mathbf{A})\psi|^2\, d\mathbf{x}$ leads to:

$$\int_\Omega |(\nabla + ih\mathbf{A}_n)\psi|^2\, d\mathbf{x} \le 2\, \|(\nabla + i\mathbf{A})\psi\|^2 + 2(\|(\mathbf{A} - h\mathbf{A}_n)|\psi|\, \|^2 ,$$

and

$$\int_\Omega |(\nabla + ih\mathbf{A}_n)\psi|^2\, d\mathbf{x} \le 2C_\Omega(1 + \kappa^{-4}) \int_\Omega |\psi(\mathbf{x})|^2\, d\mathbf{x} .$$

Since $\psi$ satisfies (106) and the Dirichlet condition on $\partial\Omega_c$, we obtain

$$\lambda_1^{DN}(h\mathbf{A}_n) \le 2C_\Omega(1 + \kappa^{-4}) , \tag{111}$$

where $\lambda_1^{DN}$ corresponds to the Dirichlet-Neumann realization of the magnetic Laplacian (Dirichlet on $\partial\Omega_c$ and Neumann on $\partial\Omega_i$). We now need an asymptotic behavior of $\lambda_1^{DN}(h\mathbf{A}_n)$ in order to get either a contradiction (if no $h$ satisfies the inequality) or an upper bound for $h$. Actually, a lower bound of $\lambda_1^{DN}$ will suffice.

Here, we observe that $\lambda_1^{DN} \ge \lambda_1^N$.

Observing that $B_n$ is continuous on $\overline{\Omega}$ and harmonic in $\Omega$, the maximum principle shows that the minimum $B_{min}$ of $B_n$ in $\Omega$ is attained on one component of $\partial\Omega_i$ and that the maximum $B_{max}$ is attained at the other component. Assume further that

$$\text{(B) } B_n^{-1}(0) = \emptyset \text{ or } B_{min} < 0 < B_{max} . \tag{112}$$

Under this assumption, one can show (see [2]) that either $B_n^{-1}(0)$ is empty or satisfy (15), Theorems 2 or 3 are consequently relevant for estimating $\lambda_1^N(h\mathbf{A}_n)$.

**Theorem 14.** *Under the assumptions of Subsection 3.1 and of this subsection, there exists, for any* $c > 0$*,* $h_r$ *and* $\widetilde{J}_r$*,* $h_0$ *such that* $h \ge h_0$*,* $\kappa \ge 1$*, and* $\kappa^2/\varsigma = c$*, any stationary solution of (TDGL1) is normal.*

Note that $\varsigma = 0$ is excluded from this last theorem. Hence the comparison with the first statement of Giorgi-Phillips is not possible.

**Acknowledgements** I would like to thank all my collaborators on this subject: Y. Almog, K. Beauchard, S. Fournais, R. Henry, X.B. Pan, L. Robbiano and J. Sjöstrand. I am also grateful to F. Hérau and A. Jeribi for their kind invitation, respectively, in Berder (July 2013) and Hammamet (December 2013) who give me the motivation for writing this survey and R. Henry for his remarks on the manuscript.

## References

1. Y. Almog, The stability of the normal state of superconductors in the presence of electric currents, SIAM Journal on Mathematical Analysis, 40 (2008), 824–850.
2. Y. Almog and B. Helffer, Global stability of the normal state of superconductors in the presence of a strong electric current, To appear in Commun. Math. Phys., **330** (2014), 1021–1094.
3. Y. Almog, B. Helffer and X. B. Pan, Superconductivity near the normal state under the action of electric currents and induced magnetic fields in $\mathbb{R}^2$, Commun. Math. Phys. 300 (2010), 147–184.
4. Y. Almog, B. Helffer and X. B. Pan, Superconductivity near the normal state in a half-plane under the action of a perpendicular electric current and an induced magnetic field, Trans. AMS 365 (2013), 1183–1217.
5. Y. Almog, B. Helffer and X. B. Pan, Superconductivity near the normal state in a half-plane under the action of a perpendicular current and an induced magnetic field II: the large conductivity limit, Siam J. Math. Anal. 44, No. 6 (2012), 3671–3733.
6. P. Bauman, H. Jadallah and D. Phillips, Classical solutions to the time-dependent Ginzburg-Landau equations for a bounded superconducting body in a vacuum, J. Math. Phys., **46** (2005), 095104.
7. K. Beauchard, B. Helffer, R. Henry, and L. Robbiano, Degenerate parabolic operators of Kolmogorov type with a geometric control condition, Submitted (2013) accepted in ESAIM:COCV, 2014.
8. W. Bordeaux-Montrieux, Estimation de résolvante et construction de quasimode près du bord du pseudospectre, Math arXiv:1301.3102 (Jan. 2013).
9. Z. M. Chen, K.-H. Hoffmann, and J. Liang, On a nonstationary Ginzburg-Landau superconductivity model, Math. Methods Appl. Sci., 16 (1993), 855–875.
10. M. Dauge, *Elliptic boundary Value Problems in Corner domains,* Lecture Notes in Mathematics 1341. Springer Verlag (1988).
11. E. B. Davies, Semi-classical states for the non-self-adjoint Schrödinger operators, Commun. Math. Phys. 200, 35–41 (1999).
12. E. B. Davies, *Linear operators and their spectra*, vol. 106 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2007.
13. N. Dencker, J. Sjöstrand, and M. Zworski, Pseudospectra of semiclassical pseudo-differential operators, Comm. Pure and Applied Mathematics, Vol. LVII (2004) 384–0415.
14. Q. Du, Global existence and uniqueness of solutions of the time-dependent Ginzburg-Landau model for superconductivity, Appl. Anal., 53 (1994), 1–17.
15. K.J. Engel and R. Nagel, *One-parameter semigroups for linear evolution equations.* Graduate texts in Mathematics 194, Springer.
16. E. Feireisl and P. Takáč, Long-time stabilization of solutions to the Ginzburg-Landau equations of superconductivity, Monatsh. Math. 133 (2001),197–221.
17. J. Fleckinger-Pellé, H. G. Kaper, and P. Takáč, Dynamics of the Ginzburg-Landau equations of superconductivity, Nonlinear Anal., 32 (1998), 647–665.
18. S. Fournais and B. Helffer, *Spectral Methods in Surface Superconductivity*, Birkhäuser, (2009).
19. T. Giorgi and D. Phillips, The breakdown of superconductivity due to strong fields for the Ginzburg-Landau model, SIAM J. Math. Anal., 30 (1999), 341–359.
20. P. Grisvard, *Singularities in boundary value problems,* Springer (1992).
21. B. Helffer, *Introduction to semi-classical analysis for the Schrödinger operator and applications,* Springer Lecture Notes 1336 (1988).
22. B. Helffer, On pseudo-spectral problems related to a time dependent model in superconductivity with electric current, Confluentes Math. 3, No. 2, 237–251 (2011).
23. B. Helffer, *Spectral theory and its applications,* Cambridge University Press 2013.
24. B. Helffer and Y. Kordyukov, Semiclassical spectral asymptotics for a magnetic Schrödinger operator with non-vanishing magnetic field, *Proceedings of Bialowiesza (XXXIII Workshop on Geometric Methods in Physics 2013, Trends in mathematics*, 259–278 (Birkhäuser).

25. B. Helffer and J. Sjöstrand, From resolvent bounds to semigroup bounds, Preprint : arXiv:1001.4171v1, (2010).
26. R. Henry, Master's memoir (2010).
27. R. Henry, Spectral instability of some non-selfadjoint anharmonic oscillators, C. R. Acad. Sci. Paris, Ser. I 350 (2012) 1043–1046.
28. R. Henry, Spectral instability for the complex Airy operator and even non-selfadjoint anharmonic oscillators, J. Spectr. Theory 4 (2014), 349–364.
29. R. Henry, On the semiclassical analysis of Schrödinger operators with purely imaginary electric potentials in a bounded domain, Submitted (2013).
30. B. I. Ivlev and N. B. Kopnin, Electric currents and resistive states in thin superconductors, Advances in Physics, 33 (1984), 47–114.
31. V.A. Kondratiev, Boundary Value Problems for elliptic equations in domain with conical or angular points, Trans. Moscow Math Soc (1967), 227–313.
32. K. Lu and X.-B. Pan, Estimates of the upper critical field for the Ginzburg-Landau equations of superconductivity, Phys. D, 127 (1999), 73–104.
33. X.-B. Pan and K.-H. Kwek, Schrödinger operators with non-degenerately vanishing magnetic fields in bounded domains, Trans. Amer. Math. Soc., 354 (2002), 4201–4227.
34. N. Raymond, *On the semiclassical magnetic Laplacian and connected topics,* Lecture notes for a course in Tunisia (August 2012). hal-00725923, version 2 - 24 Dec 2012.
35. J. Rubinstein, P. Sternberg, and J. Kim, On the behavior of a superconducting wire subjected to a constant voltage difference, SIAM Journal on Applied Mathematics, 70 (2010), 1739–1760.
36. J. Rubinstein, P. Sternberg, and K. Zumbrun, The Resistive State in a Superconducting Wire: Bifurcation from the Normal State, Archive for Rational Mechanics and Analysis, 195 (2010), 117–158.
37. E. Sandier and S. Serfaty, *Vortices in the magnetic Ginzburg-Landau model,* Progress in nonlinear differential Equations and their applications. Birkhäuser (2007).
38. J. Sjöstrand, Resolvent estimates for non-self-adjoint operators via semi-groups, http://arxiv.org/abs/0906.0094.

# Fixed Point Theory for 1-Set Contractions: a Survey

**Smaïl Djebali**

*(To the memory of Boris Nikolaevich Sadovskii)*

**Abstract** In this report, we first review some classical results concerning the fixed point theory for an important class of mappings for which the Banach contraction principle fails, namely nonexpansive mappings. Both metric and topological fixed point theory will be surveyed. We will also discuss some known results regarding the extension to nonlinear contractions and to $\alpha$-contractive mappings with respect to some measure of noncompactness $\alpha$. The second part of this survey paper will be devoted to some recent progress and development of the fixed point theory of 1-set contractions that have been achieved during the last couple of years. The theory for different boundary conditions and when the corresponding space is endowed with the weak topology are also discussed. Finally, some applications to equations of Krasnosels'kĭi type and to the solvability of nonlinear integral equations of Volterra type are presented.

**Keywords** fixed point • normal structure • expansive • nonexpansive • uniformly convex • 1-set contraction • $\psi$-expansive • $\alpha$-$\psi$-expansive • retraction • weakly compact • weakly continuous • MNC • boundary condition • weak topology • sum of operators integral equation

**AMS (MOS) Subject Classifications:** 47H10, 54C15, 54C20, 54C55, 55M15

## 1 Introduction

The fixed point theory is of fundamental importance in almost all branches of mathematics for many applied problems stem from mechanics, chemistry and ecological problems may be formulated as nonlinear equations of the form $u = Tu$,

S. Djebali (✉)
Laboratoire "Théorie du point Fixe et Applications", E.N.S., BP 92, Kouba 16050, Algiers, Algeria
e-mail: djebali@ens-kouba.dz



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_3

where $T$ stands for some nonlinear operator modeling the given equation or system and obeys some physical laws. For these reasons, fixed point theory has attracted many researchers for centuries. Presently, the theory has developed in several directions and has become a wide domain and a very rich area of nonlinear analysis, in theory as well as in applications. But still many questions remain open to the specialists. The starting point is the case where $T$ is a $k$-contraction mapping and is related to the classical Banach fixed point (Theorem 1 below). Unfortunately, this theorem fails in the limit case $k = 1$, i.e., for nonexpansive mappings. Indeed a *nonexpansive* mapping on a Banach space need not have a fixed point as shows the translation mapping $T(x) = x + x_0$ for some $x_0 \in X \setminus \{0\}$. Moreover, the identity operator shows that in general uniqueness does not hold for *nonexpansive* mappings.

The aim of this article is to present first a brief account on the fixed point theory for such mappings; we will focus on the main properties of the domains of mappings as well as on the functional spaces under consideration. As a generalization of nonexpansive mappings, we will be concerned in the second part of this work with an important class of mappings, namely that of the 1-set contractions with respect to some measure of noncompactness. The essential elements which have played a key role in the development of the fixed point theory for these mappings during the last four decades are surveyed. For further studies, the interested author may find most important results in the rich literature given in the bibliography.

The plan of the paper is as follows. After an introductory section, we recall in Sect. 2 the fixed point theory for nonexpansive mappings (generalities, approximation, structure of domain, geometry of space, nonlinear alternative, classical and recent results, boundary conditions). The more general case of 1-set contractions (MNC, first results, historical review, recent developments) is discussed in Sect. 3. Sect. 4 is devoted to presenting the theory when the Banach space is endowed with a weak topology and then some fixed point theorems are derived. We close this survey paper with some applications to the sum of operators in Sect. 5 and to the solvability of two nonlinear integral equations in Sect. 6. The paper ends with a concluding remark given in a short section.

A basic and a very important tool in the fixed point theory is the Banach contraction principle ((1922) see e.g., [66]):

**Theorem 1.** *Let* $(X, d)$ *be a complete metric space and let* $T : X \to X$ *be a contraction mapping, i.e., there exists* $0 < k < 1$ *such that*

$$d(T(x), T(y)) \leq kd(x, y), \quad \forall x, y \in X.$$

*Then there exists exactly one point* $\tilde{x} \in X$ *such that* $T(\tilde{x}) = \tilde{x}$.

The fixed point $\tilde{x}$ is obtained as the limit of the iterative sequence defined recurrently by $x_{n+1} = T(x_n)$ and $x_0 \in X$. The speed of convergence of this sequence to its limit can also be estimated. Indeed, it is easy to check that for every $x \in X$

$$d(T^n x, \tilde{x}) \leq \frac{k^n}{1-k} d(x, Tx).$$

This result has the following extension.

**Theorem 2.** *T has a unique fixed point whenever there exists* $m \in \{1, 2, \ldots\}$ *such that* $T^m$ *is a k-contraction, where* $T^m$ *is defined recurrently by* $T^0(x) = T(x)$ *and* $T^{m+1}(x) = T(T^m)(x)$.

*Proof (Sketch of the proof).* For $x \in X$, define $\varphi(x) = d(x, T^m(x))$. Then $(\varphi(T^n(x)))_n$ is a Cauchy sequence, hence converges to some limit $x_0$. In addition, for all $x \in X$, and all $n > m$, we have

$$\varphi(T^n(x)) \leq k^l \varphi(T^p(x)),$$

where $l$ is the integer part of $n/m$ and $p = n - lm$. Hence $x_0 = 0$ and so $(T^n(x))_n$ is a Cauchy sequence, hence converges to some limit $y$. $d(y, T^m(y)) = 0$ implies $T(y) = T^m(T(y))$. Since $T^m$ has only one fixed point, we conclude that $y = T(y)$. □

*Example 1.* Let $X = C([0, b], \Re)$ and $T : X \longrightarrow X$ the mapping defined by $T(x)(t) = \int_0^t x(s)ds$. Then $T$ is not a contraction if $b > 1$; however

$$T^{(n)}(x)(t) = \frac{1}{(n-1)!} \int_0^t (t-s)^{n-1} x(s)ds$$

is a contraction for $n$ large enough.

## 2 Nonexpansive mappings

### *2.1 Generalities*

We first consider a particular case

**Theorem 3 (Compactness).** *Let* $(X, d)$ *be a compact metric space and* $T : X \to X$ *be a mapping satisfying:*

$$d(T(x), T(y)) < d(x, y), \quad \forall x, y \in X.$$

*(T is said strictly nonexpansive, contracting, or shrinking). Then T has exactly one fixed point. Also, for each* $x \in X$*, the sequence* $\{T^{(n)}(x)\}_{n \in \mathbb{N}}$ *converges to this unique fixed point.*

*Proof.* For $x \in X$, define the real continuous function $\varphi(x) = d(x, T(x))$. Since $X$ is compact, there exists $\bar{x} \in X$ such that $\varphi(\bar{x}) = \inf_{x \in X} \varphi(x)$. Then $T(\bar{x}) = \bar{x}$. On the contrary

$$\varphi(T(\bar{x})) = d(T(\bar{x}), T^2(\bar{x})) < d(\bar{x}, T(\bar{x})) = \varphi(\bar{x}),$$

contradicting the definition of the infimum (the greatest lower bound). Regarding the second part of the theorem, let $u_n(x) = \varphi(T^{(n)}(x)) = d(T^{(n)}(\bar{x}), T^{(n+1)}(x))$. Then the sequence $(u_n(x))_{n\in\mathbb{N}}$ is decreasing hence converges to some limit $u(x)$; the continuity of $T^{(k)}$ $(k \in \mathbb{N})$ implies $u(x) = 0$. By compactness, we may assume that there exists $y = \lim\limits_{n\to\infty} T^{(n)}(x)$ and again by continuity

$$0 = u(x) = \lim_{n\to\infty} u_n(x) = \varphi\left(\lim_{n\to\infty} T^{(n)}(x)\right) = \varphi(y) = d(y, T(y)).$$

Hence $T(y) = y$, i.e., $y = x$, as claimed. □

*Remark 1.* (a) The existence result still holds if $T : C \longrightarrow C$ is strictly nonexpansive and $C$ is a nonempty convex weakly compact subset of a Banach space (see [57, Theorem 1.3.19] and Remark 6).

(b) The compactness of the space $X$ is essential as shows the counterexample where $X = [1, +\infty)$ and $T(x) = x + \frac{1}{x}$. Indeed, $|T(x) - T(y)| < |x - y|$, for all $x, y \in X$ but $T$ is fixed point free.

Several examples of fixed point free mappings are provided by B. Sims in [72, Chap. 2].

**Theorem 4 (Approximation).** *Let $(X, \|.\|)$ be a Banach space, $C \subset X$ a nonempty closed convex subset containing the origin, and $T$: $C \longrightarrow C$ a bounded nonexpansive mapping. Then for any small $\delta > 0$, $T$ has a $\delta$-fixed point in $C$, that is $x_\delta \in C$ such that $\|x_\delta - T(x_\delta)\| < \delta$.*

*Proof.* Since $T$ is bounded, there exists $R > 0$ such that $T(C) \subset \overline{B}(0, R)$. For $0 < \delta < R$, the mapping $\left(1 - \frac{\delta}{R}\right) T$ is a contraction hence admits a unique fixed point $x_\delta \in C$. Then

$$0 \le \|T(x_\delta) - x_\delta\| = \left\| T(x_\delta) - \left(1 - \frac{\delta}{R}\right) T(x_\delta) \right\| = \frac{\delta}{R} \|T(x_\delta)\| \le \delta.$$

□

*Remark 2.* Since $T$ is nonexpansive, if $C$ is bounded then $T(C)$ is so and thus Theorem 4 applies. More generally, it suffices that $T$ verifies the property $(\mathscr{K})$, i.e., there exists a nonempty bounded closed convex subset $K \subset E$ such that $T(K \cap C) \subset K$.

*Remark 3.* Theorem 4 (and its proof) show that

(a) every nonexpansive mapping on a bounded closed convex subset can be approximated by a sequence of contractive mappings,
(b) $\inf\limits_{x\in C}\{\|x - Tx\|\} = 0$.

As a consequence, we have

**Corollary 1.** *Let $C$ be a nonempty bounded closed convex subset in a Banach space $X$ and $T : C \longrightarrow C$ be a nonexpansive mapping such that $(I - T)(C)$ is closed in $X$. Then $T$ has a fixed point.*

*Proof.* By Theorem 4, there exists a sequence $(x_n)_{n\in\mathbb{N}} \in C$ such that $\lim_{n\to\infty} \|T(x_n) - x_n\| = 0$. Since $(I - T)(C)$ is closed, we deduce that $0 \in (I - T)(C)$; hence, $f$ has a fixed point in $C$. □

In fact, instead of nonexpansiveness of $T$, it is sufficient to assume the existence of approximate fixed points, in which case it is not necessary for $T$ to self-map $C$. We have

**Corollary 2 (Approximation+compactness).** *Let $C \subset E$ be a closed subset of a Banach space and $T$: $C \longrightarrow E$ a continuous mapping. Assume that*

*(a) $T(C)$ is compact.*
*(b) $T$ has a $\delta$-fixed point in $C$ for each $\delta > 0$.*
*Then $T$ has a fixed point in $C$.*

*Proof.* From (b), there exists a sequence $(x_n)_{n\in\mathbb{N}} \in C$ such that $\lim_{n\to\infty} \|T(x_n) - x_n\| = 0$ (one may take $\delta = 1/n$ for $n \in \{1, 2, \ldots\}$). Since $T(C)$ is compact, there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty} T(x_{n_k}) = y \in T(C)$. Therefore $y = \lim_{k\to\infty} x_{n_k}$, $y \in C$, and $T(y) = y$. □

*Remark 4.* The compactness of $C$ cannot be relaxed to the weak-compactness as shows the following counterexample provided by Alspach in 1981 [7]. Let $X = L^1[0, 1]$,

$$C = \left\{ h \in X | \int_0^1 h(t)dt = 1 \text{ and } 0 \le h(t) \le 2, \text{ for a.e. } t \in (0, 1) \right\},$$

and let $T : C \longrightarrow C$ be defined by

$$T(h)(t) = \begin{cases} \min\{2, 2h(2t)\}, & \text{if } 0 \le t \le 1/2 \\ \max\{0, 2h(2t - 1) - 2\}, & \text{if } 1/2 < t \le 1. \end{cases}$$

Then $C$ is a nonempty convex *weakly compact* subset and $T$ is an *isometry* which is *fixed point free*.

Since $T$ compact implies $(I - T)(C)$ closed, the following result can be seen as a consequence of either Theorem 4 or Corollary 2.

**Corollary 3.** *Let $C$ be a nonempty bounded closed convex subset of a Banach space $X$ and $T : C \longrightarrow C$ a nonexpansive mapping such that $T(C)$ is compact. Then $T$ has a fixed point.*

Two long surveys on nonexpansive mappings were provided by Ivanov in 1976 [61] and by Gulevich [57] in 1996, where the stress was put on the geometry of the space and the structure of the domain (see also the handbook [72, Chaps. 3, 4, 12], the book by Agarwal *et al.* [4], the survey paper [70], and [95]).

Historically, a basic result of great importance concerning the fixed point theory for nonexpansive mappings was proved in 1965. This result was proved for Hilbert spaces by Browder [16], for uniformly convex spaces by Browder [17] and Göhde [54], and for reflexive spaces with normal structure by Kirk [67] (see also [3, Theorem 2.1], [20, Theorem 4], [56, Theorem 1.3], [50, 51], [60, Theorem 3.4.4], [98]). In the next two subsections, these existence results are reviewed.

## *2.2 Structure of domain*

**Definition 1.** (a) Let $X$ be a Banach space and $C \subset X$ a nonempty bounded subset. A point $c \in C$ is said to be *diametral* if $\sup_{x \in C} \|x - c\| = \operatorname{diam} C$.

(b) We say that a set $A$ has *normal structure*, if for any given bounded convex subset $C \subset A$ containing more than one point, there exists a *nondiametral* $c \in C$, i.e. $C$ is contained in a ball whose center is a point of $C$ and whose radius is less than the diameter of $C$.

(c) A bounded sequence $(x_n)_{n \in \mathbb{N}} \subset X$ is called diametrical if $\operatorname{diam}\{x_n\} > 0$ and

$$\lim_{n \to \infty} dist\,(x_{n+1}, \operatorname{conv}\{x_1, x_2, \ldots, x_n\}) = \operatorname{diam}(x_n).$$

In 1948, Brodskii and Milman [15] proved the following characterization (see also [4, Proposition 3.3.9], [51, Lemma 4.1], [57, Theorem 1.1.3]).

**Proposition 1.** *A subset $A \subset X$ has a normal structure if and only if it has no diametrical sequences.*

In particular, every convex compact subset of an arbitrary Banach space has normal structure [4, Proposition 3.3.1]. Moreover, the result still holds true if $C$ is any bounded closed convex subset (not necessarily compact) of a uniformly convex space (see Definition 3) [4, Proposition 3.3.3].

**Definition 2.** (a) A mapping $F : C \longrightarrow C$ is called demi-closed if for any $y \in X$ and any sequence $(x_n)_{n \in \mathbb{N}} \subset C$, the condition $(x_n)$ converges weakly to $x$ and $\|F(x_n) - y\| \to 0$ imply that $x \in C$ and $F(x) = y$.

(b) We say that a Banach space $X$ satisfies *the Opial condition* if given any sequence $x_n \rightharpoonup x$ and $y \neq x$ in $X$, we have

$$\liminf_{n \to \infty} \|x_n - x\| < \liminf_{n \to \infty} \|x_n - y\|.$$

**Proposition 2.** *Let $X$ be a reflexive Banach space satisfying the Opial condition. Then*

*(a) $X$ has a normal structure.*
*(b) If $T : C \longrightarrow C$ is nonexpansive with $C \subset X$ closed and convex subset, then $I - T$ is demi-closed.*

*Proof.* (a) Arguing by contradiction and using Proposition 1, assume that $X$ contains a sequence of diametral points $(x_n)_{n\in\mathbb{N}}$. Then

$$\lim_{n\to\infty} \operatorname{dist}(x_{n+1}, \operatorname{conv}\{x_1, x_2, \ldots, x_n\}) = \operatorname{diam}(x_n).$$

Without loss of generality, assume that $x_n \rightharpoonup 0$ weakly, as $n \to \infty$. Then, for any $y \in \overline{\operatorname{conv}}\{x_1, x_2, \ldots, x_n\}$, we have $\lim_{n\to\infty} \|y - x_n\| = \operatorname{diam}(x_n)$, hence $y = 0$ and, as a consequence, $\lim_{n\to\infty} \|x_1 - x_n\| = \operatorname{diam}(x_n)$, contradicting Opial's condition.

(b) Let $(x_n)$ converge weakly to $x$ and $\|(I - T)(x_n) - y\| \to 0$. Set $T_y x = Tx + y$; then $T_y$ is nonexpansive and $\lim_{n\to\infty} \|x_n - T_y x_n\| = 0$. Since

$$\|T_y x - x_n\| \le \|T_y x - T_y x_n\| + \|T_y x_n - x_n\|,$$

then

$$\liminf_{n\to\infty} \|T_y x - x_n\| \le \liminf_{n\to\infty} \|x - x_n\|$$

which yields, by Opial's condition, $T_y x = x$, i.e., $(I - T)x = y$.

□

Observe that in part (b), the reflexivity of the Banach space is not needed, but only $C$ weakly compact is required.

*Example 2.* (a) The sequence spaces $l^p$ $(1 < p < \infty)$ satisfy the Opial condition (the weak and strong convergence coincide) while the Lebesgue spaces $L^p$ for $1 < p < \infty,\ p \neq 2$ do not (see, e.g., [61]).

(b) Every Hilbert space $H$ satisfies the Opial condition because of the parallelogram identity:

$$\|x_n - y\|^2 = \|x_n - x\|^2 + 2 < x_n - x, x - y > + \|x - y\|^2,$$

for $x_n, x, y \in H$.

Now, we consider the fixed point theory for nonexpansive mappings in some special situations. For the proof of the first result, we refer to [51, Theorem 10.2] and [72, Theorem 2.1, Chap. 3], [89, Theorem 3.4.3].

**Theorem 5.** *Let $C \subset X$ be a nonempty convex with normal structure subset of a reflexive Banach space $X$. Then every nonexpansive mapping $T : C \longrightarrow C$ has a fixed point.*

In fact, we only need $C \subset X$ to be weakly compact with normal structure, as noticed by Kirk in 1965 (see [49, Theorem 7.1.23] or [51, Theorem 4.1] or [57, Theorem 1.1.4]):

**Theorem 6.** *Let $C \subset X$ be a nonempty convex weakly compact with normal structure subset of a Banach space $X$. Then every nonexpansive mapping $T : C \longrightarrow C$ has a fixed point.*

*Proof.* By Zorn's Lemma, there is a nonempty minimal $T$-invariant closed convex subset $K \subset C$ (in the sense that it contains no proper closed convex subset invariant under $T$). Arguing by contradiction, assume that $d = \operatorname{diam} K > 0$. Since $C$ has a normal structure, there exists $r \in (0, d)$ such that

$$E = \{x \in K \mid K \subset B(x, r)\} \neq \emptyset.$$

Since $T$ is nonexpansive, for $x \in E$, $T(K) \subset B(Tx, r)$. Then $\overline{\operatorname{Conv}}\, T(K) \subset B(Tx, r)$. But $\overline{\operatorname{Conv}}\, T(K)$ is invariant by $T$, then also $K \subset B(Tx, r)$. Therefore $Tx \in E$ and $E$ is invariant by $T$. In addition, $E$ is closed convex hence $E = \bigcap\{B(y, r),\ y \in K\}$. The minimality of $K$ guarantees that $E = K$ and so $\|x - y\| \leq r$, for all $x, y \in E$, proving that $d \leq r < d$, a contradiction; as a consequence, $\operatorname{diam} K = 0$. This means that $K$ consists of a single point, a fixed point for $T$. □

Since, by Proposition 2, in a reflexive Banach space, Opial's condition implies the normal structure of the domain, we deduce

**Corollary 4.** *Let $C$ be a nonempty closed convex subset of a reflexive Banach space with the Opial condition satisfied. Then every nonexpansive mapping $T : C \longrightarrow C$ has a fixed point.*

*Remark 5.* The weak-compactness of $C$ cannot be dispensed as the following counterexample shows. Let $X = C[0, 1]$,

$$C = \{x \in X |\ 0 \leq x(t) \leq 1 \text{ for all } t \in (0, 1) \text{ and } x(0) = 0,\ x(1) = 1\},$$

and let $T : C \longrightarrow C$ be defined by $f(x)(t) = tx(t),\ t \in [0, 1]$. Then $C$ is a nonempty weakly compact convex subset but has no normal structure. Indeed, $\operatorname{diam} K = 1$. In addition, if $x \in K$, then for a given $\varepsilon > 0$, there is $\delta > 0$ such that $x(t) < \varepsilon$, for $0 < t < \delta$. Select a function $y \in K$ such that $y(t) = 1$ for $t \geq \delta/2$. Then $\|x - y\| \geq 1 - \varepsilon$. This proves that the point $x$ is diametrical and our claim follows. However $T$ is a fixed point free *nonexpansive* mapping (see [7]).

*Example 3.* (Sadovskii) In the Banach space $X = c_0$ of null sequences, let $C$ be the closed unit ball. Then $X$ is not reflexive and $C$ has no normal structure. However the mapping $T : C \longrightarrow C$ defined by the shift operator $T(x_1, x_2, x_3, \ldots) = (1, x_1, x_2, x_3, \ldots)$ is fixed point free.

## 2.3 Geometry of space

The following concept was first introduced by Clarkson in 1936 [26].

**Definition 3.** A space $X$ is said to be *uniformly convex* if it satisfies the following geometric condition: $\forall\, \varepsilon > 0,\ \exists\, \delta > 0,\ \forall\, (x, y) \in X^2$,

$$\|x - y\| \geq \varepsilon,\ \|x\| \leq 1, \|y\| \leq 1 \ \Rightarrow \|(x + y)/2\| \leq 1 - \delta.$$

We have the characterization:

**Proposition 3.** *(a) $X$ is uniformly convex if and only if for any two sequences $x_n, y_n \in B(0, 1)$, if $\|x_n + y_n\| \to 2$, then $\|x_n - y_n\| \to 0$, as $n \to \infty$.*
*(b) $X$ is uniformly convex if and only if for all $\varepsilon \in [0, 2]$, $\delta_X(\varepsilon) > 0$, where $\delta_X(\varepsilon)$ is the modulus of continuity defined by*

$$\delta_X(\varepsilon) = \inf\{1 - \|(x + y)/2\| :\ x, y \in B(0, 1),\ \|x - y\| \leq \varepsilon\}.$$

*Remark 6.* The uniform convexity is a geometric property of the unit ball: if we slide a rule of length $\varepsilon > 0$ in the unit ball, then its midpoint must stay within a ball of radius $1 - \delta$ for some $\delta > 0$. In particular, the unit sphere must be "round" and cannot include any line segment. For instance $(\Re^2, \|.\|_2)$ is uniformly convex while $(\Re^2, \|.\|_1)$ is not, where for $i = 1, 2$, $\|(x_1, x_2)\|_i = \left(x_1^i + x_2^i\right)^{1/i}$.

*Example 4.* Hilbert spaces and Lebesgue spaces $L^p(\Omega)$ $(1 < p < \infty)$ are uniformly convex spaces (see, e.g., [41]).

**Proposition 4.** *Any uniformly convex Banach space is reflexive (see, e.g., [5, 25, 41, 101]).*

*Remark 7.* Every nonempty bounded closed convex subset $C$ of a *uniformly convex* Banach space $X$ has *normal structure* (Edelstein, [39], 1974). Indeed, if $d = \operatorname{diam} C > 0$, then there exist $a, b \in C$ such that $\|a - b\| \geq d/2$. Then for every $x \in C$, we have

$$\|x - a\| \leq d,\ \|x - b\| \leq d,\ \text{ and }\ \|(x - a) - (x - b)\| \geq d/2.$$

Hence $\|(x - a) - (x - b)\| \leq 2\,(1 - \delta(1/2))\, d$, where $\delta$ is the modulus of continuity, i.e., $\|x - \frac{a+b}{2}\| \leq 2\,(1 - \delta(1/2))\, d$ and thus $C \subset B(\frac{a+b}{2}, (1 - \delta(1/2))\, d)$.

Basic references for the geometry of Banach spaces are [25, 29, 53]. The following result is a direct consequence of Theorem 5, Proposition 4, and Remark 7 (see also Subsection 2.5).

**Corollary 5.** *Let $C$ be a nonempty bounded closed convex subset in a uniformly convex space $X$. Then each nonexpansive mapping $T : C \longrightarrow C$ has a fixed point.*

## 2.4 Recent results

The property that a mapping is $\psi$-expansive is related to the closedness of the range.

**Definition 4.** An operator $F : \mathscr{D}(F) \subset X \longrightarrow X$ is said to be $\psi$-expansive if there exists a function $\psi : [0,\infty) \longrightarrow [0,\infty)$ with $\psi(0) = 0$, $\psi(r) > 0, \ \forall\, r > 0$, $\psi$ is either continuous or nondecreasing, and

$$\|Fx - Fy\| \geq \psi(\|x-y\|), \ \forall\, x,\, y \in \mathscr{D}(F).$$

Then, we have (see [1, Corollary 2.25], [45, Lemma 3.3, Proposition 3.4])

**Theorem 7.** *Let X be a Banach space, $C \ni 0$ a bounded closed convex subset of X, and $T : C \longrightarrow C$ a nonexpansive mapping such that $I - T$ is $\psi$-expansive. Then T has a unique fixed point in C.*

*Proof.* By Theorem 4, $T$ has a sequence $(x_n)$ of approximate fixed points, i.e., $\lim\limits_{n\to\infty} \|x_n - Tx_n\| = 0$. We claim that $(Tx_n)$ is a Cauchy sequence. We may assume that $\psi$ is a nondecreasing function and that $(Tx_n)$ is not a Cauchy sequence. Then, there exist $\varepsilon_0 > 0$ and two increasing sequences $(n_k)$ and $(m_k)$ of positive integers such that, for all $k \in \mathbb{N}$, the inequality $\varepsilon_0 < \|Tx_{m_k} - Tx_{n_k}\|$ holds. Since $(x_n - Tx_n)$ is a null sequence, it is a Cauchy sequence. Given $\varepsilon = \psi(\varepsilon_0) > 0$, we have

$$\psi(\varepsilon_0) < \psi(\|Tx_{m_k} - Tx_{n_k}\|) \leq \|(x_{n_k} - Tx_{n_k}) - (x_{m_k} - Tx_{m_k})\|,$$

which is a contradiction; therefore, $(Tx_n)$ is a Cauchy sequence, as claimed. Finally, since T is a continuous mapping it is clear that the limit of $(Tx_n)$ is the unique fixed point of $T$. □

For the proof, see also [92, Theorem 8]. The following result, due to Garcia-Falset and Muñiz-Pérez, 2013, clarifies the relationship between the closedness of the range $\mathscr{R}(I-T)$ and the fact that $I - T$ is $\psi$-expansive (see [48, Proposition 3.1, Lemma 3.1]).

**Proposition 5.** *Let C be a nonempty bounded closed subset of a Banach space X and $F : C \longrightarrow C$ a mapping.*

*(a) Assume that F is continuous and injective with $F^{-1} : \mathscr{R}(F) \longrightarrow C$ uniformly continuous. Then the range $\mathscr{R}(F)$ is a closed subset of X.*
*(b) Assume that $F : C \longrightarrow X$ is $\psi$-expansive. Then F is injective and $F^{-1}$ is uniformly continuous.*

*Remark 8.* Combining parts (a) and (b) of Proposition 5, we can see that if $I - T$ is $\psi$-expansive, then $\mathscr{R}(I-T)$ is a closed subset of $X$ and thus Theorem 7 falls into the scope of Theorem 23. Indeed, Petryshyn proved that if $T : X \longrightarrow X$ is a 1-set contraction (see Section 3) and $I - T$ is $c$-expansive (i.e., with $\psi(s) = cs, \ c > 0$), then $I - T$ is a bijection (see [92, Theorem 8]). If further $T$ is linear, then $I - T$-$c$-expansive is also a necessary condition for the existence and the uniqueness of a fixed point (see [92, Corollary 9]).

## *2.5 Demi-closedness and closedness of the range*

**(1)** In case of a nonexpansive mapping $T$ defined on a bounded closed convex subset of *a uniformly convex Banach space* $X$, F. Browder's demi-closedness principle [19, Theorem 3] (see also [4, Theorem 5.2.12] or [97, Lemma 3.4] or [104, Prop. 10.9]), proved in 1968, states that $I - T$ is demi-closed and has a closed range; this fact follows from Proposition 2, 4 and Remark 7. In other words, Browder [19, Theorem 1] proved Theorem 23 for the class of nonexpansive mappings (see also [17]). The proof is outlined. Since $T$ is nonexpansive, $T$ possesses by Theorem 4 a sequence of approximate fixed points $(x_n)_{n\in\mathbb{N}} \subset \overline{U}$. Now since $X$ is a uniformly convex space, it is reflexive and then $(x_n)_{n\in\mathbb{N}}$ converges weakly. According to Browder's principle, $0 \in (I - T)(\overline{U})$, proving Corollary 5. More generally, in a uniformly convex space, every weak limit of an approximate fixed point of an operator $T$ is a fixed point of this operator. Notice finally that Browder's demi-closedness principle has been also proved by Opial for Hilbert spaces in 1967 (see [87, Lemma 2]).

**(2)** Moreover, the result still holds in the wider class of nonexpansive mappings satisfying Opial's condition (see [51, Theorem 10.3]); more generally, in a Banach space $X$ with $C \subset X$ a nonempty weakly compact subset and $T : X \longrightarrow X$ a nonexpansive mapping, if $\lim_{n\to\infty} x_n = x$ weakly and $\lim_{n\to\infty} (I - T)x_n = y$ ($y \in X$) strongly, then $y = (I - T)(x)$; otherwise,

$$\|x_n - Tx - y\| \leq \|x_n - Tx_n - y\| + \|Tx_n - Tx\|$$

implies

$$\liminf_{n\to\infty} \|x_n - Tx - y\| \leq \liminf_{n\to\infty} \|x_n - x\|.$$

Then Opial's condition yields

$$\liminf_{n\to\infty} \|x_n - x\| < \liminf_{n\to\infty} \|x_n - (Tx + y)\|,$$

a contradiction.

**(3)** When $X$ is reflexive -in particular when $X$ is uniformly convex- every bounded subset $\Omega \subset X$ is weakly compact and thus $I - T$ demi-closed implies that $(I - T)(\overline{\Omega})$ is closed.

**(4)** When $C$ is a nonempty bounded closed convex subset of a uniformly convex space, Nussbaum [85, Lemma 3] proved the demi-closedness principle for the larger class of locally almost nonexpansive mappings $f : C \longrightarrow C$ (LANE, for short). These are mappings that satisfy: for all $x \in C$ and all $\varepsilon > 0$, there exists a neighborhood $U \in \mathscr{V}(x)$ such that for all $u, v \in X$

$$\|f(u) - f(v)\| \leq \|u - v\| + \varepsilon.$$

A report on the fixed theory for LANE mappings can be found in [69].

**(5)** Notice further that the demi-closedness principle is still valid for the sum of a nonexpansive mapping $T$ and a strongly continuous one, say $S$ (i.e., $Sx_n \longrightarrow Sx$ whenever $x_n \rightharpoonup x$ in $C$) provided $C$ is a bounded closed subset of a uniformly convex Banach space (see [104, Proposition 11.14(4)]).

## *2.6 Back to approximation methods*

Let $C \subset X$ be a nonempty closed subset of a Banach space $X$ and $T : \Omega \longrightarrow X$ a mapping. For some positive $\delta$ and a bounded subset $\Omega \subset C$, consider the following sets (see [51, 65, 101]):

$$F_\delta(T, \Omega) = \{x \in \Omega : \|x - Tx\| \leq \delta\}, \tag{1}$$

the set of the $\delta$-fixed points of $T$ in $\Omega$,

$$S = \{(x_n)_{n\in\mathbb{N}} \subset \Omega \mid \ x_n = \left(1 - \frac{1}{n}\right) Tx_n, \ \forall\, n = 1, 2, \ldots\}, \tag{2}$$

the set of approximate fixed points, and let

$$S_K = S \cap K, \tag{3}$$

where $K$ is a bounded closed convex subset. For some real parameters $\varepsilon > 0$ and $c > 0$ such that $0 < c < \alpha(\Omega) + \varepsilon$, define the sets:

$$N_\varepsilon(\Omega) = \{(x, y) \in \Omega^2 \mid \alpha(\Omega) - \varepsilon \leq \|x - y\| \leq \alpha(\Omega) + \varepsilon\}, \tag{4}$$

$$N_\varepsilon^c(\Omega) = \{(x, y) \in \Omega^2 \mid c \leq \|x - y\| \leq \alpha(\Omega) + \varepsilon\}, \tag{5}$$

where $\alpha$ is the measure of noncompactness of Kuratowski (see Definition 11 below). In 1974, Bruck [22] proved that $F_\delta(T, \Omega)$ is path-wise connected. The sets $F_\delta$ have been recently used to remove the condition of boundedness of the closed convex domain to prove fixed point theorems for nonexpansive mappings. In 2003, Penot [90, Corollary 3] showed that the boundedness of the subset $C$ can be dispensed if $T : C \longrightarrow C$ is not only *nonexpansive* but also *asymptotically contractive*, that is

$$\exists\, x_0 \in C, \quad \limsup_{x\in C,\, \|x\|\to\infty} \frac{\|Tx - Tx_0\|}{\|x - x_0\|} < 1.$$

Penot's result was recently generalized in [65, Theorem 2.4] for the case where $F_\delta$ is bounded:

**Theorem 8.** *Let X be a Banach space with the FPP, C be a closed convex subset of X (not necessarily bounded), and $T : C \longrightarrow C$ a nonexpansive mapping such that $F_\delta(T, C)$ is nonempty and bounded for some $\delta > 0$. Then T has a fixed point in C.*

Recall that $X$ has the FPP (fixed point property) if each of its bounded closed convex subsets has the fixed point property for nonexpansive self-mappings.

The following technical lemma (see [31, Lemma 3.1]) has been recently used to show the compactness of the set $S_K$ and then the convergence of a sequence of approximate fixed points.

**Lemma 1.** *Let $C \ni 0$ a closed convex subset of a Banach space X and $T : C \longrightarrow C$ a nonexpansive mapping satisfying the property ($\mathscr{K}$). Assume that there exist $\delta_0, \varepsilon_0 > 0$ such that for all $c \in (0, \alpha(S_K) + \varepsilon_0)$, we have*

$$[F_{\delta_0}(T, S_K) \times F_{\delta_0}(T, S_K)] \cap N^c_{\varepsilon_0}(T, S_K) = \emptyset. \tag{6}$$

*Then $\alpha(S_K) = 0$.*

The next proposition provides a sufficient condition for (6) be satisfied (see [33, Proposition 5.2]).

**Proposition 6.** *Let $\Omega$ be a nonempty bounded subset of X such that $\alpha(\Omega) > 0$ and C a closed convex subset of X. Suppose that $T : C \longrightarrow C$ is a nonexpansive mapping such that $I - T$ is $\psi$-expansive. Then for every $\varepsilon > 0$, $0 < c < \alpha(\Omega) + \varepsilon$, and all $\delta, \delta' > 0$ with $0 < \delta + \delta' < \psi(c)$, we have*

$$[F_\delta(T, \Omega) \times F_{\delta'}(T, \Omega)] \cap N^c_\varepsilon(\Omega) = \emptyset.$$

*Proof.* Let $\varepsilon > 0$ and $c > 0$ be such that $0 < c < \alpha(\Omega) + \varepsilon$. Arguing by contradiction, suppose that there exist $\delta, \delta' > 0$ such that $0 < \delta + \delta' < \psi(c)$ and

$$[F_\delta(T, \Omega) \times F_{\delta'}(T, \Omega)] \cap N^c_\varepsilon(\Omega) \neq \emptyset.$$

Since $I - T$ is $\psi$-expansive, for $(x, y) \in [F_{\delta'}(T, \Omega) \times F_\delta(T, \Omega)] \cap N^c_\varepsilon(\Omega)$, we have

$$\psi(\|x - y\|) \leq \|(x - Tx) - (y - Ty)\| \leq \|x - Tx\| + \|y - Ty\| \leq \delta + \delta'. \tag{7}$$

(a) *If $\psi$ is nondecreasing,* then since $(x, y) \in N^c_\varepsilon(\Omega)$, we have $c \leq \|x - y\|$ which implies that $\psi(c) \leq \psi(\|x - y\|) \leq \delta + \delta'$, leading to a contradiction with $0 < \delta + \delta' < \psi(c)$.

*(b)* *If $\psi$ is continuous,* then let $(x,y) \in N^c_\varepsilon(\Omega)$ be such that $\|x-y\| = c + 1/n$ for large enough $n \in \{1,2,\ldots\}$. By continuity of $\psi$, for every $\eta > 0$, there is $n_0 \in \{1,2,\ldots\}$ such that for all $n \geq n_0$

$$\psi(c) - \eta < \psi(\|x-y\|) < \psi(c) + \eta.$$

This with (7) yields $\psi(c)-\eta \leq \delta+\delta'$ and a contradiction is reached by choosing $\eta > \psi(c) - (\delta + \delta')$.

□

As a consequence, we recapture Theorem 7 (see [31, 33]).

## 2.7 *Nonlinear alternatives*

An important tool to prove nonlinear alternatives is the following concept due to K. Borsuk (1930) [13]:

**Definition 5.** Let $X$ be a Hausdorff space and $C \subset X$ a nonempty subset. We say that $C$ is *a retract* of $X$ if there exists a continuous mapping $r : X \longrightarrow C$ such that $r_{|\,C} = Id_{|\,C}$ where $Id$ is the identity operator. Then the mapping $r$ is called *a retraction*.

*Example 5.* Let $(X, \|.\|)$ be a normed space. Then every closed ball $C = \overline{B}_R$ is a retract of $X$ via *the radial retraction* $r : X \longrightarrow C$ defined by

$$r(x) = \begin{cases} x, & \text{if } x \in C \\ Rx/\|x\|, & \text{if } x \notin C. \end{cases}$$

In a Hilbert space, the nearest point mapping $r$ is a nonexpansive mapping (see, e.g., [94, Page 795], [20, Lemma 1], [56, Lemma (1.4)], [60, Theorem 6.1.4]). In the general case of a Banach space, it has a Lipschitz constant 2. Example 8 will provide a construction of a retraction on closed convex subsets with nonempty interior. We also recall that the fixed point set of a nonexpansive mapping $T : C \longrightarrow C$ is a nonempty nonexpansive retract of $C$ [21, 22] and that every uniformly convex Banach space has the nonexpansive retract property [25, 29]. Moreover, we have

$$\text{Retract} \implies \text{Closed (in any topological space)}$$

$$\text{Closed + Convex} \implies \text{Retract (in any locally convex topological vector space).}$$

The second fact is an immediate consequence of Dugundji's extension theorem:

**Theorem 9 ([37]).** *Let $X$ be a locally convex topological vector space, $A$ a closed subset of a metric space $E$, and $f : A \to X$ a continuous mapping. Then there exists a continuous extension $\tilde{f} : X \to X$ such that $\tilde{f}(E) \subset$* conv $(f(A))$.

The following results are commonly called nonlinear alternatives.

**Theorem 10.** *Let $H$ be a Hilbert space, $C = \overline{B}_R$, and $T : C \longrightarrow X$ a nonexpansive map. Then at least one of the following properties holds: either*

*(a) $T$ has a fixed point, or*
*(b) there is an $x \in \partial C$ and $\lambda \in (0,1)$ such that $x = \lambda T(x)$.*

*Proof.* Let $r_C$ be the radial retraction and $\varphi = r_C \circ T$. Then $\varphi : C \longrightarrow C$ is *nonexpansive*. Moreover $H$ is *a Hilbert space, hence uniformly convex and $C$ has normal structure*. By Theorem 5, $\varphi$ has a fixed point $x \in C$. If $T(x) \in C$, then $x = T(x)$ and we are done; otherwise, $x = \frac{RT(x)}{\|T(x)\|} \in \partial C$ in which case for $\lambda = \frac{R}{\|T(x)\|} < 1$ we have $x = \lambda T(x)$, as claimed. □

Theorem 10 is extended to *uniformly convex spaces*; for the proof we refer to [3, Theorem 3.3]] or [97, Theorems 3.2, 3.5]].

**Theorem 11.** *Let $X$ be a uniformly convex space, $\Omega \subset X$ a bounded open subset with $0 \in \Omega$, and $T : \overline{\Omega} \longrightarrow X$ a nonexpansive mapping. Then at least one of the following properties holds:*

*(a) $T$ has a fixed point,*
*(b) there is an $x \in \partial\Omega$ and $\lambda \in (0,1)$ such that $x = \lambda T(x)$.*

Each one of the conditions in the following corollary guarantees that the condition (b) in the above theorem does not occur:

**Corollary 6.** *Let $X$ be a uniformly convex space, $\Omega \subset X$ a bounded open subset with $0 \in \Omega$, and $T : \overline{\Omega} \longrightarrow X$ a nonexpansive mapping. Assume that, for $x \in \partial\Omega$, one of the following conditions holds:*

*(a) $\|T(x)\| \leq \|x\|$,*
*(b) $\|T(x)\| \leq \|x - f(x)\|$ (Petryshyn's condition),*
*(c) $\|T(x)\|^2 \leq \|x\|^2 + \|x - T(x)\|^2$ (Altman's condition),*
*(c) $< T(x), x > \leq \|x\|^2$ and $X$ Hilbert (Krasnoselskii's condition)*
*Then $T$ has a fixed point in $\overline{\Omega}$.*

Further existence results can be found in [31–33, 44, 48, 79–81] and the references therein.

## 2.8 Boundary conditions

Let $C \subset X$ be a subset of a Banach space. The main boundary conditions (Leray-Schauder condition, inward condition, Furi-Pera type condition, and interior condition) are particularly useful and often necessary when a mapping $T$ that does not self-map $C$. We are going to prospect them.

### 2.8.1 The Leray-Schauder and the inward conditions

Let $X$ be a Banach space, $\Omega$ a bounded open subset and let $T : \overline{\Omega} \longrightarrow X$ be a completely continuous mapping satisfying the so-called Leray-Schauder boundary condition:

$$(LS) \quad Tx \neq x_0 + \lambda(x - x_0), \ \forall\, x \in \partial\Omega, \ \lambda > 1,$$

$x_0 \in \Omega$. The geometric interpretation of this condition is that there is no $x \in \Omega$ such that $T(x)$ lies on the continuation of the segment $[x_0, x]$ beyond $x$. It enables one to define the Leray-Schauder topological degree for the class of compact perturbations of the identity [82]. The Leray-Schauder principle states that $T$ has at least one fixed point in $\overline{\Omega}$. Indeed, assume that $T(x) \neq x$ for all $x \in \Omega$ for otherwise we are done. Now define the homotopy $T_t(x) = x - x_0 - t(T(x) - x_0)$ for $t \in [0, 1]$ and $x \in \overline{\Omega}$. If there exist $t \in (0, 1)$ and $x \in \partial\Omega$ such that $T_t(x) = 0$. Then $T(x) - x_0 = \frac{x - x_0}{t}$, contradicting $(LS)$ for $1/t > 1$. By the homotopy invariance property of the Schauder topological degree [30, 82, 104]

$$\deg(I - T, \Omega, 0) = \deg(I - x_0, \Omega, 0) = \deg(I, \Omega, x_0) = 1,$$

for $x_0 \in \Omega$; hence, there exists an $x \in \Omega$ satisfying $T(x) = x$, as claimed. It should be emphasized that $(LS)$ is not needed when $\Omega$ is convex and $T$ leaves $\overline{\Omega}$ invariant.

**Definition 6.** (a) Let $X$ be a Banach space, $C \subset X$ is nonempty closed convex subset, and $x \in C$. Then

$$I_C(x) = \{(1 - \lambda)x + \lambda y |\ \lambda \geq 0, \ y \in C\}$$

is called *the inner set* of $x$ with respect to $C$.

(b) A mapping $T : C \longrightarrow C$ is said to be *inward* if $T(x) \in I_C(x)$, for every $x \in C$ and *weakly inward* if $T(x) \in \overline{I_C(x)}$, for every $x \in C$.

The set $I_C(x)$ is the union of all rays emanating from $x$ and passing through some other point $y \in C$. Indeed, if $y \in \overline{I_C(x)}$, then $(1 - \lambda)x + \lambda y \in \overline{I_C(x)}$ for all positive $\lambda$ [75, Lemma 2.1, Remark 4.1]. The other interesting feature of the weakly inward condition is that if $T(x) \in \overline{I_C(x)}$ (and $T(x) \notin C$), then the distance between $T(x)$ and $C$ is controlled by $\operatorname{dist}(T(x), C) < \|T(x) - x\|$.

Moreover since $C$ is convex, $C \subset I_C(x)$ and

$$\overline{I_C(x)} = x + \{y \in X |\ \lim_{h \to 0^+} h^{-1} \operatorname{dist}(x + hy, C) = 0\}.$$

The sets $I_C(x)$ were first used by Halpern and Bergman [58]. In [75, Proposition 4.1, Remark 4.1], it is proved that $C = \bigcap_{x \in C} I_C(x)$. The following fixed point theorem was proved for LANE mappings by R. Nussbaum in 1972 [86, Theorem 1] (when $T$ self-maps $C$ or only when $T(\partial C) \subset C$ and $C$ has nonempty interior) and by S. Reich in 1973 [98, Corollary 4].

**Theorem 12 ([4, Theorem 5.2.26]).** *Let $X$ be a uniformly Banach space, $C \subset X$ a nonempty bounded closed convex subset, and $T : C \longrightarrow X$ a weakly inward nonexpansive mapping. Then $T$ has a fixed point.*

When $C$ is a convex subset with nonempty interior, it easy to see that the inwardness condition is more general than the Rothe condition $f(\partial C) \subset C$ or the condition that $(x, T(x)] \cap C \neq \emptyset$, for every $x \in \partial C$, where

$$(x, T(x)] = \{x + t(T(x) - x) |\ 0 < t \leq 1\}.$$

More precisely, we have (see [98, Remark 1] for the second implication).

$$\text{Rothe's cond.} \implies \text{inwardness cond.} \implies \text{weak inwardness cond.} \implies (LS)$$

Then Theorem 12 takes the following more general form (see [72, Chap. 10, Theorem 2.8]), a result proved by Browder in 1968 [19]:

**Theorem 13.** *Let $X$ be a uniformly Banach space, $C \subset X$ a nonempty bounded closed convex subset with nonempty interior, and $T : C \longrightarrow X$ a nonexpansive mappings satisfying $(LS)$. Then $T$ has a fixed point.*

This result was improved by J. Caristi in 1976 [24] (see also [89, Theorem 3.4.38]). In fact, Caristi observed that if $C$ is a nonempty closed convex subset of $X$ and $T : C \longrightarrow X$, then $I - T$ weakly inward is equivalent to

$$\lim_{h \to 0^+} h^{-1} \operatorname{dist}(x - hT(x), C) = 0,$$

for every $x \in C$ (see [4, Proposition 5.1.1], [24, Theorem 1.2], or [30]). The latter condition is known as the Nagumo or Brezis' boundary condition. Then, Caristi proved the following result for Lipschitz, pseudo-contractive mappings [24, Theorem 2.6 ]:

**Theorem 14.** *Let $X$ be a Banach space and $C \subset X$ a nonempty closed convex subset which has the f.p.p. with respect to nonexpansive mappings. If $T : C \longrightarrow X$ is a nonexpansive weakly inward mapping, then $T$ has a fixed point.*

Notice that the result also holds (with uniqueness) for weakly inward contraction mappings; for the proof, see, e.g., [4, Proposition 5.1.2]. Then, for every $x_0 \in C$ and $t \in (0, 1)$, the mapping $T_t(x) = (1 - t)x_0 + tT(x)$ has exactly one fixed point whenever $T : C \longrightarrow X$ is weakly inward. If further $C$ is bounded, then $T$ has a sequence of approximate fixed points. Finally $T$ has a fixed point if $I - T$ is closed and the latter condition turns out to be sufficient too.

**Theorem 15.** *Let $X$ be a Banach space and $C \subset X$ a nonempty bounded closed convex subset. If $T : C \longrightarrow X$ is a nonexpansive weakly inward mapping with $I - T$ closed, then $T$ has a fixed point.*

When $\inf\{\|x - Tx\| : x \in \partial C, Tx \notin C\} > 0$, $C \subset X$ is a closed convex subset with the f.p.p., and $X$ is any Banach space, Kirk [68, Theorem 3.1] proved that $T$ has a fixed point. As a consequence, a fixed point result is obtained under the sufficient condition that there exists $x_0 \in \operatorname{int} C$ such that $\|x_0 - Tx_0\| < \|x - Tx\|$, for all $x \in \partial C$. In the general case of a bounded open subset of an arbitrary Banach space for which $(LS)$ holds, one can only assert that $\inf\{\|x - Tx\| : x \in C\} = 0$ (see, e.g., [72, Chap. 10, Proposition 2.14]). We close this subsection with a result analogous to the one in Theorem 11 but presented here in the more general framework of pseudo-contractive mappings (for the definition, see Sect. 3). The proof can be found in [4, Theorem 5.7.20]:

**Theorem 16.** *Let $X$ be a uniformly convex space, $C \subset X$ a nonempty closed convex subset with $0 \in C$, and $T : C \longrightarrow X$ a weakly inward continuous pseudo-contractive mapping. Then $T$ has a fixed point if and only if the set $\{x \in C|\ Tx = \lambda x,\ \lambda > 1\}$ is bounded.*

### 2.8.2 The interior condition

We begin with some definitions:

**Definition 7.** Let $X$ be a Banach space and $\Omega \ni 0$ a bounded open subset. We say that $\Omega$ is *strictly star-shaped* with respect to the origin if for each $x \in \partial\Omega$, we have $\{tx|\ t > 0\} \cap \partial\Omega = \{x\}$.

**Definition 8.** $\Omega$ is said to be *star-shaped* if there exists $x_0 \in \Omega$ such that for each $x \in \Omega$, it holds that $[x_0, x] \subset \Omega$, where $[x_0, x]$ refers to the closed line segment $\{tx + (1-t)x_0,\ 0 \le t \le 1\}$ joining the two points $x_0, x$, i.e., $[x_0, x] = \overline{co}(\{x_0, x\})$.

Then a convex set is a set which is star-shaped with respect to each of its points. When $\Omega$ is a bounded open neighborhood of the origin, the following strict inclusions hold (see [64, Proposition 1]):

$$\text{Convex} \subset \text{Strictly Star-shaped} \subset \text{Star-shaped}.$$

Notice further that the condition $F(\overline{\Omega}) \subset \overline{\Omega}$ implies $(LS)$ if $\Omega$ is strictly star-shaped. The proof employs the Minkowski functional introduced in Proposition 8 (see the proof of Theorem 28).

**Definition 9.** A mapping $T : \overline{\Omega} \longrightarrow X$ satisfies *the interior condition* if there exists $\delta > 0$ such that

$$(IC) \quad Tx \neq \lambda x, \ \text{ for } x \in \Omega_\delta,\ \lambda > 1 \ \text{ and } \ T(x) \notin \overline{\Omega},$$

where $\Omega_\delta = \{x \in \Omega : \operatorname{dist}(x, \partial\Omega) < \delta\}$ is the set of $\delta$-interior points.

In [64, Theorem 2], the authors proved that if $T : B \longrightarrow X$ is a *nonexpansive* mapping, where $B$ is a closed ball in a Hilbert space $X$, then $(IC)$ does imply $(LS)$. We have

**Theorem 17.** *Let $X$ be a Banach space, $\Omega \subset X$ is a nonempty bounded open subset, and $T : \overline{\Omega} \longrightarrow X$ a nonexpansive mapping satisfying $(IC)$.*

*(a) If $\Omega$ is strictly star-shaped, then $T$ has a sequence of approximate fixed points. $T$ has a fixed point if further $I - T$ has a closed range.*
*(b) If $\Omega$ is convex and $X$ is uniformly convex, then $T$ has a fixed point.*

*Proof.* (a) The first claim follows from [64, Theorem 1] already proved for condensing mappings. Indeed, it suffices to approximate $T$ by a sequence of condensing mappings $T_n(x) = (1 - 1/n)T(x)$, $(n = 1, 2, \ldots)$. The second claim is straightforward.
(b) Regarding the second part, we only notice that the Browder demi-closedness principle applies and then the sequence of approximate fixed points $(x_n)_n$ obtained in part (a) converges weakly in the reflexive Banach space to a sought fixed point.

□

In [55, Corollary 3], González *et al.* proved, as in part (a), that if $X$ is *a Hilbert space*, a nonexpansive mapping $T : \overline{\Omega} \longrightarrow X$ has a fixed point provided either $(LS)$ or $(IC)$ holds and $\Omega$ is a bounded strictly star-shaped open neighborhood of the origin. In fact, the authors proved the continuity of a radial projection on the interior of strictly star-shaped sets (see [55, Proposition 2]):

**Proposition 7.** *Let $X$ be a Banach space and $U$ a bounded strictly star-shaped open neighborhood of the origin. Let $k = k(U) = \inf\{\|x\|, x \in \partial U\} = \text{dist}(0, \partial U) > 0$, $K = K(U) = \sup\{\|x\| : x \in \partial U\}$, $\delta \in (0, k]$, and $r = \frac{\delta}{K}$. Then the function $P : X \longrightarrow \overline{U}$ defined by*

$$Px = \begin{cases} \frac{r+(1-r)g(x)}{g^2(x)}\, x, & \text{if } x \in X \setminus \overline{U}, \\ x, & \text{if } x \in \overline{U} \end{cases}$$

*is continuous on $X$, is the identity on $\overline{\Omega}$ and $P(X \setminus \overline{U}) \subset U_\delta$. Moreover $Px \in \overline{co}(\{0\} \cup \{x\})$ for all $x \in X$, where $g$ is the Minkowski functional, as recalled below.*

Let $X$ be a normed space and $\Omega \subset X$ a nonempty subset. The Minkowski functional is the function $g_\Omega = g : X \longrightarrow [0, +\infty)$ defined by (see [11])

$$g(x) = \inf\{\lambda > 0 : \lambda^{-1}x \in \overline{\Omega}\} = \inf\{\lambda > 0 : x \in \lambda\overline{\Omega}\}.$$

**Proposition 8 ([101, Lemma 4.2.5]).** *We have*

*(a) $g(\lambda x) = |\lambda| g(x)$, for $\lambda \in \Re$ and $x \in X$,*
*(b) $0 \le g(x) \le 1$, if $x \in \Omega$.*

Moreover, $g(x) < \infty$ if and only if $\Omega$ is an absorbing set. $\Omega$ is said to be an absorbing set if for each $x \in \Omega$ there is some $\lambda_0 > 0$ such that $x \in \lambda\Omega$ for all $|\lambda| \geq \lambda_0$. If $\Omega$ is convex, then $g(x) \geq 1,\ \forall\, x \notin \Omega$ and also

$$g(x + y) \leq g(x) + g(y),\ \forall\, x, y \in X.$$

If further $\Omega$ is open, then

$$g(x) = 1,\ \forall\, x \in \partial\Omega,\ g(x) < 1,\ \forall\, x\imath\Omega,\ \text{ and }\ g(x) > 1,\ \forall\, x \notin \overline{\Omega}.$$

### 2.8.3 The Furi-Pera condition

Contrarily to the interior condition, we now present a result when the interior of the open subset $\Omega$ may be empty. In 1987, Furi and Pera introduced a new condition and proved the following fixed point theorem in the general framework of Fréchet spaces:

**Theorem 18 (See [3, Theorem 8.5] or [43]).** *Let E be a Fréchet space, C a closed convex subset of E, $0 \in C$, and $T : C \to E$ a continuous compact mapping. Assume further that*

$$(FP) \begin{cases} \text{if } \{(x_j, \lambda_j)\}_{j\geq 1} \text{ is a sequence in } \partial C \times [0, 1] \\ \text{converging to } (x, \lambda) \text{ with } x = \lambda T(x) \text{ and } 0 \leq \lambda < 1, \\ \text{then } \lambda_j T(x_j) \in C, \text{ for } j \text{ sufficiently large.} \end{cases}$$

*Then T has a fixed point in C.*

In order to present some recent results with the $(FP)$ condition, let us give

**Definition 10.** (a) A subset $A \subset$ of a Banach space is a nonexpansive retract of $X$ if there exists a nonexpansive mapping $r : X \longrightarrow A$ such that $rx = x$ for all $x \in A$. The mapping $r$ is called a nonexpansive retraction.
(b) We say that a Banach space $E$ has the nonexpansive retract property (NRP for short) if each of its nonempty closed convex subsets is a nonexpansive retract of $X$. For instance, a Hilbert space enjoys the NRP.

We have (see [32, Theorem 3.1]).

**Theorem 19.** *Let X be a Banach space satisfying the NRP and $C \ni 0$ a closed convex subset of X (not necessarily bounded). Let $T : C \longrightarrow E$ be a nonexpansive mapping satisfying (6) and the property $(\mathscr{K})$. Assume that the Furi-Pera condition hold. Then T has a fixed point in C.*

*Proof.* **Step 1. Approximate fixed points for $Tr_K$.** Let $r : X \longrightarrow C \cap K$ be a nonexpansive retraction where $K$ is a closed convex subset and, for each $n \in \{1, 2, \ldots\}$, consider the nonlinear equation

$$x = (1 - 1/n)Tr_K(x), \tag{8}$$

where $r_K = r_{|K}$ is the restriction of $r$ on the set $K$ and $Tr_k = T \circ r_K : K \longrightarrow K$. Without loss of generality, assume that $0 \in K \cap C$. Indeed, in the case $0 \notin K$, one may take any $p \in K \cap C$ and instead of equation (8) rather consider the equation $x = (1 - 1/n)Tr_K(x) + p/n,\ \ n \in \{1, 2, \ldots\}$. Now, since $f(K \cap C) \subset K$ and $r_K : K \longrightarrow K \cap C$, we have that $Tr_K : K \longrightarrow K$. By convexity of $K$ and the fact that $p \in K \cap C \subset K$, we deduce that for every $x \in K,\ (1-1/n)Tr_K(x)+p/n \in K$. Also, since $T$ and $r$ are nonexpansive mappings, then for each $n \in \{1, 2, \ldots\}$, the mapping $T_n : K \longrightarrow K$ defined by $T_n(x) = (1-1/n)Tr_K(x)+p/n$ is a contraction. By the Banach fixed point theorem, for each $n \in \{1, 2, \ldots\}$, $T_n$ admits a unique fixed point $x_n \in K$. This implies that equation (8) has a unique solution $x_n$ for each $n \in \{1, 2, \ldots\}$.

**Step 2. Approximate fixed points for $T$.** We show that, for each $n \in \{1, 2, \ldots\}$, the following equation:

$$x = (1 - 1/n)T(x) \tag{9}$$

has a solution. For this, it suffices to prove that the sequence $(x_n)_n$ lies in $C$ where, for each $n \in \{1, 2, \ldots\}$, $x_n$ is a solution of the equation (8). Arguing by contradiction, assume that $(x_n)_n \not\subset C$ and let $x_{n_0} \notin C$ for some $n_0 \in \{1, 2, \ldots\}$. Since $C$ is closed, there exists $0 < \delta < dist(x_{n_0}, C)$. Following the proof of [3, Theorem 5.10], choose an integer $m \in \{1, 2, \ldots\}$ such that $m > 1/\delta$; then, for all integer $i \geq m$, consider the open set $U_i = \{x \in X |\, d(x, C) < 1/i\}$. It is clear that $dist(x_{n_0}, C) > \delta$ and $1/i < \delta$ imply that $x_{n_0} \notin \overline{U}_i$. In addition, for each $i \geq m$, $U_i \cap K \neq \emptyset$ because $C \cap K \subset U_i \cap K$ and, by definition, $C \cap K \neq \emptyset$. Thus, the mapping $(1 - 1/n_0)Tr_K : \overline{U}_i \longrightarrow K$ is well defined and it is further a contraction; in addition, $(1 - 1/n_0)Tr_K(\overline{U}_i)$ is bounded for $r_K(\overline{U}_i \cap K) \subset C \cap K$ and, by the property $(\mathscr{K})$, $f(C \cap K) \subset K$ where $K$ is a bounded subset. Since $x_{n_0} \notin \overline{U}_i$, a nonlinear alternative [3, Theorem 3.2] guarantees the existence $(y_i, \mu_i) \in \partial U_i \times (0, 1)$ such that $y_i = \mu_i(1-1/n_0)Tr_K(y_i)$. Note that since $x_{n_0} \notin \overline{U}_i$, the mapping $(1 - 1/n_0)Tr$ has no fixed points in $\overline{U}_i$. As a consequence

$$\mu_i(1 - 1/n_0)Tr_K(y_i) \notin C,\ \ \forall\, i \geq m. \tag{10}$$

The set

$$D_{n_0} = \{x \in X : \exists\, \mu \in [0, 1],\ x = \mu(1 - 1/n_0)Tr_K(x)\}$$

is nonempty for it contains $0, x_{n_0}$ and $y_i$, for all $i \geq m$. Moreover $D_{n_0}$ is compact. Indeed

$$D_{n_0} \subseteq \overline{co}\,((1 - 1/n_0)Tr_K(D_{n_0}) \cup \{0\})$$

implies

$$\alpha(D_{n_0}) \leq \alpha\left(\overline{co}((1-1/n_0)Tr_K(D_{n_0}) \cup \{0\})\right),$$

where $\alpha$ is the Kuratowski MNC. However, since $T$ and $r$ are nonexpansive mappings, we have

$$\begin{aligned}\alpha(D_{n_0}) &\leq \alpha(\overline{co}\big((1-1/n_0)Tr_K(D_{n_0})\big)\\ &\leq (1-1/n_0)\alpha(r_K(D_{n_0}))\\ &\leq (1-1/n_0)\alpha(D_{n_0}).\end{aligned}$$

Then $\alpha(D_{n_0}) = 0$ proving that $D_{n_0}$ is compact for it is closed. Now, for each $i \geq m$ and $0 \leq \mu_i \leq 1$, we have that $d(y_i, C) = 1/i$ for $y_i \in \partial U_i \cap D_{n_0}$. Then, up to a subsequence, $\mu_i \longrightarrow \mu^* \in [0,1]$ and, by the compactness of $D_{n_0}$, $y_i \longrightarrow y^* \in C$, as $i \to +\infty$. Moreover $y_i = \mu_i(1-1/n_0)Tr_K(y_i)$ tends to $\mu^*(1-1/n_0)Tr_K(y^*)$ by continuity. Hence $y^* = \mu^*(1-1/n_0)Tr_K(y^*)$. In addition $x_{n_0} \notin C$ implies $\mu^* \neq 1$, otherwise we get by uniqueness $y_* = x_{n_0}$, which is a contradiction. Therefore $0 \leq \mu^* < 1$. Finally $r_K(y_i) \in \partial C$ follows from $y_i \notin C$ and the definition of the retraction $r$. In addition $y^* = r_K(y^*)$, $\mu_i' = (1-1/n_0)\mu_j$, and $\mu' = (1-1/n_0)\mu^*$. Since $T$ satisfies the Furi-Pera condition, we infer that $\mu_i(1-1/n_0)Tr_K(y_i) \in C$, for $i$ sufficiently large. This contradicts (10) and the fact that $y_i \notin C$, for $i \geq m$. Thus, for each $n \in \{1, 2, \ldots\}$, $x_n \in C \cap K$. Hence $r_K(x_n) = x_n$ and $x_n = (1-1/n)Tr_K(x_n) = (1-1/n)T(x_n)$. To sum up, we have proved that the equation $x_n = (1-1/n)T(x_n)$ has a solution for each $n \in \{1, 2, \ldots\}$.

**Step 3. A fixed point.** It remains to prove that the sequence $(x_n)_n$, where $x_n$ is a solution of equation (9), is convergent. Let

$$S_K = \{x_n \in C \cap K |\ x_n = (1-1/n)T(x_n),\ \forall\, n \in \mathbb{N}\} = S \cap K.$$

By Steps 1, 2, the set $S_K$ is a nonempty bounded set. Lemma 1 and (6) imply that the set $\overline{S}_K$ is compact, hence sequentially compact. Therefore we can extract a sequence converging to $x$. Finally, by continuity of $T$, $x$ is a fixed point of $T$. □

*Remark 9.* With the condition (6), Theorem 19 extends a result obtained in [3, Theorem 5.11] for a Hilbert space $X$ and a bounded subset $C \subset X$. Moreover (6) enables us to recover [3, Theorem 5.10] for nonexpansive mappings instead of strict $k$-set contractions.

A survey on boundary conditions for nonexpansive mappings is provided by Kirk in [72, Chap. 10].

# 3 1-set contractions

## 3.1 Measure of noncompactness and related mappings

**Definition 11.** Let $(X, d)$ be a metric space and $\mathsf{B} \subset \mathscr{P}(X)$ the set of all bounded subsets of $X$. For a subset $A \in \mathsf{B}$, define $\alpha(A) = \inf D$ where

$$D = \{\varepsilon > 0 : A = \cup_{i=1}^{i=n} A_i, \ \text{diam } (A_i) \leq \varepsilon, \ \forall\, i = 1, \ldots, n\}.$$

$\alpha$ is called the Kuratowski measure of noncompactness, $\alpha$-MNC for short (see, e.g., [10]).

It satisfies the following properties.

**Proposition 9.** *For any $A$, $B \in \mathsf{B}$, we have*

*(a)* $0 \leq \alpha(A) \leq diam(A)$.
*(b)* $A \subseteq B \Longrightarrow \alpha(A) \leq \alpha(B)$ *($\alpha$ is nondecreasing).*
*(c)* $\alpha(A \cup B) = \max(\alpha(A), \alpha(B))$.
*(d)* $\alpha(\overline{A}) = \alpha(A)$.
*(e)* $\alpha(A) = 0 \iff A$ *is relatively compact.*
*(f)* *If $(A_n)_{n\in\mathbb{N}}$ is a decreasing sequence of nonempty closed subsets such that $\lim\limits_{n\to\infty} \alpha(A_n) = 0$, then $A_\infty = \bigcap_{n\geq 0} A_n$ is a nonempty compact subset (Cantor's property).*

*If, further, $X$ is a Banach space, then*

*(g)* $\alpha(Conv\, A) = \alpha(A)$.
*(h)* $\alpha(A + B) \leq \alpha(A) + \alpha(B)$ *($\alpha$ is lower-additive).*
*(i)* $\alpha(\lambda A) = |\lambda|\alpha(A), \ \forall\, \lambda \in \Re$.

**Definition 12.** Let $X$, $Y$ be two Banach spaces and $T : X \longrightarrow Y$ a continuous mapping which maps bounded subsets of $X$ into bounded subsets of $Y$.

(a) $T$ is called a $k$-set contraction (or $k$-set contractive) (or $k$-$\alpha$-Lipschitz) if there exists some $k \geq 0$ such that $\alpha(T(A)) \leq k\alpha(A)$, for every bounded subset $A \subset X$. It is a 1-set contraction whenever $k = 1$.
(b) $T$ is a strict $k$-set contraction (or strict $\alpha$-contraction) when $k < 1$.
(c) $T$ is said to be $\alpha$-condensing if $\alpha(T(A)) < \alpha(A)$, for every bounded subset $A \subset X$ with $\alpha(A) \neq 0$.

*Example 7.* (a) LANE mappings are 1-set contractions [85, Lemma 1].
(b) The sum of a nonexpansive mapping and a compact one is a 1-set contraction.
(c) The sum of a contraction and a compact mapping is a strict $k$-set contraction, hence a condensing and a 1-set contraction.
(d) More generally, the sum of a nonlinear contraction and a compact mapping is $\alpha$-condensing (see, e.g., the proof of [88, Theorem 2.1]) hence a 1-set contraction.

Recall that $T : X \longrightarrow X$ is called nonlinear contraction if there exists a continuous nondecreasing function $\phi : \Re^+ \longrightarrow \Re^+$ such that $\phi(0) = 0$, $\phi(r) < r,\ \forall\, r > 0$, and

$$\|T(x) - T(y)\| \leq \phi(\|x - y\|),\ \forall\, (x, y) \in X^2.$$

*Example 8.* The ball retraction in a Banach space is a 1-set contraction $X$. More generally, if $C \subset X$ is a bounded closed convex subset containing the origin, then making use of the Minkowski functional $g$, one can define a retraction $r$ by $r(x) = \frac{x}{\max\{1, g(x)\}}$. By Proposition 8, we know that $r(x) = x$ for every $x \in C$ and that, since $C$ is convex, $g(x) \geq 1$ for every $x \notin C$ and thus $g(r(x)) = g\left(\frac{x}{g(x)}\right) = 1$ implies that $r(x) \in \partial C$, i.e. $r(X \setminus C) \subset \partial C$. Moreover, it is easy to see that, for every bounded subset $A \subset X$

$$r(A) \subset \overline{\text{co}}\,(A \cup \{0\}) \implies r(\alpha(A)) \leq \alpha(A),$$

proving our claim without appealing to Dugundgi's extension Theorem 9.

*Example 9.* In [48, Example 3.3], it is showed that, if $r$ is the unit ball retraction in an infinite Banach space $X$, then $T = -r$ is a 1-set contraction and $I - T$ is $\psi$-expansive.

Also, we know that if $T$ is expansive, i.e. $\|Tx - Ty\| \geq h\|x - y\|$, $(h \geq 1)$, for all $x, y \in C$, then $T^{-1}$ is $\frac{1}{h-1}$-Lipschitz and $I - T^{-1}$ is $\psi$-expansive with $\psi(r) = (1 - \frac{1}{h})r$.

## 3.2 *First results with boundary conditions*

In the fixed point theory for $\alpha$-mappings with respect to some measure of non-compactness $\alpha$, a classical result was obtained by Darbo in 1955 for strict $k$-set contraction self-mappings [27] defined on bounded closed convex subsets of a Banach space $X$ (see also [3, Theorem 4.16]).

**Theorem 21 ([27]).** *Let $C$ be a bounded closed convex subset of a Banach space $X$ and $T : C \longrightarrow C$ a strict $k$-set contraction mapping. Then $T$ has a fixed point.*

*Proof.* Define a decreasing sequence of nonempty closed convex $T$-invariant sets by $C_0 = C$ and $C_{n+1} = \overline{\text{Conv}}(T(C_n))$, for $n \geq 0$. Then

$$\alpha(C_{n+1}) = \alpha(T(C_n)) \leq k\alpha(C_n),$$

and then $\lim\limits_{n \to \infty} \alpha(C_n) = 0$. Property (f) of Proposition 9 guarantees that $C_\infty \neq \emptyset$, where $C_\infty = \bigcap_{n=0}^{\infty} C_n$ is also a closed convex $T$-invariant subset. The Schauder fixed point theorem then concludes the proof. □

For 1-set contractions, strict $k$-set contractions play the role of contractions for nonexpansive mappings, whence the importance of Theorem 21. By contrast to the Banach principle, this theorem does not provide uniqueness, but its proof is very elegant for it is a melting of the iterative method (the metric approach) and the Schauder fixed point theorem (the topological approach). In 1967, this theorem was extended to the class of condensing mappings by Sadovskii [99].

Starting from the seventieth, several authors have been involved in the extension of the Darbo and Sadovskii fixed point theorems to the broader class of 1-set contractions, which encompass nonexpansive, LANE, condensing mappings, and sums of contractive and compact mappings. Some of these fixed point theorems have been obtained as application of the theory of the fixed point index for 1-set contractions, which is an extension of the Nussbaum local degree for condensing mappings [86].

Moreover, several papers have been concerned with $k$-set contractions, condensing mappings, and 1-set contractions when the well-known Leray-Schauder boundary condition ($LS$) holds [78] (see also [102]). We start with a result proved by Nussbaum in 1969 (see [86, Proposition 4]):

**Theorem 22.** *Let $X$ be a Banach space, $C \subset X$ a bounded closed convex subset with nonempty interior, and $T : C \longrightarrow X$ a strict $k$-set contraction such that $T$ satisfies the Leray-Schauder condition ($LS$). Then $T$ has at least one fixed point in $C$.*

In 1968, Browder proved this theorem for nonexpansive mappings defined on bounded closed convex subsets of a uniformly convex space [19] (see Theorem 13). In 1972, Kirk and Schöneberg [71] extended this theorem to the class of pseudo-contractive mappings in uniformly convex spaces. This class of mappings is connected with accretive operators. Recall that an operator $A : X \longrightarrow X$ on a Banach space $X$ is said to be *accretive* (monotone in Hilbert spaces) if the inequality $\|x-y\| \le \|x-y+\lambda(Ax-Ay)\|$ holds for all $\lambda \ge 0$ and all $x, y \in D(A)$ ($I+\lambda A$ is injective and $(I+\lambda A)^{-1}$ is nonexpansive on its domain for all $\lambda \ge 0$). An operator $B$ is said to be *dissipative* if $-B$ is accretive. For an accretive operator $A : D(A) \longrightarrow X$ and $\lambda > 0$, denote by

$$J_\lambda^A = (I+\lambda A)^{-1} : \mathscr{R}(I+\lambda A) \longrightarrow D(A) \text{ and } A_\lambda = \frac{I - J_\lambda^A}{\lambda}$$

the resolvent and the Yoshida approximant of $A$, respectively. Hence $A$ is accretive if and only if $T = I - A$ is pseudo-contractive, i.e.

$$\|x-y\| \le \|(1+r)(x-y) + r(Ty-Tx)\|,$$

for every $x, y \in D(T)$ and all $r > 0$. The class of pseudo-contractive mappings encompasses the one of nonexpansive mappings; it was first introduced by Browder in 1967 [18]. Several refinements have been obtained so far; we quote, e.g., a recent work by C.H. Morales [84], and those in the references therein.

However, in the more general context of 1-set contraction mappings, we quote the following result proved by Petryshyn in 1971 (see [91, Theorem 7'] or [80, 85, 92, 93]):

**Theorem 23.** *Let $X$ be a Banach space, $U \subset X$ a bounded open subset, and $T : \overline{U} \longrightarrow X$ a 1-set contraction such that $T$ satisfies the Leray-Schauder boundary condition (LS). If $(I - T)(\overline{U})$ is closed, then $T$ has at least one fixed point in $\overline{U}$.*

When $C$ is a bounded closed convex subset of a Banach space, the following existence result involving the Furi-Pera condition was proved by O'Regan in [88, Theorem 2.9].

**Theorem 24.** *Let $X$ be a Banach space and $C \ni 0$ a bounded closed convex subset. Let $T : C \longrightarrow X$ be a 1-set contraction with $(I - T)(C)$ closed. If the Furi-Pera condition holds, then $T$ has a fixed point.*

It is worth noting that the existence of fixed points for a weakly inward semi-closed 1-set contraction $T : \Omega \longrightarrow X$ where $\Omega$ is a bounded open subset is obtained in [36, Theorem 2] when (LS) holds while the assumption that $T$ is weakly inward is not needed in [92, Theorem 7]. Finally notice that [100, Theorem 2] is exactly [91, Theorem 7'].

## *3.3 Historical review*

If $T : X \longrightarrow X$ is a compact linear operator, we know by the basic Riesz-Schauder theory that $I - T$ is closed. We now discuss the nonlinear case.

**(1)** The condition that $(I - T)(\overline{U})$ is closed, which has already been considered for nonexpansive mappings in Theorem 3 (see Subsection 2.5), is crucial for this class of mappings. We remind that

  **(a)** if $T$ is a $k$-contraction ($0 < k < 1$) (even when it is a nonlinear contraction), then the mapping $I - T$ is a homeomorphism on the range. Indeed, the mapping $I - T$ is continuous for $\|(I - T)x - (I - T)y\| \leq (1 + k)\|x - y\|$. Since

$$\|(I - T)x - (I - T)y\| \geq \|x - y\| - \|T(x) - T(y)\| \geq (1 - k)\|x - y\|,$$

  then the mapping $(I - T)^{-1}$ is continuous and thus $(I - T)$ is one-to-one.

  **(b)** If $D \subset X$ is a nonempty bounded closed subset and $T : D \longrightarrow X$ is continuous condensing (or a strict $k$-set contraction), then $I - T$ is proper (the preimage of a compact subset is also a compact subset); hence, $(I - T)(D)$ is closed (see, e.g., [85, Lemma 1], [98, Lemma 1], [104, Proposition 4.44]). Reich [98, Theorem C] proved a result similar to Theorem 23 for continuous condensing mappings with bounded range.

**(2)** Every injective operator $X \longrightarrow Y$ with continuous inverse and closed range is proper (compare with Proposition 5).

**(3)** The interest in the closedness of $(I - T)(D)$ lies in the procedure employed to deal with 1-set contractions, more precisely in the approximation methods (sequential approximation techniques) used to construct fixed points. Indeed, if $T$ is a 1-set contraction, then $t_nT$ is a strict $k$-set contraction where $t_n \in (0, 1)$ is any positive sequence increasing to unity. In 1973, Petryshyn [93, Theorem 1] observed that the following condition is necessary and sufficient for $T$ have a fixed point:

"for every $(x_n)_{n\in\mathbb{N}}$ with $x_n - T(x_n) \to 0$ as $n \to \infty$, there is some $x \in \overline{U}$ with $(I - T)x = 0$".

In [81, Theorem 2.1]), when this condition is verified by a 1-set contraction nonself mapping $T : K \longrightarrow X$, with $K$ a bounded closed convex subset, then without boundary conditions, the author proved that there exists some $x \in K$ such that

$$g(Tx - x) = \inf\{g(Tx - y),\ y \in K\},$$

where $g$ is the Minkowski functional. Then some fixed point theorems are derived in [81, Theorem 3.1].

**(4)** In [91, Theorem 7], the author even proved that the set of fixed points is compact if further $T$ is demi-compact, that is for every bounded sequence $(x_n)_{n\in\mathbb{N}}$ such that $(I - T)(x_n)$ is convergent in $X$, $(x_n)_{n\in\mathbb{N}}$ has a convergent subsequence. Notice also that the demi-compactness implies the demi-closedness provided $T$ is continuous and $C$ is a bounded closed subset of a Banach space (see [104, Proposition 11.14(3)]).

The topological degree for 1-set contractions, introduced by Petryshyn in [92], is also developed in [79] and [80] where Theorem 23 is recovered (see [79, Theorem 3] and [80, Corollary 1]).

In the recent literature, several fixed point theorems are proved for 1-set contractions with $I - T$ closed under different boundary conditions such as the inward condition or the Furi-Pera boundary condition, as introduced in [43] (see [32, Theorem 3.4] and [88, Theorem 2.9]). Some of them are presented in the next section.

## *3.4 Recent development*

Let $T : C \subset X \longrightarrow X$ be a mapping, where $C$ is a nonempty subset of a Banach space $X$. It is not difficult to see that the boundedness of $C$ can be replaced with the one of $T(\overline{C})$.

The following existence and uniqueness result was proved by Garcia-Falset in 2010 for nonexpansive mappings (see Theorem 7 and [45, Lemma 3.3]). When

$T : X \longrightarrow X$ and $\psi(s) = cs$, $c > 0$, this is nothing but [92, Theorem 8] (see Subsection 2.5 and Remark 10). It was recently extended independently by Garcia-Falset and Muñiz-Pérez [48, Theorem 3.2] and K. Hammache and the author to the larger class of 1-set contractions [33, Proposition 3.1].

**Theorem 25.** *Let $X$ be a Banach space, $C \ni 0$ a closed convex subset (not necessarily bounded), and $T : C \longrightarrow C$ a 1-set contraction satisfying the property $(\mathscr{K})$ and such that $I - T$ is $\psi$-expansive. Then $T$ has a unique fixed point in $C$.*

The existence part of this theorem has been recently extended by K. Hammache and the author to the broader class of mappings $T$ such that $I - T$ is $\alpha$-$\psi$-expansive according to the following sense:

**Definition 13.** $T$ is called $\alpha$-$\psi$-expansive if there exists a function $\psi : [0, \infty) \longrightarrow [0, \infty)$ with $\psi(0) = 0$, $\psi(r) > 0$, $\forall\, r > 0$ and such that for every bounded subset $\Omega \subset X$, $\alpha(T(\Omega)) \geq \psi(\alpha(\Omega))$.

Then we have (see [33, Proposition 3.3]). The proof is reproduced for the sake of completeness.

**Theorem 26.** *Let $X$ be a Banach space, $C \ni 0$ a convex closed subset of $X$, and $T : C \longrightarrow C$ a 1-set contraction satisfying the property $(\mathscr{K})$. If $I - T$ is $\alpha$-$\psi$-expansive, then $T$ has a fixed point in $C$.*

*Proof.* **Step 1.** Let $K$ be a bounded closed convex subset such that $C \cap K$ is self-mapped by $T$ and let $T_n = (1 - 1/n)T$, for $n \in \{1, 2, \ldots\}$ be a sequence of approximate mappings. Without loss of generality, assume that $0 \in K$. Since $T$ is a 1-set contraction, $T_n$ is a $(1 - 1/n)$-set contraction. By Darbo's fixed point theorem (Theorem 21), for every $n \in \{1, 2, \ldots\}$, $T_n$ has at least one solution $x_n \in C \cap K$. Let $S = \{x_n \mid n = 1, 2, \ldots\}$ be such a bounded sequence. To prove its relative compactness, we show that $\alpha(S) = 0$. First, for some given $n_0 \in \{1, 2, \ldots\}$, we have $\alpha(S) = \alpha(S_0)$, where $S_0 = \{x_n \mid n \geq n_0\}$). Arguing by contradiction, assume that $\alpha(S) > 0$ and let $x_p, x_q \in S_0$ $(x_p \neq x_q)$. Then

$$\|(I - T)x_p - (I - T)x_q\| = \left\|\frac{1}{p}Tx_p - \frac{1}{q}Tx_q\right\| \leq \frac{1}{n_0}(\|Tx_p\| + \|Tx_q\|).$$

Let $\gamma > 0$ be such that $K \subset B(0, \gamma/2)$. Then

$$\|(I - T)x_p - (I - T)x_q\| \leq \frac{\gamma}{n_0}.$$

$x_p, x_q$ being arbitrary in $S_0$, diam $(I - T)(S_0) \leq \frac{\gamma}{n_0}$. Since $(I - T)$ is $\alpha$-$\psi$-expansive, we get

$$\psi\,(\alpha(S_0)) \leq \alpha\,((I - T)(S_0)) \leq \frac{\gamma}{n_0}.$$

A contradiction is then reached by choosing $n_0 > \frac{\gamma}{\psi(\alpha(S))}$. As a consequence and taking if need be a subsequence, $(x_n)_{n\geq n_0}$ converges to some limit $x \in \overline{C} = C$.

□

The importance of the class of mappings covered by Theorem 26 is justified by the following result [33, Lemma 3.2].

**Proposition 10.** *Let a mapping $T : X \longrightarrow X$ be $\psi$-expansive and assume that $\psi : [0,\infty) \to [0,\infty)$ is either continuous or nondecreasing and invertible. Then $T$ is $\alpha$-$\psi$-expansive.*

*Proof.* Let $\varepsilon > 0$ be fixed and $\Omega \subset X$ a nonempty bounded subset. Then there exist bounded subsets $(Y_i)_{1\leq i\leq n}$ and $\varepsilon_0 > 0$ such that

$$T(\Omega) = \bigcup_{i=1}^{n} Y_i \text{ with } \operatorname{diam}(Y_i) \leq \varepsilon_0,\ \forall\, i \in \{1,\ldots,n\}$$

and $\alpha(T(\Omega)) \leq \varepsilon_0 < \alpha(f(\Omega)) + \varepsilon$; hence

$$\operatorname{diam}(Y_i) \leq \alpha(T(\Omega)) + \varepsilon,\ \forall\, i \in \{1,\ldots,n\}.$$

Moreover

$$\Omega \subset T^{-1}(T(\Omega)) \subset T^{-1}\left(\bigcup_{i=1}^{n} Y_i\right) = \bigcup_{i=1}^{n} T^{-1}(Y_i) = \bigcup_{i=1}^{n} \Omega_i. \tag{11}$$

For every $i \in \{1,\ldots,n\}$, let $x_1^i, x_2^i \in \Omega_i$; then, there exist $y_1^i, y_2^i \in Y_i$ such that $T(x_1^i) = y_1^i$ and $f(x_2^i) = y_2^i$. Since $T$ is $\psi$-expansive

$$\psi(\|x_1^i - x_2^i\|) \leq \|y_1^i - y_2^i\| \leq \operatorname{diam} Y_i \leq \alpha(T(\Omega) + \varepsilon).$$

Two cases need to be distinguished separately:

(a) *$\psi$ is invertible and nondecreasing* (and so is $\psi^{-1}$). Then, for all $i \in \{1,\ldots,n\}$

$$\|x_1^i - x_2^i\| \leq \operatorname{diam}(\Omega_i) \leq \psi^{-1}(\alpha(T(\Omega) + \varepsilon)).$$

(11) implies

$$\alpha(\Omega) \leq \max_{1\leq i\leq n} \alpha(\Omega_i) \leq \psi^{-1}(\alpha(T(\Omega) + \varepsilon)).$$

$\varepsilon > 0$ being arbitrary, we deduce that $\psi(\alpha(\Omega)) \leq \alpha(T(\Omega))$, for all bounded subsets $\Omega \subset X$, as claimed.

(b) *$\psi$ is continuous.* Hence $\alpha(\Omega) = \max\{\alpha(\Theta_i),\ i \in \{1,\ldots,n\}\} = \alpha(\Theta_l)$ for some $l \in [1,n]$. By the property of the least upper bound, there exist $y_l, z_l \in \Theta_l$ such that

$$\mathrm{diam}(\Theta_l) - \varepsilon \le \|y_l - z_l\| \le \mathrm{diam}(\Theta_l).$$

Hence

$$\alpha(\Omega) - \varepsilon = \alpha(\Theta_l) - \varepsilon \le \mathrm{diam}(\Theta_l) - \varepsilon \le \|y_l - z_l\| \le \mathrm{diam}(\Theta_l) \le \alpha(\Omega) + \varepsilon.$$

This implies that

$$|\|y_l - z_l\| - \alpha(\Omega)| \le \varepsilon. \tag{12}$$

Let $y_1^l = f(y_l)$ and $y_2^l = f(z_l)$. Since $T$ is $\psi$-expansive, we obtain as in case (a):

$$\psi(\|y_l - z_l\|) \le \|y_1^l - y_2^l\| \le \mathrm{diam}\, Y_l \le \alpha(T(\Omega)) + \varepsilon. \tag{13}$$

$\psi$ being continuous, for all positive $\eta$, there exists $\delta > 0$ such that

$$\forall\, r > 0,\ (|r - \alpha(\Omega)| \le \delta \implies |\psi(r) - \psi(\alpha(S))| \le \eta)\,.$$

Taking $\eta > 0$, choosing $0 < \varepsilon \le \delta$ and using (12), (13), we finally get

$$|\psi(\alpha(\Omega)) - \psi(\|y_l - z_l\|)| \le \eta$$

and

$$0 < \psi(\alpha(\Omega)) \le \psi(\|y_l - z_l\|)| + \eta \le \alpha(T(\Omega)) + \eta + \varepsilon.$$

Since $\eta$ and $\varepsilon > 0$ are arbitrary constants, we conclude that $\psi(\alpha(\Omega)) \le \alpha(T(\Omega))$. The proof of the lemma is completed. □

When $T$ does not self-map $C$, we still have [34]:

**Theorem 27.** *Let $X$ be a Banach space, $C \subset X$ a closed convex subset, and $U \subset C$ an open subset with $p \in U$. Suppose that $T : \overline{U} \longrightarrow C$ is a 1-set contraction satisfying the property $(\mathscr{K})$ and such that $I - T$ is $\alpha$-$\psi$-expansive. If $(LS)$ is satisfied, then $T$ has a fixed point in $C$.*

A result obtained by J. Garcia-Falset and O. Muñiz-Pérez in 2013 (see [48, Theorem 3.2]) is now deduced.

**Corollary 7.** *Let $X$ be a Banach space, $C \subset X$ a bounded closed convex subset, and $U \subset C$ an open subset with $p \in U$. Suppose that $T : \overline{U} \longrightarrow C$ is a 1-set contraction satisfying the property $(\mathscr{K})$ and such that $I - T$ is $\psi$-expansive. If $(LS)$ is satisfied, then $T$ has a unique fixed point in $C$.*

*Remark 10.* In case of a 1-set contraction $T : X \longrightarrow X$ satisfying

$$\|(I-T)(x)-(I-T)(y)\| \geq c\|x-y\|, \ \forall\, x, y \in X,$$

for some positive constant $c$, the existence result in Corollary 7 is announced in [92, Theorem 8]. Moreover, Petryshyn and Nussbaum (see [92, Remark 6]) claimed the validity of the result whenever $I-T$ is $\psi$-expansive with $\psi$ satisfying $\psi(r) > 0$ for $r > 0$ and $\psi(r_n) \to \infty$ implies $r_n \to \infty$, as $n \to \infty$.

From Corollary 7, we can easily derive the following nonlinear alternative:

**Corollary 8.** *Let $X$ be a Banach space, $C \subset X$ a closed convex subset, and $U \subset C$ an open subset with $p \in U$. Suppose that $T : \overline{U} \longrightarrow C$ is a 1-set contraction satisfying the property $(\mathscr{K})$ and such that $I-T$ is $\alpha$-$\psi$-expansive. Then either*

*(a) $f$ has a fixed point in $C$, or*
*(b) there exist $x \in \partial U$ and $t \in (0,1)$ such that $x = tT(x) + (1-t)p$.*

The next result is concerned with the interior condition. For the proof, a Mönch type existence theorem is needed:

**Lemma 2 ([3, Theorem 5.5]).** *Let $X$ be a Banach space, $C \subset X$ a closed convex subset, and $U \subset C$ an open subset with $p \in U$. Suppose that $T : \overline{U} \longrightarrow C$ is a continuous mapping satisfying the so-called Mönch condition:*

$$(\mathscr{M}) \quad \big(D \subset \overline{U}, D \text{ countable}, D \subset \overline{co}(\{0\} \cup T(D))\big) \Longrightarrow \overline{D} \text{ compact}$$

*both with the boundary condition $(LS)$. Then $T$ has a fixed point in $\overline{U}$.*

**Theorem 28.** *Let $X$ be a Banach space and $C \ni 0$ a closed convex subset of $X$. Suppose that $T : C \longrightarrow C$ is a 1-set contraction and $U$ is a bounded strictly star-shaped open neighborhood of the origin such that*

$$(IC)' \quad \text{for all } \lambda > 1 \text{ and all } x \in U_\delta \cap C,\ T(x) \neq \lambda x \text{ and } T(x) \notin \overline{U} \cap C,$$

*with* $0 < \delta \leq \operatorname{dist}\{0, \partial U\}$.

*If further $I-T$ is $\alpha$-$\psi$-expansive, then $T$ has a fixed point in $C$.*

*Proof.* The mapping $T_n \equiv (1-\frac{1}{n})T : C \longrightarrow C$ is a $(1-\frac{1}{n})$-set contraction, thus satisfies the Mönch condition. Define the function $P : C \longrightarrow \overline{U} \cap C$ by

$$Px = \begin{cases} \frac{r+(1-r)g(x)}{g^2(x)}\, x, & \text{if } x \in C \setminus \overline{U} \cap C, \\ x, & \text{if } x \in \overline{U} \cap C, \end{cases}$$

where $g$ is the Minkowski functional and $r$ is as defined in Proposition 7. Let $P \circ T_n : \overline{U} \cap C \longrightarrow \overline{U} \cap C$. Then $\overline{U} \cap C$ is strictly star-shaped for $C$ is convex and, by definition of $P$, $P \circ T_n$ satisfies $(LS)$. If we assume on the contrary that there exist some $x_0 \in \partial(\overline{U} \cap C)$ and $\lambda_0 > 1$ such that $(P \circ T_n)(x_0) = \lambda x_0$, then $g(x_0) = 1$

and $g(\lambda x_0) = \lambda_0 g(x_0) = \lambda_0 > 1$ which implies that $\lambda x_0 \notin \overline{U} \cap C$, leading to a contradiction with $(P \circ T_n)(\overline{U} \cap C) \subset \overline{U} \cap C$. In fact, we have proved that if $\Omega$ is strictly star-shaped and $T(\partial \Omega) \subset \overline{\Omega}$, then $(LS)$ holds.

Now $P \circ T_n$ is further a $(1 - \frac{1}{n})$-set contraction. Indeed, notice that if $A$ is a nonempty bounded subset of $X$, then from Proposition 7, we have $PA \subset \overline{co}(\{0\} \cup A)$. As a consequence, if $A$ is any bounded subset of $X$, then

$$\alpha((P \circ T_n)(A)) \leq \alpha(T_n(A)) = \left(1 - \frac{1}{n}\right) \alpha(T(A)) \leq \left(1 - \frac{1}{n}\right) \alpha(A).$$

By Lemma 2, $P \circ T_n$ has a fixed point $x_n$ for each $n \in \{1, 2, \ldots\}$, i.e., $(P \circ T_n)(x_n) = x_n$. Now, we show that $T_n$ has a fixed point. According to the definition of the generalized projection mapping $P$, we must discuss separately two cases:

[i) If $T_n(x_n) \in \overline{U} \cap C$, then Proposition 7 guarantees that $T_n(x_n) = (P \circ T_n)(x_n) = x_n$.

[ii) Otherwise $T_n(x_n) \notin \overline{U} \cap C$ implies, by Proposition 7, $g(T_n(x_n)) > 1$. Hence $g((1 - 1/n)Tx_n) > 1$ and $g(Tx_n) > 1$, showing that $Tx_n \notin \overline{U} \cap C$ which is a contradiction with $(IC)'$.

Therefore there exists $x_n \in \overline{U} \cap C$ such that $x_n = T_n(x_n)$, for all $n \in \{1, 2, \ldots\}$. Let $S = \{x_n \subset \overline{U} | \ T_n(x_n) = x_n, \ n = 1, 2, \ldots\}$ be such a bounded sequence of approximate fixed points. Arguing as in the proof of Theorem 27, we obtain that $\alpha(S) = 0$, showing that $\overline{S}$ is compact. Therefore $(x_n)$ converges to $x$ in $\overline{U} \cap C$ and since $\overline{U} \cap C$ is closed, then $x$ lies in $\overline{U} \cap C$. Finally, $T$ being continuous, we conclude that $T$ has a fixed point in $\overline{U} \cap C$. □

We now state a final existence result in which again $C$ is an arbitrary closed convex subset while $(I - T)(C)$ closed is replaced with $I - T$ demi-closed (for the proof, we refer to [32, Theorem 3.4]).

**Theorem 29.** *Let $(E, \| \cdot \|)$ be a reflexive Banach space satisfying the NRP and $C \ni 0$ a convex closed subset of $E$. Let $T : C \longrightarrow E$ be a 1-set contraction with $I - T$ demi-closed. If the Furi-Pera condition and the property $(\mathscr{K})$ are fulfilled, then $T$ has a fixed point.*

# 4 The weak topology

## *4.1 Introduction*

Let $X$ be a Banach space and $X^*$ its topological dual. The weak topology is generated by the bounded linear functionals on $X$. $\Omega \subset X$ is open in the weak topology if for every $x \in \Omega$, there are bounded linear functions $\{f_i\}, \ i \in \mathbb{N}$ and positive real numbers $\{\varepsilon_i\}, \ i \in \mathbb{N}$ such that

$$\{y \in X : |f_i(x) - f_i(y)| < \varepsilon_i, \ i = 1, 2, \dots, n\} \subset \Omega.$$

Then, instead of balls centered at the origin in case of the strong topology, a base of neighborhoods of the origin is given, in the weak topology, by

$$V(f_1, \dots, f_n; \varepsilon_i\} = \{x \in X : | < x, f_i > | < \varepsilon_i, \ i = 1, 2, \dots, n\}.$$

So by the weak convergence of a sequence $x_n$ to a limit $x$ (in the $\sigma(X, X^*)$ topology), it is meant that $< x_n, f > \longrightarrow < x, f >$, for all $f \in X^*$. We shall write $x_n \rightharpoonup x$, as $n \to +\infty$. We recall that the weak topology is Hausdorff locally convex, hence can be defined by the family of semi-norms:

$$\{\rho_f(x) = | < x, f > | : f \in X^* \text{ and } \|f\|_{f^*} \leq 1\}.$$

However this topology is not metrizable (see, e.g., [38, 41]).

An operator $T : X \to X$ is said to be weakly continuous if for every $f \in X^*$, the mapping $f \circ T : X \to \Re$ is continuous.

$T$ is weakly sequentially continuous if for each sequence $x_n \in X$, we have $Tx_n \rightharpoonup Tx$ whenever $x_n \rightharpoonup x$. $T$ is weakly strongly continuous if $Tx_n \longrightarrow Tx$ whenever $x_n \rightharpoonup x$ ($T$ is a Dunford-Pettis operator). Then $T$ weakly strongly continuous implies $T$ weakly sequentially continuous which in turn implies that $T$ is demi-closed.

A subset $C \subset X$ is said to be weakly closed (resp. compact) if it is closed (resp. compact) in the weak topology $\sigma(X, X^*)$. The following results are standard results in functional analysis (see, e.g., [14, 38, 41, 83]).

**Theorem 30.** *A convex subset of a normed space is closed if and only if it is weakly closed.*

**Theorem 31.** *A subset of a reflexive Banach space is weakly compact if and only if it is closed in the weak topology and bounded in the norm topology.*

**Theorem 32.** *Let X be a Banach space. Then the following are equivalent:*

*(a) X is reflexive.*
*(b) The unit ball $B_X$ is weakly compact.*
*(c) Every bounded sequence in X (in the strong topology) has a weakly convergent subsequence.*
*(d) $X^*$ is reflexive.*

**Theorem 33 (Eberlein-Šmulian Theorem).** *Let K be a weakly closed subset of a Banach space X. Then the following are equivalent:*

*(a) K is weakly compact.*
*(b) K is weakly sequentially compact.*

## 4.2 The weak MNC

Let $B(X)$ be the collection of all nonempty bounded subsets of a Banach space $X$ and $W(X)$ the subset of $B(X)$ consisting of all weakly compact subsets of $X$. Let $B_r$ denote the closed ball in $X$ centered at 0 with radius $r > 0$. In [28], De Blasi introduced the mapping $\omega : B(X) \to [0, +\infty)$ defined, for all $M \in B(X)$ by

$$\omega(M) = \inf\{r > 0,\ \exists N \in W(X) :\ M \subseteq N + B_r\}.$$

For the sake of completeness, we recall some important properties of $\omega$ we need in this section; for further details and proofs, we refer the reader to [28].

**Lemma 3.** *Let $M_1, M_2 \in B(X)$. Then*

(a) $\omega(M_1) \leq \omega(M_2)$ *whenever* $M_1 \subseteq M_2$.
(b) $\omega(M) = 0$ *if and only if $M$ is relatively weakly compact.*
(c) $\omega(\overline{M}^w) = \omega(M)$ *where* $\overline{M}^w$ *is the weak closure of $M$.*
(d) $\omega(co(M)) = \omega(M)$ *where $co(M)$ refers to the convex hull of $M$.*
(e) $\omega(M_1 + M_2) \leq \omega(M_1) + \omega(M_2)$.
(f) $\omega(M_1 \cup M_2) = \max(\omega(M_1), \omega(M_2))$.
(g) *(Cantor intersection condition) If $\{X_n\}_1^\infty$ is a decreasing sequence of nonempty bounded weakly closed subsets of $E$ with $\lim_{n\to\infty} \omega(X_n) = 0$, then the set $\bigcap_{n=1}^\infty X_n$ is nonempty and weakly compact.*

The mapping $\omega$ is called the De Blasi measure of weak non-compactness. In [8], Appel and De Pascale gave to $\omega$ the following simple form in $L^1$ spaces, also called measure of nonequiabsolute continuity:

$$\omega(M) = \limsup_{\varepsilon\to 0} \left\{ \sup_{\psi\in M} \left[ \int_D \|\psi(t)\|_X \, dt,\ D \subset \Omega,\ meas(D) \leq \varepsilon \right] \right\}, \tag{14}$$

for all bounded $M \subset L^1(\Omega, X)$, where $X$ is a finite dimensional Banach space, $\Omega \subset \Re^n$ and $meas(.)$ denotes the Lebesgue measure. This formula is usually employed to prove the relative weak-compactness of the set $M$. We point out however that, for this purpose the relation (14) can be dispensed. Indeed, by the Dunford-Pettis compactness criterion in $L^1$ spaces (see, e.g., [14, Theorem 4.3.] or [38, Corollary 11]), it can be directly proved that

$$\limsup_{\varepsilon\to 0} \left\{ \sup_{\psi\in M} \left[ \int_D \|\psi(t)\|_X \, dt,\ D \subset \Omega,\ meas(D) \leq \varepsilon \right] \right\} = 0$$

implies that $M$ is weakly relatively compact, i.e., $\omega(M) = 0$.

**Definition 14.** (a) A mapping $T : M \subset X \to X$ is said to be $\omega$-$k$-set-contraction if it maps bounded sets into bounded sets, and there exists some $k > 0$ such that $\omega(T(V)) \leq k\omega(V)$ for all bounded subsets $V \subseteq M$. It is a strict $\omega$-set-contraction whenever it is a $\omega$-$k$-set-contraction with $0 \leq k < 1$ and $\omega$-1-set-contraction if $k = 1$.

(b) $T$ is said to be $\omega$-condensing if its maps bounded sets into bounded sets and for all $V \in B(M)$ we have $\omega(T(V)) < \omega(V)$ whenever $\omega(V) > 0$.

(c) $T$ is said to satisfy condition ($\mathscr{A}1$) if $Tx_n)_{n\in\mathbb{N}}$ has a strongly convergent subsequence in $X$ whenever $(x_n)_{n\in\mathbb{N}}$ is a weakly convergent sequence in $X$ ($T$ is weakly strongly compact, ws-compact for short).

(d) $T$ is said to satisfy condition ($\mathscr{A}2$) if $Tx_n)_{n\in\mathbb{N}}$ has a weakly convergent subsequence in $X$ whenever $(x_n)_{n\in\mathbb{N}}$ is a weakly convergent sequence in $X$ ($T$ is weakly compact, ww-compact for short).

## *4.3 Fixed point theorems*

We begin with an existence result which appears in [96, Theorem 2.2] for weakly sequentially continuous $\omega$-$k$-set contraction mappings, in [42] for weakly continuous $\omega$-condensing mappings and in [12, Theorem 3.2], [46, Lemma 3.2] for weakly sequentially continuous $\omega$-condensing mappings.

**Theorem 34.** *Let $C$ be a nonempty bounded closed convex subset of a Banach space $X$ and let $T : C \longrightarrow C$ be either a weakly continuous or a weakly sequentially continuous mapping. If $T$ is $\omega$-condensing (or $\omega$-$k$-set contractive), then it has a fixed point in $C$.*

This result can be easily derived from Arino, Gautier, and Penot fixed point theorem:

**Theorem 35 ([9, Theorem 1]).** *Let $E$ be a metrizable locally convex topological vector space and let $C$ be a weakly compact convex subset of $E$. Then every weakly sequentially continuous mapping $f : C \longrightarrow C$ has a fixed point.*

This theorem in turn follows from the Schauder-Tychonoff theorem. Now, as a consequence of Theorem 34, we get

**Theorem 36.** *Let $X$ be a Banach space, $C \subset X$ a nonempty bounded closed convex subset, and $T : C \longrightarrow X$ an $\omega$-1-set-contraction which is either weakly continuous or weakly sequentially continuous. If $(I - T)(C)$ is closed, then $T$ has a fixed point in $C$.*

*Proof.* Arguing as in Theorem 4 for nonexpansive mappings, and appealing to Theorem 34, we obtain that $T$ has a sequence of fixed points $(x_n)_{n\in\mathbb{N}}$; indeed, it is sufficient to consider an $\omega$-$(1-1/n)$-set-contraction mapping $T_n(x) = (1-\frac{1}{n})T+\frac{1}{n}x_0$ for some $x_0 \in C$; this is a convex homotopy between $T$ and the constant $x_0$; hence, they have the same fixed point index. Since $C$ is bounded, so is $T(C)$ and then $\lim_{n\to\infty}(x_n - Tx_n) = 0$. Finally $(I - T)(C)$ is closed, which completes the proof. □

The result still holds if $C$ is weakly compact and $I - T$ is demi-closed. Owing to [44, Theorem 8], the result of Theorem 36 remains true if the weak (resp. the weak sequential) continuity is replaced with $T$ continuous and satisfies $(\mathscr{A}1)$. Next, we recall some recent results without proofs.

**Theorem 37 ([48, Theorem 2.3]).** *Let $C$ be a nonempty closed convex subset of a Banach space $X$ and suppose that $T : C \longrightarrow C$ is an $\omega$-condensing mapping satisfying $(\mathscr{A}1)$. If there exists $x_0 \in C$ and $R > 0$ such that $T(x) - x_0 \neq \lambda(x - x_0)$ for every $\lambda > 1$ and for every $x \in C \cap S_R(x_0)$, then $T$ has a fixed point.*

**Theorem 38 ([48, Theorem 3.1]).** *Let $X$ be a Banach space and $C \ni 0$ be a closed convex of $X$. Assume that $T : C \longrightarrow C$ is a continuous mapping which satisfies $(\mathscr{A}1)$ and*

*(a) $T$ is an $\omega$-1-set contraction.*
*(b) $T$ satisfies $(LS)$.*
*(c) $I - T$ is $\psi$-expansive. Then $T$ has a unique fixed point in $C$.*

**Proposition 11 ([48, Corollary 3.2]).** *Let $C$ be a bounded closed subset of a Banach space $X$ and $T : C \longrightarrow C$ a continuous mapping such that $I - T : C \longrightarrow X$ is $\psi$-expansive. If there exists an almost fixed point sequence $(x_n)_n$ of $T$ in $C$, then $T$ has a unique fixed point in $C$.*

This result is interesting for we know that a sequence of fixed points can be obtained when, e.g., $T$ is nonexpansive as in Theorem 17. Next, we present a nonlinear alternative in the weak topology.

**Theorem 39 ([35, Theorem 2.3]).** *Let $C$ be a nonempty bounded closed convex subset of a Banach space $X$ and $U \subset C$ an open subset containing some $x_0$. Let $T : \overline{U} \to C$ be a continuous $\omega$-contraction satisfying $(\mathscr{A}1)$. Then*

*(a) either the equation $Tu = u$ has a solution in $\overline{U}$, or*
*(b) there exists an element $u \in \partial U$ such that $u = \lambda Tu + (1 - \lambda)x_0$ for some $\lambda \in (0, 1)$.*

Our final existence result deals with $\omega$-1-set contractions defined on strictly star-shaped sets and satisfying the interior condition. We will need

**Lemma 4 ([48, Corollary 3.2]).** *Let $C$ be a bounded closed subset of a Banach space $X$ and $T : C \longrightarrow C$ a continuous mapping such that $I - T : C \longrightarrow X$ is $\psi$-expansive. If there exists an almost fixed point sequence $(x_n)_n$ of $T$ in $C$, then $T$ has a unique fixed point $x_0$ in $C$.*

**Theorem 40.** *Let $X$ be a Banach space and $C \ni 0$ a closed convex subset of $X$. Assume that $T : C \longrightarrow C$ is a continuous mapping satisfying $(\mathscr{A}1)$ and such that*

*(a) $T$ is an $\omega$-1-set contraction.*
*(b) There exists a bounded strictly star-shaped open neighborhood of the origin $U$ such that for all $\lambda > 1$ and $x \in U_\delta \cap C$ $(0 < \delta \leq \text{dist}\{0, \partial U\})$, we have that $T(x) \neq \lambda x$ and $T(x) \notin \overline{U} \cap C$.*

*If further $I - T$ is $\psi$-expansive, then $T$ has a unique fixed point in $C$.*

*Proof.* Arguing as in the proof of Theorem 28, define $P \circ T_n : \overline{U} \cap C \longrightarrow \overline{U} \cap C$. Since $P \circ T_n$ is a $(1 - \frac{1}{n})$-$\omega$-set contraction, Theorem 39 guarantees that $P \circ T_n$ has a fixed point. We prove that $(x_n)_n$ is an almost fixed point sequence. We have $T_n(x_n) = (1 - \frac{1}{n})T(x_n)$, for all $n \in \{1, 2, \ldots\}$. Then

$$\|T(x_n)\| \leq \|T(x_n) - T_n(x_n)\| + \|T_n(x_n)\|.$$

Since $\|T_n(x_n)\| = \|x_n\| \leq K$, where $K = \sup\{\|x\|, \; x \in \partial U\}$, then $\|T(x_n)\| \leq K + \frac{1}{n}\|T(x_n)\|$. Therefore $\|T(x_n)\| \leq \frac{n}{n-1}K$ and

$$\|x_n - T(x_n)\| = \frac{1}{n}\|T(x_n)\| \leq \frac{1}{n}\frac{n}{n-1}K = \frac{K}{n-1}.$$

Finally $(x_n)_n$ is a bounded almost fixed point sequence for $T$. By Lemma 4, $T$ has a unique fixed point. □

In the next section, further fixed point theorems are proved for the sum of operators including recent results both in the weak and the strong topologies.

# 5 Sum of operators

## *5.1 Introduction*

Many problems in physical sciences are modeled by equations of the form:

$$Ax + Bx = x, \quad x \in Q, \tag{15}$$

where $Q$ is a closed convex subset of a Banach space $X$ and $A, B$ are two nonlinear operators, one of which can be of integral type. A classical and useful tool to solve Problem (15) is the following celebrated fixed point theorem proved by Krasnozels'kĭi in 1958 (see [73]).

**Theorem 41.** *Let $Q$ be a nonempty closed convex subset of a Banach space $X$ and $A, B$ be two mappings from $Q$ to $X$ such that*

*(a) $A$ is compact and continuous.*
*(b) $B$ is a contraction.*
*(c) $AQ + BQ \subset Q$.*
*Then $A + B$ has at least one fixed point in $Q$.*

The proof of Theorem 41 combines the Banach contraction principle and the Schauder fixed point theorem [104] and uses the fact that if $T\colon D \longrightarrow X$ is a contraction, where $D$ is a subset of a Banach space $X$, then the mapping $I - T\colon D \longrightarrow (I - T)(D)$ is a homeomorphism (see Subsect. 3.3). Let $y \in Q$ be fixed. From the Banach fixed point theorem, the mapping $\varphi\colon Q \longrightarrow Q$ defined by $\varphi(x) = Bx + Ay$ has a unique fixed point in $Q$. Thus $(I - B)^{-1} \circ A(y) \in Q$, that is $F = (I - B)^{-1} \circ A$

maps $Q$ into $Q$. Since $F$ is continuous, compact, and sends $Q$ into itself, the Schauder fixed point theorem guarantees that $F$ has a fixed point in $Q$.

In 1998, Burton [23] observed that the Krasnozels'kĭi fixed point theorem remains true if condition (c) is replaced with the weaker one:

$$\forall\, y \in M, \quad (x = Ay + Bx) \implies x \in M.$$

However, under the weak topology, the first known result is due to Edmunds and Zabreiko *et al.*, 1967:

**Theorem 42 (see [40, Theorem 1], [103, Theorem]).**
*Let $Q$ be a nonempty closed convex subset of a Hilbert space $X$ and $A, B$ be two mappings from $Q$ to $X$ such that*

*(a) $A$ is weakly strongly continuous.*
*(b) $B$ is nonexpansive.*
*(c) $AQ + BQ \subset Q$.*
*Then $A + B$ has at least one fixed point in $Q$.*

## 5.2 Recent contributions

The following result is of Krasnosel'skii type (see [74]) and it concerns the sum of a compact and a nonexpansive mappings. Condition (c) replaces the demi-closedness of $I - B$.

**Theorem 43 ([1, Theorem 2.19]).** *Let $X$ be a Banach space and $Q \subset X$ a nonempty bounded closed convex subset. Let $A, B : Q \longrightarrow X$ be two operators such that*

*(a) $A$ is completely continuous.*
*(b) $B$ is nonexpansive.*
*(c) If $(x_n)$ is a sequence of $Q$ such that $(I - B)x_n$ is strongly convergent, then $(x_n)$ has a strongly convergent subsequence.*
*(d) $Ax + By \in Q$, for all $x, y \in Q$.*
*Then $A + B$ has a fixed point $x \in Q$.*

In case where the mapping $T$ is a 1-set-contraction, we have (see [33, Theorem 3.7]):

**Theorem 44.** *Let $X$ be a Banach space and $Q \subset X$ a nonempty bounded closed convex subset. Let $A, B : Q \longrightarrow X$ be two operators such that*

*(a) $A$ is completely continuous.*
*(b) $B$ is a 1-set contraction and $I - B$ is $\psi$-expansive.*
*(c) $x \in Q$ whenever $x = Bx + Ay \in Q$, for some $y \in Q$.*
*Then $A + B$ has a fixed point $x \in Q$.*

*Proof.* For fixed $x \in Q$, let $A_x : Q \longrightarrow Q$ be the mapping defined by $A_x(y) = Ax + By$. Since $B$ is a 1-set contraction, $A_x$ is a 1-set contraction too. Since $I - B$ is $\psi$-expansive, then $I - A_x$ is so. By Theorem 25, $A_x$ admits a unique fixed point $z$ in $Q$, i.e., $z = Ax + Bz$. Consequently $Ax = (I - B)z$ and thus the mapping $(I - B) : Q \longrightarrow (I - B)(Q)$ is bijective; indeed, by definition it is surjective and since $(I - B)$ is $\psi$-expansive, then it is one-to-one. Let $J^B = (I - B)^{-1} : (I - B)(Q) \longrightarrow Q$. Then $J^B Ax = z \in Q$. By Assumption $(c)$, $T = J^B \circ A$ maps $Q$ into itself. According to the proof of [45, Theorem 3.7], we know that $J^B$ is continuous and since $A$ is completely continuous, we deduce that $T$ is completely continuous too. By the Schauder fixed point theorem, there exists $x \in Q$ such that $Tx = x$. Hence $x = Bx + Ax$, proving the theorem. □

As a consequence, we recover a result due to Garcia-Falset [45, Theorem 3.7]:

**Corollary 9.** *Let $X$ be a Banach space and $Q \subset$ a nonempty closed bounded convex subset. Let $A, B : X \longrightarrow X$ be two operators such that*

*(a) $A$ is completely continuous.*
*(b) $B$ is nonexpansive and $I - B$ is $\psi$-expansive.*
*(c) $x, y \in Q \Longrightarrow Bx + Ay \in Q$.*
*Then there exists $x \in Q$ such that $x = Ax + Bx$.*

Now, we derive two direct consequences from Theorem 26:

**Corollary 10.** *Let $X$ be a Banach space and $Q \subset X$ be a nonempty bounded closed convex subset. Let $A, B : Q \longrightarrow X$ be two operators such that*

*(a) $A$ is completely continuous.*
*(b) $B$ is a 1-set contraction and $I - (A + B)$ is $\alpha$-$\psi$-expansive.*
*(c) $x, y \in Q \Longrightarrow Bx + Ay \in Q$.*
*Then $A + B$ has a fixed point $x \in Q$.*

*Proof.* By condition (c), $A + B$ maps $Q$ into itself. Since $A$ is completely continuous and $B$ is a 1-set contraction, $A + B$ is a 1-set contraction too. Moreover $I - (A + B)$ is $\alpha$-$\psi$-expansive. By Theorem 26, $A + B$ has a fixed point in $Q$. □

**Corollary 11.** *Let $X$ be a Banach space and $Q \subset X$ a nonempty bounded closed convex subset. Let $A, B : Q \longrightarrow X$ be two operators such that*

*(a) $A$ is a 1-set contraction.*
*(b) $B$ is dissipative and $I - (J_1^{-B} \circ A)$ is $\alpha$-$\psi$-expansive.*
*(c) $x, y \in Q \Longrightarrow Bx + Ay \in Q$.*
*Then $A + B$ has a fixed point $x \in Q$.*

*Proof.* Since $B$ is dissipative, then $J^{-B}$ is nonexpansive and so $J_1^{-B} \circ A$ is a 1-set contraction. Moreover $I - (J_1^{-B} \circ A)$ is $\alpha$-$\psi$-expansive. By Theorem 25, $J_1^{-B} \circ A$ has a fixed point and so $A + B$ has a fixed point. □

Our second existence result is (see [33, Theorem 3.9])

**Theorem 45.** *Let $X$ be a Banach space and $Q \subset X$ a nonempty bounded closed convex subset. Let $A, B : Q \longrightarrow X$ be two mappings such that*

*(a) $B$ is continuous, $J_1^{-B}$ exists, and $I - B$ is $\alpha$-$\psi$-expansive.*
*(b) $A$ is completely continuous.*
*(c) $x, y \in Q \Longrightarrow Bx + Ay \in Q$.*
*Then $A + B$ has a fixed point $x \in Q$.*

*Proof.* Since $J_1^{-B}$ exists and $A$ is completely continuous, $J_1^{-A} \circ B$ is completely continuous. Indeed $(I - B) \circ J_1^{-B} \circ A(Q) = A(Q)$ and thus

$$\alpha\left((I - B) \circ J_1^{-B} \circ A(Q)\right) = \alpha(A(Q)).$$

Now, since $I - B$ is $\alpha$-$\psi$-expansive and $A$ is completely continuous, we have $\alpha(A(Q)) = 0$ and then $\psi\left(\alpha\left(J_1^{-B} \circ A(Q)\right)\right) \leq 0$ which implies that

$$\psi\left(\alpha\left(J_1^{-B} \circ A(Q)\right)\right) = 0.$$

Since $\psi(0) = 0$, then $\alpha\left(J_1^{-B} \circ A(Q)\right) = 0$ and thus the mapping $(J_1^{-B} \circ A)(Q)$ is completely continuous. By the Schauder fixed point theorem, $A+B$ has a fixed point in $Q$. □

The following nonlinear alternative was recently proved by Garcia-Falset and Muñiz-Pérez in [48, Theorem 4.2]. Condition (c) is a substitute to the closedness of the range $\mathscr{R}(I - B)$.

**Theorem 46.** *Let $Q$ be a closed convex subset of a Banach space $X$ and $A : Q \longrightarrow X$, $B : X \longrightarrow X$ two continuous mappings such that*

*(a) $A$ is completely continuous.*
*(b) $B$ is a 1-set contraction.*
*(c) $(I - B)^{-1} : \mathscr{R}(I - B) \longrightarrow X$ exists and is uniformly continuous.*
*(d) If $x = Ay + Bx$ for some $y \in Q$, then $x \in Q$.*

*Then either*

*(i) the equation $x = Ax + Bx$ has a solution, or*
*(ii) the set $\{x \in Q|\ x = \lambda B(x/\lambda) + \lambda A(x),\ \lambda \in (0, 1)\}$ is unbounded.*

We close this subsection with some Krasnosels'kii fixed point theorems under the weak topology. First a nonlinear alternative due to O'Regan (see [88, Theorems 2.4, 2.5]) is presented:

**Theorem 47.** *Let $\Omega$ be an open subset of a Banach space $X$ and $A : \overline{\Omega} \longrightarrow X$, $B : X \longrightarrow X$ two continuous mappings such that*

*(a) $A$ is completely continuous.*
*(b) $B$ is nonexpansive.*
*(c) Either $X$ is uniformly convex and $A$ is strongly continuous, or $\overline{\Omega}$ is weakly compact and $I - F$ is demi-closed.*

*Then at least one of the following properties holds:*

*(i)* *the equation $x = Ax + Bx$ has a solution;*
*(ii)* *there is $x \in \partial\Omega$ and $\lambda \in (0,1)$ with $x = \lambda(A+B)(x)$.*

The following result can be thought of as a weak version of Theorem 43:

**Theorem 48.** *[1, Theorem 2.1] Let $Q$ be a bounded closed convex subset of a Banach space $X$ and $A : Q \longrightarrow X$, $B : X \longrightarrow X$ two mappings such that*

*(a)* *$A$ is weakly strongly continuous and $A(Q)$ is relatively weakly compact.*
*(b)* *$B$ is nonexpansive and satisfies $(\mathscr{A}2)$.*
*(c)* *If $(x_n)$ is a sequence of $Q$ such that $(I-B)x_n$ is weakly convergent, then $(x_n)$ has a weakly convergent subsequence.*
*(d)* *$I - B$ is demi-closed.*
*(e)* *$Ax + By \in Q$ for all $x, y \in Q$.*
*Then $A + B$ has a fixed point $x \in Q$.*

To round off this section, three recent results obtained by Garcia-falset *et al.* (2009, 2012) under the weak topology are presented without proofs:

**Theorem 49 ([44, Theorem 15]).** *Let $Q$ be a bounded closed convex subset of a Banach space $X$ and $A, B : Q \longrightarrow X$ two mappings such that*

*(a)* *$A$ is continuous and satisfies $(\mathscr{A}1)$.*
*(b)* *$B$ is nonexpansive and $I - B$ is $\phi$-expansive.*
*(c)* *$A + B$ is $\omega$-$k$-set contraction.*
*(d)* *$A(Q) + B(Q) \subset Q$.*
*Then $A + B$ has a fixed point $x \in Q$.*

**Theorem 50 ([47, Theorem 3.2]).** *Let $Q$ be a bounded closed convex subset of a Banach space $X$ and $A, B : Q \longrightarrow X$ two continuous mappings such that*

*(a)* *$A(Q)$ is relatively weakly compact.*
*(b)* *$A$ satisfies $(\mathscr{A}1)$.*
*(c)* *$B$ is nonexpansive and $\omega$-condensing.*
*(d)* *$A(Q) + B(Q) \subset Q$.*
*Then $A + B$ has a fixed point $x \in Q$.*

**Theorem 51 ([47, Theorem 3.4]).** *Let $X$ be a Banach space $X$ and $A, B : X \longrightarrow X$ two mappings such that*

*(a)* *$A$ maps bounded sets into relatively weakly compact ones.*
*(b)* *$A$ satisfies $(\mathscr{A}1)$.*
*(c)* *$B$ is nonexpansive and $\omega$-condensing.*
*(d)* *$I - B$ is $\psi$-expansive, where $\psi$ is either strictly increasing or $\lim_{r\to\infty} \psi(r) = \infty$.*

*Then at least one of the following properties holds:*

*(i)* *the equation $x = Ax + Bx$ has a solution,*
*(ii)* *the set $\{x \in Q : x = \lambda B(x/\lambda) + \lambda A(x),\ \lambda \in (0,1)\}$ is unbounded.*

Theorem 51 is an extension of [35, Theorem 2.5] where $B$ is an $\omega$-contraction. This theorem will be next applied to solve a nonlinear integral equation which extends the one considered in [35] (see also [76]).

## 6 Applications

To illustrate some of the fixed point theory developed in this paper, we have selected two examples dealing with nonlinear integral equations. Further applications to nonlinear Volterra and Hammerstein equations can be found, e.g., in [2, 45, 62, 63].

*Example 10.* Let $X = C([0,T],\Re)$ be the Banach space of real continuous functions endowed with the sup-norm. Consider the following integral equation (see also [48, Theorem 4.4] where the equation is considered in $C([0,T],E)$, $E$ being any Banach space):

$$u(t) = f(t,u(t)) + \int_0^t g(s,u(s))ds, \ t \in [0,T]. \tag{16}$$

The functions $f : [0,T] \times X \longrightarrow X$ and $g : [0,T] \times X \longrightarrow X$ are continuous. Let the integral and superposition operators $A,\ B : X \longrightarrow X$ be defined by

$$A(u)(t) = \int_0^t g(s,u(s))ds \ \text{ and } \ B(u)(t) = f(t,u(t)).$$

Assume that the nonlinear functions $f$ and $g$ satisfy the following conditions:

($H_1$) $f(t,\cdot)$ is a 1-set contraction and $I - f(t,\cdot)$ is $\psi$-expansive, for $t \in [0,T]$.
($H_2$) $\|u\| \leq \|u - f(t,u\|$, for all $(t,u) \in [0,T] \times X$.
($H_3$) $A(Q) \subset Q$.

To prove existence of solution to Equation (16), Theorem 45 is applied. We just check the hypotheses. Since $f$ is continuous, it is clear that $A$ is completely continuous. Assumption ($H_1$) implies that

$$\|(u - B(u)) - (v - B(v))\| \geq \psi(\|u(t) - v(t)\|), \quad \text{for all } t \in [0,T].$$

Hence the operator $I - B$ is $\psi$-expansive. Moreover, from Assumption ($H_1$), the operator $B$ is a 1-set contraction. Finally suppose that $u = B(u) + A(y)$ holds for some $y \in Q$. The fact that $u \in Q$ follows from ($H_2$) and ($H_3$); indeed,

$$\|u\| \leq \|u(t) - B(u)(t)\| = \|A(y)(t)\|.$$

Since $A(Q) \subset Q$, we conclude that $u \in Q$.

*Example 11.* Consider the nonlinear generalized Hammerstein equation (see [47, Example 4.1]):

$$y(t) = \eta \int_0^1 \varsigma(t,s)g(y(s))\,ds + \int_0^1 \nu(t,s)f(s,y(s))ds \tag{17}$$

posed on $L^1((0,1),\Re)$, the space of Lebesgue integrable functions on $(0,1)$. The function $\varsigma, g, \nu, f$ satisfy the following assumptions:

(1) $\varsigma(.,.)$ is essentially bounded.
(2) $f : [0,1]\times\Re \longrightarrow \Re$ is of Carathéodory type and there exist $\rho > 0$ and $\gamma \in L_1^+(0,1)$ such that $|f(t,x) \leq \gamma(t) + \rho|x|$.
(3) $g : \Re \longrightarrow \Re$ is nonexpansive.
(4) $\nu : [0,1]\times[0,1] \longrightarrow \Re$ is strongly measurable, $\int_0^1 \nu(\cdot,s)y(s)ds \in L^1(0,1)$ whenever $y \in L^1(0,1)$, and there exists $\mu : [0,1] \longrightarrow \Re$ such that $\nu(t,s) \leq \mu(t)$ for all $(t,s) \in [0,1]^2$.
(5) $\rho\|\nu\| < 1$.
(6) $0 < \eta|\varsigma\|_\infty + \rho\|\mu\| < 1$.

Define the nonlinear operators $A, B : L^1(0,1) \longrightarrow L^1(0,1)$ by:

$$Ay(t) = \int_0^1 \nu(t,s)f(s,y(s))ds,\ By(t) = \eta\int_0^1 \varsigma(t,s)g(y(s))ds.$$

Then, it can be proved that $A$ is continuous, satisfies ($\mathscr{A}1$), and maps bounded subsets of $L^1(0,1)$ into relatively compact subsets of $L^1(0,1)$. This has already been proved in [35, Theorem 3.1] and in [77, Theorem 3.1]. However, we point out that $A$ is not weakly strongly continuous unless $f$ is linear in the second argument (see [59, Theorem 5.1]). Regarding $B$, it is a separate contraction and satisfies ($\mathscr{A}2$); this follows from Assumptions (3) and (5). Then $B$ is nonexpansive and $I - B$ is $\psi$-expansive where $\psi$ is strictly increasing. Finally Hypothesis (6) provides a priori estimates so that condition (ii) in Theorem 51 is fulfilled, proving that Equation (17) has at least one solution in $L^1((0,1),\Re))$.

## 7 Conclusion

In this paper, the fixed point theory for nonexpansive mappings and for the more general class of 1-set contractions was surveyed. The stress was put on the difficulties encountered when dealing with these mappings as well as on the methods developed and the techniques that have been used to overcome such difficulties during the late years. A special attention and emphasis was made on the historical development of such theories. However, it was not the ambition of the paper to investigate with very deep details the numerous results obtained by a great number of authors since the sixties and especially the great progress recently achieved on the subject. In particular, and for the sake of limitation of the subject discussed herein, we have not considered in detail some questions -and even some extensions- such that those relating to the investigation of the more general classes of pseudo-contractive mappings -which we discussed briefly- and those concerning the so-called convex-power mappings. Also, the approximation methods in the spirit of recursive Halpern type algorithms, which require particular investigation, have not been covered by this paper (see, e.g., [6, Chapter 8]). Concerning the fixed point theory for multi-valued mappings, we recommend [6, Chapter 7] and [72, Chapter 19].

We hope however that this survey article highlights the development of the fixed point theory for 1-set contractions and also reports in a satisfactory way some recent advances in this theory. Note finally that some open problems, in connection with metric fixed point theory can be found in [52] while a recent survey is given in [6, Chapter 1].

## References


1. Agarwal, R.P., O'Regan, D., Taoudi, M.A.: Browder-Krasnoselskii-type fixed point theorems in Banach spaces, Fixed Point Th. Appl. **vol. 2010** ID 243716, 20 pages (2010) doi:10.1155/2010/243716
2. Agarwal, R.P., O'Regan, D., Taoudi, M.A.: Fixed point theorems for ws-compact mappings in Banach spaces, Fixed Point Th. Appl. **vol. 2010** ID 183596, 13 pages (2010) doi:10.1155/2010/183596
3. Agarwal, R.P., Meehan, M., O'Regan, D.: Fixed Point Theory and Applications, Cambridge Tracts in Math. **141**, Cambridge Univ. Press (2001)
4. Agarwal, R.P., Meehan, M., Shu, D.R.: Fixed Point Theory for Lipschitzian-type Mappings with Applications, Fixed Point Theory and its Applications **6**, Springer (2009)
5. Aksoy A.G., Khamsi, M.A.: Nonstandard Methods in Fixed point Theory, Springer, New York, Berlin (1990)
6. Almezel, S., Ansari, Q.H., Khamsi, M.A., Edts.: Topics in Fixed Point Theory, Springer Int. Publ., Switzerland (2014)
7. Alspach, D: A fixed point free nonexpansive map, Proc. A.M.S. **82(3)**, 423–424 (1981)
8. Appell, J., De Pascale, E.: Su alcuni parametri connessi con la misura di non compatteza di Hausdorff in spazi di funzioni misurabili, Boll. Un. Mat. Ita. B **6(3)**, 497–515 (1984)
9. Arino, O., Gautier, S., Penot, J.P.: A fixed point theorem for sequentially continuous mappings with applications to ordinary differential equations, Funkcial. Ekvac. **27**, 273–279 (1984)
10. Banas̀, J., Goebel, K.: Measures of Noncompactness in Banach Spaces, In Lecture Notes in Pure and Applied Mathematics **60**, Marcel Dekker, New York (1980)
11. Barbu, V., Precupanu, T.: Convexity and Optimization in Banach Spaces, Springer Monographs in Math. (2012)
12. Ben Amar, A., Mnif, M.: Leray-Schauder alternatives for weakly sequentially continuous mappings and application to transport equation, Math. Meth. Appl. Sci. **33**, 80–90 (2010)
13. Borsuk, K.: Theory of Retracts, Polish Scientific Publishers, Warszawa (1967)
14. Brezis, H.: Functional Analysis, Sobolev Spaces, and Partial Differential Equations, Universitext, Springer (2010)
15. Brodskii, M.S., Milman, D.P.: On the center of a convex set, Dokl. Akad. Nauk SSSR. **59(5)**, 837–840 (1948)

16. Browder, F.E.: Fixed point theorems for noncompact mappings in Hilbert spaces, Proc. Nat. Acad. Sci. USA **43**, 272–6 (1965)
17. Browder, F.E.: Nonexpansive nonlinear operators in a Banach space, Proc. Nat. Acad. Sci. USA **54**, 1041–1044 (1965)
18. Browder, F.E.: Nonlinear mappings of non-expansive and accretive type in Banach spaces, Bull. A.M.S. **73**, 875–882 (1967)
19. Browder, F.E.: Semicontractive and semiaccretive nonlinear mappings in Banach spaces, Bull. A.M.S. **74**, 660–665 (1968)
20. Browder, F.E., Petryshyn, W.V.: Construction of fixed points of nonlinear mappings in Hilbert space, J. Math. Anal. Appl. **20**, 197–228 (1967)
21. Bruck, R.E.: Properties of fixed-point sets of nonexpansive mappings in Banach spaces, Trans. A.M.S. **179**, 251–262 (1973)
22. Bruck, R.E.: A common fixed point theorem for a commuting family of nonexpansive mappings, Pacific J. Math. **53**, 59–71 (1974)
23. Burton, T.A.: A fixed point theorem of Krasnozels'kĭi, Appl. Math. Lett., **11**, 85–88 (1998)
24. Caristi, J.: Fixed point theorems for mappings satisfying inwardness conditions, Trans. A.M.S. **215**, 241–251 (1976)
25. Chidume, C.: Geometric Properties of Banach Spaces and Nonlinear Iterations, L.N.M. **1965**, Springer Verlag (2009)
26. Clarkson, J.A.: Uniformly convex spaces, Trans. A.M.S. **40(3)**, 396–414 (1936)
27. Darbo, G.: Punti uniti in transformationi a condominio non-compactto, Rend. Sem. Mat. Univ. Padova **24**, 84–92 (1955)
28. De Blasi, F.S.: On a property of the unit sphere in Banach spaces, Bull. Math. Soc. Sci. Math. Roumanie **21**, 259–262 (1997)
29. Deimling, K.: Geometry of Banach Spaces, L.N.M. **485**, Springer Verlag (1975)
30. Deimling, K.: Nonlinear Functional Analysis, Springer (1985)
31. Djebali, S., Hammache, K.: Fixed point theorems for nonexpansive maps in Banach spaces, Nonlin. Anal. **73**, 3440–3449 (2010)
32. Djebali, S., Hammache, K.: Furi-Pera fixed point theorems for nonexpansive maps in Banach spaces, Fixed Point Th. (Cluj) **13(2)**, 461–474 (2012)
33. Djebali, S., Hammache, K.: Fixed point theorems for 1-set contractions in Banach spaces, Fixed Point Th. (Cluj), **16(2)**, 285–302 (2015)
34. Djebali, S., Hammache, K.: Boundary conditions for 1-set contractions, Mediterr. J. Math. DOI 10.1007/s00009-015-0589-0 Springer Basel 2015
35. Djebali, S., Sahnoune, Z.: Nonlinear alternatives of Schauder and Krasnosel'skij types with applications to Hammerstein integral equations in $L^1$ spaces, J. Diff. Eqns. **249**, 2061–2075 (2010)
36. Duan, H.-G., Xu, S.-Y., Li, G.-Z.: Fixed points of weakly inward 1-set contraction mappings, J. Korean Math. Soc. **45(6)**, 1725–1740 (2008)
37. Dugundji, J.: An extension of Tietze's Theorem, Pacific J. Math. **I**, 351–367 (1951)
38. Dunford, N., Schwartz, J.T.: Linear operators, Part I: General Theory, Interscience, New York, Reprint of the original 1958, edition (1988)
39. Edelstein, M.: Fixed point theorems in uniformly convex Banach spaces, Proc. A.M.S. **44(2)**, 369–374 (1974)
40. Edmunds, D.E.: Remarks on nonlinear functional equations, Math. Annalen **174**, 233–239 (1967)
41. Edwards, R.: Functional Analysis. Theory and Applications, Holt-Rinehart-Winston, New York (1965)
42. Emmanuele, G.: Measure of weak noncompactness and fixed point theorems, Bull. Math. Soc. Sci. Math. Roumanie **25**, 353–358 (1981)
43. Furi, M., Pera, P.: A continuation method on locally convex spaces and Applications to ode on noncompact intervals, Annal. Pol. Math. **XLVII**, 331–346 (1987)
44. Garcia-Falset, J.: Existence of fixed points and measures of weak noncompactness, Nonlin. Anal., T.M.A. **71**, 2625–2633 (2009)

45. Garcia-Falset, J.: Existence of fixed points for the sum of two operators, Math. Nachr. **283(12)**, 1736–1757 (2010)
46. Garcia-Falset, J., Latrach, K.: Krasnoselskii-type fixed-point theorems for weakly sequentially continuous mappings, Bull. London Math. Soc. **44**, 25–38 (2012)
47. Garcia-Falset, J., Latrach, K., E. Moreno-Galvez, E., Taoudi, M.A.: Schaefer-Krasnoselskii fixed point theorems using a usual measure of weak noncompactness, J. Diff. Eqns., **252**, 3436–3452 (2012)
48. Garcia-Falset, J., Muñiz-Pérez, O.: Fixed point theory for 1-set contractive and pseudocontractive mappings, Appl. Math. Comput. **219**, 6843–6855 (2013)
49. Gasinski, L., Papageorgiou, N.S.: Nonlinear Analysis, Taylor and Francis **9**, (2005)
50. Goebel, K.: Concise Course on Fixed Point Theorems, Yokohama Publish. (2002)
51. Goebel, K., Kirk, W.A.: Topics in Metric Fixed Point Theory, Cambridge Univ. Press, Cambridge (1990)
52. Goebel, K., Kirk, W.A.: Some problems in metric fixed point theory, J. Fix. Point. Th. Appl. **4**, 13–25 (2008)
53. Goebel, K., Reich, S.: Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings, Monographs and Textbooks in Pure and Applied Mathematics **83**, Marcel Dekker, New york (1984)
54. Göhde, D.: Zum prinzip der kontraktiven abbildung, Math. Nachr. **30**, 251–258 (1965)
55. González, C., Jimenéz-Melado, A., LIorens-Fuster, E.: A Mönch type fixed point theorem under the interior condition, J. Math. Anal. Appl. **352**, 816–821 (2009)
56. Granas A., Dugundji, J.: Fixed Point Theory, Springer-Verlag, New York (2003)
57. Gulevich, N.M.: Fixed points of nonexpansive mappings, J. Math. Sc. **79(1)**, 755–815 (1996)
58. Halpern B.R., Bergman, G.M.: A fixed-point theorem for inward and outward maps, Trans. A.M.S. **130**, 353–358 (1968)
59. Hencl, S., Kolář, J., Pangrác, O.: Integral functionals that are continuous with respect to the weak topology on $W^{1,p}(0,1)$, Nonlin. Anal. T.M.A. **63**, 81–87 (2005)
60. Istratescu, V.I.: Fixed point Theory. An Introduction, D. Reidel Publishing Company (1981)
61. Ivanov, A.A.: Fixed points of mappings of metric spaces, Translated from Zapiski Nauchnykh Seminarov Leningradskogo Otdeleniya Matematicheskogo Instituta im. V.A. Steklova AN SSSR **66**, 5–102 (1976)
62. Jeribi, A., Krichen, B., Mefteh, B.: Existence of solutions of a nonlinear Hammerstein integral equation, Num. Funct. Anal. Optim. **35**, 1328–1339 (2014)
63. Jeribi, A., Krichen, B., Mefteh, B.: Existence of solutions of a two-dimensional boundary value problem for a system of nonlinear equations arising in growing cell populations, J. Biol. Dyn. **7**, 218–232 (2013)
64. Jimenéz-Melado A., Morales, C.H.: Fixed point theorems under the interior condition, Proc. A.M.S. **134(2)**, 501–507 (2005)
65. Kaewcharoen A., Kirk, W.A.: Nonexpansive mappings defined on unbounded domains, Fixed Point Th. Appl. **82080**, 1–13 (2006)
66. Khamsi, M.A., Kirk, W.A.: An Introduction to Metric Spaces and Fixed Point Theory, John Willey & Sons (2001)
67. Kirk, W.A.: A fixed point theorem for mappings which do not increase distance, Amer. Math. Month. **72**, 1004–1006 (1965)
68. Kirk, W.A.: Fixed point theorems for nonexpansive mappings satisfying certain boundary conditions, Proc. A.M.S. **50**, 143–149 (1975)
69. Kirk, W.A.: Locally nonexpansive mappings in Banach spaces, Lect. Notes in Math., Springer, Berlin-New York **886**, 178–198 (1981)
70. Kirk, W.A.:, Fixed point theory for nonexpansive mappings, II. Fixed points and nonexpansive mappings (Cincinnati, Ohio, 1982), Contemp. Math., Amer. Math. Soc., Providence, RI **18**, 121–140 (1983)
71. W.A. Kirk, Schöneberg, R.: Some results on pseudo-contractive mappings, Pacific J. Math. **71**, 89–100 (1977)
72. Kirk, W.A., Sims, B., Edts.: Handbook of Metric Fixed Point Theory, Springer (2001)

73. Krasnozels'kĭi, M.A.: Positive Solutions of Operator Equations, Noordhoff, Groningen, The Netherlands (1964)
74. Krasnozels'kĭi, M.A.: Integral Operators in Spaces of Summable Functions, Noordhoff, Leyden, The Netherlands (1976)
75. Lan, K.Q., Webb, J.R.L.: A fixed point index for weakly inward $A$-proper maps, Nonlin. Anal., T.M.A. **28(2)**, 315–327 (1997)
76. Latrach, K., Taoudi, M.A., Zeghal, A.: Some fixed point theorems of the Schauder and the Krasnosel'skij type and application to nonlinear transport equations, J. Diff. Equ. **221**, 256–271 (2006)
77. Latrach, K., Zeghal, A.: Existence results for a generalized nonlinear Hammerstein equation in $L^1$ spaces, Nonlin. Anal. T.M.A. **66**, 2325–2333 (2007)
78. Leray J., Schauder, J.: Topologie et équations fonctionnelles, Ann. Sci. Éc. Norm. Sup. **51**, 45–78 (1934)
79. Li, G.-Z.: The fixed-point index and the fixed-point theorems of 1-set-contraction mappings, Proc. A.M.S. **104(4)**, 1163–1170 (1988)
80. Li, G.-Z., Xu, S.-Y., Duan, H.-G.: Fixed point theorems of 1-set contractive operators in Banach spaces, Appl. Math. Lett. **19**, 403–412 (2006)
81. Liu, L.-S.: Approximation theorems and fixed point theorems for various classes of 1-set contractive mappings in Banach spaces, Acta Math. Sinica, Eng. Ser. **17(1)**, 103–112 (2001)
82. Lloyd, N.G : Degree Theory, Cambridge University Press (1978)
83. Megginson, R.E.: An Introduction to Banach Space Theory, Graduate Texts in Math. **183**, Springer Verlag, New York (1998)
84. Morales, C.H.: The Leray-Schauder condition for continuous pseudo-contractive mappings, Proc. A.M.S. **137(3)**, 1013–1020 (2009)
85. Nussbaum, R.D.: The fixed point index and asymptotic fixed point theorems for $k$-set contractions, Bull. A.M.S. **75**, 90–495 (1969)
86. Nussbaum, R.D.: Degree theory for local condensing maps, J. Math. Anal. Appl. **37**, 741–766 (1972)
87. Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings, Bull. A.M.S. **73**, 591–597 (1967)
88. O'Regan, D.: Fixed point theory for the sum of two operators, Appl. Math Lett. **9(1)**, 1–8 (1996)
89. Papageorgiou N.S., Kyrtsi-Yiallourou, S.Th. Edts.: Handbook of Applied Analysis, **vol. 19**, Springer (2009)
90. Penot, J.-P.: A fixed point theorem for asymptotically contractive mappings, Proc. A.M.S. **131(8)**, 2371–2377 (2003)
91. Petryshyn, W.V.: Structure of the fixed point sets of $k$-set contractions, Arch. Rat. Mech. Anal. **40**, 312–328 (1971)
92. Petryshyn, W.V.: Remarks on condensing and $k$-set-contractive mappings, J. Math. Anal. Appl. **39**, 717–741 (1972)
93. Petryshyn, W.V.: Fixed point theorems for various classes of 1-set-contractive and 1-ball-contractive mappings in Banach spaces, Trans. A.M.S. **182**, 323–3582 (1973)
94. Phelps, R.R.: Convex sets and nearest points, Proc. A.M.S. **8**, 790–797 (1957)
95. Lin, P.-K.: Fixed point theory and nonexpansive mappings, Arab J. Math. **1**, 495–509 (2012)
96. O'Regan, D.: Fixed-point theory for weakly sequentially continuous mappings, Math. Comput. Model. **27**, 1–14 (1998)
97. O'Regan D., Precup, R.: Theorems of Leray-Schauder type and Applications, Ser. Math. Anal. Appl. **3**, Gordon and Breach Sci. Publish (2001)
98. Reich, S.: Fixed points of condensing functions, J. Math. Anal. Appl. **41**, 460–467 (1973)
99. Sadovskiĭ, B.N.: A fixed point principle, Func. Anal. Appl. **1(2)**, 151–153 (1967)
100. Shaini, P.: Fixed points for 1-set contractions, Gen. Math. **19(2)**, 59–64 (2011)
101. Smart, D.R.: Fixed Point Theorems, Cambridge Univ. Press (1974)
102. Xu, S.: New fixed point theorems for 1-set-contractive operators in Banach spaces, Nonlin. Anal., T.M.A. **67**, 938–944 (2007)

103. Zabreĭko, P.P., Kačurovskiĭ, R.I., Krasnosel'skiĭ, M.A.: On a fixed point principle for operators in a Hilbert space, (Russian) Funkcional. Anal. i Priložen **1(2)**, 93–94 (1967)
104. Zeidler, E.: Nonlinear Functional Analysis and its Applications. I: Fixed Point Theorems, Springer (1985)

# Spectral results on quantum waveguides

**Hatem Najar**

**Abstract** In this document we review some results dealing with the study of the spectral properties of quantum waveguide. Precisely we are interested in the bound states of the Hamiltonian describing a quantum particle living on three dimensional straight strip of width $d$. We impose the Neumann boundary condition on a disc window of radius $a$ and Dirichlet boundary conditions on the remained part of the boundary of the strip (Najar et al., Math Phys Anal Geom 13:19–28, 2010).

We study the case when we destroy the plan symmetry, i.e. we impose the Neumann boundary condition on the two concentric disc windows of the radii $a$ and $b$ located on the opposite walls and the Dirichlet boundary condition on the remaining part of the boundary of the strip (Najar and Olendski, J Phys A Math Theor 44, 2011).

The effect of a magnetic field of Aharonov-Bohm type when the magnetic field is turned on this system is considered (Najar and Raissi, On the spectrum of the Schrodinger Operator with Aharonov-Bohm Magnetic Field in quantum waveguide with Neumann window, Math. Meth. App. Sci. (2015)).

## 1 Introduction

The study of quantum waves on quantum waveguide has gained much interest and has been intensively studied during the last years for their important physical consequences. The main reason is that they represent an interesting physical effect with important applications not only in nanophysical devices, but also in flat electromagnetic waveguide. See the monograph [15] and the references therein.

Exner et al. have done seminal works in this field. They obtained results in different contexts, we quote [6, 10, 13, 14]. Also in [16, 20, 21] research has been

H. Najar (✉)
Faculté des Sciences de Moanstir, Département de Mathématiques,
Université de Monastir, Monastir, Tunisia

Laboratoire de recherche: Algèbre Géométrie et Théorie Spectrale : LR11ES53
e-mail: hatem.najar@ipeim.rnu.tn

 
A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings
in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_4

conducted in this area; the first is about the discrete case and the two others for deals with the random quantum waveguide.

It should be noticed that the spectral properties essentially depend on the geometry of the waveguide, in particular, the existence of a bound states induced by curvature [9, 10, 12, 13] or by coupling of straight waveguides through windows [13, 15] was shown. The waveguide with Neumann boundary condition was also investigated in several papers [22, 24]. A possible next generalization are waveguides with combined Dirichlet and Neumann boundary conditions on different parts of the boundary. The presence of different boundary conditions also gives rise to nontrivial properties like the existence of bound states.

## 2 The model

The system we are going to study is given in Fig. 1. We consider a Schrödinger particle whose motion is confined to a pair of parallel plans of width $d$. For simplicity, we assume that they are placed at $z = 0$ and $z = d$. We shall denote this configuration space by $\Omega . \Omega = \mathbb{R}^2 \times [0, d]$. Let $\gamma(a)$ be a disc of radius $a$, without loss of generality we assume that the center of $\gamma(a)$ is the point $(0, 0, 0)$;

$$\gamma(a) = \{(x, y, 0) \in \mathbb{R}^3;\ x^2 + y^2 \le a^2\}. \tag{1}$$

We set $\Gamma = \partial\Omega \setminus \gamma(a)$. We consider Dirichlet boundary condition on $\Gamma$ and Neumann boundary condition in $\gamma(a)$.

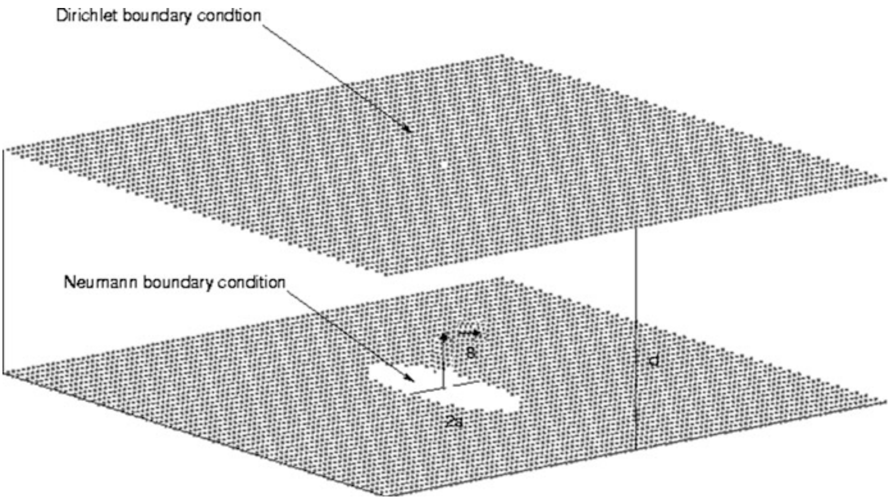


**Fig. 1** The waveguide with a disc window and two different boundary conditions

## 2.1 The Hamiltonian

Let us define the self-adjoint operator on $L^2(\Omega)$ corresponding to the particle Hamiltonian $H$. This will be done by the means of quadratic forms. Precisely, let $q_0$ be the quadratic form

$$q_0(f,g) = \int_\Omega \nabla f \cdot \overline{\nabla g} d^3x, \text{ with domain } \mathcal{Q}(q_0) = \{f \in H^1(\Omega);\, f\lceil \Gamma = 0\}, \tag{2}$$

where $H^1(\Omega) = \{f \in L^2(\Omega) | \nabla f \in L^2(\Omega)\}$ is the standard Sobolev space and we denote by $f\lceil\Gamma$, the trace of the function $f$ on $\Gamma$. It follows that $q_0$ is a densely defined, symmetric, positive, and closed quadratic form. We denote the unique self-adjoint operator associated with $q_0$ by $H$ and its domain by $D(\Omega)$. It is the hamiltonian describing our system. From [25] (page 276), we infer that the domain $D(\Omega)$ of $H$ is

$$D(\Omega) = \Big\{f \in H^1(\Omega);\ -\Delta f \in L^2(\Omega), f\lceil\Gamma = 0, \frac{\partial f}{\partial z}\lceil\gamma(a) = 0\Big\}$$

and

$$Hf = -\Delta f, \quad \forall f \in D(\Omega).$$

## 2.2 Some known facts

Let us start this subsection by recalling that in the particular case when $a = 0$, we get $H^0$, the Dirichlet Laplacian, and $a = +\infty$ we get $H^\infty$, the Dirichlet-Neumann Laplacian. Since

$$H = (-\Delta_{\mathbb{R}^2}) \otimes I \oplus I \otimes (-\Delta_{[0,d]}), \text{on } L^2(\mathbb{R}^2) \otimes L^2([0,d]),$$

(see [25]) we get that the spectrum of $H^0$ is $[(\frac{\pi}{2d})^2, +\infty[$. Consequently, we have

$$\Big[(\frac{\pi}{d})^2, +\infty\Big[ \subset \sigma(H) \subset \Big[(\frac{\pi}{2d})^2, +\infty\Big[.$$

Using the property that the essential spectra are preserved under compact perturbation, we deduce that the essential spectrum of $H$ is

$$\sigma_{ess}(H) = \Big[(\frac{\pi}{d})^2, +\infty\Big].$$

An immediate consequence is the discrete spectrum lies in $\Big[(\frac{\pi}{2d})^2, (\frac{\pi}{d})^2\Big]$.

## 2.3 Preliminary: Cylindrical coordinates

Let us notice that the system has a cylindrical symmetry, therefore, it is natural to consider the cylindrical coordinates system $(r, \theta, z)$. Indeed, we have that

$$L^2(\Omega, dxdydz) = L^2(]0, +\infty[\times[0, 2\pi[\times[0, d], rdrd\theta dz),$$

We note by $\cdot\langle\cdot,\rangle_r$, the scaler product in $L^2(\Omega, dxdydz) = L^2(]0, +\infty[\times[0, 2\pi[\times[0, d], rdrd\theta dz)$ given by

$$\langle f, g\rangle_r = \int_{]0,+\infty[\times[0,2\pi[\times[0,d]} fgrdrd\theta dz.$$

We denote the gradient in cylindrical coordinates by $\nabla_r$. The Laplacian operator in cylindrical coordinates is given by

$$\Delta_{r,\theta,z} = \frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial}{\partial r}) + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2} + \frac{d^2}{dz^2}. \tag{3}$$

Therefore, the eigenvalue equation is given by

$$-\Delta_{r,\theta,z} f(r, \theta, z) = Ef(r, \theta, z). \tag{4}$$

Since the operator is positive, we set $E = k^2$. The equation (4) is solved by separating variables and considering $f(r, \theta, z) = \varphi(r) \cdot \psi(\theta)\chi(z)$. We replace $f$ by its expression in equation (4) and separate $\chi$ by putting all the $z$ dependence in one term so that $\frac{\chi''}{\chi}$ can only be constant. The constant is taken as $-s^2$ for convenience. Second, we separate the term $\frac{\psi''}{\psi}$ which has all the $\theta$ dependence. Using the fact that the problem has an axial symmetry and the solution has to be $2\pi$ periodic and single value in $\theta$, we obtain $\frac{\psi''}{\psi}$ should be a constant $-n^2$ for $n \in \mathbb{Z}$. Finally, we get the following equation for $\varphi$

$$\varphi''(r) + \frac{1}{r}\varphi'(r) + [k^2 - s^2 - \frac{n^2}{r^2}]\varphi(r) = 0. \tag{5}$$

We notice that the equation (5) is the Bessel equation and its solutions could be expressed in terms of Bessel functions. More explicit solutions could be given by considering boundary conditions.

# 3 Results on discrete spectrum

## 3.1 One Neumann Window

The first result we give is the following theorem.

**Theorem 1 ([17]).** *The operator $H$ has at least one isolated eigenvalue in $\big[(\frac{\pi}{2d})^2, (\frac{\pi}{d})^2\big]$ for any $a > 0$.*

*Moreover for a big enough, if $\lambda(a)$ is an eigenvalue of $H$ less than $\dfrac{\pi^2}{d^2}$, then we have*

$$\lambda(a) = \left(\frac{\pi}{2d}\right)^2 + o\left(\frac{1}{a^2}\right). \tag{6}$$

*Proof.* Let us start by proving the first claim of the theorem. To do so, we define the quadratic form $\mathcal{Q}_0$,

$$\mathcal{Q}_0(f,g) = \langle \nabla f, \nabla g\rangle_r = \int_{]0,+\infty[\times[0,2\pi[\times[0,d]} (\partial_r f\overline{\partial_r g} + \frac{1}{r^2}\partial_\theta f\overline{\partial_\theta g} + \partial_z f\overline{\partial_z g})rdrd\theta dz, \tag{7}$$

with domain

$$\mathcal{D}_0(\Omega) = \left\{f \in L^2(\Omega, rdrd\theta dz); \nabla_r f \in L^2(\Omega, rdrd\theta dz); f\lceil\Gamma = 0\right\}.$$

Consider the functional $q$ defined by

$$q[\Phi] = \mathcal{Q}_0[\Phi] - (\frac{\pi}{d})^2\|\Phi\|^2_{L^2(\Omega,rdrd\theta dz)}. \tag{8}$$

Since the essential spectrum of $H$ starts at $(\dfrac{\pi}{d})^2$, if we construct a trial function $\Phi \in \mathcal{D}_0(\Omega)$ such that $q[\Phi]$ has a negative value, then the task is achieved. Using the quadratic form domain, $\Phi$ must be continuous inside $\Omega$ but not necessarily smooth. Let $\chi$ be the first transverse mode, i.e.

$$\chi(z) = \begin{cases} \sqrt{\frac{2}{d}}\sin(\frac{\pi}{d}z) & \text{if} \quad z \in (0,d) \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

For $\Phi(r,\theta,z) = \varphi(r)\chi(z)$, we compute

$$\begin{aligned} q[\Phi] &= \langle \nabla_r\varphi\chi, \nabla_r\varphi\chi\rangle - (\frac{\pi}{d})^2\|\varphi\chi\|^2_{L^2(\Omega,rdrd\theta dz)}, \\ &= 2\pi\|\varphi'\|^2_{L^2([0,+\infty[,rdr)} \end{aligned}$$

Now let us consider an interval $J = [0, b]$ for a positive $b > a$ and a function $\varphi \in \mathscr{S}([0, +\infty[)$ such that $\varphi(r) = 1$ for $r \in J$. We also define a family $\{\varphi_\tau : \tau > 0\}$ by

$$\varphi_\tau(r) = \begin{cases} \varphi(r) & \text{if } r \in (0, b) \\ \varphi(b + \tau(\ln r - \ln b)) & \text{if } \quad r \geq b. \end{cases} \tag{10}$$

Let us write

$$\begin{aligned} \|\varphi_\tau'\|_{L^2([0,+\infty),rdr)} &= \int_{(0,\infty)} |\varphi_\tau'(r)|^2 rdr, \\ &= \int_{(b,+\infty)} \tau^2 |\varphi'(b + \tau(\ln r - \ln b))|^2 rdr, \\ &= \tau \int_{(0,+\infty)} |\varphi'(s)|^2 ds = \tau \|\varphi'\|^2_{L^2((0,+\infty))}. \end{aligned} \tag{11}$$

Let $j$ be a localization function from $C_0^\infty(0, a)$ and for $\tau, \varepsilon > 0$ we define

$$\Phi_{\tau,\varepsilon}(r, z) = \varphi_\tau(r)[\chi(z) + \varepsilon j(r)^2] = \varphi_\tau(r)\chi(z) + \varphi_\tau \varepsilon j^2(r) = \Phi_{1,\tau,\varepsilon}(r, z) + \Phi_{2,\tau,\varepsilon}(r). \tag{12}$$

$$\begin{aligned} q[\Phi] &= q[\Phi_{1,\tau,\varepsilon} + \Phi_{2,\tau,\varepsilon}] \\ &= \mathscr{Q}_0[\Phi_{1,\tau,\varepsilon} + \Phi_{2,\tau,\varepsilon}] - (\frac{\pi}{d})^2 \|\Phi_{1,\tau,\varepsilon} + \Phi_{2,\tau,\varepsilon}\|^2_{L^2(\Omega,rdrd\theta dz)}. \\ &= \mathscr{Q}_0[\Phi_{1,\tau,\varepsilon}] - (\frac{\pi}{d})^2 \|\Phi_{1,\tau,\varepsilon}\|^2_{L^2(\Omega,rdrd\theta dz)} + \mathscr{Q}_0[\Phi_{2,\tau,\varepsilon}] - (\frac{\pi}{d})^2 \|\Phi_{2,\tau,\varepsilon}\|^2_{L^2(\Omega,rdrd\theta dz)} \\ &+ 2\langle \nabla_r \Phi_{1,\tau,\varepsilon}, \nabla_r \Phi_{2,\tau,\varepsilon}\rangle_r - (\frac{\pi}{d})^2 \langle \Phi_{1,\tau,\varepsilon}, \Phi_{2,\tau,\varepsilon}\rangle_r. \end{aligned}$$

Using the properties of $\chi$, noting that the supports of $\varphi$ and $j$ are disjoints and taking into account equation (11), we get

$$\begin{aligned} q[\Phi] &= 2\pi\tau \|\varphi'\|_{L^2(0,+\infty)} - 8\pi d\varepsilon \|j^2\|^2_{L^2(0,+\infty)} \\ &\quad + 2\varepsilon^2 \pi \{2\|jj'\|^2_{(L^2(0,\infty),rdr)} - (\frac{\pi}{d})^2 \|j^2\|^2_{(L^2(0,\infty),rdr)}\}. \end{aligned} \tag{13}$$

Firstly, we notice that only the first term of the last equation depends on $\tau$. Secondly, the linear term in $\varepsilon$ is negative and could be chosen sufficiently small so that it dominates over the quadratic one. Fixing this $\varepsilon$ and then choosing $\tau$ sufficiently small the right-hand side of (13) is negative. This ends the proof of the first claim.

The proof of the second claim is based on bracketing argument. Let us split $L^2(\Omega, rdrd\theta dz)$ as follows: $L^2(\Omega, rdrd\theta dz) = L^2(\Omega_a^-, rdrd\theta dz) \oplus L^2(\Omega_a^+, rdrd\theta dz)$, with

$$\Omega_a^- = \{(r, \theta, z) \in [0, a] \times [0, 2\pi[\times[0, d]\},$$
$$\Omega_a^+ = \Omega \backslash \Omega_a^-.$$

Therefore

$$H_a^{-,N} \oplus H_a^{+,N} \leq H \leq H_a^{-,D} \oplus H_a^{+,D}.$$

Here we index by $D$ and $N$ depending on the boundary conditions considered on the surface $r = a$. The min-max principle leads to

$$\sigma_{ess}(H) = \sigma_{ess}(H_a^{+,N}) = \sigma_{ess}(H_r^{+,D}) = \Big[(\frac{\pi}{d})^2, +\infty\Big[.$$

Hence if $H_r^{-,D}$ exhibits a discrete spectrum below $\frac{\pi^2}{d^2}$, then $H$ does as well. We mention that this is not a necessary condition. If we denote by $\lambda_j(H_a^{-,D})$, $\lambda_j(H_a^{-,N})$ and $\lambda_j(H)$, the $j$-th eigenvalue of $H_a^{-,D}$, $H_a^{-,N}$ and $H$, respectively, then, again the minimax principle yields the following

$$\lambda_j(H_a^{-,N}) \leq \lambda_j(H) \leq \lambda_j(H_a^{-,D}) \tag{14}$$

and for $2 \geq j$

$$\lambda_{j-1}(H_a^{-,D}) \leq \lambda_j(H) \leq \lambda_j(H_a^{-,D}). \tag{15}$$

$H_a^{-,D}$ has a sequence of eigenvalues [2, 27], given by

$$\lambda_{k,n,l} = \left(\frac{(2k+1)\pi}{2d}\right)^2 + \left(\frac{x_{n,l}}{a}\right)^2,$$

where $x_{n,l}$ is the $l$-th positive zero of Bessel function of order $n$ ( see [2, 27]) . The condition

$$\lambda_{k,n,l} < \frac{\pi^2}{d^2}, \tag{16}$$

yields that $k = 0$, so we get

$$\lambda_{0,n,l} = \left(\frac{\pi}{2d}\right)^2 + \left(\frac{x_{n,l}}{a}\right)^2.$$

This yields that the condition (16) to be fulfilled, will depend on the value of $\left(\frac{x_{n,l}}{a}\right)^2$.

We recall that $x_{n,l}$ are the positive zeros of the Bessel function $J_n$. So, for any $\lambda(a)$, eigenvalue of $H$, there exists $n, l, n', l' \in \mathbb{N}$, such that

$$\frac{\pi^2}{4d^2} + \frac{x_{n,l}^2}{a^2} \leq \lambda(a) \leq \frac{\pi^2}{4d^2} + \frac{x_{n',l'}^2}{a^2}. \tag{17}$$

The proof of (6) is completed by observing that $x_{n,l}$ and $x_{n',l'}$ are independent from $a$. ⊡

## 3.2 Two Neumann Windows

We consider a Schrödinger particle whose motion is confined to a pair of parallel planes separated by the width $d$. For simplicity, we assume that they are placed at $z = 0$ and $z = d$. We shall denote this configuration space by $\Omega$

Let $\gamma(a)$ be a disc of radius $a$ with its center at $(0, 0, 0)$ and $\gamma(b)$ be a disc of radius $b$ centered at $(0, 0, d)$; without loss of generality, we assume that $0 \leq b \leq a$.

$$\gamma(a) = \{(x, y, 0) \in \mathbb{R}^3;\ x^2 + y^2 \leq a^2\};\ \gamma(b) = \{(x, y, d) \in \mathbb{R}^3;\ x^2 + y^2 \leq b^2\}. \tag{18}$$

We set $\Gamma = \partial\Omega \diagdown (\gamma(a) \cup \gamma(b))$. We consider Dirichlet boundary condition on $\Gamma$ and Neumann boundary condition in $\gamma(a)$ and $\gamma(b)$.

**Theorem 2 ([18]).** *The operator $H$ has at least one isolated eigenvalue in $\left[0, (\frac{\pi}{d})^2\right]$ for any $a$ and $b$ such that $a + b > 0$.*

*Moreover for $a$ big enough, if $\lambda(a)$ is an eigenvalue of $H$ less than $\frac{\pi^2}{d^2}$, then we have*

$$\lambda(a, b) \in \left(\frac{1}{a^2}, \frac{1}{b^2}\right). \tag{19}$$

*Remark 1.* 1. The first claim of Theorem 2 is valid for more general shape of bounded surface $\mathscr{S}$, with Neumann boundary condition, not necessarily a disc; (see Figure 2) it suffices that the surface contains a disc of radius $a > 0$.

2. For more general shape $\mathscr{S}$ using discs of radius $a$ and $a'$, such that

$$\gamma(a) \subset \mathscr{S} \subset \gamma(a'); \tag{20}$$

In [1] Assel and Ben Salah considered the case of square window.

$$\Omega = \mathbb{R}^2 \times [0, d].$$

When $b$ is big enough, we get the result.

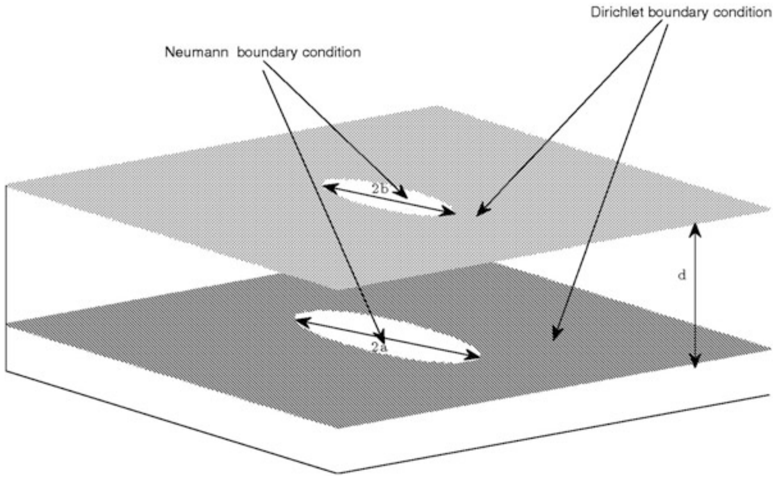


**Fig. 2** Dirichlet wave guide with two concentric Neumann disc windows on the opposite walls with (in general) different radii $a$ and $b$.

**Proposition 1 ([18]).** *When the radius $a$ is equal to a critical value $a_l$ at which a new bound state emerges from the continuum, equation (4) with $E = 1$ has a solution $f_l^{(0)}(r, \theta, z)$, unique to a multiplicative constant which at infinity behaves like (valid for both configurations of the boundary conditions)*

$$f_l^{(0)}(r, \theta, z) = \frac{e^{im\theta}}{\sqrt{2\pi}} \left[ \frac{\sqrt{2} \sin \pi z}{r^{|m|}} + \beta_l \frac{e^{-\pi\sqrt{3}r}}{\sqrt{r}} \sin 2\pi z + \mathcal{O}\left( \frac{e^{-\pi\sqrt{8}r}}{\sqrt{r}} \right) \right], \quad r \to \infty \tag{21}$$

*with some constants $\beta_l$. Here the two quantum numbers $n$ and $m$ are compacted into the single index $l$: $l \equiv (n, m)$.*

A proof of this statement will be given in the next section.

*Remark.* Compared to the corresponding equation for the quasi-one-dimensional wave guide [5, 6, 14], this asymptotics has a different form which is explained by the additional degree of the in-plane motion.

## 3.3 Effect of Aharonov-Bohm filed

Results on the discrete spectrum of a magnetic Schrödinger operator in waveguide-type domains are scarce. A planar quantum waveguide with constant magnetic field and a potential well is studied in [11], where it was proved that if the potential well is purely attractive, then at least one bound state will appear for any value of the magnetic field. Stability of the bottom of the spectrum of a magnetic Schrödinger operator was also studied in [26]. Magnetic field influence on the Dirichlet-Neumann structures was analyzed in [7, 23], the first dealing with a smooth compactly supported field as well as with the Aharonov-Bohm field in a two dimensional strip and second with perpendicular homogeneous magnetic field in the quasi-dimensional.

Despite numerous investigations of quantum waveguides during the last few years, many questions remain to be answered, this concerns, in particular, effects of external fields. Most attention has been paid to magnetic fields, either perpendicular to the waveguides plane or threaded through the tube, while the influence of the Aharonov-Bohm magnetic field alone remained mostly untreated.

In their celebrated 1959 paper [4] Aharonov and Bohm pointed out that while the fundamental equations of motion in classical mechanics can always be expressed in terms of field alone, in quantum mechanics the canonical formalism is necessary, and as a result, the potentials cannot be eliminated from the basic equations. They proposed several experiments and showed that an electron can be influenced by the potentials even if no field acts upon it. More precisely, in a field-free multiply-connected region of space, the physical properties of a system depend on the potentials through the gauge-invariant quantity $\oint \mathbf{A} dl$, where **A** represents the vector potential. Moreover, the Aharonov-Bohm effect only exists in **the multiply connected region of space**. The Aharonov-Bohm experiment allows in principle to measure the decomposition into homotopy classes of the quantum mechanical propagator.

A possible next generalization are waveguides with combined Dirichlet and Neumann boundary conditions on different parts of the boundary with an Aharonov-Bohm magnetic field with the flux $2\pi\alpha$. The presence of different boundary conditions and Aharonov-Bohm magnetic field also gives rise to nontrivial properties like the existence of bound states. This question is the main object of the paper. The rest of the paper is organized as follows: in Section 2, we define the model and recall some known results. In section 3, we present the main result of this note followed by a discussion. Section 4 is devoted to numerical computations.

### 3.3.1 The model

Let $H_{AB}$ be the Aharonov-Bohm Schrödinger operator in $\mathrm{L}^2(\Omega)$, defined initially on the domain $\mathrm{C}_0^\infty(\Omega)$, and given by the expression

$$H_{AB} = (i\nabla + \mathbf{A})^2, \tag{22}$$

where **A** is a magnetic vector potential for the Aharonov-Bohm magnetic field **B**, and given by

$$\mathbf{A}(x,y,z) = (A_1, A_2, A_3) = \alpha\left(\frac{y}{x^2+y^2}, \frac{-x}{x^2+y^2}, 0\right), \quad \alpha \in \mathbb{R} \setminus \mathbb{Z}. \tag{23}$$

The magnetic field $\mathbf{B} : \mathbb{R}^3 \to \mathbb{R}^3$ is given by

$$\mathbf{B}(x, y, z) = curl\mathbf{A} = 0 \tag{24}$$

outside the $z$-axis and

$$\int_{\varrho} \mathbf{A} = 2\pi\alpha, \tag{25}$$

where $\varrho$ is a properly oriented closed curve which encloses the $z$-axis. It can be shown that $H_{AB}$ has a four-parameter family of self-adjoint extensions which is constructed by means of von Neumann 's extension theory [8]. Here we are only interested in the Friedrichs extension of $H_{AB}$ on $\mathrm{L}^2(\Omega)$ which can be constructed by means of quadratic forms. We get that the domain $D(\Omega)$ of $H$ is

$$D(\Omega) = \{u \in \mathrm{H}^1(\Omega); \quad (i\nabla + \mathbf{A})^2 u \in \mathrm{L}^2(\Omega), u\lceil\Gamma = 0, \nu.(i\nabla + \mathbf{A})u\lceil\gamma(a) = 0\},$$

where $\nu$ the normal vector and

$$Hu = (i\nabla + \mathbf{A})^2 u, \quad \forall u \in D(\Omega). \tag{26}$$

Let's start by recalling that in the particular case when $a = 0$, we get $H^0$, the magnetic Dirichlet Laplacian, and when $a = +\infty$ we get $H^\infty$, the magnetic Dirichlet-Neumann Laplacian.

**Proposition 2.** *The spectrum of* $H^0$ *is* $[(\frac{\pi}{d})^2, +\infty[$, *and the spectrum of* $H^\infty$ *coincides with* $[(\frac{\pi}{2d})^2, +\infty[$.

*Proof.* We have

$$H = (i\nabla + \tilde{\mathbf{A}})^2 \otimes I \oplus I \otimes (-\Delta_{[0,d]}), \quad \text{on} \quad \mathrm{L}^2(\mathbb{R}^2 \setminus \{0\}) \otimes \mathrm{L}^2([0, d]),$$

where $\tilde{\mathbf{A}} := \alpha\left(\frac{y}{x^2+y^2}, \frac{-x}{x^2+y^2}\right)$. Consider the quadratic form

$$\begin{aligned}\tilde{q}[u] &= \int_{\mathbb{R}^2} \mid (i\nabla + \tilde{\mathbf{A}})u \mid^2 dxdy \\ &= \int_{\mathbb{R}^2} \left|\left(i\partial_x + \alpha\frac{y}{x^2 + y^2}\right)u\right|^2 dxdy + \int_{\mathbb{R}^2} \left|\left(i\partial_y - \alpha\frac{x}{x^2 + y^2}\right)u\right|^2 dxdy.\end{aligned} \tag{27}$$

By introducing polar coordinates we get

$$r = \sqrt{x^2 + y^2}; \qquad \frac{x}{r} = \cos\theta, \qquad \frac{y}{r} = \sin\theta,$$

and

$$\frac{\partial\theta}{\partial x} = \frac{-y}{r^2}, \quad \frac{\partial\theta}{\partial y} = \frac{x}{r^2}, \quad \partial_x = \cos\theta\frac{\partial}{\partial r} - \frac{y}{r^2}\frac{\partial}{\partial\theta}, \quad \partial_y = \sin\theta\frac{\partial}{\partial r} + \frac{x}{r^2}\frac{\partial}{\partial\theta}.$$

Hence (27) becomes

$$\tilde{q}[u] = \int \left( \mid \partial_r u \mid^2 + \frac{1}{r^2} \mid (i\partial_\theta u - \alpha u) \mid^2 \right) r dr d\theta. \tag{28}$$

Expanding $u$ into Fourier series with respect to $\theta$

$$u(r,\theta) = \sum_{k=-\infty}^{\infty} u_k(r)\frac{e^{ik\theta}}{\sqrt{2\pi}}.$$

we get

$$\int_{\mathbb{R}^2} \mid (i\nabla + \tilde{\mathbf{A}})u \mid^2 dxdy \geq \min_k \mid k+\alpha \mid^2 \int \frac{1}{x^2+y^2} \mid u(x,y) \mid^2 dxdy. \tag{29}$$

Here the form in the right-hand side is considered on the function class $\mathrm{H}^1(\mathbb{R}^2)$, obtained by the completion of the class $\mathscr{C}_0^\infty(\mathbb{R}^2\backslash\{0\})$. Inequality (29) is the Hardy inequality in two dimensions with Aharonov-Bohm vector potential [3]. This yields that $\sigma\left((i\nabla + \tilde{\mathbf{A}})^2\right) \subset [0, +\infty[$

Since $\sigma(-\Delta) = \sigma_{ess}(-\Delta) = [0,+\infty[$, then there exists a Weyl sequences $\{h_n\}_{n=1}^\infty$ for the operator $-\Delta$ in $\mathrm{L}^2(\mathbb{R}^2)$ at $\lambda \geq 0$. Construct the functions

$$\varphi_n(x,y) = \begin{cases} h_n \text{ if } x > n \text{ and } y > n, \\ 0 \quad \text{if not.} \end{cases}$$

$$\begin{aligned} \parallel \left((i\nabla + \tilde{\mathbf{A}})^2 - \lambda\right)\varphi_n \parallel &\leq \parallel (\Delta - \lambda)\varphi_n \parallel + \parallel \tilde{\mathbf{A}}^2\varphi_n \parallel + \parallel \tilde{\mathbf{A}}\nabla\varphi_n \parallel \\ &\leq \parallel (\Delta - \lambda)\varphi_n \parallel + \frac{c}{n}, \end{aligned}$$

where $c$ is positive.

Therefore, the functions $\psi_n = \dfrac{\varphi_n}{\parallel \varphi_n \parallel}$ is Weyl sequence for $(i\nabla + \tilde{\mathbf{A}})^2$ at $\lambda \geq 0$, thus $[0,+\infty[\subset \sigma_{ess}\left((i\nabla + \tilde{\mathbf{A}})^2\right) \subset \sigma\left((i\nabla + \tilde{\mathbf{A}})^2\right)$.

Then we get that the spectrum of $(i\nabla + \tilde{\mathbf{A}})^2$ is $[0,+\infty[$, we know that the spectrum of $-\Delta^0_{[0,d]}$ and $-\Delta^\infty_{[0,d]}$ is $\{(\frac{j\pi}{d})^2, \quad j \in \mathbb{N}^\star\}$ and $\{(\frac{(2j+1)\pi}{2d})^2, \quad j \in \mathbb{N}\}$, respectively. Therefore we have the spectrum of $H^0$ is $[(\frac{\pi}{d})^2, +\infty[$. And the spectrum of $H^\infty$ coincides with $[(\frac{\pi}{2d})^2, +\infty[$. ■

Consequently, we have

$$\Big[(\frac{\pi}{d})^2, +\infty\Big[ \subset \sigma(H) \subset \Big[(\frac{\pi}{2d})^2, +\infty\Big[.$$

Using the property that the essential spectra are preserved under compact perturbation, we deduce that the essential spectrum of $H$ is

$$\sigma_{ess}(H) = \Big[(\frac{\pi}{d})^2, +\infty\Big[.$$

**Theorem 3 ([19]).** *Let $H$ be the operator defined on (26) and $\alpha \in \mathbb{R} \setminus \mathbb{Z}$. There exist $a_0 > 0$ such that for any $0 < \frac{a}{d} < a_0$, we have*

$$\sigma_d(H) = \emptyset.$$

*There exist $a_1 > 0$, such that $\frac{a}{d} > a_1$, we have*

$$\sigma_d(H) \neq \emptyset.$$

*Remark.* The presence of magnetic field in three dimensional straight strip of width d with the Neumann boundary condition on a disc window of radius $0 < \frac{a}{d} < a_0$ and Dirichlet boundary conditions on the remained part of the boundary destroys the creation of discrete eigenvalues below the essential spectrum. If $\frac{a}{d} > a_1$, the effect of the magnetic field is reduced.

*Remark.* This result is still true for more general Neumann window containing some disc. To get the optimal result of $a_0$ and $a_1$, we need explicit calculation.

*Proof.* The proof follows the same steps as in the previous two subsections. The main difference is by introducing the magnetic field we get a new Bessel equation we obtain $\frac{1}{P}(i\frac{\partial}{\partial\theta} + \alpha)^2 P$ should be a constant $-(m-\alpha)^2 = -\nu^2$ for $m \in \mathbb{Z}$.

Finally, we get the new equation for $R$

$$R''(r) + \frac{1}{r}R'(r) + [\lambda - k_z^2 - \frac{\nu^2}{r^2}]R(r) = 0. \tag{30}$$

We notice that the equation (30) is a Bessel equation which by the introduction of the term $\alpha$ is different from equation (5). The solutions of (30) could be expressed in terms of Bessel functions. More explicit solutions could be given by considering boundary conditions.

The solution of the equation (30) is given by $R(r) = cJ_\nu(\beta r)$, where $c \in \mathbb{R}^\star$, $\beta^2 = \lambda - k_z^2$ and $J_\nu$ is the Bessel function of first kind of order $\nu$.

We assume that

$$\begin{aligned} R'(a) = 0 &\Leftrightarrow J_\nu(\beta a) = 0 \\ &\Leftrightarrow a\beta = x'_{\nu,n}, \end{aligned} \tag{31}$$

where $x'_{\nu,n}$ is the $n-$th positive zero of the Bessel function $J'_\nu$.

Consequently to equation (31), $H_a^{-,N}$ has a sequence of eigenvalues given by

$$\begin{aligned} \lambda_{j,\nu,n} &= \frac{x'^2_{\nu,n}}{a^2} + k_z^2 \\ &= \frac{x'^2_{\nu,n}}{a^2} + \left(\frac{(2j+1)\pi}{2d}\right)^2. \end{aligned}$$

As we are interested in discrete eigenvalues which belong to $[\left(\frac{\pi}{2d}\right)^2, \left(\frac{\pi}{d}\right)^2)$ only $\lambda_{0,\nu,n}$ intervenes.

If

$$\left(\frac{\pi}{d}\right)^2 \leq \lambda_{0,\nu,n}, \tag{32}$$

then there $H$ does not have a discrete spectrum. We recall that $\nu^2 = (m-\alpha)^2$ and it is related to magnetic flux, also recall that $x'_{\nu,n}$ are the positive zeros of the Bessel function $J'_n$ and $\forall \nu > 0$, $\forall n \in \mathbb{N}^\star$; $0 < x'_{\nu,n} < x'_{\nu,n+1}$ ( see [2]). So, for any eigenvalue of $H_a^{-,N}$,

$$\frac{x'^2_{\nu,1}}{a^2} + \left(\frac{\pi}{2d}\right)^2 < \frac{x'^2_{\nu,n}}{a^2} + \left(\frac{\pi}{2d}\right)^2 = \lambda_{0,\nu,n}.$$

An immediate consequence of the last inequality is that to satisfy (32) it is sufficient to have

$$3\left(\frac{\pi}{2d}\right)^2 < \frac{x'^2_{\nu,1}}{a^2},$$

therefore

$$\frac{\sqrt{3}\pi}{2d} < \frac{x'_{\nu,1}}{a},$$

then

$$\frac{a}{d} < \frac{2x'_{\nu,1}}{\sqrt{3}\pi}.$$

We have( see [2, 27])

$$\nu + \alpha_n \nu^{1/3} < x'_{\nu,n},$$

where $\alpha_n = 2^{-1/3}\beta_n$ and $\beta_n$ is the $n-$th positive root of the equation

$$J_{\frac{2}{3}}\left(\frac{2}{3}x^{3/2}\right) - J_{\frac{-2}{3}}\left(\frac{2}{3}x^{3/2}\right) = 0.$$

For $n = 1$, we have $\alpha_n \nu^{1/3} \approx 0.6538$ (see [2]), then

$$c_0 := 0.6538 + \alpha < 0.6538 + \nu < x'_{\nu,1}. \tag{33}$$

Then we get that for $d$ and $a$ positives such that $\frac{a}{d} < a_0 := \frac{2c_0}{\sqrt{3}\pi}$,

$$\sigma_d(H) = \emptyset.$$

This ends the proof of the first result of the theorem (3).

By the min-max principle and (14), we know that if $H_a^{-,D}$ exhibits a discrete spectrum below $(\frac{\pi}{d})^2$, then $H$ does as well.

$H_a^{-,D}$ has a sequence of eigenvalues [17, 18, 27], given by

$$\lambda_{j,\nu,n} = \left(\frac{x_{\nu,n}}{a}\right)^2 + \left(\frac{(2j+1)\pi}{2d}\right)^2,$$

here $x_{\nu,n}$ is the $n-$th positive zero of Bessel function of order $\nu$ (see [2]). As we are interested in discrete eigenvalues which belong to $[(\frac{\pi}{2d})^2, (\frac{\pi}{d})^2)$ only for $\lambda_{0,\nu,n}$.

If the following condition

$$\lambda_{0,\nu,n} < \left(\frac{\pi}{d}\right)^2 \tag{34}$$

is satisfied, then $H$ has a discrete spectrum.

We recall that $0 < x_{\nu,n} < x_{\nu,n+1}$ for any $\nu > 0$ and any $n \in \mathbb{N}^\star$ ( see [2]). So, for any eigenvalue of $H_a^{-,D}$,

$$\frac{x_{\nu,1}^2}{a^2} + \left(\frac{\pi}{2d}\right)^2 < \frac{x_{\nu,n}^2}{a^2} + \left(\frac{\pi}{2d}\right)^2 = \lambda_{0,\nu,n}.$$

An immediate consequence of the last inequality is that to satisfy (34) it is sufficient to set then

$$\frac{2x_{\nu,1}}{\sqrt{3}\pi} < \frac{a}{d}.$$

We have

$$\sqrt{\left(n - \frac{1}{4}\right)^2 \pi^2 + \nu^2} < x_{\nu,n},$$

For $n = 1$, we have

$$c_1 := \sqrt{\left(\frac{3\pi}{4}\right)^2 + \alpha^2} < \sqrt{\left(\frac{3\pi}{4}\right)^2 + \nu^2} < x_{\nu,1}. \tag{35}$$

Then we get that for $d$ and $a$ positives such that $\frac{a}{d} > a_1 := \frac{2c_1}{\sqrt{3}\pi}$,

$$\sigma_d(H) \neq \emptyset.$$

■

# References


1. R. Assel, M. Ben Salah *Spectral properties of the Dirichlet wave guide with square Neumann window* prepint.
2. M. Abramowitz and I. A. Stegun *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables* New York: Dover, (1972).
3. L. Aermark: *Spectral and Hardy Inequalities for some Sub-Elliptic Operators.* Thesis.
4. Y. Aharonov and D. Bohm: *Significance of Electromagnetic Potentials in the Quantum Theory.Phys. Rev.***115**, 485–491(1959).
5. D. Borisov, P. Exner and R. Gadyl'shin *Geometric coupling thresholds in a two-dimensional strip* Jour. Math. Phy. **43** 6265 (2002)
6. D. Borisov and P. Exner: *Exponential splitting of bound states in a waveguide with a pair of distant windows.* J. Phys. A 37 n° 10, p3411–3428 (2004).
7. D. Borisov, T. Ekholm and H. Kovařík: *Spectrum of the magnetic Schrödinger operator in a waveguide with combined boundary conditions. Ann. Henri Poincaré* **6**, 327–342(2005).

8. J. F. Brasche and M . Melgaard: *The Friedrichs extension of the Aharonov-Bohm Hamiltonian on a disk. Integral Equations and Operator Theory* **52**, 419–436(2005).
9. W. Bulla, F. Gesztesy, W. Renger, and B. Simon: *Weakly coupled Bound States in Quantum Waveguides.* Proc. Amer. Math. Soc. **125** , no. 5, 1487–1495 (1997).
10. P. Duclos and P. Exner: *Curvature-induced Bound States in Quantum waveguides in two and three dimensions* Rev. Math. Phy. (**37**) p 4867–4887 (1989).
11. P. Duclos, P. Exner and B. Meller: *Resonances from perturbed symmetry in open quantum dots. Rep. Math. Phys.* **47**, no. 2, 253–267 (2001).
12. P. Exner, P. Šeba: *Bound states and scattering in quantum waveguides coupled laterally through a boundary window.* J. Math. Phys. **(30)** n° 10, p 2574 (1989).
13. P. Exner, P. Šeba, M. Tater, and D. Vaněk: *Bound states and scattering in quantum waveguides coupled laterally through a boundary window.* J. Math. Phys. **(37)** n° 10, p4867–4887 (1996).
14. P. Exner, S. A Vugalter: *Asymptotic Estimates for Bound States in Quantum Waveguide Coupled laterally through a boundary window.*
15. N. E. Hurt: *Mathematical Physics Of Quantum Wires and Devices* Mathematics and its Application **(506)** Kluer Academic, Dordrecht, (2000)
16. H. Najar: *Lifshitz tails for acoustic waves in random quantum waveguide* Jour. Stat. Phy. Vol 128 No 4, p 1093–1112 (2007).
17. H. Najar, S. Ben Hariz, M. Ben Salah: *On the Discrete Spectrum of a Spatial Quantum Waveguide with a Disc Window* Math. Phys. Anal. Geom. (2010) 13:19–28.
18. H. Najar, O. Olendski : *Spectral and localization properties of the Dirichlet wave guide with two concentric Neumann discs* J. Phys. A: Math. Theor. 44 (2011).
19. H. Najar, M. Raissi : *On the spectrum of the Schrodinger Operator with Aharonov-Bohm Magnetic Field in quantum waveguide with Neumann window* Math. Meth. Appl. Sci. (2015).
20. A. Klein; J. Lacroix, and A. Speis, Athanasios: *Localization for the Anderson model on a strip with singular potentials.* J. Funct. Anal. **(94)** n° 1, p135–155 (1990).
21. F. Kleespies and P. Stollmann: *Lifshitz Asymptotics and Localization for random quantum waveguides.* Rev. Math. Phy. **(12)** p 1345–1365 (2000).
22. D. Krejcirik and J. Kriz: *On the spectrum of curved quantum waveguides* Publ. RIMS, Kyoto University, **(41)**, no. 3 p 757–791, (2005).
23. L. Mikhailovska and O. Olendski: *A straight quantum wave guide with mixed Dirichlet and Neumann boundary conditions in uniform magnetic fields.* Jour. Phy. A. **40**, 4609–4633(2007).
24. S. A. Nazarov and M. Specovius-Neugebauer: *Selfadjoint extensions of the Neumann Laplacian in domains with cylindrical outlets.* Commu. Math. Phy. **185** p 689–707 (1997).
25. M. Reed and B. Simon: *Methods of Modern Mathematical Physics* Vol. IV: Analysis of Operators. Academic, Press, (1978).
26. T. Weidl: *Remarks on virtual bound states for semi-bounded operators. Comm. in Part. Diff. Eq.* **24**, 25–60(1999).
27. G. N. Watson: *A Treatise On The Theory of Bessel Functions* Cambridge University Press.

# Multi-field Modeling of Nonsmooth Problems of Continuum Mechanics, Differential Mixed Variational Inequalities and Their Stability

**Joachim Gwinner**

**Abstract** This paper surveys various nonsmooth problems in continuum mechanics, presents multi-field variational models for these problems in the form of mixed variational inequalities and differential mixed variational inequalities, and exhibits stability results for differential variational inequalities with respect to perturbations of the data.

## 1 Introduction

In this survey paper we consider various nonsmooth problems in continuum mechanics, present multi-field variational models for these problems in the form of mixed variational inequalities and differential mixed variational inequalities (DVIs), and exhibit stability results for DVIs with respect to perturbations of the data.

Herewith nonsmoothness is not understood as nonsmoothness of domains with corners and edges what is investigated in the mathematical regularity theory of partial differential equations. Instead nonsmoothness comes from the nonsmooth data of the problems itself, in particular from constraints and from functionals that are classically not differentiable. Thus for such problems the classical approach of characterization of solutions by variational equations has to be abandoned. Similar to finite dimensional optimization with implicit set constraints we have to work with *variational inequalities* instead. Also similar to finite dimensional optimization with inequality constraints, Lagrange multipliers and - in their physical interpretations - dual variables become signed.

Since it is not a priori known where those dual variables are positive or vanish, respectively where the associated inequality constraint to the Lagrange multiplier becomes active and where not, the boundary between those regions is unknown too. So these problems belong to the class of *free boundary value problems*. Moreover these problems are *nonlinear*.

---

J. Gwinner (✉)
Fakultät für Luft- und Raumfahrttechnik, Institut für Mathematik und Rechneranwendung, Universität der Bundeswehr München, 85577 Neubiberg/München, Germany
e-mail: joachim.gwinner@unibw.de

 
A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_5

On the other hand, with a linear regime, e.g. with linear elasticity in solid mechanics, we arrive at convex variational problems, i.e. the energy becomes a convex functional. This allows to apply methods of convex analysis. We also can treat nonlinear material behavior. Then monotonicity methods of nonlinear analysis come into play. But we do not enter solid mechanic problems with geometric nonlinearity and large deformations.

To introduce into multiple field modeling in continuum mechanics let us shortly review the literature. The classical Babuška-Brezzi theory [5] addresses variational problems with equality constraints as saddle point problems and establishes unique solvability by means of inf-sup conditions what are the basis of the mixed finite element method (mixed FEM) for the numerical treatment of these problems. A prominent instance of this two-field modeling is incompressible Stokes flow with the velocity and pressure as coupled unknown fields. This classical saddle point theory linked with two-field modeling has been extended by Gatica [17, 18] to some classes of nonlinear variational problems and to three-field variational models that can be understood as dual-dual mixed variational models or as two-fold saddle point formulations. Such augmented variational models are well adapted for multi-physics problems with different coupled unknown quantities and in particular for engineering problems, where in terms of solid mechanics, strains and stresses are often of more interest than the displacements.

The present paper extends this novel modeling approach to nonsmooth boundary value problems with inequality constraints. This has to be distinguished from the standard duality approach which hinges on the Lagrange duality theory of convex analysis in calculus of variations (see [13] for a systematic study) and which is employed in the numerical FEM analysis of various unilateral boundary and obstacle problems as pioneered by Haslinger and Lovíšek [28, 29], see also the monograph [31]. More recent work in this direction is [10] on error estimates of mixed FEM applied to Signorini elastic contact and [47] on mixed formulations and mixed FEM for a class of variational inequalities in a more general framework; see also the recent exhaustive exposition [48] on variationally consistent discretization schemes for the frictional contact problem. More related to, but different from the present paper is [11] that investigates nonconvex unilateral contact problems and analyzes a three-field augmented Lagrangian formulation in the triple of the bulk displacement, normal boundary displacement, and normal boundary stress fields.

In addition to steady-state problems we also consider time-dependent problems. Here we focus on problems that are first order in time neglecting mass effects and that are thus of parabolic type. Because of the presence of constraints these problems fit to the frame of differential mixed variational inequalities (DMVIs). For such nonsmooth evolution problems we present very recent stability results. Herewith stability is not understood as stability over an infinite time horizon like Liapunov stability. Instead we are concerned with stability of solutions in a fixed time interval with respect to perturbations of the data. These stability results can be interpreted as set convergence results for the solution set and are very much linked to the stability of numerical schemes that arise from the application of numerical discretization techniques like boundary element, finite difference, finite element, and finite volume methods.

The paper has the following outline. In the subsequent section we introduce to the multi-field modeling approach and consider a scalar nonlinear boundary problem from heat conduction and its time-dependent extension what leads to a first example of a (DMVI). Then in section 3 we extend the multi-field modeling of a boundary value problem in nonlinear material elasticity given in [4] to a frictionless unilateral contact problem. Section 4 addresses nonsmooth problems in elastoplasticity. First we recall from [26, 27] the variational formulation as a multi-field evolutionary variational inequality, which can be readily rewritten as a (DMVI). Then we expand further this formulation, using the modeling approach of [4] and of section 3, and arrive at another (DMVI). The bulk of the more analytic results is in section 5. There we survey the recent stability results from [25], discuss the nonlinear heat conduction problem as an example, and provide more applications. In the final section we delineate some further directions of research.

## 2 Multiple field modelling of a nonsmooth heat conduction problem

To demonstrate this novel approach of multiple field modelling, let us consider a nonlinear boundary value problem with Signorini boundary conditions that arises from nonlinear heat conduction [8] with semi-permeable walls [12]. We first show how the steady-state problem can be variationally formulated as a variational inequality in mixed form. Then we turn to the transient problem and derive the associated differential mixed variational inequality.

To describe the problem of interest, let $\Omega$ be a bounded simply connected domain in $\mathbb{R}^2$ with the Lipschitz boundary $\partial\Omega = \Gamma$. Then $\boldsymbol{\nu}$ the outward normal to $\Gamma$, exists almost everywhere and $\boldsymbol{\nu} \in [L^\infty(\Gamma)]^2$. Let $\Gamma_D$ and $\Gamma_S$ be parts of $\Gamma$ such that $|\Gamma_D| > 0, |\Gamma_S| > 0, \Gamma = \overline{\Gamma}_D \cup \overline{\Gamma}_S$, and $\Gamma_D \cap \Gamma_S = \emptyset$. Also let $a_i : \Omega \times \mathbb{R}^2 \to \mathbb{R}, i = 1, 2$ be nonlinear functions satisfying certain conditions (specified in what follows) and write

$$\mathbf{a}(x, \xi) := \big(a_1(x, \xi), a_2(x, \xi)\big)^T \text{ for all } (x, \xi) \in \Omega \times \mathbb{R}^2.$$

Then, for a given right-hand side $f : \Omega \to \mathbb{R}$, given function $g : \partial\Omega \to \mathbb{R}$ defining the boundary conditions, we look for a function $u : \overline{\Omega} \to \mathbb{R}$ such that

$$\left.\begin{aligned} -\text{div}\, \mathbf{a}(\cdot, \nabla u(\cdot)) &= f && \text{on } \Omega \\ u &= g && \text{on } \Gamma_D \\ u \geq g, \mathbf{a}(\cdot, \nabla u(\cdot)) \cdot \boldsymbol{\nu} \geq 0, &\quad (u - g)\, \mathbf{a}\big(\cdot, \nabla u(\cdot)\big) \cdot \boldsymbol{\nu} = 0 && \text{on } \Gamma_S\,, \end{aligned}\right\} \tag{1}$$

where div is the usual divergence operator. As in [19] in the case of the Dirichlet problem, let us introduce the gradient $\mathbf{p} := \nabla u$ in $\Omega$ and the flux $\boldsymbol{\sigma} := \mathbf{a}(\cdot, \mathbf{p})$ in $\Omega$ as additional unknowns. In this way, the elliptic pde $(1_1)$ in (1) writes as the three equations

$$\left.\begin{aligned} -\text{div}\,\boldsymbol{\sigma} &= f \\ \mathbf{a}(\cdot,\mathbf{p}) &= \boldsymbol{\sigma} \\ \mathbf{p} &= \nabla u \end{aligned}\right\} \tag{2}$$

that should hold in the distributional sense in $\Omega$. By this reformulation we can relax the regularity of the unknown $u$. We require that $u \in L^2(\Omega), \mathbf{p} \in L^2(\Omega, \mathbb{R}^2)$ and $\boldsymbol{\sigma} \in H(\text{div}\,, \Omega) := \{\boldsymbol{\sigma} \in L^2(\Omega, \mathbb{R}^2) \,|\, \text{div}\,\boldsymbol{\sigma} \in L^2(\Omega)\}$. Thus testing $(1_1)$, (2) with $v \in L^2(\Omega), \mathbf{q} \in L^2(\Omega, \mathbb{R}^2)$, respectively, gives

$$-\int_\Omega \text{div}\,\boldsymbol{\sigma}\; v\, d\mathbf{x} = \int_\Omega f\, v\, d\mathbf{x} =: \langle f, v\rangle_{L^2(\Omega)}\,,$$

$$\int_\Omega \mathbf{a}(\cdot,\mathbf{p}) \cdot \mathbf{q}\, d\mathbf{x} = \int_\Omega \boldsymbol{\sigma} \cdot \mathbf{q}\, d\mathbf{x} =: \langle \boldsymbol{\sigma}, \mathbf{q}\rangle_{L^2(\Omega,\mathbb{R}^2)}\,.$$

For the last equation $(2)_3$ in (2) we incorporate the boundary conditions and use Green's formula (see [20])

$$\langle \nabla\varphi, \boldsymbol{\tau}\rangle_{L^2(\Omega,\mathbb{R}^2)} + \langle \varphi, \text{div}\,\boldsymbol{\tau}\rangle_{L^2(\Omega)} = \langle \gamma_0\varphi, \gamma_{\boldsymbol{\nu}}\boldsymbol{\tau}\rangle_{L^2(\Gamma)} \tag{3}$$

for $\varphi \in H^1(\Omega), \boldsymbol{\tau} \in H(\text{div}\,, \Omega)$, where $\gamma_0\varphi$ and $\gamma_{\boldsymbol{\nu}}\boldsymbol{\tau} = \boldsymbol{\tau} \cdot \boldsymbol{\nu}$ denote the traces. Testing $(2)_3$ with $\boldsymbol{\tau} - \boldsymbol{\sigma} \in H(\text{div}\,, \Omega)$ results in

$$\begin{aligned} &\langle \mathbf{p}, \boldsymbol{\tau} - \boldsymbol{\sigma}\rangle_{L^2(\Omega,\mathbb{R}^2)} + \langle u, \text{div}\,\boldsymbol{\tau} - \text{div}\,\boldsymbol{\sigma}\rangle_{L^2(\Omega)} \\ &= \langle \gamma_0 u, \gamma_{\boldsymbol{\nu}}(\boldsymbol{\tau} - \boldsymbol{\sigma})\rangle_{L^2(\Gamma)} \\ &= \langle \gamma_0 u - g, \gamma_{\boldsymbol{\nu}}\boldsymbol{\tau}\rangle - \langle \gamma_0 u - g, \gamma_{\boldsymbol{\nu}}\boldsymbol{\sigma}\rangle + \langle g, \gamma_{\boldsymbol{\nu}}(\boldsymbol{\tau} - \boldsymbol{\sigma})\rangle\,. \end{aligned}$$

By the boundary conditions, the second term above vanishes. The second inequality in the Signorini boundary condition $(1)_3$ tells us that we have to require that $\boldsymbol{\sigma}$ belongs to the closed convex cone

$$H_+ := H_+(\text{div}\,, \Omega, \Gamma_S) := \left\{\boldsymbol{\tau} \in H(\text{div}\,, \Omega) : \gamma_{\boldsymbol{\nu}}\boldsymbol{\tau} | \Gamma_S \geq 0\right\},$$

where "$\geq 0$" means that $\langle \gamma_0\varrho, \gamma_{\boldsymbol{\nu}}\boldsymbol{\tau}\rangle \geq 0$ for any smooth function $\varrho$ on $\overline{\Omega}$ with $\varrho = 0$ on $\Gamma_D$ and $\varrho \geq 0$ on $\Gamma_S$. Thus we obtain for any $\boldsymbol{\tau} \in H_+(\text{div}, \Omega, \Gamma_S)$,

$$\begin{aligned} &\langle \mathbf{p}, \boldsymbol{\tau} - \boldsymbol{\sigma}\rangle_{L^2(\Omega,\mathbb{R}^2)} + \langle u, \text{div}(\boldsymbol{\tau} - \boldsymbol{\sigma})\rangle_{L^2(\Omega)} \\ &\geq \langle g, \gamma_{\boldsymbol{\nu}}(\boldsymbol{\tau} - \boldsymbol{\sigma})\rangle_{L^2(\Gamma_S)}\,. \end{aligned}$$

Altogether we arrive at the following variational problem in mixed form: Find $[u, \boldsymbol{\sigma}, \mathbf{p}] \in L^2(\Omega) \times H_+(\text{div}\,, \Omega, \Gamma_S) \times L^2(\Omega, \mathbb{R}^2)$ such that for all $[v, \boldsymbol{\tau}, \mathbf{q}] \in L^2(\Omega) \times H_+(\text{div}\,, \Omega, \Gamma_S) \times L^2(\Omega, \mathbb{R}^2)$,

$$\left. \begin{aligned} -\langle \operatorname{div} \boldsymbol{\sigma}, v\rangle &= \langle f, v\rangle\,, \\ \langle \mathbf{p}, \boldsymbol{\tau}-\boldsymbol{\sigma}\rangle + \langle u, \operatorname{div}(\boldsymbol{\tau}-\boldsymbol{\sigma})\rangle &\geq \langle g, \gamma_{\boldsymbol{\nu}}(\boldsymbol{\tau}-\boldsymbol{\sigma})\rangle_{\Gamma_S}\,, \\ \langle \mathbf{a}(\cdot,\mathbf{p}), \mathbf{q}\rangle \qquad\qquad -\langle \boldsymbol{\sigma}, \mathbf{q}\rangle &= 0\,. \end{aligned} \right\} \tag{4}$$

Note that if the functions $a_i$ are Caratheodory functions and satisfy the growth conditions : $\exists\, C > 0, \varphi_i \in L^2(\Omega), i = 1, 2$ such that

$$|a_i(x,\xi)| \leq C\{1+|\xi|\} + |\varphi_i(x)|, \;\; \forall\; \xi \in \mathbb{R}^2, \;\text{a.e. } x \in \Omega,$$

then we obtain the Nemytskii operator $\mathbf{A} : \mathbf{q} \in L^2(\Omega, \mathbb{R}^2) \mapsto \mathbf{a}(\cdot, \mathbf{q}) \in L^2(\Omega, \mathbb{R}^2)$. Clearly, this nonlinear operator $\mathbf{A}$ inherits monotonicity from the vector field $\mathbf{a}$, see, e.g., [35]. To reveal the separable structure in (4), (1), introduce the linear operator $B : H(\operatorname{div}, \Omega) \to L^2(\Omega), B\boldsymbol{\tau} := \operatorname{div} \boldsymbol{\tau}$ and the embedding $E : H(\operatorname{div}, \Omega) \to L^2(\Omega, \mathbb{R}^2), E\boldsymbol{\tau} := \boldsymbol{\tau}$. Then with the zero operator 0 at appropriate places, the left-hand side of (4), (1) is given by the following operator in block form:

$$\begin{pmatrix} 0 & -B & 0 \\ B^* & 0 & E^* \\ 0 & -E & \mathbf{A}(\cdot) \end{pmatrix} \tag{5}$$

Now we turn to the time-dependent case. Let $I = (0, T)$ be the given time interval. Then instead of the elliptic pde $(1)_1$ we have the parabolic pde

$$\partial_t u - \operatorname{div} \mathbf{a}\big(\cdot, \nabla u(\cdot)\big) = f \text{ in } \Omega \times I.$$

Also the boundary conditions have to hold in $\Omega \times I$, where we assume the gap function $g$ to be time-independent. In addition, we have the initial condition $u = u_0$ in $\Omega \times \{0\}$ with some given $u_0$.

As the solution space for the unknown $u$ we introduce

$$W := \big\{u \in L^2(\Omega \times I) : \partial_t u \in L^2(\Omega \times I)\big\}\,.$$

Then the three-field variational formulation of the considered initial boundary value problem reads as the following differential mixed variational inequality: Find $[u, \boldsymbol{\sigma}, \mathbf{p}] \in W \times L^2\big(I, H_+(\operatorname{div}, \Omega, \Gamma_S)\big) \times L^2(\Omega \times I, \mathbb{R}^2) =: \mathscr{X}$ such that for all $[v, \boldsymbol{\tau}, \mathbf{q}] \in \mathscr{X}$ there holds

$$\left. \begin{aligned} \int_I \int_\Omega (\partial_t u - \operatorname{div} \boldsymbol{\sigma})\, v \; dx\, dt &= \int_I \int_\Omega f\, v\; dx\, dt, \\ \int_I \int_\Omega u \operatorname{div}(\boldsymbol{\tau}-\boldsymbol{\sigma})\, dx\, dt + \int_I \int_\Omega \mathbf{p}\cdot(\boldsymbol{\tau}-\boldsymbol{\sigma})\, dx\, dt & \\ \geq \int_I \int_{\Gamma_S} g(\boldsymbol{\tau}-\boldsymbol{\sigma})\cdot\boldsymbol{\nu}\; ds\, dt & \\ -\int_I \int_\Omega \boldsymbol{\sigma}\cdot\mathbf{q}\; dx\, dt + \int_I \int_\Omega \mathbf{a}(\cdot,\mathbf{p})\cdot\mathbf{q}\; dx\, dt &= 0, \,. \end{aligned} \right\} \tag{6}$$

## 3 Multiple field modelling of frictionless unilateral contact problems in nonlinear elasticity

In this section we extend the two-fold saddle point approach of [4] from the nonlinear elasticity boundary value problem with Dirichlet/Neumann boundary conditions to frictionless unilateral contact problems including Signorini type boundary conditions. We show how the steady-state problem can be variationally formulated as a variational inequality in mixed form.

Let a hyperelastic body occupy a bounded simply connected domain $\Omega$ in $\mathbb{R}^2$ with the Lipschitz boundary $\partial\Omega = \Gamma$. Then the outward normal $\boldsymbol{\nu}$ to $\Gamma$ is in $[L^\infty(\Gamma)]^2$. Let $\Gamma_D, \Gamma_N$, and $\Gamma_S$ be parts of $\Gamma$ such that $|\Gamma_D| > 0, |\Gamma_C| > 0$, $\Gamma = \overline{\Gamma}_D \cup \overline{\Gamma}_N \cup \overline{\Gamma}_C$. and $\Gamma_D \cap \Gamma_C = \emptyset$.

The hyperelastic behavior of the body is assumed to be described by a Hencky-von Mises stress-strain relation as discussed in [9, 41]. Then the Cauchy stress tensor $\boldsymbol{\sigma} \in \mathbb{R}^{2\times 2}$ depends on the displacement field $\mathbf{u} \in \mathbb{R}^2$ via the kinematic relation

$$\mathbf{e} = \frac{1}{2}\,[\mathbf{grad\,u} + \mathbf{grad\,u}^T] \tag{7}$$

with the linearized strain tensor $\mathbf{e}$ and the constitutive equation

$$\boldsymbol{\sigma} = \tilde{\lambda}(\mathrm{dev}(\mathbf{e}))\,\mathrm{tr}(\mathbf{e})\,\mathbf{I} + \tilde{\mu}(\mathrm{dev}(\mathbf{e}))\,\mathbf{e}\,. \tag{8}$$

Here $\tilde{\lambda}, \tilde{\mu}$ are nonlinear Lamé functions that under appropriate assumptions on the stored energy function give rise to a monotone stress-strain relation. Moreover, $\mathbf{I}$ is the identity matrix in $\mathbb{R}^{2\times 2}$, $\mathrm{tr}(\boldsymbol{\tau}) = \boldsymbol{\tau}_{jj}$ denotes the trace, $\mathrm{dev} : \mathbb{R}^{2\times 2} \to \mathbb{R}^+$ is defined by $\mathrm{dev}(\boldsymbol{\tau}) = (\boldsymbol{\tau} - \frac{1}{2}\,\mathrm{tr}(\boldsymbol{\tau})\,\mathbf{I}) : (\boldsymbol{\tau} - \frac{1}{2}\,\mathrm{tr}(\boldsymbol{\tau})\,\mathbf{I})$, where $\boldsymbol{\tau} : \boldsymbol{\sigma} = \boldsymbol{\tau}_{ij}\boldsymbol{\sigma}_{ij}$ is the scalar product for matrices in $\mathbb{R}^{2\times 2}$.

Then, for given right-hand sides $\mathbf{f} \in [L^2\Omega)]^2, \mathbf{g} \in [H^{\frac{1}{2}}(\Gamma_D)]^2, h \in H^{\frac{1}{2}}(\Gamma_C)$ our nonlinear boundary value problem reads as follows: Find a tensor field $\boldsymbol{\sigma}$ and a vector field $\mathbf{u}$ that satisfy (7), (8), the equilibrium equation

$$-\,\mathbf{div}\boldsymbol{\sigma} = \mathbf{f} \;\text{ in } \Omega \tag{9}$$

and the boundary conditions

$$\left.\begin{aligned} \mathbf{u} = \mathbf{g} \;\; &\text{on } \Gamma_D \\ \boldsymbol{\sigma}\,\boldsymbol{\nu} = \mathbf{0} \;\; &\text{on } \Gamma_N \\ u_{\nu} \le h, \boldsymbol{\sigma}_{\nu} \le 0, & \\ (u_{\nu} - h)\sigma_{\nu} = 0, \boldsymbol{\sigma}_{t} = \mathbf{0} \;\; &\text{on } \Gamma_C\,, \end{aligned}\right\} \tag{10}$$

where $u_\nu := \mathbf{u} \cdot \boldsymbol{\nu}$, $\sigma_\nu := \boldsymbol{\nu} \cdot \boldsymbol{\sigma}\,\boldsymbol{\nu}$, $\boldsymbol{\sigma}_t := \boldsymbol{\sigma}\,\boldsymbol{\nu} - \sigma_\nu \boldsymbol{\nu}$, and for any $\boldsymbol{\tau} = (\tau_{ij}) \in \mathbb{R}^{2\times 2}$,

$$\mathbf{div}\,\boldsymbol{\tau} := \begin{pmatrix} \mathrm{div}\,(\tau_{11}\tau_{12}) \\ \mathrm{div}\,(\tau_{21}\tau_{22}) \end{pmatrix}.$$

To derive a multi-field variational formulation of the above boundary value problem, introduce the tensor spaces

$$H(\mathbf{div}\,,\Omega) := \{\boldsymbol{\tau} \in L^2(\Omega,\mathbb{R}^{2\times 2}) :\ \mathbf{div}\,\boldsymbol{\tau} \in L^2(\Omega,\mathbb{R}^2)\}$$
$$H_0(\mathbf{div}\,,\Omega) := \{\boldsymbol{\tau} \in H(\mathbf{div}\,,\Omega) :\ \boldsymbol{\tau}\,\boldsymbol{\nu} = \mathbf{0} \text{ on } \Gamma_N, \boldsymbol{\tau}_t = \mathbf{0} \text{ on } \Gamma_C\}$$

with scalar product

$$\langle \boldsymbol{\tau}, \boldsymbol{\sigma} \rangle_{H(\mathbf{div}\,,\Omega)} := \langle \boldsymbol{\tau}, \sigma \rangle_{L^2(\Omega,\mathbb{R}^{2\times 2})} + \langle \mathbf{div}\,\boldsymbol{\tau}, \mathbf{div}\,\sigma \rangle_{L^2(\Omega,\mathbb{R}^2)}$$

where

$$\langle \boldsymbol{\tau}, \boldsymbol{\sigma} \rangle_{L^2(\Omega,\mathbb{R}^{2\times 2})} := \int_\Omega \boldsymbol{\tau} : \boldsymbol{\sigma}\, dx \quad \forall \boldsymbol{\tau}, \boldsymbol{\sigma} \in L^2(\Omega,\mathbb{R}^{2\times 2})$$
$$\langle \mathbf{v}, \mathbf{w} \rangle_{L^2(\Omega,\mathbb{R}^2)} := \int_\Omega \mathbf{v} \cdot \mathbf{w}\, dx \quad \forall \mathbf{v}, \mathbf{w} \in L^2(\Omega,\mathbb{R}^2)\,.$$

Moreover we need the convex closed cone

$$H_0^-(\mathbf{div}\,,\Omega) := \{\boldsymbol{\tau} \in H_0(\mathbf{div}\,,\Omega) :\ \tau_\nu \le 0 \text{ on } \Gamma_C\}$$

As in [4] in the case of a mixed boundary value problem, the strain field **e** will be an extra unknown. Here (7) will be rewritten to

$$\mathbf{e} = \mathbf{grad}\,\mathbf{u} - \boldsymbol{\rho} \tag{11}$$

where the further unknown

$$\boldsymbol{\rho} := \frac{1}{2}(\mathbf{grad}\,\mathbf{u} - (\mathbf{grad}\,\mathbf{u})^T)\,,$$

is the skew-symmetric part of the deformation tensor **grad u**, represents rotations, and lies in the space

$$\mathscr{R} := \{\boldsymbol{\rho} \in L^2(\Omega,\mathbb{R}^{2\times 2}) :\ \boldsymbol{\rho} + \boldsymbol{\rho}^T = \mathbf{0}\}\,.$$

Then in view of the boundary condition $(10)_3$ we impose $\boldsymbol{\sigma} \in H_0^-(\mathbf{div}\,,\Omega)$, multiply (11) by $\boldsymbol{\tau} - \boldsymbol{\sigma}$ with a test function $\boldsymbol{\tau} \in H_0^-(\mathbf{div}\,,\Omega)$, integrate by part on $\Omega$, use the boundary conditions $(10)_1$, $(10)_2$ to get

$$\int_\Omega \mathbf{e} : (\boldsymbol{\tau} - \boldsymbol{\sigma})\,dx + \int_\Omega \mathbf{u} \cdot \mathbf{div}\,(\boldsymbol{\tau} - \boldsymbol{\sigma})\,dx + \int_\Omega \boldsymbol{\rho} : (\boldsymbol{\tau} - \boldsymbol{\sigma})\,dx$$
$$= \int_{\Gamma_D} \mathbf{g} \cdot (\boldsymbol{\tau} - \boldsymbol{\sigma})\boldsymbol{\nu}\,dx + \int_{\Gamma_C} \mathbf{u} \cdot (\boldsymbol{\tau} - \boldsymbol{\sigma})\boldsymbol{\nu}\,dx$$

and by the boundary conditions $(10)_1$, $(10)_2$ estimate the latter boundary integral,

$$\int_{\Gamma_C} \mathbf{u} \cdot (\boldsymbol{\tau} - \boldsymbol{\sigma})\boldsymbol{\nu}\,dx$$
$$= \int_{\Gamma_C} h\,\boldsymbol{\nu} \cdot (\boldsymbol{\tau} - \boldsymbol{\sigma})\boldsymbol{\nu}\,dx + \int_{\Gamma_C} (u_\nu - h)\boldsymbol{\nu} \cdot \boldsymbol{\tau}\,\boldsymbol{\nu}\,dx$$
$$\geq \int_{\Gamma_C} h\,\boldsymbol{\nu} \cdot (\boldsymbol{\tau} - \boldsymbol{\sigma})\boldsymbol{\nu}\,dx$$

Altogether we obtain the following variational inequality for $\boldsymbol{\sigma} \in H_0^-(\mathbf{div}\,, \Omega)$:

$$\int_\Omega \mathbf{e} : (\boldsymbol{\tau} - \boldsymbol{\sigma})\,dx + \int_\Omega \mathbf{u} \cdot \mathbf{div}\,(\boldsymbol{\tau} - \boldsymbol{\sigma})\,dx + \int_\Omega \boldsymbol{\rho} : (\boldsymbol{\tau} - \boldsymbol{\sigma})\,dx \tag{12}$$
$$\geq \int_{\Gamma_D} \mathbf{g} \cdot (\boldsymbol{\tau} - \boldsymbol{\sigma})\boldsymbol{\nu}\,dx + \int_{\Gamma_C} h\,\boldsymbol{\nu} \cdot (\boldsymbol{\tau} - \boldsymbol{\sigma})\boldsymbol{\nu}\,dx\,, \forall \boldsymbol{\tau} \in H_0^-(\mathbf{div}\,, \Omega)\,.$$

Further with the shorthands

$$\hat{\lambda}(\mathbf{e}) := \tilde{\lambda}(\mathrm{dev}(\mathbf{e}))\,, \quad \hat{\mu}(\mathbf{e}) := \tilde{\mu}(\mathrm{dev}(\mathbf{e}))$$

the constitutive equation (8) and the equilibrium equation (9), respectively, yield

$$\textstyle\int_\Omega \hat{\lambda}(\mathbf{e})\mathrm{tr}(\mathbf{e})\mathrm{tr}(\mathbf{d}) + \hat{\mu}(\mathbf{e})\mathbf{e} : \mathbf{d}\,dx - \int_\Omega \boldsymbol{\sigma} : \mathbf{d}\,dx = 0\,, \forall \mathbf{d} \in L^2(\Omega, \mathbb{R}^{2\times 2}) \tag{13}$$
$$\textstyle -\int_\Omega \mathbf{v} \cdot \mathbf{div}\,\boldsymbol{\sigma}\,dx = \int_\Omega \mathbf{f} \cdot \mathbf{v}\,dx\,, \forall \mathbf{v} \in L^2(\Omega, \mathbb{R}^2) \tag{14}$$

Finally, the symmetry of $\boldsymbol{\sigma}$ is weakly required by

$$\int_\Omega \boldsymbol{\sigma} : \boldsymbol{\rho}\,dx = 0\,, \forall \boldsymbol{\rho} \in \mathscr{R}\,. \tag{15}$$

Consequently, collecting (12), (13), (14), and (15) we arrive at the following multi-field variational problem: Find $[\mathbf{e}, \boldsymbol{\sigma}, \mathbf{u}, \boldsymbol{\rho}] \in L^2(\Omega, \mathbb{R}^{2\times 2}) \times H_0^-(\mathbf{div}\,, \Omega) \times L^2(\Omega, \mathbb{R}^2) \times \mathscr{R}$ such that for all $[\mathbf{d}, \boldsymbol{\tau}, \mathbf{v}, \boldsymbol{\pi}] \in L^2(\Omega, \mathbb{R}^{2\times 2}) \times H_0^-(\mathbf{div}\,, \Omega) \times L^2(\Omega, \mathbb{R}^2) \times \mathscr{R}$,

$$\langle \hat{\lambda}(\mathbf{e})\mathrm{tr}(\mathbf{e}), \mathrm{tr}(\mathbf{d})\rangle_{L^2(\Omega)} + \langle \hat{\mu}(\mathbf{e})\mathbf{e}, \mathbf{d}\rangle_{L^2(\Omega,\mathbb{R}^{2\times 2})} - \langle \boldsymbol{\sigma}, \mathbf{d}\rangle_{L^2(\Omega,\mathbb{R}^{2\times 2})} = 0\,,$$
$$\langle \mathbf{e}, \boldsymbol{\tau} - \boldsymbol{\sigma}\rangle_{L^2(\Omega,\mathbb{R}^{2\times 2})} + \langle \mathbf{u}, \mathbf{div}\,(\boldsymbol{\tau} - \boldsymbol{\sigma})\rangle_{L^2(\Omega,\mathbb{R}^2)} + \langle \boldsymbol{\rho}, \boldsymbol{\tau} - \boldsymbol{\sigma}\rangle_{L^2(\Omega,\mathbb{R}^{2\times 2})}$$
$$\geq \langle \mathbf{g}, (\boldsymbol{\tau} - \boldsymbol{\sigma})\boldsymbol{\nu}\rangle_{L^2(\Gamma_D,\mathbb{R}^2)}\ , + \langle h, \boldsymbol{\nu}\cdot(\boldsymbol{\tau} - \boldsymbol{\sigma})\boldsymbol{\nu}\rangle_{L^2(\Gamma_C)}\,,$$
$$\langle \mathbf{v}, \mathbf{div}\,\boldsymbol{\sigma}\rangle_{L^2(\Omega,\mathbb{R}^2)} + \langle \boldsymbol{\sigma}, \boldsymbol{\pi}\rangle_{L^2(\Omega,\mathbb{R}^{2\times 2})} = -\langle \mathbf{f}, \mathbf{v}\rangle_{L^2(\Omega,\mathbb{R}^2)}\ .$$

## 4 Multiple field modelling in quasistatic elastoplasticity and differential mixed variational inequalities

In this section we first recall from [26, 27] the variational formulation of quasistatic problems in elastoplasticity as an evolutionary variational inequality. This evolutionary variational inequality can be readily cast as a (DMVI). Moreover, by a procedure similar to [4] exploited in the previous section, we expand this multi-field model by adding the elastic stress and elastic strain fields as further unknown variable. This involves the $H(\mathbf{div}, \Omega)$ space and is more related to the classical mixed approach of the Babuška-Brezzi theory. Also this new expanded formulation fits into the (DMVI) setting, which is investigated in a more formal way in the subsequent section.

### *4.1 Primal formulation*

Let the elastoplastic body occupy a bounded simply connected domain $\Omega$ in $\mathbb{R}^3$ with the Lipschitz boundary $\partial\Omega = \Gamma$. Following [26, 27], the unknowns are the displacement $\mathbf{u}$, the plastic strain $\mathbf{p}$, and the internal hardening variable $\boldsymbol{\xi}$ that are required to satisfy in $\Omega$ the equilibrium equation

$$-\,\mathbf{div}\boldsymbol{\sigma} = \mathbf{f} \tag{16}$$

for some given volume load vector $\mathbf{f}$, the kinematic relation

$$\mathbf{e} = \frac{1}{2}\,[\mathbf{grad}\,\mathbf{u} + \mathbf{grad}\,\mathbf{u}^T] \tag{17}$$

with the linearized strain tensor $\mathbf{e}$ and the constitutive equation

$$\boldsymbol{\sigma} = \mathbf{C}(\mathbf{e} - \mathbf{p}) \tag{18}$$

with the fourth order elasticity tensor $\mathbf{C}$ and the plastic strain $\mathbf{p}$, and moreover, the flow law,

$$\left.\begin{array}{l} [\dot{\mathbf{p}}, \dot{\boldsymbol{\xi}}] \in K_p \\ D(\mathbf{q}, \boldsymbol{\eta}) \geq D(\dot{\mathbf{p}}, \dot{\boldsymbol{\xi}}) + \boldsymbol{\sigma} : (\mathbf{q} - \dot{\mathbf{p}}) - (\mathbf{H}\boldsymbol{\xi})\cdot(\boldsymbol{\eta} - \dot{\boldsymbol{\xi}})\,, \forall [\mathbf{q}, \boldsymbol{\eta}] \in K_p \end{array}\right\} \tag{19}$$

with the hardening modulus $\mathbf{H} \in \mathbb{R}^{m \times m}$ and with $K_p = \text{dom}\, D$. There the dissipation function $D$ is assumed to be nonnegative, convex, positively homogeneous with $D(\mathbf{0}) = 0$ and $K_p$ closed.

The initial condition is $\mathbf{u}(\cdot, 0) = \mathbf{0}$ and for simplicity zero Dirichlet boundary conditions are prescribed, i.e. $\mathbf{u} = \mathbf{0}$ on $\Gamma$ and in the time interval $[0, T]$.

For the variational formulation introduce the following function spaces. The displacement, plastic strain, internal variable live in

$$V = [H_0^1(\Omega)]^3, Q_0 := \{\mathbf{q} \in L^2_{\text{symm}}(\Omega, \mathbb{R}^3) : \text{tr}\, \mathbf{q} = 0\}, M = [L^2(\Omega)]^m,$$

respectively. This gives the product space $Y := V \times Q_0 \times M$. Corresponding to the set $K_p = \text{dom}\, D$ is the closed, convex cone

$$Y_p = \{y = [\mathbf{v}, \mathbf{q}, \boldsymbol{\eta}] \in Y \ : \ [\mathbf{q}, \boldsymbol{\eta}] \in K_p \text{ a.e. in } \Omega\}.$$

With [26, 27], (17) is understood as a function $\mathbf{e} = \mathbf{e}(\mathbf{u})$. Integrating (19) and using (17), (18) gives for $[\dot{\mathbf{p}}, \dot{\boldsymbol{\xi}}] \in K_p$,

$$\left.\begin{aligned} &\textstyle\int_\Omega D(\mathbf{q}, \boldsymbol{\eta})\, dx - \int_\Omega D(\dot{\mathbf{p}}, \dot{\boldsymbol{\xi}})\, dx \\ &\geq \textstyle\int_\Omega \big(\mathbf{C}(\mathbf{e}(\mathbf{u}) - \mathbf{p}) : (\mathbf{q} - \dot{\mathbf{p}}) - (\mathbf{H}\boldsymbol{\xi}) \cdot (\boldsymbol{\eta} - \dot{\boldsymbol{\xi}})\big)\, dx\,, \forall [\mathbf{q}, \boldsymbol{\eta}] \in K_p\,. \end{aligned}\right\} \tag{20}$$

Testing (16) with $\mathbf{v} - \dot{\mathbf{u}}$ for arbitrary $\mathbf{v} \in V$, integrating by parts, and using (18) yield

$$\int_\Omega \mathbf{C}(\mathbf{e}(\mathbf{u}) - \mathbf{p}) : (\mathbf{e}(\mathbf{v}) - \mathbf{e}(\dot{\mathbf{u}}))\, dx = \int_\Omega \mathbf{f} \cdot (\mathbf{v} - \dot{\mathbf{u}})\, dx\,, \forall \mathbf{v} \in V\,. \tag{21}$$

Adding (20) and (21) we arrive at the differential mixed variational inequality: Find $y = [\mathbf{u}, \mathbf{p}, \boldsymbol{\xi}]$ such that $y(0) = 0$, for almost all $t \in (0, T)$, $y(t) \in Y$ and

$$\left.\begin{aligned} &\dot{y}(t) = w(t) \in Y_p \\ &a(y(t), z - w(t)) + j(z) - j(w(t)) \geq \langle l(t), z - w(t)\rangle\,, \forall z \in Y_p\,, \end{aligned}\right\} \tag{22}$$

where as in [27] the bilinear form $a : Y \times Y \to \mathbb{R}$, the linear form $l(t) : Y \to \mathbb{R}$, the functional $j : Y_p \to \mathbb{R}$ are defined, respectively, by

$$\begin{aligned} a(y, z) &= \int_\Omega \big(\mathbf{C}(\mathbf{e}(\mathbf{u}) - \mathbf{p}) : (\mathbf{e}(\mathbf{v}) - \mathbf{q}) + \boldsymbol{\xi} : \mathbf{H}\boldsymbol{\eta}\big)\, dx\,, \\ \langle l(t), z\rangle &= \int_\Omega \mathbf{f}(t) \cdot \mathbf{v}\, dx\,, \\ j(z) &= \int_\Omega D(\mathbf{q}, \boldsymbol{\eta})\, dx \end{aligned}$$

for $y = [\mathbf{u}, \mathbf{p}, \boldsymbol{\xi}], z = [\mathbf{v}, \mathbf{q}, \boldsymbol{\eta}]$.

## 4.2 A mixed formulation

Similar to [4] in the case of a mixed boundary value problem of nonlinear elasticity and similar to the previous section, both the stress field $\boldsymbol{\sigma}$ and the strain field $\mathbf{e}$ can be considered as extra unknowns. This will lead to a new mixed formulation in elastoplasticity.

Again (17) is rewritten to

$$\mathbf{e} = \mathbf{grad}\,\mathbf{u} - \boldsymbol{\rho} \tag{23}$$

with the further unknown $\boldsymbol{\rho} := \frac{1}{2}(\mathbf{grad}\,\mathbf{u} - (\mathbf{grad}\,\mathbf{u})^T)$ in $\mathscr{R}$ as defined in the previous section. Then multiply (23) by $\boldsymbol{\tau} \in H(\mathbf{div}, \Omega)$ and integrate by part on $\Omega$, to get

$$\int_\Omega \mathbf{e} : \boldsymbol{\tau}\, dx + \int_\Omega \mathbf{u} \cdot \mathbf{div}(\boldsymbol{\tau})\, dx + \int_\Omega \boldsymbol{\rho} : \boldsymbol{\tau}\, dx = 0\,. \tag{24}$$

Further the constitutive equation (18) and the equilibrium equation (16), respectively, yield

$$\int_\Omega \mathbf{C}(\mathbf{e} - \mathbf{p}) : \mathbf{d}\, dx - \int_\Omega \boldsymbol{\sigma} : \mathbf{d}\, dx = 0\,,\ \forall \mathbf{d} \in L^2(\Omega, \mathbb{R}^{2\times 2})\,; \tag{25}$$

$$-\int_\Omega \mathbf{v} \cdot \mathbf{div}\boldsymbol{\sigma}\, dx = \int_\Omega \mathbf{f} \cdot \mathbf{v}\, dx\,, \quad \forall \mathbf{v} \in L^2(\Omega, \mathbb{R}^2)\,. \tag{26}$$

Again, the symmetry of $\boldsymbol{\sigma}$ is weakly required by

$$\int_\Omega \boldsymbol{\sigma} : \boldsymbol{\rho}\, dx = 0\,, \forall \boldsymbol{\rho} \in \mathscr{R}\,. \tag{27}$$

Observe that $\dot{\mathbf{e}}$ is in $[L^2(\Omega)]^{2\times 2} := L^2(\Omega, \mathbb{R}^{2\times 2})$. Hence, (25) writes for all $d \in [L^2(\Omega)]^{2\times 2}$ as

$$\int_\Omega \mathbf{C}(\mathbf{e} - \mathbf{p}) : (\mathbf{d} - \dot{\mathbf{e}})\, dx - \int_\Omega \boldsymbol{\sigma} : (\mathbf{d} - \dot{\mathbf{e}})\, dx = 0\,. \tag{28}$$

Add (28) and (20) to obtain

$$\left.\begin{array}{l} \int_\Omega D(\mathbf{q}, \boldsymbol{\eta})\, dx - \int_\Omega D(\dot{\mathbf{p}}, \dot{\boldsymbol{\xi}})\, dx \\ + \int_\Omega \mathbf{C}(\mathbf{e}(\mathbf{u}) - \mathbf{p}) : \big((\mathbf{d} - \mathbf{q}) - (\dot{\mathbf{e}} - \dot{\mathbf{p}})\big) \geq \\ \int_\Omega \big(\boldsymbol{\sigma} : (\mathbf{d} - \dot{\mathbf{e}}) - (\mathbf{H}\boldsymbol{\xi}) \cdot (\boldsymbol{\eta} - \dot{\boldsymbol{\xi}})\big)\, dx\,, \forall [\mathbf{q}, \boldsymbol{\eta}] \in K_p\,. \end{array}\right\} \tag{29}$$

Consequently, collecting the variational equalities (24), (26) and (27), and the variational inequality (29) we arrive at another DMVI: Find $y = [\mathbf{e}, \mathbf{p}, \boldsymbol{\xi}]$, $z = [\boldsymbol{\sigma}, \mathbf{u}, \mathbf{p}]$ such that $y(0) = 0$, for almost all $t \in (0, T)$, $y(t) \in [L^2(\Omega)]^{2\times 2} \times Q_0 \times M$, $z(t) \in H(\mathbf{div}, \Omega) \times V \times Q_0$ satisfies (24), (26), and (27), and moreover

$$\left.\begin{aligned} &\dot{y}(t) = w(t) \in Z_p, \\ &A\big([y, w](t), \zeta - w(t)\big) + J(\zeta) - J(w(t)) \\ &\geq \langle L(t), \zeta - w(t)\rangle \,, \forall \zeta = [\mathbf{v}, \mathbf{q}, \boldsymbol{\eta}] \in Z_p \,, \end{aligned}\right\} \tag{30}$$

where the closed convex cone $Z_p$, the bilinear form $A$, the linear form $L(t)$, and the convex positively homogeneous functional $J$ are appropriately defined. Since these definitions are obvious, details are omitted.

# 5 Differential mixed variational inequalities and their stability

Motivated by the nonsmooth boundary value problems and their variational formulation in the previous sections, we deal in this section with a general class of differential mixed variational inequalities. Indeed, with some change of notation, all the concrete time-dependent variational inequalities of the previous sections can be subsumed in this class, when introducing some appropriate product spaces. This will be elaborated with the nonsmooth transient heat conduction problem of section 2.

Since in our stability analysis, we permit perturbations in the nonsmooth convex functionals and in the convex constraint set, we provide auxiliary results on epiconvergence and Mosco convergence. We sketch how the monotonicity method of Browder and Minty can be used to establish a general stability result under weak convergence assumptions.

## *5.1 The general setting of differential mixed variational inequalities*

Let $X, V$ be two real, separable Hilbert spaces that are endowed with norms $\|\cdot\|_X$, $\|\cdot\|_V$, respectively and with scalar products denoted by $\langle\cdot,\cdot\rangle$, $(\cdot,\cdot)$ respectively. Further let there be given $T > 0$, a convex closed subset $K \subset V$, a convex, lower semicontinuous (lsc) proper functional $\phi : V \to \mathbb{R} \cup +\infty$, maps $F : [0, T] \times X \times V \to X$, and $G : [0, T] \times X \times V \to V$, and some fixed $x_0 \in X$. Then we consider the following problem: Find an $X$ - valued function $y$ and a $V$ - valued function $w$ both defined on $[0, T]$ that satisfy for a.a. (almost all) $t \in [0, T]$

$$(\text{DMVI})(F,G,K,\phi;x_0) \quad \begin{cases} \dot{y}(t) &= F(t,y(t),w(t)) \\ w(t) &\in \Sigma(K,\phi,G(t,y(t),\cdot)) \end{cases}$$

complemented by the initial condition $y(0) = y_0$. Here $\dot{y}(t)$ denotes the time derivative of $y(t)$. $\Sigma(K,\phi,G(t,y(t),\cdot))$ stands for the solution set of the mixed variational inequality defined by $K$, $\phi$ and $G(t,y(t),\cdot)$, that is, $w(t)$ has to satisfy

$$w(t) \in K, \quad \big(G(t,y(t),w(t)), z - w(t)\big) + \phi(z) - \phi(w(t)) \geq 0, \ \forall z \in K .$$

To give a precise meaning to a DMVI we have to introduce appropriate function spaces and impose some hypotheses on the data.

The fixed finite time interval $[0,T]$ gives rise to the Hilbert space $L^2(0,T;V)$ endowed with the scalar product

$$[v,w] := \int_0^T \Big(v(t), w(t)\Big)\, dt .$$

We consider weak solutions of the differential equation in a DMVI in the sense of Caratheodory. In particular, the $X$-valued function $y$ has to be absolutely continuous with derivative $\dot{y}(t)$ defined almost everywhere. Moreover to define the initial condition, the "trace" $y(0)$ is needed. Therefore we are led to the function space

$$\mathscr{Y}(0,T;X) := \{y \mid y \in L^2(0,T;X),\ \dot{y}(t) \in L^2(0,T;X)\} ,$$

a Hilbert space endowed with the scalar product $[x,y] + [\dot{x},\dot{y}]$.

We assume that the map $G$ satisfies the following growth condition: There exist $g_0 \in L^\infty(0,T)$ and $g^0 \in L^2(0,T)$ such that $\forall t \in (0,T)$, $\forall (y,w) \in X \times V$ there holds

$$\|G(t,y,w)\|_V \leq g_0(t)\, (\|y\|_X + \|w\|_V) + g^0(t) . \tag{31}$$

Hence the Nemytskii operator $\mathscr{G}$ that acts from $L^2(0,T;X) \times L^2(0,T;V)$ to $L^2(0,T;V)$ derives from $G$ by

$$\mathscr{G}(y,w)(t) := G(t,y(t),w(t)) , \quad t \in (0,T) .$$

We introduce the closed convex subset

$$\mathscr{K} := L^2(0,T;K) := \{w \in L^2(0,T;V) \mid w(t) \in K, \forall a.a.\ t \in (0,T)\} \tag{32}$$

For the time-independent functional $\phi$, we simply require that the functional $\Phi$ given by

$$\Phi(w) := \int_0^T \phi(w(t))\,dt, \quad w \in L^2(0,T;V) \tag{33}$$

is real-valued on $L^2(0,T;K)$.

Then it makes sense to replace the above pointwise formulation of the mixed variational inequality in a DMVI by its integrated counterpart,

$$w \in \mathscr{K}, \quad \Big[\mathscr{G}(y,w), v-w\Big] + \Phi(v) - \Phi(w) \geq 0, \;\; \forall v \in \mathscr{K}\,.$$

Concerning the map $F$, we assume the following growth condition similar to (31): There exist $f_0 \in L^\infty(0,T)$ and $f^0 \in L^2(0,T)$ such that $\forall t \in (0,T)$, $\forall (y,w) \in X \times V$ there holds

$$\|F(t,y,w)\|_X \leq f_0(t)\,(\|y\|_X + \|w\|_V) \; + \; f^0(t)\,. \tag{34}$$

Hence the Nemitskii operator $\mathscr{F}$ derived from $F$ by

$$\mathscr{F}(y,w)(t) := F(t,y(t),w(t))\,, \quad t \in (0,T)$$

acts from $L^2(0,T;X) \times L^2(0,T;V)$ to $L^2(0,T;X)$.

Using a standard device in dynamical systems (see, e.g., [36]), we can introduce the unknown $\tilde{y} := (y,t)$ and write the above DMVI as

$$\begin{cases} \dfrac{d\tilde{y}}{d\tau} = \dfrac{d}{d\tau}\begin{pmatrix} y \\ t \end{pmatrix} = \tilde{F}(\tilde{y},w) := \begin{pmatrix} F(\tilde{y}(t),w(t)) \\ 1 \end{pmatrix} \\ w(t) \in \; \Sigma(K,\phi,G(\tilde{y}(t),\cdot))\,, \end{cases}$$

complemented by the initial condition $\tilde{y}(0) = (y_0,0)$. Therefore in the following we can consider the autonomous problem without any loss of generality and drop in DMVI the dependence on $t$.

*Example 1.* To subsume the nonsmooth transient heat conduction problem of section 2 in the framework of a DMVI define the Hilbert spaces $X := L^2(\Omega)$, $V := H(\mathrm{div},\Omega) \times L^2(\Omega,\mathbb{R}^2)$, hence $\mathscr{Y} = W$. Further $K := H_+(\mathrm{div},\Omega,\Gamma_S) \times L^2(\Omega,\mathbb{R}^2)$ is a convex cone such that the variational inequality $(6)_2$ and the nonlinear variational equation $(6)_3$ can be written as a single variational inequality. Here for $z = [\boldsymbol{\sigma},\boldsymbol{p}]$, simply $\phi(z) := \langle g, \gamma_{\boldsymbol{\nu}}\boldsymbol{\sigma}\rangle_{L^2(\Gamma)}$ is a linear form. Then for $y = u$, $w = [\boldsymbol{\sigma},\mathbf{p}]$ we let $F(y,w) := \mathrm{div}\,\boldsymbol{\sigma} + f$. To define the operator $G$ we use the Riesz representation theorem as follows. For any $u \in L^2(\Omega), \mathbf{p} \in L^2(\Omega,\mathbb{R}^2)$, $\boldsymbol{\sigma} \mapsto \langle \mathbf{p},\boldsymbol{\sigma}\rangle_{L^2(\Omega,\mathbb{R}^2)} + \langle u, \mathrm{div}\,\boldsymbol{\sigma}\rangle_{L^2(\Omega)}$ is a continuous linear form on $H(\mathrm{div},\Omega)$, hence can be represented by some $\boldsymbol{\rho} \in H(\mathrm{div},\Omega)$. This leads to the continuous linear operator $R : [u,\mathbf{p}] \mapsto \boldsymbol{\rho}$. Thus finally we define $G(y,w) = G(u,\boldsymbol{\sigma},\mathbf{p}) := \big(R(u,\mathbf{p}), \mathbf{a}(\cdot,\mathbf{p}) - \boldsymbol{\sigma}\big)$ and arrive at a concrete example of a DMVI.

In what follows we study stability of differential mixed variational inequalities formulated as DMVI and admit perturbations $y_{0,n}$ of $y_0$ in the initial condition $y(0) = y_0$, $F^n, G^n$ of the maps $F : X \times V \to X, G : X \times V \to V$, $K_n$ of the convex closed subset $K \subset V$. and $\phi^n$ of the convex, lower semicontinuous proper functional $\phi : V \to \mathbb{R} \cup +\infty$. Suppose that $(y^n, w^n)$ solves (DMVI)$(F^n, G^n, K_n, \phi^n; y_{0,n})$ and assume that $(y^n, w^n) \to (y, w)$ with respect to an appropriate convergence for $X$-valued, respectively $V$-valued functions on $[0, T]$. Then we are interested in conditions on $F^n \to F, G^n \to G, K_n \to K, \phi^n \to \phi, y_{0,n} \to y_0$ that guarantee that $(y, w)$ solves the limit problem (DMVI)$(F, G, K, \phi; y_0)$. Such a stability result can be understood as a result of upper set convergence for the solution set of (DMVI)$(F, G, K, \phi; y_0)$.

## *5.2 Preliminaries; Mosco convergence of sets; epiconvergence of functions*

As the convergence of choice in variational analysis we employ Mosco set convergence for a sequence $\{K_n\}$ of closed convex subsets which is defined as follows. A sequence $\{K_n\}$ of closed convex nonvoid subsets of the Hilbert space $V$ is called Mosco convergent to a closed convex nonvoid subset $K$ of $V$, written $K_n \overset{M}{\longrightarrow} K$, if and only if

$$\sigma - \limsup_{n\to\infty} K_n \subset K \subset s - \liminf_{n\to\infty} K_n.$$

Here the prefix $\sigma$ and $\overset{\sigma}{\to}$ mean sequentially weak convergence in contrast to strong convergence denoted by the prefix $s$ and by $\overset{s}{\to}$. Further, lim sup, respectively lim inf are in the sense of Kuratowski upper, resp. lower limits of sequences of sets (see [2] for more information on Mosco convergence). Here we note that for the nonempty set $K$ the second inclusion provides the existence of $g_n \in K_n$ such that $g_n \overset{s}{\to} g$ for some given $g \in K$. Clearly, $K_n \overset{M}{\longrightarrow} K$, if and only if $C_n := K_n - g_n \overset{M}{\longrightarrow} C := K - g$, This simple translation argument shows there is no loss of generality to assume later that $0 \in K_n, K$.

As a preliminary result we need that Mosco convergence of convex closed sets $K_n$ inherits to Mosco convergence of the associated sets $\mathcal{K}_n = L^2(0, T; K_n)$, derived from $K_n$ similar to (32). To prove this convergence we first show Mosco convergence of the polars $K_n^0$ to $K^0 := \{\zeta \in V^* \ : \ (\zeta|z) \le 1\,, \forall z \in K\}$ using the duality $(\cdot|\cdot)$ on $V^* \times V$.

**Lemma 1.** *Let $K_n \overset{M}{\longrightarrow} K$. Then (a) $K_n^0 \overset{M}{\longrightarrow} K^0$; (b) $\mathcal{K}_n \overset{M}{\longrightarrow} \mathcal{K}$ in $L^2(0, T; V)$.*

Let us sketch the *proof*. To verify $\sigma - \limsup_{n\to\infty} K_n^0 \subset K^0$ let $\zeta = \sigma - \lim\limits_{n\to\infty} \zeta_n$ with $\zeta_n \in K_n^0$. Choose $z \in K$ arbitrarily. Then by assumption, there exist (eventually for a subsequence) $z_n \in K_n$ with $z = s - \lim_{n\to\infty} z_n$. By definition of the polar $K_n^0$, $(\zeta_n|z_n) \leq 1$, $\forall n$, hence in the limit $(\zeta|z) \leq 1$, $\forall z \in K$ what gives $\zeta \in K^0$. To show $K^0 \subset s - \liminf_{n\to\infty} K_n^0$, we use a result in [2] on the convergence of $s(K_n)$ to $s(K)$, where $s(K)(\zeta) := \sup\{(\zeta|z) \ : \ z \in K\}, \zeta \in V$ is the support function of $K$.

To verify $\sigma - \limsup_{n\to\infty} L^2(0,T;K_n) \subset L^2(0,T;K)$ let $w = \sigma - \lim_{n\to\infty} w_n$ with $w_n \in L^2(0,T;K_n)$. By the bipolar theorem ($K^{00} = K$) it is enough to show that $\forall \zeta \in K^0$, for a.a. $t \in (0,T)$ there holds $(\zeta|w(t)) \leq 1$. This follows from an indirect argument. To show $\mathscr{K} \subset s - \liminf_{n\to\infty} \mathscr{K}_n$ it is enough to verify the claim for the subset of $K$ - valued simple functions on $(0,T)$ that is dense in $\mathscr{K}$. For more details, see [25].

As a further preliminary result we next need that epiconvergence [2] of convex lsc functions $\phi_n$ inherits to epiconvergence of the associated functionals $\Phi_n$, derived from $\phi_n$ similar to (33).

**Lemma 2.** *Let the convex lsc proper functionals $\phi_n$ epiconverge to a convex lsc proper functional $\phi$ on $V$. Suppose, the functionals $\phi_n$ are equi-lower bounded in the sense that there exist $c_o \in \mathbb{R}, w_0 \in V$ such that*

$$\phi_n(w) \geq c_0 + (w_0, w), \quad \forall n \in \mathbb{N}, w \in V\,.$$

*Then the associated functionals $\Phi_n$ epiconverge to $\Phi$ on $L^2(0,T;V)$.*

For the *proof* of the claimed convergence properties of the integral functions we use the Lemma of Fatou; for details, see [25].

By combination of the previous lemmas we obtain the following auxiliary result.

**Lemma 3.** *Let $K_n \xrightarrow{M} K$ in $V$. Let the convex lsc proper functionals $\phi_n : K_n \to \mathbb{R}$ epiconverge to $\phi : K \to \mathbb{R}$ on $V$. Suppose, the functionals $\phi_n$ are equi-lower bounded in the sense that there exist $c_o \in \mathbb{R}, w_0 \in V$ such that*

$$\phi_n(w_n) \geq c_0 + (w_0, w_n), \quad \forall n \in \mathbb{N}, w_n \in K_n\,. \tag{35}$$

*Then the associated functionals $\Phi_n : L^2(0,T;K_n) \to \mathbb{R}$ epiconverge to $\Phi : L^2(0,T;K) \to \mathbb{R}$ in $L^2(0,T;V)$.*

## 5.3 The stability result

Before stating the result, some remarks are in order. In view of the existence theory of variational inequalities in infinite dimensional spaces (see, e.g., [35]) the best one can hope for is weak convergence of the perturbations $u^n$ in the general case of nonunique solutions of the underlying variational inequalities in (DMVI). Weak convergence can namely be readily derived from a posteriori estimates. However, continuity of a nonlinear map (here $G$, $F$) with respect to weak convergence is a hard requirement. To circumvent these weak convergence difficulties we apply the monotonicity method of Browder and Minty. Then as we shall see below, a stability condition on the maps $G^n$ with respect to the basic Hilbert space norm suffices.

These weak convergence difficulties also affect $F^n$. Therefore we have to impose a generally strong stability condition on the nonlinear maps $F^n$. In the situation of linear operators this condition can be drastically simplified to a stability condition with respect to convergence in the operator norm, see [24, Theorem 4.1] in the case $\phi = \phi^n = 0$.

On the other hand, stronger assumptions on $G^n$, like uniform monotonicity, imply that the solution sets $\Sigma(K_n, \phi^n, \mathscr{G}(y^n, \cdot))$ are single-valued. Uniform monotonicity with respect to $n$ moreover entails that the sequence $w^n$ strongly converges. Then the stability assumption for $F^n$ can be relaxed.

Since our stability assumptions pertain the given maps $G^n, G$, not the derived maps $\mathscr{G}^n, \mathscr{G}$, we have a delicate interplay between the pointwise almost everywhere formulation and the integrated formulation of the variational inequality in the perturbed DMVI and in the limit DMVI.

We need the following hypotheses on the convergence of $(F^n, G^n)$ to $(F, G)$:

(H1) Let $z_n \overset{s}{\to} z$ in $X$ and $v_n \overset{\sigma}{\to} v$ in $V$. Moreover, let $F^n(z_n, v_n) \overset{s}{\to} p$ in $X$. Then $p = F(z, v)$.

(H2) All maps $G^n(z, \cdot)$ for any $z \in X$ are monotone. If $z_n \overset{s}{\to} z, v_n \overset{s}{\to} v$ in $X$, respectively in $V$, then $G^n(z_n, v_n) \overset{s}{\to} G(z, v)$ in $V$. $G$ is hemicontinuous in the sense that for any $z \in X; v, w \in V$ the real-valued function $r \in \mathbb{R} \mapsto (g(z, v + rw), rw)$ is lower semicontinuous.

Now we can state the following stability result.

**Theorem 1.** *Let* $(y^n, w^n)$ *solve (DMVI)*$(F^n, G^n, K_n, \phi^n; y_{0,n})$. *Suppose,* $F^n, F,$ *respectively* $G^n, G$ *satisfy (H1), (H2) respectively. Let* $y_{0,n} \overset{s}{\to} y_0$. *Let the convex closed nonvoid sets* $K_n$ *Mosco-converge to* $K$ *in* $V$ *and let the convex lsc proper functions* $\phi^n : K_n \to \mathbb{R}$ *epiconverge to* $\phi : K \to \mathbb{R}$ *on* $V$. *Suppose, the functions* $\phi^n$ *are equi-lower bounded in the sense of (35). Assume that* $y^n \overset{s}{\to} y$ *in* $\mathscr{Y}(0, T; X)$ *and that* $w^n \in L^2(0, T; V)$ *converges weakly to* $w$ *pointwise in* $V$ *for a.a.* $t \in (0, T)$ *with* $\|w^n(t)\|_V \le m(t),\ \forall$ *a.a.t* $\in (0, T)$ *for some* $m \in L^2(0, T)$. *Then* $(y, w)$ *is a solution to (DMVI)*$(F, G, K, \phi; y_0)$.

For the *proof*, we refer to [25].

When the DMVI has a separable structure, the hypotheses (H1) and (H2) can be expressed more explicitly and thus simplified. An instance are dynamical systems $(31)_1$ that are affine in $w$. More precisely let as in the paper [38] on differential mixed variational inequalities in finite dimensions,

$$F(z, w) = F_1(z) + B(z)\, w,\ F_n(z, w) = F_{1,n}(z) + B_n(z)\, w\,.$$

Then obviously (H1) splits in conditions on $F_{1,n} \to F_1$ and $B_n \to B$, separately. In the important case of bilinear dynamical systems, $B$ and $B_n$ become linear operators in $\mathscr{L}(X, \mathscr{L}(V, X))$. Then $B_n \to B$ in the operator norm along with $F_{1,n}(z_n) \overset{s}{\to} F_1(z)$ for $z_n \overset{s}{\to} z$ implies (H1).

Also similarly to [38], let

$$G(z,w) = G_1(z) + G_2(w),\; G_n(z,w) = G_{1,n}(z) + G_{2,n}(w)\,.$$

Then (H2) is satisfied, if the $G_{2,n}$ are monotone, $G_2$ is hemicontinuous, $G_{1,n}(z_n) \overset{s}{\to} G_1(z)$ for $z_n \overset{s}{\to} z$, and $G_{2,n}(w_n) \overset{s}{\to} G_2(w)$ for $w_n \overset{s}{\to} w$.

Another instance of a separable structure are linear differential variational inequalities (see [24] and the references given there) which are of the form

$$\begin{cases} \begin{pmatrix} \dot{y}(t) \\ q(t) \end{pmatrix} = \mathscr{A} \begin{pmatrix} y(t) \\ w(t) \end{pmatrix} + \begin{pmatrix} f(t) \\ g(t) \end{pmatrix} \\ w(t) \in K,\;\; \big(q(t), z - w(t)\big) \geq 0,\;\; \forall z \in K \end{cases}$$

and where $\mathscr{A} : X \times V \to X \times V$ is a given linear continuous operator defined by

$$\mathscr{A} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

with appropriate linear operators $A, B, C, D$. Then the hypotheses (H1) and (H2) are satisfied and a stability result with respect to perturbations $\mathscr{A}_n, f_n, g_n$ can be proved, provided the operators $D_n = \mathrm{proj}_V \mathscr{A}_n(0,\cdot)$ are monotone, $\mathscr{A}_n \to \mathscr{A}$ in the operator norm, and $f_n \to f, g_n \to g$ in $L^2(0,T;X)$, respectively in $L^2(0,T;V)$; see [24] for details.

In the multi-field formulation of the nonsmooth problems considered we find this separable structure, too. Here in particular for the nonsmooth heat conduction problem of section 2, see in particular (5) and Example 1, we have for $w = [\boldsymbol{\sigma}, \mathbf{p}]$,

$$G_2(w) = \begin{pmatrix} 0 & E^* \\ -E & \mathbf{A}(\cdot) \end{pmatrix} \begin{pmatrix} \boldsymbol{\sigma} \\ \mathbf{p} \end{pmatrix} - \begin{pmatrix} g \\ 0 \end{pmatrix}.$$

Then $G_2$ inherits monotonicity from the nonlinear operator $\mathbf{A}$.

We refrain from deriving a stability result for stationary problems from the present Theorem 1. Instead we can refer to [23, Theorem 3] for a much stronger result.

## 6 Some concluding remarks: An outlook

Finally let us shortly outline some main streams of potential applications of multiple field modelling to nonsmooth multiphysics problems and nonsmooth nonconvex variational problems.

Let us first note that the present paper confines in section 3 to frictionless contact of a deformable body with a rigid foundation. Unilateral contact [14] or bilateral multi-body contact [33] with friction can be similarly described by a multi-field variational formulation. There in the model of Tresca friction (given friction model), a nonsmooth convex integral functional on the boundary part $\Gamma_C$ appears in addition; see [25] for a multi-field variational formulation of a simplified scalar frictional unilateral contact problem. Similar boundary value problems involving a nonsmooth functional appear in micropolar hemitropic contact [16], moreover in Stokes flow and non-Newtonian flow with friction or with leak boundary conditions; see [3, 34, 45, 46]. For *quasistatic* contact with friction, we refer to the monograph [44].

In section 4 we focused on the multi-field modeling approach to the simplest quasi-static elastoplasticity problem. More involved problems arise in multi-surface elastoplasticity [6], elastoplasticity with hysteresis, see, e.g., [7], and in viscoplasticity [21].

Nowadays nonsmooth multiphysics problems where differential equations of different type are coupled receive much attention. A prominent example are thermoelastic contact problems with frictional heating (see, e.g., [1]) or even thermoelectroconductive problems with Signorini contact [32]. Intelligent “smart” devices use piezoceramic material, here the coupling of the electricity field with solid mechanics is of interest (see, e.g., [15, 37]).

Another line of research are nonsmooth nonconvex variational problems that use the theory of nonconvex *hemivariational inequalities* as coined by Panagiotopoulos [43]. Here instead of convex analysis, the Clarke generalized differential calculus comes into play. This theory [22, 39, 40] and also their recent numerical treatment [30, 42] by regularization techniques and finite element methods allow to tackle nonconvex contact problems, like adhesion and delamination problems.

## References


1. Andersson, L.-E., Barber, J. R., Ciavarella, M.: On the existence and uniqueness of steady state solutions in thermoelastic contact with frictional heating. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **461** no. 2057, 1261–1282 (2005)
2. Attouch,H.: *Variational convergence for functions and operators*, Pitman, Boston, London, Melbourne (1984)
3. Ayadi,M., Gdoura,M.K., Sassi,T.: Mixed formulation for Stokes problem with Tresca friction. C. R. Math. Acad. Sci. Paris **348** 19–20 (2010)
4. Barrientos,M.A., Gatica,G.N., Stephan,E.P.: A mixed finite element method for nonlinear elasticity: two-fold saddle point approach and a-posteriori error estimate. Numer. Math. **91** 197–222 (2002)

5. F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer, New York (1991)
6. Brokate, M., Carstensen, C., Valdman, J.: A quasi-static boundary value problem in multi-surface elastoplasticity. I. Analysis, Math. Methods Appl. Sci. **27** 1697–1710 (2004)
7. Brokate, M., Krejčí, P., Rachinskii, D.: Some analytical properties of the multidimensional continuous Mróz model of plasticity. Control Cybernet., **27**, 199—215 (1998)
8. Carslaw,H.C, Jaeger,J.C.: *Conduction of Heat in Solids*, Clarendon, New York (1960, 1988)
9. Ciarlet,P.G.: *Mathematical Elasticity*, Vol. I, North-Holland, Amsterdam (1988)
10. Chen,Zhiming: On the augmented Lagrangian approach to Signorini elastic contact problems. Numer. Math. **88** 323–346 (2001)
11. Doyen,D., Ern,A., Piperno,S.: A three-field augmented Lagrangian formulation of unilateral contact problems with cohesive forces. ESAIM: M2AN **44** 323–346 (2010)
12. Duvaut, G, Lions, J.L.: *Les inéquations en méchanique et en physique*, Dunod, Paris (1972)
13. Ekeland,I., Temam,R.: *Convex Analysis and Variational Problems,* North-Holland, Amsterdam (1976)
14. Eck, C., Jarušek, J., Krbec, M.: *Unilateral Contact problems: Variational Methods and Existence Theorems*, Chapman & Hall/CRC, Boca Raton (2005)
15. Fernane, K., Dalah, M., Ayadi, A.: Existence and uniqueness solution of a quasistatic electro-elastic antiplane contact problem with Tresca friction law. Appl. Sci. **14** 45–59 (2012).
16. Gachechiladze, A., Gachechiladze, R., Gwinner, J., Natroshvili, D: A boundary variational inequality approach to unilateral contact problems with friction for micropolar hemitropic solids. Math. Methods Appl. Sci., **33**, No. 18, 2145–2161 (2010)
17. Gatica,G.N.: An application of Babuška-Brezzi's theory to a class of variational problems. Appl. Anal. **75** no.3–4, 297–303 (2000)
18. Gatica,G.N.: Solvability and Galerkin approximations of a class of nonlinear operator equations. Z. Anal. Anwendungen **21** no.3, 761–781 (2002)
19. Gatica,G.N., Heuer,N., Meddahi,S.: On the numerical analysis of nonlinear twofold saddle point problems. IMA J. Numer. Anal. **23** No.2, 301–330 (2003)
20. Girault,V., Raviart,P.-A.: *Finite element approximation of the Navier-Stokes equations*, Revised reprint of the 1st edition, Lecture Notes in Mathematics, vol. 749, Springer, Berlin (1981)
21. R. Glowinski, Wachs,A.: On the numerical simulation of viscoplastic fluid flow, Handbook of Numerical Analysis XVI, P.G. Ciarlet (Ed.) pp. 483–717, Elsevier, Amsterdam (2011)
22. Goeleven, D., Motreanu, D., Dumont, Y., and Rochdi, M.: *Variational and hemivariational inequalities: theory, methods and applications. Volume I: Unilateral analysis and unilateral mechanics*, Kluwer, Boston (2003)
23. Gwinner,J.: Stability of monotone variational inequalities with various applications,Variational Inequalities and Network Equilibrium Problems, F. Giannessi and A. Maugeri (Eds.), pp. 123–142, Plenum Press, New York (1995)
24. Gwinner, J.: A note on linear differential variational inequalities in Hilbert space, Modeling and Optimization, D. Hömberg and F. Tröltzsch (Eds.), pp. 85–91, Springer, Heidelberg (2013)
25. Gwinner, J.: Three-field modelling of nonlinear nonsmooth boundary value problems and stability of differential mixed variational inequalities. Abstr. Appl. Anal. **2013** Article ID 108043, 10 pages (2013)
26. Han, W., Reddy, B.D., Schroeder,G.: Qualitative and numerical analysis of quasi-static problems in elastoplasticity. SIAM J. Numer. Anal. **34** 143–177. (1997)
27. Han W., Reddy, B.D.: *Plasticity: Mathematical Theory and Numerical Analysis*, Springer, New York (1999)
28. Haslinger,J.: Mixed formulation of variational inequalities and its approximation. Apl. Mat. **26** 462–475 (1981)
29. Haslinger,J., Lovíšek,J.: Mixed variational formulation of unilateral problems. Comment. Math. Univ. Carolinae **21** 231–246 (1980)
30. Haslinger, J., Miettinen, M., Panagiotopoulos, P.D.: *Finite Element Methods for Hemivariational Inequalities*, Kluwer Academic Publishers (1999)

31. Hlaváček, I., Haslinger, J., Nečas, J., Lovišek, J.: *Solution of Variational Inequalities in Mechanics*, SNTL, Prague (1983)
32. Hömberg, D., Khludnev; A.M.: Equilibrium problem for thermoelectroconductive body with the Signorini condition on the boundary. Math. Methods Appl. Sci. **24** 233–244 (2001)
33. Hüeber, S., Wohlmuth, B.: Equilibration techniques for solving contact problems with Coulomb friction. Comput. Methods Appl. Mech. Engrg. **205/208** 29–45 (2012)
34. Kashiwabara, T.: On a strong solution of the non-stationary Navier-Stokes equations under slip or leak boundary conditions of friction type. J. Differential Equations **254** 756–778 (2013).
35. Kinderlehrer, D., Stampacchia,G.: *An Introduction to Variational Inequalities and Their Applications*, Academic Press, New York (1984)
36. Kunze,M.: *Non-Smooth Dynamical Systems*, Lecture Notes in Mathematics, vol. 1744 Springer, Berlin, 2000.
37. Kutter, M., Sändig, A.-M.: Modeling of ferroelectric hysteresis as variational inequality. GAMM-Mitteilungen,**34** 84–89 (2011)
38. Li, X.S., Huang,N., O'Regan,D.: Differential mixed variational inequalities in finite dimensional spaces. Nonlinear Anal. **72**, no. 9–10, 3875–3886 (2010)
39. Migórski, S., Ochal, A., Sofonea, M.: *Nonlinear Inclusions and Hemivariational Inequalities: Models and Analysis of Contact Problems*, Springer, New York (2013)
40. Naniewicz, Z., Panagiotopoulos, P. D.: *Mathematical Theory of Hemivariational Inequalities and Applications*, Marcel Dekker, New York (1995)
41. Nečas, J., Hlaváček, I.: *Mathematical theory of elastic and elasto-plastic bodies: An introduction*, Elsevier, Amsterdam (1981)
42. Ovcharova, N.: Regularization Methods and Finite Element Approximation of Hemivariational Inequalities with Applications to Nonmonotone Contact Problems, PhD Thesis, Universität der Bundeswehr München, Cuvillier, Göttingen (2012)
43. Panagiotopoulos, P.D.: *Inequality problems in mechanics and application. Convex and nonconvex energy functions*, Birkhäuser, Basel (1998)
44. Shillor, M., Sofonea, M., Telega, J.J.: *Models and Analysis of Quasistatic Contact: Variational Methods*, Springer, Berlin (2004)
45. Saidi, F.: Non-Newtonian flow in a thin film with boundary conditions of Coulomb's type. ZAMM Z. Angew. Math. Mech. **86** 702–721 (2006)
46. Saidi, F.: On the Navier-Stokes equation with slip boundary conditions of friction type. Math. Model. Anal. **12** 389–398 (2007).
47. Simane, L., Bendali, A., Laborde, P.: Mixed formulations for a class of variational inequalities. ESAIM: M2AN **38** 177–201 (2004)
48. Wohlmuth,B.: Variationally consistent discretization schemes and numerical algorithms. Acta Numer. **20** 569–734 (2011)

# $\mathscr{I}_\theta$-statistical convergence of order $\alpha$ in topological groups

**Ekrem Savaş and Rahmet Savaş Eren**

**Abstract** In this paper, we introduce and study $\mathscr{I}$-lacunary statistical convergence of order $\alpha$ in topological groups and we shall also present some inclusion theorems.

## 1 Introduction

Before continuing with this paper we present some definitions and preliminaries:

The notion of statistical convergence, which is an extension of the idea of usual convergence, was introduced by Fast [9] and Schoenberg [28] and its topological consequences were studied first by Fridy [10] and Šalát [19]. Di Maio and Kočinac [17] introduced the concept of statistical convergence in topological spaces and statistical Cauchy condition in uniform spaces and established the topological nature of this convergence. The notion has also been defined and studied in different steps, for example, in the locally convex space [16]; in intuitionistic fuzzy normed spaces[18]. In [1] Albayrak and Pehlivan studied this notion in locally solid Riesz spaces. Quite recently, Das and Savas [6] introduced the ideas of $\mathscr{I}_\tau$-convergence, $\mathscr{I}_\tau$-boundedness, and $\mathscr{I}_\tau$-Cauchy condition of nets in a locally solid Riesz space and also $\mathscr{I}_\lambda$-statistical convergence in a locally solid Riesz space was introduced by Das and Savas [8]. Savas [26] introduced and studied $\mathscr{I}$- double lacunary statistical convergence in a locally solid Riesz space.

If $\mathbb{N}$ denotes the set of natural numbers and $K \subset \mathbb{N}$, then $K(m,n)$ denotes the cardinality of the set $K \cap [m,n]$. The upper and lower natural density of the subset $K$ is defined by

$$\overline{d}(K) = \lim_{n\to\infty} \sup \frac{K(1,n)}{n} \text{ and } \underline{d}(K) = \lim_{n\to\infty} \inf \frac{K(1,n)}{n}.$$

E. Savaş (✉)
Department of Mathematics, Istanbul Commerce University, Uskudar, Istanbul, Turkey
e-mail: ekremsavas@yahoo.com

R.S. Eren
Department of Mathematics, Istanbul Medeniyet University, Istanbul, Turkey
e-mail: rahmetsavas@gmail.com



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_6

If $\overline{d}(K) = \underline{d}(K)$, then we say that the natural density of $K$ exists and it is denoted simply by $d(K)$. Clearly $d(K) = \lim_{n\to\infty} \frac{K(1,n)}{n}$. A sequence $x = (x_k)$ of real numbers is said to be statistically convergent to $L$ if for arbitrary $\epsilon > 0$, the set $K(\epsilon) = \{n \in \mathbb{N} : |x_k - L| \geq \epsilon\}$ has natural density zero. In this case, we write $st - \lim_k x_k = L$ and we denote the set of all statistical convergent sequences by $st$.

By a lacunary sequence, we mean an increasing sequence $\theta = (k_r)$ of positive integers such that $k_0 = 0$ and $h_r : k_r - k_{r-1} \to \infty$ as $r \to \infty$. Throughout this paper, the intervals determined by $\theta$ will be denoted by $I_r = (k_{r-1}, k_r]$, and the ratio $(k_r)\,(k_{r-1})^{-1}$ will be abbreviated by $q_r$.

In another direction in [11], a new type of convergence called lacunary statistical convergence was introduced as follows: A sequence $(x_k)$ of real numbers is said to be lacunary statistically convergent to $L$ (or, $S_\theta$-convergent to $L$ ) if for any $\epsilon > 0$,

$$\lim_{r\to\infty} \frac{1}{h_r} |\{k \in I_r : |x_k - L| \geq \epsilon\}| = 0$$

where $|A|$ denotes the cardinality of $A \subset \mathbb{N}$. In [11] the relation between lacunary statistical convergence and statistical convergence was established among other things.

The order of statistical convergence of a sequence of numbers was given by Gadjiev and Orhan in [12] and later on statistical convergence of order $\alpha$ and strongly $p$- Cesàro summability of order $\alpha$ studied by Çolak [4].

In [13], P. Kostyrko et al. introduced the concept of $\mathscr{I}$-convergence of sequences in a metric space and studied some properties of such convergence. Note that $\mathscr{I}$-convergence is an interesting generalization of statistical convergence. More investigations in this direction and more applications of ideals can be found in [5, 14, 15, 20–25].

Recently in [5] we used ideals to introduce the concepts of $\mathscr{I}$-statistical convergence and $\mathscr{I}$-lacunary statistical convergence. Also Das and Savas[7] extended the concepts of $\mathscr{I}$-statistical convergence and $\mathscr{I}$-lacunary statistical convergence to the concepts of $\mathscr{I}$-statistical convergence and $\mathscr{I}$-lacunary statistical convergence of order $\alpha$, $0 < \alpha \leq 1$.

The purpose of this paper is to study $\mathscr{I}$-lacunary statistical convergence of order $\alpha$, $0 < \alpha \leq 1$ in topological groups and to give some important inclusion theorems.

## 2 Definitions and Notations

The following definitions and notions will be needed.

**Definition 1.** A family $\mathscr{I} \subset 2^{\mathbb{N}}$ is said to be an ideal of $\mathbb{N}$ if the following conditions hold:

(a) $A, B \in \mathscr{I}$ implies $A \cup B \in \mathscr{I}$,
(b) $A \in \mathscr{I}$, $B \subset A$ implies $B \in \mathscr{I}$,

**Definition 2.** A non-empty family $F \subset 2^{\mathbb{N}}$ is said to be a filter of $\mathbb{N}$ if the following conditions hold:

(a) $\phi \notin F$,
(b) $A, B \in F$ implies $A \cap B \in F$,
(c) $A \in F,\ A \subset B$ implies $B \in F$,

If $\mathscr{I}$ is a proper ideal of $\mathbb{N}$ (i.e., $\mathbb{N} \notin \mathscr{I}$ ), then the family of sets $F(\mathscr{I}) = \{M \subset \mathbb{N} : \exists A \in \mathscr{I} : M = \mathbb{N} \setminus A\}$ is a filter of $\mathbb{N}$. It is called the filter associated with the ideal.

**Definition 3.** A proper ideal $\mathscr{I}$ is said to be admissible if $\{n\} \in \mathscr{I}$ for each $n \in \mathbb{N}$.

Throughout $\mathscr{I}$ will stand for a proper admissible ideal of $\mathbb{N}$.

**Definition 4 (see [13]).** Let $\mathscr{I} \subset 2^{\mathbb{N}}$ be a proper admissible ideal in $\mathbb{N}$. The sequence $x = (x_k)$ of elements of $\mathbb{R}$ is said to be $\mathscr{I}$-convergent to $L \in \mathbb{R}$ if for each $\epsilon > 0$ the set $A(\epsilon) = \{n \in \mathbb{N} : |x_k - L| \geq \epsilon\} \in \mathscr{I}$.

In [7], Das and Savaş defined $\mathscr{I}$- statistical convergence and $\mathscr{I}$- lacunary statistical convergence of order $\alpha$ as follows:

**Definition 5.** A sequence $x = (x_k)$ is said to be $\mathscr{I}$-statistically convergent of order $\alpha$ to $L$ or $S(\mathscr{I})^\alpha$-convergent to $L$, where $0 < \alpha \leq 1$, if for each $\epsilon > 0$ and $\delta > 0$

$$\{n \in \mathbb{N} : \frac{1}{n^\alpha}|\{k \leq n : |x_k - L| \geq \epsilon\}| \geq \delta\} \in \mathscr{I}.$$

In this case we write $x_k \to L(S(\mathscr{I})^\alpha)$. The class of all $\mathscr{I}$-statistically convergent of order $\alpha$ sequences will be denoted by simply $S(\mathscr{I})^\alpha$.

**Definition 6.** Let $\theta$ be a lacunary sequence. A sequence $x = (x_k)$ is said to be $\mathscr{I}$-lacunary statistically convergent of order $\alpha$ to $L$ or $S_\theta(\mathscr{I})^\alpha$-convergent to $L$ if for any $\epsilon > 0$ and $\delta > 0$

$$\{r \in \mathbb{N} : \frac{1}{h_r^\alpha}|\{k \in I_r : |x_k - L| \geq \epsilon\}| \geq \delta\} \in \mathscr{I}.$$

In this case we write $x_k \to L(S_\theta(\mathscr{I})^\alpha)$. The class of all $\mathscr{I}$-lacunary statistically convergent sequences of order $\alpha$ will be denoted by $S_\theta(\mathscr{I})^\alpha$.

By $X$, we will denote an abelian topological Hausdorff group, written additively, which satisfies the first axiom of countability. In [2], a sequence $(x_k)$ in $X$ is called to be statistically convergent to an element $L$ of $X$ if for each neighborhood $U$ of 0,

$$\lim_{n\to\infty} \frac{1}{n}|\{k \leq n : x_k - L \notin U\}| = 0,$$

where the vertical bars indicate the number of elements in the enclosed set. The set of all statistically convergent sequences in $X$ is denoted by $st(X)$.

Also, Cakalli [3] defined lacunary statistical convergence in topological groups as follows: A sequence $(x_k)$ is said to be $S_\theta$-convergent to $L$ (or lacunary statistically convergent to $L$) if for each neighborhood $U$ of 0, $lim_{r\to\infty}(h_r)^{-1}\,|k \in I_r : x_k - L \notin U\}| = 0$. In this case, we define

$$S_\theta(X) = \left\{(x_k) : \text{ for some } L,\ S_\theta - \lim_{k\to\infty} x_k = L\right\}.$$

Now we are ready to give the main definitions of $\mathscr{I}$-statistical convergence and $\mathscr{I}$- lacunary statistical convergence of order $\alpha$ in topological groups as follows:

**Definition 7.** A sequence $x = (x_k)$ in $X$ is said to be statistically convergent of order $\alpha$ to $L$ or $S(\mathscr{I})^\alpha$-convergent of order $\alpha$ to $L$ if for each $\delta > 0$ and for each neighborhood $U$ of 0,

$$\{n \in \mathbb{N} : \frac{1}{n^\alpha}|\{k \leq n : x_k - L \notin U\}| \geq \delta\} \in \mathscr{I}.$$

In this case, we write $x_k \to L(S(\mathscr{I})^\alpha)$. The class of all $S(\mathscr{I})^\alpha$-statically convergent sequences will be denoted by simply $S(\mathscr{I})^\alpha(X)$.

*Remark 1.* For $\alpha = 1$ the definition coincides with $\mathscr{I}$-statistical convergence in topological groups [27]. For $\mathscr{I} = \mathscr{I}_{fin} = \{A \subseteq \mathbf{N} : A \text{ is a finite subset}\}$, and $\alpha = 1$, $\mathscr{I}$- statistical convergence becomes statistical convergence in topological groups which is studied by Cakalli [2].

**Definition 8.** Let $\theta$ be a lacunary sequence. A sequence $x = (x_k)$ in $X$ is said to be $\mathscr{I}$-lacunary statistically convergent of order $\alpha$ to $L$ or $S_\theta(\mathscr{I})^\alpha$-convergent to $L$ if for any $\delta > 0$ and for each neighborhood $U$ of 0,

$$\{r \in \mathbb{N} : \frac{1}{h_r^\alpha}|\{k \in I_r : x_k - L \notin U\}| \geq \delta\} \in \mathscr{I}.$$

In this case, we write

$$S_\theta(\mathscr{I})^\alpha - \lim_{k\to\infty} x_k = L \quad \text{or} \quad x_k \to L(S_\theta(\mathscr{I})^\alpha)$$

and define

$$S_\theta(\mathscr{I})^\alpha(X) = \left\{(x_k) : \text{ for some } L, S_\theta(\mathscr{I})^\alpha - \lim_{k\to\infty} x_k = L\right\}$$

and in particular,

$$S_\theta(\mathscr{I})^\alpha(X)_0 = \left\{(x_k) : S_\theta(\mathscr{I})^\alpha - \lim_{k\to\infty} x_k = 0\right\}.$$

*Remark 2.* For $\alpha = 1$, the definition coincides with $\mathscr{I}$-lacunary statistical convergence in topological groups [27]. Further it must be noted in this context that lacunary statistical convergence of order $\alpha$ in topological groups has not been studied till now. Obviously lacunary statistical convergence of order $\alpha$ in topological groups is a special case of $\mathscr{I}$-lacunary statistical convergence of order $\alpha$ in topological groups when we take $\mathscr{I} = \mathscr{I}_{fin}$. Also, for $\mathscr{I} = \mathscr{I}_{fin}$, and $\alpha = 1$, $\mathscr{I}$-lacunary statistical convergence of order $\alpha$ becomes lacunary statistical convergence in topological groups which is studied by Cakalli [3].

It is obvious that every $\mathscr{I}$-lacunary statistically convergent of order $\alpha$ has only one limit, that is, if a sequence is $\mathscr{I}$-lacunary statistically convergent of order $\alpha$ to $L_1$ and $L_2$, then $L_1 = L_2$.

## 3 Inclusion Theorems

In this section, we prove the following theorems.

**Theorem 1.** *Let* $0 < \alpha \leq \beta \leq 1$. *Then* $S_\theta(\mathscr{I})^\alpha(X) \subset S_\theta(\mathscr{I})^\beta(X)$.

*Proof.* Let $0 < \alpha \leq \beta \leq 1$. Then

$$\frac{|\{k \in I_r : x_k - L \notin U\}|}{h_r^\beta} \leq \frac{|\{k \in I_r : x_k - L \notin U\}|}{h_r^\alpha}$$

and so for any $\delta > 0$ and for each neighborhood $U$ of 0,

$$\{n \in \mathbb{N} : \frac{|\{k \in I_r : x_k - L \notin U\}|}{h_r^\beta} \geq \delta\} \subset \{n \in \mathbb{N} : \frac{|\{k \in I_r : x_k - L \notin U\}|}{h_r^\alpha} \geq \delta\}.$$

Hence, if the set on the right-hand side belongs to the ideal $\mathscr{I}$, then obviously the set on the left-hand side also belongs to $\mathscr{I}$. This shows that $S_\theta(\mathscr{I})^\alpha(X) \subset S_\theta(\mathscr{I})^\beta(X)$.

**Corollary 1.** *If a sequence is $\mathscr{I}$-lacunary statistically convergent of order $\alpha$ to $L$ for some $0 < \alpha \leq 1$, then it is $\mathscr{I}$-lacunary statistically convergent to $L$, i.e.* $S_\theta(\mathscr{I})^\alpha(X) \subset S_\theta(\mathscr{I})(X)$.

Similarly we can show that

**Theorem 2.** *Let* $0 < \alpha \leq \beta \leq 1$. *Then*

*(i)* $S(\mathscr{I})^\alpha(X) \subset S(\mathscr{I})^\beta(X)$.
*(ii) In particular* $S(\mathscr{I})^\alpha(X) \subset S(\mathscr{I})(X)$.

**Theorem 3.** *For any lacunary sequence $\theta$, $\mathscr{I}$-statistical convergence of order $\alpha$ implies $\mathscr{I}$-lacunary statistical convergence of order $\alpha$, that is* $S(\mathscr{I})^\alpha(X) \subset S_\theta(\mathscr{I})^\alpha(X)$ *if* $\liminf_r q_r^\alpha > 1$.

*Proof.* Suppose first that $\liminf_r q_r^\alpha > 1$. Then there exists $\sigma > 0$ such that $q_r^\alpha \geq 1 + \sigma$ for sufficiently large $r$ which implies that

$$\frac{h_r^\alpha}{k_r^\alpha} \geq \frac{\sigma}{1+\sigma}.$$

Since $x_k \to L(S(I)^\alpha(X))$, then for any neighborhood $U$ of 0 and for sufficiently large $r$, we have

$$\frac{1}{k_r^\alpha}|\{k \leq k_r : x_k - L \notin U\}| \geq \frac{1}{k_r^\alpha}|\{k \in I_r : x_k - L \notin U\}|$$

$$\geq \frac{\sigma}{1+\sigma}.\frac{1}{h_r^\alpha}|\{k \in I_r : x_k - L \notin U\}|.$$

Then for any $\delta > 0$, we get

$$\{r \in \mathbb{N} : \frac{1}{h_r^\alpha}|\{k \in I_r : x_k - L \notin U\}| \geq \delta\}$$

$$\subseteq \{r \in \mathbb{N} : \frac{1}{k_r^\alpha}|\{k \leq k_r : x_k - L \notin U\}| \geq \frac{\delta\sigma}{(1+\sigma)}\} \in \mathscr{I}.$$

This proves the result.

*Remark 3.* The converse of this result is true for $\alpha = 1$ (see Theorem 1 [27]). However for $\alpha < 1$ it is not clear and we leave it as an open problem.

For the next result we assume that the lacunary sequence $\theta$ satisfies the condition that for any set $C \in F(\mathscr{I})$, $\bigcup\{n : k_{r-1} < n < k_r, r \in C\} \in F(\mathscr{I})$.

**Theorem 4.** *For a lacunary sequence $\theta$ satisfying the above condition, $\mathscr{I}$-lacunary statistical convergence of order $\alpha$ implies $\mathscr{I}$-statistical convergence of order $\alpha$, $0 < \alpha \leq 1$, that is, $S_\theta(\mathscr{I})^\alpha(X) \subset S(\mathscr{I})^\alpha(X)$ if $\sup_r \sum_{i=0}^{r-1} \frac{h_{i+1}^\alpha}{(k_{r-1})^\alpha} = B(say) < \infty$.*

*Proof.* Suppose that $x_k \to L(S_\theta(\mathscr{I})^\alpha(X))$. Take any neighborhood $U$ of 0. For $\delta, \delta_1 > 0$ define the sets

$$C = \{r \in \mathbb{N} : \frac{1}{{h_r}^\alpha}|\{k \in I_r : x_k - L \notin U\}| < \delta\}$$

and

$$T = \{n \in \mathbb{N} : \frac{1}{n^\alpha}|\{k \leq n : x_k - L \notin U\}| < \delta_1\}.$$

It is obvious from our assumption that $C \in F(\mathscr{I})$, the filter associated with the ideal $\mathscr{I}$. Further observe that

$$A_j = \frac{1}{h_j^{\alpha}}|\{k \in I_j : x_k - L \notin U\}| < \delta$$

for all $j \in C$. Let $n \in \mathbb{N}$ be such that $k_{r-1} < n < k_r$ for some $r \in C$. Now

$$\frac{1}{n^{\alpha}}|\{k \leq n : x_k - L \notin U\}| \leq \frac{1}{k_{r-1}^{\alpha}}|\{k \leq k_r : x_k - L \notin U\}|$$

$$= \frac{1}{k_{r-1}^{\alpha}}|\{k \in I_1 : x_k - L \notin U\}| + \cdots + \frac{1}{k_{r-1}^{\alpha}}|\{k \in I_r : x_k - L \notin U\}|$$

$$= \frac{k_1^{\alpha}}{k_{r-1}^{\alpha}}\frac{1}{h_1^{\alpha}}|\{k \in I_1 : x_k - L \notin U\}| + \frac{(k_2 - k_1)^{\alpha}}{k_{r-1}^{\alpha}}\frac{1}{h_2^{\alpha}}|\{k \in I_2 : x_k - L \notin U\}| + \cdots +$$

$$+ \frac{(k_r - k_{r-1})^{\alpha}}{k_{r-1}^{\alpha}}\frac{1}{h_r^{\alpha}}|\{k \in I_r : x_k - L \notin U\}|$$

$$= \frac{k_1^{\alpha}}{k_{r-1}^{\alpha}}A_1 + \frac{(k_2 - k_1)^{\alpha}}{k_{r-1}^{\alpha}}A_2 + \cdots + \frac{(k_r - k_{r-1})^{\alpha}}{k_{r-1}^{\alpha}}A_r$$

$$\leq \sup_{j \in C} A_j . \sup_r \sum_{i=0}^{r-1} \frac{(k_{i+1} - k_i)^{\alpha}}{k_{r-1}^{\alpha}} < B\delta.$$

Choosing $\delta_1 = \frac{\delta}{B}$ and in view of the fact that $\bigcup\{n : k_{r-1} < n < k_r, r \in C\} \subset T$ where $C \in F(\mathscr{I})$ it follows from our assumption on $\theta$ that the set $T$ also belongs to $F(\mathscr{I})$ and this completes the proof of the theorem.

**Corollary 2.** *Let $\theta = \{(k_r)\}$ be a lacunary sequence, then $S(\mathscr{I})^{\alpha}(X) = S_\theta(\mathscr{I})^{\alpha}(X)$ iff*

$$1 < \liminf_r q_r \leq \limsup_r q_r < \infty.$$

## References


1. H. Albayrak and S. Pehlivan, Statistical convergence and statistical continuity on locally solid Riesz spaces, Topology Appl. 159, 1887–1893 (2012)
2. H. Çakalli, On Statistical Convergence in topological groups, Pure and Appl. Math. Sci. **43**, No. 1-2, 27–31 (1996)
3. ———-, Lacunary statistical convergence in topological groups, Indian J. Pure Appl. Math. **26**, (2), 113–119 (1995)

4. R. Colak, Statistical convergence of order $\alpha$, Modern methods in Analysis and its Applications, New Delhi, India, Anamaya Pub. 121–129 (2010)
5. P. Das and E. Savaş, S. K. Ghosal, On generalizations of certain summability methods using ideals, Appl. Math. Letters. 24, 1509–1514 (2011)
6. P. Das, E. Savaş, On $\mathscr{I}$-convergence of nets in locally solid Riesz spaces, Filomat. 27:1, 84–89 (2013)
7. P. Das and E. Savaş, On $\mathscr{I}$-statistical and $\mathscr{I}$-lacunary statistical convergence of order $\alpha$, Bull. Iranian Math. Soc. (Preprint)
8. P. Das and E. Savaş, On $\mathscr{I}_\lambda$-statistical convergence in locally solid Riesz spaces, Math. Slovaca (Preprint)
9. H. Fast, Sur la convergence statistique, Colloq Math. 2, 241–244 (1951)
10. J. A. Fridy, On ststistical convergence, Analysis. 5, 301–313 (1985)
11. J. A. Fridy and C. Orhan, Lacunary statistical convergence, Pacific J. Math. 160, 43–51 (1993)
12. A. D. Gadjiev and C. Orhan, some approximation theorems via statistical convergence, Rocky Mountain J. Math. 32(1), 508–520 (2002)
13. P. Kostyrko, T. Šalát, W.Wilczynki, $\mathscr{I}$-convergence, Real Anal. Exchange. 26 (2), 669–685 (2000/2001)
14. P. Kostyrko, M. Macaj, T. Šalát, and M. Sleziak, $\mathscr{I}$-convergence and extremal $\mathscr{I}$-limit points, Math. Slovaca. 55, 443–464 (2005)
15. B.K. Lahiri and Pratulananda Das, On $\mathscr{I}$ and $\mathscr{I}*$ convergence of nets, Real Anal.Exchange, 33(2), 431–442 (2007–2008)
16. I. J Maddox, Statistical convergence in locally convex spaces, Math Proc. Camb. Phil. Soc. 104, 141–145 (1988)
17. G. D. Maio and L. D. R. Kocinac, Statistical convergence in topology, Topology Appl. 156, 28–45 (2008)
18. M. Mursaleen and S. A. Mohiuddine, On lacunary statistical convergence with respect to the intuitionistic fuzzy normed space, Journal of Computational and Applied Mathematics. 233(2), 142–149 (2009)
19. T. Šalát, On statistically convergent sequences of real numbers, Math. Slovaca. 30, 139–150 (1980)
20. E. Savaş and Pratulananda Das, A generalized statistical convergence via ideals, Appl. Math. Lett. 24, 826–830 (2011)
21. E. Savaş, $\Delta^m$-strongly summable sequence spaces in 2-normed spaces defined by ideal convergence and an Orlicz function, Appl. Math. Comput. 217 271–276 (2010)
22. E. Savaş, A sequence spaces in 2-normed space defined by ideal convergence and an Orlicz function, Abst. Appl. Anal., Vol. (2011), Article ID 741382
23. E. Savaş, On some new sequence spaces in 2-normed spaces using Ideal convergence and an Orlicz function, J. Ineq. Appl. Article Number: 482392 DOI: 10.1155/2010/482392 Published: (2010)
24. E. Savaş, On generalized double statistical convergence via ideals, The Fifth Saudi Science Conference,16–18 April (2012)
25. E. Savaş and Pratulananda Das, A generalized statistical convergence via ideals, Appl. Math. Letters. 24, 826–830 (2011)
26. E. Savaş, On Lacunary double statistical convergence in locally solid Riesz spaces J. Ineq. Appl. (2013)
27. E. Savaş, $\mathscr{I}_\theta$ -statistically convergent sequences in topological groups, the Revista de la Unión Matemática Argentina. (Preprint)
28. I.J. Schoenberg, The integrability methods, Amer. Math. Monthly. 66, 361–375 (1959)

# Periodic Solutions of Cohen-Grossberg type model of Neural Networks with Delay and Impulses

**Zehour Nedjraoui Benbouziane and Nadira Boukli-Hacene**

**Abstract** An iterative method and some analysis techniques are applied to study the existence of periodic solutions of Cohen-Grossberg type model of neural networks with delay and impulses. Our result extends those existing ones.

**Keywords** Cohen-Grossberg type neural networks (CGNN) • Periodic solution • Impulses • upper and lower solutions

## 1 Introduction

We consider the differential system from CGNN model

$$\begin{cases} \dfrac{dx_i}{dt} = -\, d_i\left(x_i(t)\right)\Big[\, a_i\, x_i(t) - \displaystyle\sum_{j=1}^{m} b_{ij}\, f_j\big(x_i(t)\big) - \sum_{j=1}^{m} c_{ij}\, f_j\big(x_j(t-\tau_{ij})\big) + J_i(t)\Big] \quad t > t_0,\;\; t \neq t_k \\ \\ \Delta\, x_i(t_k) \;=\; x_i(t_k^+) - x_i(t_k^-) \;=\; -\,\gamma_{ik}\left(x_i(t_k)\right); \quad i \;=\; 1, \cdots, m\,, \quad k \in \mathbb{N}^* \end{cases} \tag{1}$$

where $m \geq 2$ is the number of neurons in the networks; $\Delta x_i(t_k)$ are the impulses at moments $t_k$ and $t_1 < t_2 < \cdots$ is a strictly increasing sequence such that $t_k \to +\infty$ when $k \to +\infty$, $x_i(t)$ denote the potential (or voltage ) of cell $i$ at time $t$; $d_i(t)$ represents an amplification function; $a_i(t)$ is the rate with which the unit self-regulates or rests its potential when isolated from others and inputs; $b_{ij}(t), c_{ij}(t)$ denote the strengths of connectivity between cell $i$ and $j$ time $t$, respectively. The activation function $f_j(\cdot)$ shows how the $i$ neuron reacts to the inputs time delay $\tau_{ij}$ is nonnegative constant, which corresponds to the finite speed of the axonal signal transmission; $J_i(t)$ is the external bias of cell $i$ at time $t$. The system (1) is supplemented with initial values given by:

$$x_i(s) = \varphi_i(s)\; s \in [-\tau, 0] \tag{2}$$

Z.N. Benbouziane (✉) • N. Boukli-Hacene
Département de Mathématiques, Université de Tlemcen, B.P.119, Tlemcen, Algérie
e-mail: Zehour.Benbouziane@yahoo.fr



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_7

$\tau = max\{\tau_{ij} \; 1 \leq i, \; j \leq m\}$
$\varphi_i \in C([-\tau, 0], \mathbb{R}) \quad i = 1, \cdots, m.$

In this communication, we consider the system (1), (2) and apply a different method from the ones in the literature; our method is based on upper and lower solutions and an iterative technique; our assumptions on the activation function and $a_i(t), b_{ij}(t), c_{ij}(t), J_i(t)$ are less restrictive, we note that we don't need the periodicity for all the functions.

## 2 Notations and Definitions

We suppose that there exists $T > 0$ and $q \in \mathbb{N}^* : t_{k+q}=t_k + T \quad \forall k=1, 2, \cdots$ Set that :

- $A_i\big(x_i(t)\big) = d_i \; \big(x_i(t)\big) \Big[ \; a_i \; x_i(t) - \sum_{j=1}^{m} \; b_{ij}(t) \; f_j\big(x_j(t)\big) + \sum_{j=1}^{m} \; c_{ij} \; f_j \big(x_j(t-\tau_{ij})\big) - J_i(t)\Big]$
- $D = \{t_1, \cdots, t_q\}$ and the Banach space $PC_D\Big([-\tau, T]\Big)$ of piece-wise continuous functions on $[-\tau, T]$

And we study the auxiliary system

$$\begin{cases} \dfrac{dx_i}{dt} = -A_i\big(x_i(t)\big) \quad t \neq t_k, \;\; t \in ]0, T] \\ \Delta \, x_i(t_k) = - \, \gamma_{ik} \, \big(x_i(t_k)\big); \quad k = 1, \; 2, \cdots, q \\ x_i(s) = \varphi_i(s) \; s \in [-\tau, 0], \;\; x_i(0) = x_i(T) \end{cases} \tag{3}$$

The lower solution of CGNN model satisfies differential system

$$\begin{cases} \dfrac{d\alpha_i}{dt} \leq -A_i\big(\alpha_i(t)\big) \quad t \neq t_k, \;\; t \in ]0, T] \\ \Delta \, \alpha_i(t_k) \leq - \, \gamma_{ik} \, \big(\alpha_i(t_k)\big); \quad k = 1, \; 2, \cdots, q \\ \alpha_i(s) \leq \varphi_i(s) \; s \in [-\tau, 0], \;\; \alpha_i(0) = \alpha_i(T) \end{cases}$$

The upper solution $\beta_i$ is defined by:

$$\begin{cases} \dfrac{d\beta_i}{dt} \geq -A_i\big(\beta_i(t)\big) \quad t \neq t_k, \;\; t \in ]0, T] \\ \Delta \, \beta_i(t_k) \geq - \, \gamma_{ik} \, \big(\beta_i(t_k)\big); \quad k = 1, \; 2, \cdots, q \\ \beta_i(s) \geq \varphi_i(s) \; s \in [-\tau, 0], \;\; \beta_i(0) = \beta_i(T) \end{cases}$$

**Proposition 1.** *If $x = (x_i)_{i=1,\cdots,m}$, is a solution for the system (3), then the piecewise continuous functions defined on $[-\tau, +\infty[$ by :*

$$\tilde{x}(t) = \begin{cases} \sum_{n \geq 0} \chi_{[nT,(n+1)T]} \quad x(t - nT) & \text{if} \quad t > 0 \\ \\ \varphi(t) & \text{if} \quad t \in [-\tau, 0] \end{cases}$$

*where*

$$\chi(t) = \begin{cases} 1 & \text{if} \quad t \in [nT, (n+1)T] \\ \\ 0 & \text{if not.} \end{cases}$$

*is a periodic solution for the system (1), (2).*

## 3 The Main Result

Through this paper we assume that:

- **($H_1$)** $a_i,\ b_{ij},\ c_{ij} \in C(\mathbb{R}, \mathbb{R}^+),\ \ d_i, J_i \in C(\mathbb{R}, \mathbb{R})\ \ \forall\ \ i\ ,j = 1, \cdots, m.$
- **($H_2$)** $d_i(\cdot)\ \ J_i(\cdot)$ is decreasing $\forall\ \ i = 1, \cdots, m.$
- **($H_3$)** $\forall u_1,\ u_2,\ v_1,\ v_2\ \in \mathbb{R}, \exists l_{ij} > 0 : d_i(u_1)f_j(v_1) - d_i(u_2)f_j(v_2) \leq l_{ij}(u_1 - u_2)(v_1 - v_2)\ \ \forall\ \ i\ ,j = 1, \cdots, m.$
- **($H_4$)** $\gamma_{ik}(u) - \gamma_{ik}(v) \leq \delta_{ik}(u - v)$ with $(1 - \delta_{ik}) > 0;\ i = 1, \cdots, m;\ k = 1, \cdots, q.$
- **($H_5$)** $\forall u,\ v\ \in \mathbb{R}, u \geq v,\ \ d_i(u)u - d_i(v)v \leq v - u.$

**Theorem 1.** *Let $(\alpha_i)_{i=1,\cdots,m}$, $(\beta_i)_{i=1,\cdots,m}$ be the lower and the upper solutions of the system (3), such that $\alpha_i \leq \beta_i$, $i = 1, \cdots, m$. Under conditions **(H1)–(H5)** the system (3) admits at last one solution $x = (x_i)_{i=1,\cdots,m}$, such that : $\alpha_i \leq x_i \leq \beta_i, i = 1, \cdots, m$.*

*Proof.* The proof is based on an iterative method and Ascoli-Arzela Theorem. We construct a recurrent suite of functions $(y_{n,i})_i$ in $PC_D(I)$ such that $y_{0,i} = \alpha_i$ and verifies the system:

$$\begin{cases} y'_{n,i}(t) \ = \ A_i\big(y_{n-1,i}(t)\big) \quad t \neq t_k, \ \ t \in ]0, T] \\ \\ y_{n,i}(t_k^+) \ = y_{n-1,i}(t_k^-) - \gamma_{ik}\big(y_{n-1}(t_k)\big) \quad k = 1, \cdots, q. \\ \\ y_{n,i}(0) \ = y_{n-1,i}(T)\ , \qquad y_{n,i}(T) \ = y_{n-1,i}(0). \end{cases}$$

We show that there exists a subsequence which converges to the solution of the system (3).

# Some Fixed Point Theorems for Orbitally-$(p, q)$-Quasi-contraction Mappings

**Wajdi Chaker, Abdelaziz Ghribi, Aref Jeribi, and Bilel Krichen**

**Abstract** In this paper, we provide some existence and uniqueness results for a $(p, q)$-quasi-contraction mapping acting on an orbitally-complete cone metric space. These results generalize several fixed point theorems, in particular those due to Ilić and Rakočević's for quasi-contraction mappings (Ilić and Rakočević, Appl Math Lett 22(5):728–731, 2009), convex contraction mapping, and two-sided convex contraction of order 2.

## 1 Introduction

The Banach contraction principle is the simplest and one of the most versatile elementary results in fixed point theory [3, 11]. It was introduced by Banach in [2] and remained a fundamental tool in nonlinear analysis. Especially for nonlinear mappings, this principle has incited several authors to extend it. At this level, we can mention the contraction type of Kannan [14], Chatterjea [5], Zamfirescu [19], Reich [16], and Ćirić [6] who gave one of the most general contraction conditions called Quasi-contraction.

W. Chaker (✉)
Higher institute of informatics and multimedia of Gabes., BP 122, 6033 Cité El Amel 4, Gabès, Tunisie
e-mail: wajdi_chaker@yahoo.fr

A. Ghribi
Higher Institute of Business Administration of Sfax, Airport Road Km 4 B.P. 1013, 3018 B.P. 1013, Sfax, Tunisia
e-mail: ghribi_ abdelaziz@yahoo.fr

A. Jeribi
Faculty of Science of Sfax, University of Sfax, Soukra Road Km 3.5, B.P. 1171, 3000, Sfax, Tunisia
e-mail: Aref.Jeribi@fss.rnu.tn

B. Krichen
Preparatory Engineering Institute, Menzel Chaker Road Km 0.5, B.P. 1172 - 3018, Sfax, Tunisia
e-mail: krichen.bilel.ipeis@gmail.com



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_8

Inspired by the paper of Huang and Zhang [8], Ilić and Rakočević [9] have extended the concept of quasi-contraction mappings to the cone metric spaces and have also generalized Theorem 1 in [8] to quasi-contraction mappings in complete cone metric spaces.

Recently, many authors have studied some variants of contraction conditions and proved some fixed point theorems in cone metric space whether the underlying cone is normal or not normal. We mention for examples [4, 12, 13, 15, 17, 18, 20].

This paper is organized as follows. In Section 2, we give some definitions and preliminary results needed in the sequel. In Section 3, we extend the concept of $(p, q)$-quasi-contraction mappings [7] in cone metric space. This mapping type extends the Ilić and Rakočević's quasi-contraction mappings, convex contraction mappings of order $n$ (see, for instance, [1, 10]) and the two-sided convex contraction mappings [10]. The main result of this section is that every continuous $(p, q)$-quasi-contraction mapping in complete cone metric space has a unique fixed point and the Piccard iteration converges to this point. Moreover, we obtain fixed point theorems for certain classes of not continuous mappings which generalize many known results.

## 2 Preliminaries

Let $E$ be a real Banach space. A nonempty subset $P$ of $E$ is called a cone if, and only if, we have

(i) $P$ is closed and $P \neq \{0\}$.
(ii) For every positive real $a$, $aP \subset P$.
(iii) $P + P \subset P$ and $P \cap (-P) = \{0\}$.

Given a cone $P \subset E$, we can define a partial ordering $\preccurlyeq$ on $E$ with respect to $P$ by $x \preccurlyeq y$ if, and only if, $y - x \in P$. We will indicate by $x \prec y$ that $x \preccurlyeq y$ but $x \neq y$, and by $x \ll y$ that $y - x \in intP$, where $intP$ denotes the interior of $P$. The cone $P$ is called normal, if there is a number $K > 0$ such that, for all $x, y \in E$, we have

$$0 \preccurlyeq x \preccurlyeq y \text{ implies } \|x\| \preccurlyeq K \|y\| . \tag{1}$$

The least positive number satisfying the above inequality is called the normal constant of $P$. In [8], Huang and Zhang introduce the notion of cone metric space which generalizes the metric space.

**Definition 1 ([8]).** Let $P$ be a cone in a Banach space such that $intP \neq \varnothing$ and $\preccurlyeq$ is a partial ordering in $E$ with respect to $P$. A cone metric on a nonempty set $X$ is a function $d : X \times X \longrightarrow E$ such that, for all $x, y, z \in X$, we have

(a) $x = y$ if, and only if, $d(x, y) = 0$,
(b) $0 \preccurlyeq d(x, y) = d(y, x)$,
(c) $d(x, y) \preccurlyeq d(x, z) + d(z, y)$.

A cone metric space is a pair $(X,d)$ such that $X$ is a nonempty set and $d$ is a cone metric on $X$. We recall some definitions:

**Definition 2 ([8]).** Let $(X,d)$ be a cone metric space. Let $\{x_n\}$ be a sequence in $X$. Then,

(i) $\{x_n\}$ converges to $x \in X$ if, for every $c \in E$ with $0 \ll c$, there exists a natural number $N$ such that, for all $n \geq N$, we have $d(x_n,x) \ll c$. We denote this by $x_n \to x (n \to \infty)$ or $\lim_{n\to\infty} x_n = x$.
(ii) $\{x_n\}$ is a Cauchy sequence if, for every $c \in E$ with $0 \ll c$, there exists a natural number $N$ such that, for all $n,m \geq N$, we have $d(x_n,x_m) \ll c$.
(iii) $(X,d)$ is a complete cone metric space, if every Cauchy sequence is convergent.

In the case where $P$ is a normal cone, we have the following lemmas.

**Lemma 1 ([8]).** *Let $(X,d)$ be a cone metric space, let $P$ be a normal cone with a normal constant $K$, and let $\{x_n\}$ be a sequence in $X$.*

*(i) If the limit of $\{x_n\}$ exists, then it is unique.*
*(ii) Every convergent sequence in $X$ is a Cauchy sequence.*
*(iii) $\{x_n\}$ is a Cauchy sequence if, and only if, $d(x_n,x_m) \to 0$ $(n,m \to \infty)$.*

**Lemma 2 ([8]).** *Let $(X,d)$ be a cone metric space and let $P$ be a normal cone with a normal constant $K$. If $\{x_n\}$ and $\{y_n\}$ are two sequences in $X$ such that $x_n \to x$ and $y_n \to y$, then*

$$d(x_n,y_n) \to d(x,y) \ (n \to \infty). \tag{2}$$

## 3 Main results

In the sequel, we suppose that $E$ is a Banach space, $P$ is a normal cone in $E$ with $intP \neq \varnothing$, $K$ is the normal constant of $P$, and $\preccurlyeq$ is a partial ordering in $E$ with respect to $P$. We generalize the $(p,q)$-quasi-contraction mapping on a cone metric space as follows. Such a mapping is a generalization of Ilić and Rakočević's quasi-contraction mapping on a cone metric space [9].

**Definition 3.** Let $(X,d)$ be a cone metric space and let $p,q$ be two natural numbers such that $0 < p \leq q$. The mapping $T : X \longrightarrow X$ is said a $(p,q)$-quasi-contraction, if there exists a number $c \in [0,1)$ such that, for every $x,y \in X$, there is $u \in C_{p,q}(x,y)$, such that

$$d(T^p x, T^q y) \preccurlyeq c.u, \tag{3}$$

where $C_{p,q}(x,y) = \{d(T^r x, T^s y), d(T^r x, T^{r'} x), d(T^s y, T^{s'} y) : 0 \leq r,r' \leq p$ and $0 \leq s,s' \leq q\}$.

Let $n \in \mathbb{N}$, $x \in X$ and $O(x : n) = \{x, Tx, T^2x, \ldots, T^nx\}$, the set $O(x : \infty) = \{x, Tx, T^2x, \ldots\}$ is called the orbit of $T$ at $x$. The partial cone metric space $(X, d)$ is said to be $T$-orbitally complete, if every Cauchy sequence contained in an orbit of $T$ converges in $X$. It is obvious that each complete cone metric space is $T$-orbitally complete, but the converse does not hold [Example 3.1, [15]].

Let $(X, d)$ be a cone metric space. The mapping $T : X \longrightarrow X$ is called orbitally $(p, q)$-quasi-contraction if it is $(p, q)$-quasi-contraction on any orbit of $X$.

*Example 1.* $E = \mathbb{R}^2$,

$$P = \{(x, y) \in E : x, y \geq 0\},$$

$X = \mathbb{R}$ and $d : X \times X \to E$ is defined by

$$d(x, y) = (|y - x|, \alpha|y - x|),$$

where $\alpha \geq 0$ is a constant. $(X, d)$ is a cone metric space.

The mapping $T$, which is defined in the cone metric space $(X, d)$ by $Tx = x$, is an orbitally $(p, q)$-quasi-contraction, but it is not a $(p, q)$-quasi-contraction.

Let $\Delta(x, p, n) = \{d(a, b) : a, b \in \{T^ix, p \leq i \leq n\}\}$. For more simplicity, we denote $\Delta(x, n) = \Delta(x, 0, n)$ and $\Delta(x : \infty) = \{d(a, b) : a, b \in O(x : \infty)\}$. For $F \subset E$, we define $\delta(F) = \sup\{\|x\|, x \in F\}$. We will need the following result.

**Lemma 3.** *Let $(X, d)$ be a cone metric space. Let $T : X \longrightarrow X$ be an orbitally $(p, q)$-quasi-contraction mapping. Then, $\Delta(x : \infty)$ is bounded.*

*Proof.* Let $x \in X$. If $Tx = x$, it is obvious that $\Delta(x : \infty)$ is bounded. We suppose that $Tx \neq x$, and let $n_0 \in \mathbb{N}$ such that $\max\{c^{n_0}K, c^{n_0}K^2\} < 1$ and $Tx \neq x$. Let's choose $i, j, n \in \mathbb{N}$ such that $n_0.q \leq i < j \leq n$. Since $T$ is a $(p, q)$-quasi-contraction, we deduce that

$$d(T^ix, T^jx) \preccurlyeq c.u_1,$$

where $u_1 \in \Delta(x, i - q, n)$. By using the same argument, we infer that there exists $u_2 \in \Delta(x, i - 2q, n)$, such that $u_1 \preccurlyeq c.u_2$. Hence,

$$d(T^ix, T^jx) \preccurlyeq c^2.u_2.$$

After $n_0$ iterations, we conclude that there exists $u_{n_0} \in \Delta(x, n)$, such that

$$d(T^ix, T^jx) \preccurlyeq c^{n_0}.u_{n_0}. \tag{4}$$

Since $P$ is normal and $c^{n_0}K < 1$, we have

$$\left\|d(T^ix, T^jx)\right\| < \delta(\Delta(x, n)). \tag{5}$$

We conclude that

$$\delta(\Delta(x,n)) = \max\{\left\|d(T^i x, T^l x)\right\|, \left\|d(T^i x, T^j x)\right\| : 0 \le i,j < n_0.q \le l \le n\}. \quad (6)$$

In the first case, we suppose that $\delta(\Delta(x,n)) = \left\|d(T^i x, T^l x)\right\|$ for some $0 \le i < n_0.q \le l \le n$. Since $d$ is a cone metric, we deduce that

$$d(T^i x, T^l x) \preccurlyeq d(T^i x, T^{n_0.q} x) + d(T^{n_0.q} x, T^l x)),$$

which implies that

$$\left\|d(T^i x, T^l x)\right\| \le K \left\|d(T^i x, T^{n_0.q} x)\right\| + c^{n_0} K^2 \delta(\Delta(x,n)).$$

Hence,

$$\delta(\Delta(x,n)) \le \frac{K}{1 - c^{n_0} K^2}.\delta(\Delta(x, n_0.q)). \quad (7)$$

In the second case, we assume that $\delta(\Delta(x,n)) = \left\|d(T^i x, T^j x)\right\|$ for some$1 \le i, j \le n_0$. Since $d(T^i x, T^j x) \in \Delta(x, n_0)$, we have

$$\delta(\Delta(x,n)) \le \delta(\Delta(x, n_0.q)). \quad (8)$$

The Inequalities (7) and (8) imply

$$\delta(\Delta(x,n)) \le \max\{1, \frac{K}{1 - c^{n_0} K}\}\delta(\Delta(x, n_0.q)). \quad (9)$$

Since $\delta(\Delta(x,\infty)) = \sup\{\delta(\Delta(x,n)) : n \in \mathbb{N}\}$, we conclude that $\Delta(x,\infty)$ is bounded.

**Theorem 1.** *Let $(X, d)$ be a cone metric space and let $T : X \longrightarrow X$ be a continuous and a $T$-orbitally $(p, q)$-quasi-contraction mapping. If $X$ is $T$-orbitally complete, then the sequence $\{T^n x\}$ converges to a fixed point, for every $x \in X$. Further, if $T$ is a $(p, q)$-quasi-contraction, then for any $x \in X$, the fixed point is unique.*

*Proof.* If $Tx = x$, the result holds. In the sequel, we assume that $Tx \ne x$, and we prove that $\{T^n x\}$ is a Cauchy sequence.

Let $\epsilon > 0$ and let's choose $N$ so that $K.c^N.\delta(\Delta(x,\infty)) < \epsilon$. For every two natural numbers, namely $n$ and $m$, such that $m \ge n \ge N.q$, there exists $u_1 \in \Delta\{x, n-q, m\}$, such that

$$d(T^n x, T^m x) \preccurlyeq c.u_1.$$

Since $T$ represents a $T$-orbitally $(p, q)$-quasi-contraction mapping, every $v_1 \in \Delta(x, n-q, m)$ satisfies $v_1 \preccurlyeq cv_2$ where $v_2 \in \Delta(x, n-2q, m)$, and after $N$ steps, we deduce that there exists $u_N \in \Delta(x, m)$, such that

$$d(T^n x, T^m x) \preccurlyeq c^N.u_N.$$

By applying Lemma 3, we get

$$\|d(T^n x, T^m x)\| \leq k.c^N \delta(\Delta(x:\infty)) < \epsilon,$$

and $\{T^n x\}$ is a Cauchy sequence in $(X, d)$. Since (X,d) is orbitally complete, there exists $y \in X$, such that $\{T^n x\}$ converges to $y$. By using the continuity of $T$, we conclude that $y$ is a fixed point of $T$.

Further, if $T$ is a $(p, q)$-quasi-contraction on $X$, suppose that there exists another $z \in X$ such that $Tz = z$, then we get

$$d(z, y) = d(T^p z, T^q y) \preccurlyeq c.d(z, y). \tag{10}$$

Since $c < 1$, we deduce that $d(z, y) = 0$. Hence, the fixed point of $T$ is unique.

As a corollary, when $X$ is a metric space, we obtain Fisher's main result [Theorem 2, [7]].

**Corollary 1.** *Let $T$ be a $(p, q)$-quasi-contraction on the complete metric space $X$ into itself, and let $T$ be continuous. Then, $T$ has a unique fixed point in $X$.*

In Theorem 2, we didn't need the continuity of $T$ when $p = 1$. In this case, we mention the following theorem.

**Theorem 2.** *Let $(X, d)$ be a cone metric space and let $T : X \longrightarrow X$ be a continuous and $T$-orbitally $(1, q)$-quasi-contraction mapping. If $X$ is $T$-orbitally complete, then the sequence $\{T^n x\}$ converges to a fixed point, for every $x \in X$. Further, if $T$ is a $(1, q)$-quasi-contraction, then for any $x \in X$, the fixed point is unique and $T$ is continuous in such a point.*

*Proof.* According to Theorem 2, the sequence $\{T^n x\}$ converges to an element $y \in X$. Let $n \in \mathbb{N}$ be big enough. Then, we have

$$\begin{aligned} d(y, Ty) &\preccurlyeq d(y, T^n y) + d(T^n y, Ty) \\ &\preccurlyeq d(y, T^n y) + c.u, \end{aligned}$$

where $u \in \{d(T^{n-i}y, T^{n-j}y), d(Ty, T^{n-j}y), d(y, T^{n-j}y), d(y, Ty) : 0 \leq i, j \leq q\}$. By using the fact that $P$ is a normal cone with a constant $K$, we deduce that

$$\|d(y, Ty)\| \leq K. \|d(y, T^n y)\| + K.c. \|u\| . \tag{11}$$

Since the sequence $\{T^n x\}$ converges to $y$, we infer that $d(y, Ty) = 0$. Hence, $y$ is a fixed point of $T$. Further, if $T$ is a $(1, q)$-quasi-contraction, then the uniqueness of the fixed point is similar to that of Theorem 2. We will have to prove that $T$ is continuous in the fixed point $y$. For this purpose, let $\{y_n\}$ be a sequence of points in the cone metric space $X$ which converges to $y$. Then there exists $u \in C_{1,q}(y_n, y)$, such that

$$\begin{aligned} d(Ty_n, y) = d(Ty_n, T^q y) &\preccurlyeq c.u \\ &\preccurlyeq c[d(y_n, y) + d(y, Ty_n)]. \end{aligned}$$

It follows that

$$d(Ty_n, y) \preccurlyeq \frac{c}{1-c} d(y_n, y).$$

Hence,

$$\|d(Ty_n, y)\| \leq \frac{cK}{1-c} \|d(y_n, y)\| . \tag{12}$$

We conclude that $\lim_{n\to\infty} \|d(Ty_n, y)\| = 0$, which completes the proof.

As a corollary, when $p = q = 1$, we get the main result of Ilić and Rakočević [9].

**Corollary 2.** *Let $(X, d)$ be a complete cone metric space and let $T : X \longrightarrow X$ satisfy the inequality*

$$d(Tx, Ty) \preccurlyeq c.u \tag{13}$$

*for some $u \in \{d(x, y), d(x, Tx), d(y, Ty), d(x, Ty), d(y, Tx)\}$. Then, $T$ has a unique fixed point in $X$, and for every $x \in X$, the sequence $\{T^n x\}$ converges to this fixed point.*

When $X$ is a metric space and $p = 1$, we obtain Theorem 3 in [7].

**Corollary 3.** *Let $T$ be a $(1, q)$-quasi-contraction on the complete metric space $X$ into itself. Then, $T$ has a unique fixed point in $X$.*

*Remark 1.* If $P$ is minihedral, i.e. $\sup\{x, y\}$ exists for all $x, y \in E$, and $p = q = 2$, then we obtain, as particular cases, the Istratescu's fixed point theorem for convex contraction mappings of order 2 [Theorem 1.2., [10]], the Istratescu's fixed point theorem for a two-sided convex contraction of order 2 [Theorem 2.3., [10]] in complete metric spaces, and their generalizations to the cone metric spaces obtained by Alghamdi et al. in [1].

## References

1. M.A. Alghamdi, S.H. Alnafei, S. Radenović, and N. Shahzad, *Fixed point theorems for convex contraction mappings on cone metric spaces*, Math. Comput. Modelling **54** (2011), no. 9-10, 2020–2026.
2. S. Banach, *Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales*, Fund. Math. **3** (1922), no. 3333, 133–181.
3. W. Chaker and A. Jeribi, *Common fixed point of single valued and multivalued mappings.* Afr. Mat. 25 (2014), no. 3, 617–622. DOI 10.1007/s13370-012-0120-8.
4. W. Chaker, A. Ghribi, A. Jeribi and B. Krichen, *Fixed point theorem for quasi-contractive mapping in partial cone metric space.* Afr. Mat. (2014), On line. DOI 10.1007/s13370-014-0271-x.
5. S.K. Chatterjee, *Fixed point theorems*, Rend. Acad. Bulgare Sc. **25** (1972), 727–730.

6. Lj. B. Ćirić, *A generalization of Banach's contraction principle*, Proc. Amer. Math. Soc. **45** (1974), 267–273.
7. B. Fisher, *Quasicontractions on metric spaces*, Proc. Amer. Math. Soc. **75** (1979), no. 2, 321–325.
8. L.-G. Huang and X. Zhang, *Cone metric spaces and fixed point theorems of contractive mappings*, J. Math. Anal. Appl. **332** (2007), no. 2, 1468–1476.
9. D. Ilić and V. Rakočević, *Quasi-contraction on a cone metric space*, Appl. Math. Lett. **22** (2009), no. 5, 728–731.
10. V.I. Istrăţescu, *Some fixed point theorems for convex contraction mappings and mappings with convex diminishing diameters. I*, Ann. Mat. Pura Appl. (4) **130** (1982), 89–104.
11. A. Jeribi and B. Krichen, *Nonlinear Functional Analysis in Banach Spaces and Banach Algebras: Fixed Point Theory under Weak Topology for Nonlinear Operators and Block Operator Matrices with Applications.* Monographs and Research Notes in Mathematics, CRC Press: Taylor and Francis, 2015.
12. Z. Kadelburg, M. Pavlović, and S. Radenović, *Common fixed point theorems for ordered contractions and quasicontractions in ordered cone metric spaces*, Comput. Math. Appl. **59** (2010), no. 9, 3148–3159.
13. Z. Kadelburg, S. Radenović, and V. Rakočević, *Remarks on "Quasi-contraction on a cone metric space"*, Appl. Math. Lett. **22** (2009), no. 11, 1674–1679.
14. R. Kannan, *Some results on fixed points*, Bull. Calcutta Math. Soc. **60** (1968), 71–76.
15. H. K. Pathak and N. Shahzad, *Fixed point results for generalized quasicontraction mappings in abstract metric spaces*, Nonlinear Anal. **71** (2009), no. 12, 6068–6076.
16. S. Reich, *Kannan's fixed point theorem*, Boll. Un. Mat. Ital. **4** (1971), 1–11.
17. Sh. Rezapour, R. H. Haghi, and N. Shahzad, *Some notes on fixed points of quasi-contraction maps*, Appl. Math. Lett. **23** (2010), no. 4, 498–502.
18. Sh. Rezapour and R. Hamlbarani, *Some notes on the paper: "Cone metric spaces and fixed point theorems of contractive mappings*, J. Math. Anal. Appl. **345** (2008), no. 2, 719–724.
19. T. Zamfirescu, *Fix point theorems in metric spaces*, Arch. Math. (Basel) **23** (1972), 292–298.
20. X. Zhang, *Fixed point theorem of generalized quasi-contractive mapping in cone metric space*, Comput. Math. Appl. **62** (2011), no. 4, 1627–1633.

# Third order rational ordinary differential equations with integer indices of Fuchs

**Yassine Adjabi and Arezki Kessi**

**Abstract** A series of papers was devoted to the investigation of third order ordinary differential equations of P-type. The interest in such problems is due to their applications in physics, chemistry, etc.

In the year 2007, Conte et al. in the paper entitled *"Painlevé structure of a multi-ion electrodiffusion system"* showed that the coupled nonlinear system descriptive of multi-ion electrodiffusion of the first order corresponds to a nonlinear ordinary differential equation of P-type (solutions of such equations have no movable critical singular points). We understand that the solutions of this problem are not completely known; a topic addressed in the present paper.

Some new third order rational ordinary differential equations with integer indices of Fuchs as well as recent ones are found.

**Keywords** Nonlinear differential equation • Painlevé property • fixed singular points • Painlevé test • Fuchs indices (resonances) • homographic transformations

**Mathematics Subject Classification** 34M55, 34A34, 34E20, 33E30

## 1 Introduction

The current interest in the "Painleve property (PP)" stems from the observation by Ablowitz and Segur [1] that reductions of partial differential equations of soliton type gave rise to ODEs whose movable singularities were only poles. The Painleve property and test are concerned with the singularity structure of the general

Y. Adjabi (✉)
Faculty of Sciences, Department of Mathematics, University of M'hamed Bougra, Boumerdes, Algeria
e-mail: adjabiy@yahoo.fr

A. Kessi
Faculty of Mathematics, USTHB, BP 32 El Alia, 16 311 Bab Ezzouar, Algiers, Algeria
e-mail: arkessi@yahoo.com



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_9

solution of a nonlinear differential equation, which can “spontaneously evolve” with positions depending upon constants of integration and, therefore, upon initial conditions. Such singularities are said to be movable. The Painlevé test has been a very successful tool for isolating integrable differential equations (both ODEs and PDEs). The test could be any algorithm designed to determine necessary conditions for a differential equation to have the PP.

The undergone work is motivated by the recent appearance of Painlevé equations type in the following coupled nonlinear system describing a multi-ion electrodiffusion (see [5])

$$\begin{cases} \frac{dn_i}{dx} = \nu_i n_i p - c_i, \quad \nu_i n_i \neq 0, \; i = 0, 1, \dots, m, \\ \frac{dp}{dx} = \sum_{i=1}^{m} \nu_i n_i, \end{cases} \tag{1}$$

where $x$ is the coordinate normal to the planar boundaries, $p$ is the electric field, and $n_j$ is the number of ions. On the basis of the established connection between solutions of the system (1) and solutions of nonlinear P-type equations, a particular solution of the equation for the case $m = 3$ is given by

$$p^2 p''' = a_1 p p'' p' + a_2 p'^3 + \text{subdominant terms} = 0. \tag{2}$$

In section 5 of [5, page 6], the authors reported that the coupled system (1) contains numerous other interesting questions. It should be noted that the details of (1) and (2) were not given in [5]. For this reason, we study in this article the latter equation.

Equation (2) is a differential equation of Chazy type [4]. This class has been investigated in [2, 5, 10–12], but the results therein are insufficient.

The present paper deals with investigating the nonlinear third order rational differential equations possessing the Painlevé property. This is achieved by considering the differential equations of the form

$$y''' = a_1 \frac{y'' y'}{y} + a_2 \frac{y'^3}{y^2} + B(z, y)\, y'' + C(z, y) \left(y'\right)^2 + D(z, y)\, y' + E(z, y), \tag{3}$$

where $B(z, y)$, $C(z, y)$, $D(z, y)$, and $E(z, y)$ are rational functions of $y$ with coefficients analytic in $z$.

Using the small parameter method [9], we find that if equation (3) belongs to the Painlevé type, then

$$B(z, y) = \sum_{i=-1}^{1} b_i y^i, \; C(z, y) = \sum_{i=-2}^{0} c_i y^i, \; D(z, y) = \sum_{i=-2}^{2} d_i y^i, \; E(z, y) = \sum_{i=-2}^{4} e_i y^i, \tag{4}$$

where the coefficients $b_i$, $c_i$, $d_i$, and $e_i$ are analytic functions of $z$.

A number of papers, in particular [2, 5, 8, 11, 12] deal with movable singularities of solutions of third order *rational* differential equations of the form (3) but fails to provide complete classification. In contrast, a complete classification was constructed for third order ordinary equations of *polynomial* class in [7].

As a continuation to [2], our main task is to find all the canonical (distinct) reduced equations of the form (3) that have the Painlevé property and hence all full equations that are built upon these reduced equations. The full equations necessarily take the form (3) with (4) above. We are looking for such values of constants $a_1$, $a_2$ and the expressions $b_i$, $c_i$, $d_i$, and $e_i$ so that equation (3) passes the Conte-Fordy-Pickering test.

Using the compatibility condition for the Painlevé type equations (see [6, 7]), several new third order ordinary differential equations (ODEs) were found.

## 2 Corresponding simplified equations

The method of small parameter [3, 8] plays an important role in Painlevé analysis. It permits one to justify the Painlevé property by studying a simplified equation obtained by a special procedure. For the simplified equation, the properties of movable singularities of solutions can often be established directly (either by integrating the simplified equation or by finding a family of solutions with a movable singularity).

The simplified equation corresponding to (3) is easily determined by setting $(z \mapsto z_0 + \varepsilon z)$, where $\varepsilon \neq 0$ is a parameter, then we put $\varepsilon = 0$, if the equation (3) is to be free of mobile critical points, it is necessary that $y$ be uniform, so that it must be given by an equation of the form

$$y'''y^2 = a_1 yy'y'' + a_2 y'^3. \tag{5}$$

By making the change of variable

$$y' = vy, \tag{6}$$

in the equation (5), the following equation is produced

$$v'' = (a_1 - 3)\, vv' + (a_1 + a_2 - 1)\, v^3. \tag{7}$$

Equation (7) is referred to in [3]. It is known that, for solutions of (5) to have only single-valued nonstationary polar singularities, it is necessary and sufficient that all nonstationary singularities of solutions of (7) are simple poles with integer residues.

The expansion of general solution for equation (7) in the formal Laurent series takes the form

$$v = v_0\,(z - z_0)^q + v_1\,(z - z_0)^{(q+1)} + v_2\,(z - z_0)^{(q+2)} + \ldots, \tag{8}$$

with $v_0$ nonzero and $r$ is the leading power that needs to be found. When we substitute this series into (7), we see that $q = -1$ and the residue $v_0$ of any moving pole must be an integer root of the equation

$$(a_1 + a_2 - 1)\, v_0^2 - (a_1 - 3)\, v_0 - 2 = 0. \tag{9}$$

Moreover, the roots of the indicial equation corresponding to

$$(r+1)\left(-r^2+((a_1-3)\,v_0+3)\,r+3\,(a_1+a_2-1)\,v_0^2-2\,(a_1-3)\,v_0-2\right)=0 \tag{10}$$

must be distinct integers.

Firstly, by taking into account these requirements, we find that the equation (5) is with fixed critical points only if the couple $(a_1,\ a_2)$ takes one of the following values

$$(a_1,\ a_2)\in\left\{\begin{array}{c}\left(3\frac{n-1}{n},\ -\frac{(n-1)(2n-1)}{n^2}\right),\left(\frac{3n-1}{n},\ -\frac{(n-1)(2n+1)}{n^2}\right),\\ \left(\frac{3n-2}{n},\ -2\frac{n-1}{n}\right),(3,\ -2),\left(3,\ -2\frac{n^2-1}{n^2}\right)\end{array}\right\}, \tag{11}$$

where $n$ is an integer number different from 0.

*Remark 1.* In ref. [2], it was shown that the corresponding list of reduced equations (3) that associate the coefficients $a_k,\ k=1,\ 2$ in equation (11) with $n$ is an integer number different from 0 has passed the Painlevé test.

Secondly, two equations will be regarded as equivalent if they are related by a gauge transformation of the form

$$v\,(z)=\lambda\,(z)\,V\,(t)+\mu\,(z)\,,\ t=\varphi\,(z)\,, \tag{12}$$

we observe also that to obtain canonical forms for the Painlevé-type equations, it is often most convenient to use a transformation (12), where $\lambda,\ \mu$ and $\varphi$ are analytic functions of the complex variables $z$, does not alter the form of equation (7) which becomes

$$\ddot{V}=AV\dot{V}+BV^3+C\dot{V}+EV^2+FV+G, \tag{13}$$

According to [3], the necessary condition for equation (13) satisfies the Painlevé property

$$(A,\ B)\in\{(-3,\ -1),\ (-2,\ 0),\ (-1,\ +1),\ (0,\ 2),\ (0,\ 0)\}\,,$$

we assume that this condition is satisfied; one has only to determine $\lambda,\ \mu,\ \varphi$ by

$$\begin{cases}\varphi' A=(a_1-3)\,\lambda, \\ \varphi'^2 B=(a_1+a_2-1)\,\lambda^2, \\ \varphi' C=\left((a_1-3)\,\mu-2\dfrac{\lambda'}{\lambda}-\frac{\varphi''}{\varphi'}\right), \\ \varphi'^2 E=(a_1-3)\,\lambda'+3\,(a_1+a_2-1)\,\mu, \\ \varphi'^2 F=(a_1-3)\left(\dfrac{\lambda'}{\lambda}\mu+\mu'\right)+3\,(a_1+a_2-1)\,\mu^2-\frac{\lambda''}{\lambda}, \\ \lambda\varphi'^2 G=(a_1-3)\,\mu\mu'+(a_1+a_2-1)\,\mu^3-\mu'',\end{cases} \tag{14}$$

where $C,\ E,\ F$, and $G$ are given by system (14).

Consequently one can see that one of the roots of (9) may be taken equal to 1 (i.e., $n = 1$), then it is necessary that the couple $(a_1,\ a_2)$ takes one of the following values in order to obtain solutions of (5) with no movable critical points

$$(a_1,\ a_2) \in \{(0,\ 0),\ (1,\ 0),\ (2,\ 0),\ (3,\ 0),\ (3,\ -2)\}\,.$$

The Painleve analysis of third order polynomial equation (3) was carried out in Chazy [4], Bureau [3], and Cosgrove [7] for $a_1 = 0$ and $a_2 = 0$.

## 3 Corresponding reduced equations

In this section, we need to find all reduced equations of the form (3) having the Painlevé property.

Now consider the dominant behavior in the neighborhood of a critical point $z = z_0$,

$$y = y_0\,(z - z_0)^q\,, \quad \text{as } z \to z_0,\quad q \in \mathbb{Z}, \tag{15}$$

where $q$ is the singularity order, and substitute equation (15) in equation (3). For certain values of the exponent $q$, two or more terms may balance and the remainder of the terms can be ignored as $z \to z_0$. The terms which balance for each choice of the exponent $q$ are called the dominant terms.

In the following subsections, the reduced equations that retain only leading terms as $z \to z_0$ will be considered for $q = -1, -2, -3$ with distinct Fuchs indices or $q$ is negative integer.

### *3.1 Leading order $q = -1$*

#### 3.1.1 Determination of the values of $b_1,\ c_0,\ d_2$ and $e_4$

Here, we have used the result obtained in [2]. If $b_1,\ c_0,\ d_2$, or $e_4$ is nonzero, then it may be possible to construct analogous solutions of (3) which feature a leading term containing a pole of order $q = -1$, then the Painlevé $\varepsilon-$ test,

$$z \to z_0 + \varepsilon z,\quad y \to \varepsilon^{-1} y,\quad \varepsilon \to 0,$$

reduces equation (3) to

$$y''' = a_1 \frac{y'' y'}{y} + a_2 \frac{y'^3}{y^2} + b_1 y y'' + c_0 \left(y'\right)^2 + d_2 y^2 y' + e_4 y^4. \tag{16}$$

The last equation is investigated in [8, 11] and [12], but the results obtained are not complete. Here we investigated also the situation in which $a_1 = 3$ and $a_2 = 0$ this case is not examined in [8, 11, 12]. For more details the reader is referred to [2].

*Case 1.* when $a_1 = 1$ *and* $a_2 = 0$

$$\begin{cases} I : yy''' = y'y''; \\ \quad y_{01} = arbitrary \ (r_{11}, r_{12}) = (0,\ 2), \end{cases}$$
$$\begin{cases} II : yy''' = y'y'' + y^3y'; \\ \quad y_{01} = -1, y_{02} = 1 \ (r_{11}, r_{12}) = (2, 4),\ (r_{21}, r_{22}) = (2, 4). \end{cases}$$
$$\begin{cases} III : yy''' = y'y'' + y^2y'' - \frac{2}{9}y^3y'; \\ \quad y_{01} = -3, y_{02} = -6 \ (r_{11}, r_{12}) = (2, -3),\ (r_{21}, r_{22}) = (-2, 2). \end{cases}$$
$$\begin{cases} IV : yy''' = y'y'' + y^2y'' + 2y^3y'; \\ \quad y_{01} = -1, y_{02} = 2 \ (r_{11}, r_{12}) = (2, 3),\ (r_{21}, r_{22}) = (2, 6). \end{cases}$$
$$\begin{cases} V : yy''' = y'y'' + y^2y'' + \frac{1}{4}y^3y' - \frac{1}{8}y^5; \\ \quad y_{01} = 4, y_{02} = -4, y_{03} = -2, (r_{11}, r_{12}) = (4, 6), \\ \quad (r_{21}, r_{22}) = (-2, 4), (r_{31}, r_{32}) = (1, 3), \end{cases}$$
$$\begin{cases} VI : yy''' = y'y'' + y^2y'' + 4y^3y' - 2y^5; \\ \quad y_{01} = -1, y_{02} = 1, y_{03} = -2, (r_{11}, r_{12}) = (1, 4), \\ \quad (r_{21}, r_{22}) = (3, 4), (r_{31}, r_{32}) = (-2, 6). \end{cases}$$

*Case 2.* when $a_1 = 2$ *and* $a_2 = 0$

$$\begin{cases} VII : yy''' = 2y'y''; \\ \quad y_{01} = arbitrary \ (r_{11}, r_{12}) = (0,\ 3), \end{cases}$$
$$\begin{cases} VIII : yy''' = 2y'y'' + y^3y'; \\ \quad y_{01} = -1, y_{02} = 1 \ (r_{11}, r_{12}) = (1, 4),\ (r_{21}, r_{22}) = (1, 4). \end{cases}$$
$$\begin{cases} IX : yy''' = 2y'y'' + 2y^2y'' - 2yy'^2; \\ \quad y_{01} = -2 \ (r_{11}, r_{12}) = (1,\ 2), \end{cases}$$
$$\begin{cases} X : yy''' = 2y'y'' + y^2y'' - yy'^2 + y^3y'; \\ \quad y_{01} = -1, y_{02} = 2 \ (r_{11}, r_{12}) = (1, 3),\ (r_{21}, r_{22}) = (1, 6). \end{cases}$$
$$\begin{cases} XI : yy''' = 2y'y'' + y^2y'' - \frac{3}{2}yy'^2 + 2y^3y' - \frac{1}{2}y^5; \\ \quad y_{01} = -4, y_{02} = 1, y_{03} = -1, (r_{11}, r_{12}) = (-5, 6), \\ \quad (r_{21}, r_{22}) = (1, 5), (r_{31}, r_{32}) = (1, 3), \end{cases}$$
$$\begin{cases} XII : yy''' = 2y'y'' + 2y^5; \\ \quad y_{01} = -1, y_{02} = \frac{-1+I\sqrt{3}}{2}, y_{03} = \frac{-1-I\sqrt{3}}{2}, (r_{11}, r_{12}) = (2, 3), \\ \quad (r_{21}, r_{22}) = (2, 3), (r_{31}, r_{32}) = (2, 3), \end{cases}$$
$$\begin{cases} XIII : yy''' = 2y'y'' + 2y^2y'' - 3yy'^2 + 2y^3y' - y^5; \\ \quad y_{01} = -1, y_{02} = 1, y_{03} = -2, (r_{11}, r_{12}) = (1, 2), \\ \quad (r_{21}, r_{22}) = (1, 6), (r_{31}, r_{32}) = (-2, 3). \end{cases}$$

*Case 3.* when $a_1 = 3$ *and* $a_2 = 0$

$$\begin{cases} XIV : yy''' = 3y'y''; \\ \quad y_{01} = arb.,\ (r_1, r_2) = (0, 4), \end{cases}$$

$$\begin{cases} XV : yy''' = 3y'y'' + (2-n)\, y^2y'' - 4yy'^2 + 2ny^3y'; \\ \qquad y_{01} = -1, (r_1, r_2) = (2, n)\,, \end{cases}$$

$$\begin{cases} XVI : yy''' = 3y'y'' + (4-n)\, y^2y'' - 2\,(4-n)\, yy; \\ \qquad y_{01} = arb.,\;\; (r_1, r_2) = (0, 4 + (4-n)\, y_{01}), \end{cases}$$

$$\begin{cases} XVII : yy''' = 3y'y'' + (1-n)\, y^2y'' - 2yy'^2 + ny^3y' + ny^5; \\ \qquad y_{01} = -1, y_{02} = 2,\;\; (r_{11}, r_{12}) = (3, n)\,,\;\; (r_{21}, r_{22}) = (6, -2n)\,, \end{cases}$$

$$\begin{cases} XVIII : yy''' = 3y'y'' + (3-n)\, y^2y'' - 6yy'^2 + 3ny^3y' - ny^5; \\ \qquad y_{01} = -1, y_{02} = -2,\;\; (r_{11}, r_{12}) = (1, n)\,,\;\; (r_{21}, r_{22}) = (-2, 2n)\,, \end{cases}$$

$$\begin{cases} XIX : yy''' = 3y'y'' - ny^2y'' + 2ny^5; \\ \qquad y_{01} = -1\,, y_{02} = 1,\;\; (r_{11}, r_{12}) = (4, n)\,,\;\; (r_{21}, r_{22}) = (4, -n). \end{cases}$$

where $n$ is an integer number .

*Case 4.* when $a_1 = 3$ *and* $a_2 = -2$

$$\begin{cases} XIX : \qquad y^2y''' = 3yy'y'' - 2y'^3; \\ XX : \qquad y^2y''' = 3yy'y'' - 2y'^3 + y^3y''. \end{cases}$$

### 3.1.2 Study of obtained equations

Now, we rewrite (16) as a system

$$\begin{cases} y' = vy^2, \\ v' = ty, \\ tt^*_v = H(v)t + R\,(v)\,, \left(* \text{ denotes } \frac{d}{dv}\right), \end{cases} \tag{17}$$

the last equation of the (17) is formally satisfied by the series

$$t = \sum_{j=0}^{+\infty} E_j v^j, \tag{18}$$

where $E_j$ are constants (*see* [8]).

From the first and second equation in (17) yields

$$y''y = 2y'^2 + ty^4. \tag{19}$$

According to [3], the first necessary condition for equation (17) satisfies the Painlevé property. That is the right-hand side of (19), with respect to $y_0$; is a polynomial of the second degree, takes the form

$$y''y = (E_2 + 2)\, y'^2 + E_1y'y^2 + E_0y^4, \tag{20}$$

where $E_2$, $E_1$, and $E_0$ are unknown coefficients that should be found. Substituting (18) into the third equation of the (17) and solving the system of the linear algebraic equations for coefficients $E_2$; $E_1$ and $E_0$, we conclude that all the equations (16) have the particular integrals with the Painlevé property in the form (20).

**Case** $(a)$ : Study of equations $(I - VI)$

− We can see that equations $(I, II, III, IV)$ can be presented in the form:

$$\begin{array}{ll} I: & y'' = ky; \\ II: & y'' = \frac{1}{2}y^3 + ky; \\ III: & y'' = yy' - \frac{1}{9}y^3 + ky; \\ IV: & y'' = yy' + y^3 + ky, \end{array}$$

after integration over $z$, where $k$ is an arbitrary complex constant. According to [3], all equations $(I, II, III)$ possess the Painlevé property, but $(IV)$ is an equation with critical mobile points.

− On one hand, introducing the change of variable

$$y = \alpha \frac{w'}{w}\ ,\ with\ \alpha = -1, \tag{21}$$

in $(V, VI)$ , we obtain

$$\begin{array}{ll} V: & w''' = kww', \\ VI: & w''' = kw^2w'. \end{array}$$

Therefore, equations $(V, VI)$ possess the Painlevé property (*see* [7]).

On the other hand, the first integrals of the equations $(V, VI)$ are, respectively, given by the following formulas:

$$V: \left(\frac{y''}{y^2} - y'\right)^2 = \frac{1}{4}\left(y' - \tfrac{1}{2}y^2\right)^2 + k,$$

$$VI: \left(\frac{y''}{y^2} - y'\right)^2 = 4\left(y' - \tfrac{1}{2}y^2\right)^2 + k.$$

The corresponding full equations $(V, VI)$ are, respectively, determined in section 4.

**Case** $(b)$ : Study of equations $(VII - XIII)$

− The first integrals of the equations $(VII, VIII)$ are, respectively, given by the following formulas:

$$\begin{array}{ll} VII: & y'' = ky^2,\ or\ \ yy'' = \dfrac{3}{2}y'^2 + k; \\ VIII: & yy'' = \frac{3}{2}y'^2 + \frac{1}{4}y^4 + k, \end{array}$$

these equations possess the Painlevé property [3].

– The stability conditions of the equations $(IX, X)$ are, respectively, given by the formulas:

$$\begin{array}{ll} IX: & Q_2^{(0)} \equiv -4y_1^{(0)2} = 0;\ y_1^{(0)} \text{ arbitrary.} \\ X: & Q_3^{(0)} \equiv -16y_1^{(0)3} = 0; \end{array}$$

in spite of this, the equations $(IX,\ X)$ admit the first integrals, which are, respectively, given by the following formulas:

$$\begin{array}{ll} IX: & y'' = yy' + ky^2; \\ X: & y'' = yy' + y^3 + ky^2. \end{array}$$

and so none of them have the Painlevé property [3].

– The stability condition at Fuchs indice $(r_{22} = 6)$ of the equation $(XI)$ is given by the formula:

$$XI: \qquad Q_6^{(6)} \equiv -\frac{2548813568922169729500\,7}{79330949271303634944000000}\left(y_{-5}^{(1)}\right)^6\left(y_6^{(0)}\right)^6 = 0,$$

where the coefficients $y_{-5}^{(1)}$, $y_{-1}^{(1)}$ and $y_6^{(0)}$ are arbitrary, these being obtained by using the perturbative Painlevé test (the CFP test) of Conte, Fordy, and Pikering to order $\varepsilon^6$ (see [6]) , which is a strong indication that equation $(XI)$ does not possess the Painlevé property, however, the corresponding *particular* integral is given by

$$y''y = \tfrac{3}{2}y'^2 + \tfrac{1}{2}y^4, \tag{22}$$

and this latter equation is easily integrated, which is solvable by means of elliptic functions.

– On one hand, the general solution of equation $(XII)$ possesses three families of simple poles, one can see that the necessary condition for the absence of movable critical singularities of the logarithmic type is satisfied.

On the other hand, be integrated by a process which may be considered as a method of *“ variation of parameters”* [3]. As this method will be used often later on, we shall explain its particulars on equation $(XII)$. The general solution of the equation

$$yy''' = 2y'y'',$$

is

$$y'' = wy^2,$$

where $w$ is an arbitrary constant. Suppose now that $w(z)$ is a function of $z$ and

$$y'' = 2wy^2,$$

we obtain

$$w' = y^2.$$

By direct calculation we obtain

$$w'' = 2y'y \ \ and \ \ y' = w^2 + k,$$

where $k$ is an arbitrary complex constant.

Indeed, the equation (*XII*) is equivalent to the differential system

$$\begin{cases} y' = w^2 + k, \\ w' = y^2. \end{cases}$$

By setting

$$y = -\beta u - v \ and \ \ w = \beta u - v, \ where \quad \beta^2 = -3,$$

one obtains

$$\begin{cases} u' = 2uv - k_1, where \ \ k_1 = \frac{k}{2\beta} \\ v' = -\beta^2 u^2 - v^2 - \beta k_1, \end{cases}$$

on eliminating $v$ between these relations, one finds

$$uu'' = \tfrac{1}{2}u'^2 - 2\beta^2 u^4 - 2\beta k_1 u^2 - \tfrac{k_1^2}{2}, \tag{23}$$

equation (23) is solvable by means of the elliptic functions, and equation (*XII*) has no moving critical points.

Furthermore, the differential equation (*XII*) has the first integral in the form:

$$y''^2 - 4y'y^4 + ky^4 = 0,$$

this latter possess the Painlevé property (see Cosgrove and Scoufis). Hence the reduced form of equation (*XII*) has the Painlevé property.

– The insertion of the well-known Riccati transformation (21) into (*XIII*) transforms the equation into the form

$$w''' = kw'^2 \ , \tag{24}$$

where $k$ is an arbitrary complex constant. Equation (*XIII*) has no moving critical points because the corresponding equation (24) has the Painlevé property.

**Case** ($c$) : Study of equation ($XIV - XIX$)

– The differential equation (*XIV*) has the first integral in the form:

$$XIV: \qquad y'' = ky^3, \ \ or \ \ yy'' = 2y'^2 + k,$$

where $k$ is the constant of integration. Hence equation (*XIV*) has the Painlevé property.

– It is necessary that $n = 1$, in order that the equations of type (20) will be with fixed critical points. In this case the stability condition at Fuchs indice ($r_{22} = 3$) of the equation ($XV_{n=1}$) is given by the formula:

$$XV_{n=1}: \qquad Q_3^{(0)} \equiv 128 y_1^{(0)3} = 0,$$

thus the equation exhibits logarithmic branching, which is an equation with critical mobile points.

For example, the stability conditions of the equations $\left(XV_{n\neq 1}\right)$ are given by the formulas:

$$\begin{aligned} n = +4: &\quad Q_{-8}^{(4)} \equiv y_{-8}^{(1)4} y_6^{(0)4} = 0, \\ n = -4: &\quad Q_{-1}^{(1)} \equiv 132 y_{-4}^{(1)} y_3^{(0)} = 0, \\ n = -6: &\quad Q_{-1}^{(2)} \equiv \tfrac{50}{7} y_{-1}^{(1)} y_{-6}^{(1)} y_3^{(0)2} = 0. \end{aligned}$$

thus the equation exhibits logarithmic branching, which is an equation with critical mobile points.

– It is necessary that $n = 2$, in order that the equations of type (20) will be with fixed critical points. In this case the equation ($XVI$) admits double indices, consequently does not have the Painlevé property. In particular for $n = 4$, the equation ($XVI_{n=4}$) has the first integral

$$XVI_{n=4}: \qquad yy'' = 2y'^2 - 2y^2 y' + 2y^4 + k. \tag{25}$$

For example, the stability conditions of the equations $\left(XVI_{n\neq 4}\right)$ are given by the formulas, for example :

$$\begin{aligned} n = +1: &\quad Q_2^{(0)} \equiv -8 y_1^{(0)2} = 0, \\ n = -2: &\quad Q_{-2}^{(2)} \equiv -48 y_{-2}^{(1)2} y_2^{(0)} = 0, \\ n = -3: &\quad Q_{-1}^{(1)} \equiv 42 y_{-3}^{(1)} y_2^{(0)} = 0. \end{aligned}$$

thus the equation exhibits logarithmic branching, which is an equation with critical mobile points.

– When $n = 3$, in order that the equation of type (20) will be with fixed critical points. In this case, the stability condition at Fuchs indice ($r_{22} = 6$) of the equation ($XVII_{n=3}$) is given by the following formula:

$$XVII_{n=3}: \qquad Q_6^{(3)} \equiv \tfrac{1248}{5} y_{-2}^{(1)3} y_6^{(0)2} = 0,$$

consequently, ($XVII$) is an equation with critical mobile points.

When $n = -1$, the equation ($XVII_{n=-1}$) admits double indices and so does not have the Painlevé property.

The stability conditions of the equations ($XVII_{n=-2,-3,-4,-5,-6,-9}$) are given by the formulas:

$$XVII_{n=-2,-3,-4,-5,-6,-9}: \begin{cases} Q_{-1}^{(1)} \equiv \textit{terms proportional to } y_1^{(0)(-1-n)} y_n^{(1)} = 0; \\ Q_1^{(1)} \equiv \textit{terms proportional to } y_1^{(0)(1-n)} y_n^{(1)} = 0; \\ Q_n^{(2)} \equiv \textit{terms proportional to } y_1^{(0)(-n)} y_n^{(1)2} = 0. \end{cases}$$

On the other hand, by making the change of variable (21), we obtain the third-order equation,

$$w''' = kw'^3 w^{n-2}. \tag{26}$$

If we take $n = 2$ in equation (26), we obtain the second-order equation in the form $\delta'' = k\delta^3$. Here $\delta = w'$ is a Jacobian elliptic function with simple poles. Hence a series expansion of $w$ around such a singularity $z_0$, say, must start with $\log(z - z_0)$. The remainder of the series is a power series expansion in powers of $(z-z_0)$. In such cases, the Painlevé property holds for the new variables $\delta$. However, when $n = 2$, the stability condition of equation ($XVII_{n=2}$) is given by the following formula:

$$XVII_{n=2}: \qquad Q_4^{(3)} \equiv 756 y_4^{(0)2} y_{-2}^{(1)} y_{-1}^{(1)2} - 180 y_4^{(0)2} y_{-2}^{(1)2} = 0,$$

implying the existence of a movable logarithmic branch point.

– When $n$ is in $\{-1,\ -4,\ 4\}$, the equation ($XVIII$) admits the double indices and so it is with moving critical points.

The stability conditions of the equation ($XVIII_{n=\pm1,\pm2,\pm3,\pm5,\pm6,\pm7,\ldots}$) in which we fixed the parameter $n \neq 0$ are given by the formulas:

$$\begin{aligned}
XVIII_{n=\pm1}: &\qquad Q_4^{(0)} \equiv \pm 864 y_1^{(0)4} = 0; \\
XVIII_{n=\pm2}: &\qquad Q_4^{(0)} \equiv \pm 36 y_2^{(0)2} = 0; \\
XVIII_{n=\pm3}: &\qquad Q_{-1}^{(3)} \equiv -\tfrac{66675}{26} y_{-3}^{(1)} y_4^{(0)2} = 0; \\
XVIII_{n=\pm5}: &\qquad Q_{-1}^{(1)} \equiv 300 y_4^{(0)} y_{-5}^{(1)} = 0; \\
XVIII_{n=\pm6}: &\qquad Q_4^{(6)} \equiv \mathit{Coeff}.y_4^{(0)10} y_{-6}^{(1)6} = 0; \\
XVIII_{n=\pm7}: &\qquad Q_{-1}^{(3)} \equiv \pm\tfrac{3509611875}{8704} y_4^{(0)5} y_{-7}^{(1)3} = 0, \\
XVIII_{n=\pm8}: &\qquad Q_4^{(4)} \equiv \mathit{Coeff}.y_4^{(0)9} y_{-8}^{(1)4} = 0,
\end{aligned}$$

where the coefficients $y_n^{(1)}$, $y_{-1}^{(1)}$, and $y_4^{(0)}$ are arbitrary, implying the existence of a movable logarithmic branch point, we think that the equation of type ($XVIII$) does not have the Painlevé property (*see* [6]), however, the corresponding *particular* integral is such that

$$y'' = 2y^3.$$

**Case** ($d$) : Study of equation ($XIX,\ XX$)

The first integrals of the equations ($XIX,\ XX$) are, respectively, given by the following formulas:

$$\begin{aligned} XIX: &\quad yy'' = y'^2 + ky^2; \\ XX: &\quad yy'' = y'^2 + y^2y' + ky^2, \end{aligned}$$

these equations possess the Painlevé property (*see Ref.*[3]) .

## 3.2 Leading order $q = -2$

If $d_1 \neq 0$, then it may be possible to construct analogous solutions of (16) which feature a leading term containing a pole of order $q = -2$, then the Painlevé $\varepsilon-$ test,

$$z = z_0 \to \varepsilon z, \quad y \to \varepsilon^{-2} y, \quad \varepsilon \to 0,$$

will produce a nontrivial reduced equation, which admits the particular integral

$$y = y_0 \left(z - z_0\right)^q .$$

The dominant terms arise from $y'''y^2$, $y''y'y,\ y'^3$ and $y^3y'$. Then $q = -2$ and the corresponding equation is

$$y'''y^2 = a_1 y''y'y + a_2 y'^3 + d_1 y^3 y'. \tag{27}$$

If we substitute

$$y \cong y_0 \left(z - z_0\right)^{(-2)} + \ldots + \hat{\boldsymbol{\beta}} \left(z - z_0\right)^{(r-2)}$$

into (27), we obtain the following equations for the Fuchs indices $r$ and $y_0$

$$(r+1)\left(r^2 + (2a_1 - 10)\, r - 12a_1 - 8a_2 + 24\right)(3a_1 + 2a_2 - 6)^2 = 0, \tag{28}$$

and

$$d_1 y_0 + 6a_1 + 4a_2 - 12 = 0, \tag{29}$$

equation (29) implies that there is only one branch.

Now, we determine $y_0$ and $d_1$, if $y_0 \neq 0$, then equations (28) and (29) imply that the values of Fuchs indices are:

For $a_1 = 0,\ 1,\ 3$, and $a_2 = 0$

$$y_0 = -\tfrac{6(a_1-2)}{d_1},\ r_{1,2} = (6,\ 4 - 2a_1)\,. \tag{30}$$

If $a_1 = 1$ and $a_2 = 0$, the Fuchs indices are as follows: $r_1 = 6$ and $r_2 = 2$, equation (27) has the first integral in the form:

$$y'' = y^2 + ky \text{ or } y'' = \frac{y'^2}{y} + \tfrac{1}{3}y^2 + \frac{k}{y}.$$

If $a_1 = 2$ and $a_2 = 0$, the Fuchs indices are as follows: $r_1 = 6$ and $r_2 = 0$, but $y_0$ cannot equal 0. Equation (27) has the first integral in the form:

$$y'' = \tfrac{3}{2}\frac{y'^2}{y} + \tfrac{1}{3}y^3 + k.$$

If $a_1 = 3$ and $a_2 = 0$, the Fuchs indices are as follows: $r_1 = 6$ and $r_2 = -2$, equation (27) has the first integral in the form :

$$y'' = 2\frac{y'^2}{y} + \frac{1}{3}y^2 + \frac{k}{y},$$

equation (27) does not pass the Painlevé test, since the compatibility condition at $(r_2 = -2)$ is not satisfied identically, i.e.,

$$Q_{-2}^{(4)} = -\tfrac{104}{9}\left(y_{-2}^{(1)}\right)^4\left(y_6^{(0)}\right) = 0 \text{ and } Q_{-1}^{(4)} = -\tfrac{322}{9}\left(y_{-2}^{(1)}\right)^3\left(y_{-1}^{(1)}\right)\left(y_6^{(0)}\right) = 0,$$

where $y_6^{(0)}$, $y_{-1}^{(1)}$, and $y_{-2}^{(1)}$ are arbitrary constants.

For $a_1 = 3$ and $a_2 = -2$, equation (27) admits the double indices $(r_{1,2} = (2,\ 2))$ and so does not have the Painlevé property.

## 3.3 Leading order $q = -3$

In particular if $e_2 \neq 0$ and $q = -3$, the Painlevé $\varepsilon-$ test,

$$z \to z_0 + \varepsilon z, \quad y \to \varepsilon^{-3}y, \ \ \varepsilon \to 0,$$

gives, in the limit $\varepsilon \longrightarrow 0$, the reduced equation

$$y'''y^2 = a_1y''y'y + a_2y'^3 + e_2y^4. \tag{31}$$

If we substitute

$$y \cong y_0\,(z - z_0)^{(-3)} + \ldots + \hat{\boldsymbol{\beta}}\,(z - z_0)^{(r-3)}\,,$$

into (31), we obtain the following equations for the Fuchs indices $r$ and $y_0$

$$(r + 1)\left(r^2 + (3a_1 - 13)\,r - 36a_1 - 27a_2 + 60\right) = 0, \tag{32}$$

and

$$e_2 y_0 - 36a_1 - 27a_2 + 60 = 0, \tag{33}$$

equation (33) implies that there is only one branch.

Now, we determine $y_0$ and $e_2$, if $y_0 \neq 0$, then the only values of $a_1$ that satisfy (32) and (33) are $a_1 = 1$, then the corresponding equation has the form

$$\begin{cases} y'''y = y''y' + e_2 y^3, \\ \qquad y_0 = \frac{-24}{e_2}, \text{ resonances: } (r_1, r_2) = (4, 6) \,. \end{cases} \tag{34}$$

By setting $y = w'$ and $e_2 = 1$ from (34), we obtain the following equation

$$w'\left(w'' = \tfrac{1}{2}w^2 + k_1 w + k_2\right), \tag{35}$$

The differential equation (35) has the first integral in the form:

$$w'^2 = \frac{1}{3}w^3 + k_1 w^2 + 2k_2 w + 2k_3, \tag{36}$$

where $k_1, k_2$, and $k_3$ are arbitrary constants, equation (36) is solvable by means of the elliptic functions and equation (34) has no moving critical points.

Furthermore, the differential equation (34) has the first integral in the form:

$$y''^2 + 4y'y^2 + ky^2 = 0,\ e_2 = 2,$$

this latter possess the Painlevé property (see Cosgrove and Scoufis).

## 3.4 *Leading order q : negative integer (q ≤ −4)*

If $b_1 = c_0 = d_1 = d_2 = e_2 = e_4$, then the reduced equation can be written

$$y^2 y''' = a_1 y y' y'' + a_2 y'^3 + b_{-1} y^2 y'' + c_{-2} y y'^2 + d_{-1} y^2 y' + e_0 y^3. \tag{37}$$

Putting $y' = vy$ in (37), we obtain for $v$ the differential equation

$$v'' = (a_1 - 3)\, v v' + (a_1 + a_2 - 1)\, v^3 + (b_{-1} + c_{-2})\, v^2 + b_{-1} v' + d_{-1} v + e_0. \tag{38}$$

The equation (37) occupies a position in these studies analogous to the occupied by (38) in the study of the second order case (see [3]).

*Example 1.* We consider the equation

$$\begin{cases} y^2 y''' = 3yy'y'' + d_0 y^2 y', \ w = \dfrac{y'}{y}, \\ w'' = 2w^3 + d_0(z)\, w, \\ d_0(z) = C_1 z + C_2, \begin{cases} \text{if } \ C_1 = 0 : w(z) \ \text{ elliptic function} \\ \text{if } \ C_1 \neq 0 : y = Ce^{\int w(z)dz}. \end{cases} \end{cases}$$

*Example 2.* We consider the equation

$$yy''' = 3y'y'' + 6\frac{y'^2}{z}.$$

On setting $w = \frac{y'}{y}$, one obtains

$$w'' = 2w^3 + 6\frac{w^2}{z}.$$

Thus, $y$ is the Painlevé type.

$$\begin{cases} yy''' = 3y'y'' + 3\frac{y'^2}{z}, \ w'' = 2w^3 + 3\frac{w^2}{z}, \\ yy''' = 3y'y'' + \frac{1}{z^3}y^2 + \frac{1}{z^2}yy' + \frac{1}{z}y'^2. \end{cases}$$

*Example 3.* We consider the equation

$$y^2 y''' = 3yy''y' - 2y'^3 + b_0\left(y^2 y'' - yy'^2\right) + d_0 y^2 y' + e_1 y^3. \tag{39}$$

We rewrite (39) as

$$\left(\frac{y'}{y}\right)'' = b_0\left(\frac{y'}{y}\right)' + d_0\left(\frac{y'}{y}\right) + e_1. \tag{40}$$

On setting

$$y = e^v,$$

from (40), we obtain

$$v''' = b_0 v'' + d_0 v' + e_1,$$

this latter is linear equation, therefore, $v$ and $y$ are Painlevé type.

# 4 Corresponding full equations

In this section, we complete the determination of the remaining coefficients in the full equations (3). Our approach has the following steps:

1. The first step is to use gauge freedom to simplify the problem. The form of the equation (3) is preserved under changes of variables of the form

$$y(z) = \lambda(z) V(t) + \mu(z), \ t = \varphi(z). \tag{41}$$

   We first simplify the complete equation by means of the substitution (41), here $V$ is a new function, $t$ is the independent variable, and $\lambda$, $\mu$, and $\varphi$ are chosen so that the resulting equations for $V$ become simpler.
2. The second step, we are looking for such values of constants $a_1$, $a_2$ and the expressions $b_i$, $c_i$, $d_i$ and $e_i$ so that equation (3) passes the Conte-Fordy-Pickering test.
3. The third step is to use the substitution $wy = 1$, we reduce equation (3) with coefficients (4) to the form

$$\begin{aligned} w''' = {} & (6-a_1)\frac{w''w'}{w} + (2a_1+a_2-6)\tfrac{w'^3}{w^2} + \sum_{j=-1}^{1} b_i w^{-i} w'' \\ & - \left[2\sum_{j=-1}^{1} b_i w^{-1-i} + \sum_{j=-2}^{0} c_i w^{-2-i}\right](w')^2 - \sum_{j=-2}^{2} d_i w^{-i} w' + \sum_{j=-2}^{4} e_i w^{2-i}. \end{aligned} \tag{42}$$

## 4.1 Leading order $q = -1$

In this subsection, we consider the problem of constructing all equations of the form (16), for example:

$$yy''' = y'y'' + y^2y'' + 4y^3y' - 2y^5 + B(z,y)\,y'' + C(z,y)\left(y'\right)^2 + D(z,y)\,y' + E(z,y)\,. \tag{43}$$

There are three sets of resonance numbers:

$$y_{01} = -1, \ y_{02} = 1, \ y_{03} = -2, \ (r_{11}, r_{12}) = (1, \ 4), \\ (r_{21}, \ r_{22}) = (2, \ 4), \ (r_{31}, \ r_{32}) = (-2, \ 6).$$

We show details of the calculation only for (43), since the other cases are similar. We are going to apply the Painlevé test to equation (43) in this section using the perturbative Painlevé approach presented in [6].

For the case $y_{01} = -1$, the coefficient $y_2$ in the Laurent series

$$y = y_0(z-z_0)^q + y_1(z-z_0)^{q+1} + y_2(z-z_0)^{q+2} + \ldots + y_i(z-z_0)^{q+i} + \ldots \tag{44}$$

can be set to zero by the gauge choice,

$$c_{-1} = 3e_3 + d_1,\ e_2 = -c_{-2} - 2b_{-1} - d_0.$$

Hence, the resonances $(r_{11}, r_{12}) = (1,\ 4)$ give the constraints

$$b_0 = -d_1 - 2e_3,\ d_1 = -3e_3,\ e_1 = \frac{1}{2}e_3c_{-2} - 2d_{-1},$$

$$d_0 = -3b_{-1} - \frac{5}{2}c_{-2},\ e_0 = -\frac{1}{2}c_{-2}e_3^2 + d_{-1}e_3 - d_{-2}.$$

For the case $y_{02} = 1$, the coefficient $y_1$ can be set to zero by the gauge choice,

$$e_3 = 0.$$

Hence, the resonances $(r_{21}, r_{22}) = (2,\ 4)$ give the constraints

$$d_{-1} = 0,\ d_{-2} = -\frac{45}{8}c_{-2}^2 - 13b_{-1}c_{-2} - \frac{15}{2}b_{-1}^2.$$

For the case $y_{03} = -2$, the coefficient $y_0$ can be set to zero by the gauge choice,

$$c_{-2} = 0.$$

Hence, the resonances $(r_{31}, r_{32}) = (-2, 6)$ give the constraints

$$e_{-1} = 0,\ e_{-2} = 0.$$

Because, the compatibility conditions are given by the formulas:

$$Q_6^{(2)} \equiv -\frac{1}{8}\left(e_{-1}^2\right)\left(y_{-2}^{(1)}\right)^2 = 0 \text{ and } Q_6^{(5)} \equiv \frac{15}{32}\left(e_{-2}\right)\left(y_{-2}^{(1)}\right)\left(y_{-1}^{(1)}\right)^4\left(y_6^{(0)}\right) = 0.$$

Finally, the remaining conditions at the resonances $-2$ and 6 force all recessive terms to vanish. We therefore conclude that the full equation (43) is

$$yy''' = y'y'' + y^2y'' + 4y^3y' - 2y^5. \tag{45}$$

The second example is

$$yy''' = y'y'' + y^2y'' + \frac{1}{4}y^3y' - \frac{1}{8}y^5 + B(z,y)\,y'' + C(z,y)\left(y'\right)^2 + D(z,y)\,y' + E(z,y)\,. \tag{46}$$

There are three sets of resonance numbers:

$$y_{01} = -2, y_{02} = 4, y_{03} = -4, (r_{11}, r_{12}) = (1, 3)\,,$$
$$(r_{21}, r_{22}) = (4, 6)\,, (r_{31}, r_{32}) = (-2, 4)\,,$$

For the case $y_{01} = -2$, the coefficient $y_2$ in the Laurent series (44) can be set to zero by the gauge choice,

$$c_{-1} = 12e_3 + 2d_1,\ c_{-2} = -4e_2 - 2b_{-1} - 2d_0.$$

Hence the resonances $(r_{11},\ r_{12}) = (1,\ 3)$ give the constraints

$$b_0 = -8e_3 - 2d_1,\ d_1 = -6e_3,\ b_{-1} = -8e_3 - 2d_0,\ d_{-1} = -2e_1.$$

For the case $y_{02} = 4$, the coefficient $y_1$ can be set to zero by the gauge choice,

$$e_3 = 0.$$

Hence the resonances $(r_{21}, r_{22}) = (4, 6)$ give the constraints

$$d_{-2} = 12d_0e_2 - 12e_2^2 - 3d_0^2 + 4e_0,\ e_2 = 0,\ e_{-2} = -\tfrac{27}{8}d_0^3 + 32e_1^2 + 3d_0e_0.$$

For the case $y_{03} = -4$, the coefficient $y_0$ can be set to zero by the gauge choice,

$$d_0 = 0.$$

Hence the resonances $(r_{31}, r_{32}) = (-2, 6)$ give the constraints

$$e_1 = 0,\ e_0 = 0\ \ or\ \ e_0 = \tfrac{-435}{26}\left(y_4^{(0)}\right) \equiv C,\ e_{-1} = 0.$$

Because, the compatibility conditions are given by the formulas:

$$Q_4^{(1)} \equiv 128e_1\left(y_{-2}^{(1)}\right) = 0,$$

$$Q_4^{(2)} \equiv e_0\left(-290\left(y_4^{(0)}\right) - \tfrac{52}{3}e_0\right)\left(y_{-2}^{(1)}\right)^2 = 0,$$

$$Q_4^{(3)} \equiv \tfrac{1}{128}e_{-1}^2\left(y_{-2}^{(1)}\right)^3 = 0.$$

The result is that compatibility conditions at the resonance numbers are satisfied identically. There are no tests to run.

The final form of full equation (46) is

$$yy''' = y'y'' + y^2y'' + \tfrac{1}{4}y^3y' - \tfrac{1}{8}y^5 + C\left(4\tfrac{y'}{y} + y\right). \tag{47}$$

Using the above-mentioned approach we obtained a list of equations, which can be with fixed critical points, these are the following ones:

$$\begin{cases} yy''' = y'y'' + y^3y' + C_1y^2 + C_2y'; \\ yy''' = y'y'' + y^2y'' + 2y^3y' + Cy' + \frac{C}{2}y^2; \\ yy''' = y'y'' + y^2y'' + 2y^3y' + Cyy' + \left(\frac{60}{z^3} + \frac{C}{2}\right)y^2; \\ yy''' = y'y'' + y^2y'' + 4y^3y' - 2y^5 + 4\left(C_1z + C_2\right)y^3 - 2C_1y^2 + 2C_1y'; \\ yy''' = y'y'' + y^2y'' + \frac{1}{4}y^3y' - \frac{1}{8}y^5 - 2\left(C_1z + C_2\right)y^3 - 2C_1y^2 - 4C_1y'; \\ yy''' = y'y'' + y^2y'' + \frac{1}{4}y^3y' - \frac{1}{8}y^5 - 2\left(C_1z + C_2\right)y^3 + C_1y^2 - C_1y'; \end{cases} \tag{48}$$

$$\begin{cases} yy''' = 2y'y'' - 3yy' + b_0\left(y'' - y\right), \\ yy''' = 2y'y'' + y^3y' + Cy'; \\ yy''' = 2y'y'' + y^3y' + Cyy'; \\ yy''' = 2y'y'' + 2y^2y'' - 3yy'^2 + 2y^3y' - y^5 + Cy; \end{cases} \tag{49}$$

$$yy''' = 3y'y'' - \left(2C_1z - C_2\right)yy' + C_3y' + C_2y^2; \tag{50}$$

$$\begin{aligned} y^2y''' = {} & 3yy'y'' - 2y'^3 + y^3y'' + b_0\left(y^2y'' - 2yy'^2 - y^5\right) \\ & + b_{-1}\left(yy'' - 2y'^2\right) + d_0\left(y^2y' + y^4\right) + \left(2\left(b_{-1}\right)' - 3b_0b_{-1}\right)yy' \\ & + e_1y^3 - \left(\left(b_{-1}\right)'' - b_0\left(b_{-1}\right)' - d_0b_{-1}\right)y^2 - b_0\left(b_{-1}\right)^2y. \end{aligned} \tag{51}$$

## 4.2 Leading order $q = -2$

Equation (27) with (4) can be analyzed in a similar fashion to the preceding subsection, for example:

$$yy''' = y'y'' + y^2y'' + 2y^3y' + B(z,y)\,y'' + C(z,y)\left(y'\right)^2 + D(z,y)\,y' + E(z,y). \tag{52}$$

There are two sets of resonance numbers

$$y_{01} = -1, y_{02} = 2, \left(r_{11}, r_{12}\right) = (2,3), \ \left(r_{21}, r_{22}\right) = (2,6).$$

A choice of gauge that sets $y_1 = 0$ in both corresponding Laurent series is

$$d_1 = -e_3 - 2b_0 - c_{-1}, \ b_0 = -\tfrac{3}{2}c_{-1}.$$

The resonance $r_{21} = 2$ gives the constraints

$$d_0 = -2b_{-1} - e_2 - c_{-2}, \ c_{-2} = -2b_{-1} - 2e_2.$$

The compatibility conditions at the second resonance $r_{12} = 3$ is linear in $y_2$, we obtain

$$e_1 = -d_{-1}, \ e_3 = c_{-1}.$$

and the condition at the second resonance $r_{22} = 6$ is second degree in $y_2$ and therefore gives three separate constraints, we obtain

$$\begin{aligned} b_{-1} &= -2e_2 - \tfrac{1}{4}c_{-1}^2, \\ e_0 &= -\tfrac{1}{16}c_{-1}^4 - \tfrac{5}{8}e_2c_{-1}^2 - \tfrac{1}{2}d_{-1}c_{-1} + d_{-2} + \tfrac{1}{2}e_2^2, \\ e_{-2} &= -\tfrac{1}{2}e_2^3 + \tfrac{3}{8}c_{-1}^2e_2^2 - \tfrac{1}{2}\left(d_{-2} - \tfrac{3}{4}c_{-1}^4 + 2c_{-1}d_{-1}\right)e_2 + \tfrac{1}{32}c_{-1}^6 \end{aligned}$$

$$+\tfrac{1}{4}d_{-1}c_{-1}^3-\tfrac{1}{4}d_{-2}c_{-1}^2-c_{-1}e_{-1}-4d_{-1}^2.$$

A similar calculation and choice of gauge can be used to construct the full equations (52). The complete list of the full equations is found to be:

$$\begin{cases} y'''y=y''y'+6y^2y'+C_1y', \\ y'''y=y''y'+6y^2y'+(C_1z+C_2)\,y', \\ y'''y=y''y'+6y^2y'+C_1y+\left(C_1z^2+C_2z+C_3\right)y'+(C_1z+C_2)\,y. \end{cases} \tag{53}$$

### 4.3 Leading order $q = -3$

In this subsection, we consider the problem of constructing all equations of the form (31), for example:

$$y'''y=y''y'+e_2y^3+F_3\,(x,\ y)\,, \tag{54}$$

where

$$F_3\,(x,\ y)=b_0y^2y''+b_{-1}yy''+c_{-1}yy'^2+c_{-2}y'^2+d_0y^2y'$$
$$+d_{-1}yy'+d_{-2}y'+e_1y^3+e_0y^2+e_{-1}y+e_{-2}.$$

There is only one set resonance numbers:

$$y_{01}=\tfrac{-24}{e_2},\ (r_{11},r_{12})=(4,\ 6)\,,$$

the coefficient $y_1$ in the Laurent series (44) can be set to zero by the gauge choice, by means of a transformation $T\,(\lambda,\ \mu,\ \varphi)$, one may set

$$12b_0+9c_{-1}=0.$$

The compatibility condition at the second resonance $r_{12}=6$ is linear in $y_4^{(0)}$, we obtain

$$Q_6^{(0)}\equiv Ay_4^{(0)}+B=0,\ A=B=0.$$

and hence

$$d_0=\tfrac{15}{4}c_{-1}^2,\ e_0=\tfrac{880875}{1024}c_{-1}^6+\tfrac{4455}{16}e_1c_{-1}^3+\tfrac{45}{16}b_{-1}c_{-1}^2+\tfrac{45}{4}d_{-1}c_{-1}-8e_1^2.$$

The compatibility condition at the second resonance $r_{11}=4$ is

$$c_{-2} = -2c_{-1}b_0 - \tfrac{9}{8}c_{-1}^2 d_0 - \tfrac{3}{2}d_0^2 - \tfrac{4}{3}b_{-1}.$$

By means of a transformation $T(\lambda,\ \mu,\ \varphi)$, one may set

$$b_0 = c_{-1} = 0,\ b_1 = c_0 = d_{-2} = e_{-2} = e_3 = 0.$$

On setting $wy = 1$, one obtains

$$w^2 w''' = 5ww'w'' - 4w'^3 + d_{-1}w^3 w' + F_3(x,\ w);$$
$$d_{-1}y_0 = -2,\ (r_1,\ r_2) = (-2,\ 2).$$

By means of a transformation $T(\lambda,\ \mu,\ \varphi)$, one may set

$$d_{-1} = -2,\ b_0 = c_{-1} = d_0 = 0,$$

Hence the resonances $(r_{31}, r_{32}) = (-2, 2)$ give the constraints

$$e_2 = e_0 = 0,\ e_1 = 0,\ d_1 = k \text{ (arbitrary)}.$$

The full equation (54) takes the following form:

$$yy''' = y'y'' - 24y^3 + ky'. \tag{55}$$

The complete list of full equations is found to be:

$$\begin{cases} y'''y = y''y' + y^3 + Cy'y, \\ y'''y = y''y' + y^3 + C_1y^2 + (C_2z + C_3)\,y' - C_2y, \end{cases} \tag{56}$$

*Example 4.* We consider the equation

$$yy''' = y'y'' + 6y^3 - y'. \tag{57}$$

Multiplying (57) by $(y'' - 1)$ we can rewrite it as

$$\frac{d}{dz}\left(\frac{(y''-1)^2 - 12y^2(y'-z)}{y^2}\right) = 0,$$

and absorbing the integration constant by a translation of $z$, we obtain

$$\left(y'' - 1\right)^2 - 12y^2\left(y' - z\right) = 0.$$

This latter is particular case of the $SD_v$ equation in Cosgrove's classification [1993]. It is solved, expectedly, in terms of Painlvé $I$ equations $P_I$.

## 5 Conclusion and prospects

In this paper we have used the perturbative Painlevé approach presented in [6] for equations (3) to find several new rational third order ODEs with integer Fuchs indices. Our results improve and extend various known results existing in the literature.

To conclude this article, we point out some problems that are still unresolved: It remains to examine whether the general integral equations found in section 4 are effectively a Painleve property and are general solutions of these equations.

## References

1. M. J. Ablowitz, A. Ramani and H. Segur. *"A Connection between nonlinear evolution equations and ordinary differential equations of P-type II"*, J. Math. Phys., **21**(1980), 715–721, 1006–1015.
2. Y. Adjabi and A. Kessi. *" Third order differential equations with fixed critical points",* Annals: The proceeding of the Institute of Mathematics of the National Academy of Sciences of Belarus (NASB) - Google Scholar, **12** (2004), Nr. 2, *UDC* 517.925, 12–17.
3. F. J. Bureau. *"Equations differentielles du second ordre en Y et du second degré en $\ddot{Y}$ dont l'intégrale générale est à points critiques fixes"*, Annali di Mat., **91**(1972), 163–281.
4. J. Chazy. *"Sur les équations différentielles du troisième et d'ordre supérieur dont l'intégrale générale à ses points critiques fixés"*, Acta Math., **34**(1911).
5. R. Conte, C. Rogers and W. K. Schief. *"Painlevé structure of a multi-ion electrodiffusion system"*, Journal of Phys. A: Math. and Theoretical. vol. **40** (2007), Nr. 48, *F* 1031–*F* 1040.
6. R. Conte, A. Fordy and A. Pickering. *" A perturbative Painlevé approach to nonlinear differential equations"*, Physica D, **69**(1993)**,** 33–58.
7. C.M.Cosgrove. *"Chazy classes $IX - XII$ of third-order differential equations"*. Stud. Appl. Math. **104**(2000),Nr. 3, 171–228.
8. H. Exton. *"On nonlinear ordinary differential equations with fixed critical points"*, Rendiconti di Matematica, **4**(1971), 385–448.
9. E. L. Ince. *"Ordinary differential equations"*. Dover Pub., New York, 1956.
10. A. V. Kozulin and N. A. Lukashevitch. *"Third-Order Differential Equation of a Special Form"*, V. I.Lenin Belorussian State University, Translated from Differentsial'nye Uravneniya, vol. **24** (1988)., Nr. 12, 2064–2069.
11. I. P. Martynov. *"Analytic properties of solutions of a third order equations "*, Translated from Differntsial'nye Uravneniya, vol. **21**(1985), Nr. 5, 512–517, Nr. 6, 937–946.
12. U. Mugan, F. Jrad. *"Non-polynomial third order equations which pass the Painlevé test",* Z. Naturforsch. A, **59**(2004), 163–180.

# Existence of Weighted Pseudo Almost Periodic Solutions for some Partial Differential Equations with Delay

**Nadira Boukli-Hacene and Khalil Ezzinbi**

**Abstract** In this work, sufficient conditions are derived to get the existence and uniqueness of a weighted pseudo almost periodic solutions for some partial functional differential equations in hyperbolic case. To illustrate our main result, we study the existence of a weighted pseudo almost periodic solution for some diffusion equation with delay.

**Keywords** Partial functional differential equation • Hyperbolic semigroup • Weighted pseudo almost periodic solution

## 1 Introduction

The notion of pseudo-almost periodicity was introduced in the literature in the early 90s by C. Zhang [18–20], as a natural generalization of the classical almost periodicity in the sense of Bohr. Since then, the existence of pseudo-almost periodic solutions to differential equations, partial differential equations, and functional differential equations has been of a great interest to several authors and hence generated various contributions: E. Aitdads, O. Arino, and K. Ezzinbi [3–5] obtained sufficient condition for existence of pseudo almost periodic solutions of some delay differential equations, and other contributions upon pseudo almost periodic solutions to various differential equations have recently been made in T. Diagana, E.M. Hernàndez, G.M. Mahop, G.M. N'Guérékata [8, 9, 11–13].

T. Diagana [7, 10] introduced a new generalization of the concept of pseudo almost periodicity, some new classes of functions called weighted pseudo almost periodic functions which is the central issue in this paper, to construct those new spaces, the main idea consists of enlarging the so-called ergodic component,

N. Boukli-Hacene (✉)
Département de Mathématiques, Université de Tlemcen, B.P.119, Tlemcen, Algérie
e-mail: nadboukli@yahoo.fr

K. Ezzinbi
Faculté des Sciences Semlalia, Département de Mathématiques, Université Cadi Ayyad, B.P. 2390, Marrakech, Morocco

 
A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_10

utilized in Zhang's definition of pseudo-almost periodicity, with the help of a weighted measure $d\mu(x) = \rho(x)\ dx$, where $\rho : \mathbb{R} \to (0, \infty)$ is a locally integrable function over $\mathbb{R}$, which is commonly called weight, the author obtained conditions for existence of the weighted pseudo almost periodic solutions for abstract differential equations; however, L. Zhang and Y. Xu [21] give sufficient condition for the existence of a weighted pseudo almost periodic solutions for functional differential equations.

The existence, uniqueness, and stability of almost periodic, pseudo-almost periodic, weighted pseudo-almost periodic, almost automorphic and pseudo-almost automorphic solutions are among the most attractive topics in the qualitative theory of differential equations due to their applications in several areas such as mathematical biology, physics, control theory, and others.

To study those weighted pseudo-almost periodic spaces we consider a binary equivalence relation, $\sim$, on the set of weights $\mathbb{U}_\infty$, which enables us to categorize and reorganize those weights into different equivalence classes. Among other things, if two weights $\rho_1$ and $\rho_2$ are equivalent, that is, $\rho_1 \sim \rho_2$, then their corresponding weighted pseudo-almost periodic spaces coincide. In particular, when a weight $\rho$ is bounded with $\inf\limits_{x\in\mathbb{R}} \rho(x) > 0$, it is then equivalent to the constant function 1, and hence the weighted pseudo-almost periodic space with weight $\rho$ coincides with ZhangŠs spaces (Corollary 1).

This work extends the corresponding results in [21], our objective is to show the existence of a weighted pseudo almost periodic solutions for the following delay functional differential equation

$$\frac{d}{dt}x(t) = Ax(t) + L(x_t) + f(t) \qquad \text{for } t \in \mathbb{R}, \tag{1}$$

where $A : D(A) \to \mathbb{E}$ is a linear operator (not necessarily densely defined) on a Banach space $\mathbb{E}$, we assume that A satisfies the Hille-Yoshida condition, which means that A satisfies the following spectral condition: there exists $M \geq 1$ and $\omega \in \mathbb{R}$ such that

$$(\omega, +\infty) \subset \rho(A) \text{ and } |(\lambda I - A)^{-n}| \leq \frac{M}{(\lambda - \omega)^n} \quad \text{for } n \in \mathbb{N} \text{ and } \lambda > \omega,$$

where $\rho(A)$ is the resolvent set of $A$.

$C := C([-r, 0], \mathbb{E})$ is the Banach space of continuous functions on $[-r, 0]$ into $\mathbb{E}$ provided with the uniform norm topology, $L$ is a bounded linear operator from $C([-r, 0], \mathbb{E})$ to $\mathbb{E}$ and $f$ is a weighted pseudo almost periodic $\mathbb{E}$-valued function on $\mathbb{R}$ (will be precise in the next).

For every $t \geq 0$, as usual the history function $x_t \in C$ is defined by

$$x_t(\theta) = x(t + \theta) \qquad \text{for } -r \leq \theta \leq 0.$$

There are many examples where A is nondensely defined. In particular, nondensity occurs in many situations due to restrictions on the space where the equation is considered (for example, periodic continuous functions, Hölder continuous functions) or due to boundary conditions (for example, the space $C^1$ with null value on the boundary is nondense in the space of continuous functions).

The organization of this work is as follows: In Section 2, we recall some notations and definitions of a weighted pseudo almost periodic functions. In Section 3, we give the variation of constants formula that will be the principal working tools in this work, moreover, we establish fundamental results about the spectral decomposition of solutions. in Section 4, we prove the main theorem of existence and uniqueness of a weighted pseudo almost periodic solutions. In Section 5 we apply the result in the previous section to the nonlinear functional differential equation. In particular, Section 6 is to illustrate our main result (Theorem 9), we will examine sufficient conditions for the existence and uniqueness of pseudo almost periodic solutions to the diffusion equation with delay.

$$\begin{cases} \dfrac{\partial}{\partial t} y(t,x) = \dfrac{\partial^2}{\partial x^2} y(t,x) + \displaystyle\int_{-r}^{0} q(\theta)\, y(t+\theta,x)\, d\theta + z(t,x) \quad \text{for } t \in \mathbb{R} \text{ and } x \in [0,\pi], \\ y(t,0) = y(t,\pi) = 0 \quad \text{for } t \in \mathbb{R}, \end{cases}$$

where $q,\ u$ are functions satisfying some additional assumptions.

## 2 Weighted Pseudo Almost Periodic Functions

In what follows we recall some definitions and notations needed in the sequel.

Let $(\mathbb{E}, \|\ \ \|)$ be a Banach space and $L^1_{loc}(\mathbb{R})$ denote the space of all locally integrable functions on $\mathbb{R}$.

Let $\mathbb{U}$ be defined by

$$\mathbb{U} := \left\{\rho \in L^1_{loc}(\mathbb{R}) : \rho(x) > 0 \quad \text{almost everywhere } \ x \in \mathbb{R}\right\}$$

From now, if $\rho \in \mathbb{U}$ and for $R > 0$, we then set

$$m(R,\rho) := \int_{-R}^{R} \rho(x)\, dx$$

The space of weighted functions is defined by

$$\mathbb{U}_\infty := \left\{\rho \in \mathbb{U} : \lim_{R\to\infty} m(R,\rho) = \infty\right\}$$

$$\mathbb{U}_B := \left\{\rho \in \mathbb{U}_\infty :\ \rho \text{ is bounded and } \inf_{x\in\mathbb{R}} \rho(x) > 0\right\}.$$

Throughout this work $BC(\mathbb{R},\mathbb{E})$ is the space of all $\mathbb{E}$-valued bounded continuous functions equipped with the sup norm defined by $\|\phi\|_\infty := \sup_{t\in\mathbb{R}} \|\phi(t)\|$.

**Definition 1 ([6, 15]).** A function $f : \mathbb{R} \longrightarrow \mathbb{E}$ is called almost periodic if

(i) $f$ is continuous,
(ii) for each $\varepsilon > 0$ there exists an $l(\varepsilon) > 0$, such that every interval $I$ of length $l(\varepsilon)$ contains a number $\tau$ with the property that $\|f(t+\tau) - f(t)\| < \varepsilon$ for all $t \in \mathbb{R}$.

The number $\tau$ above is called $\varepsilon$-translation number of $f$.

Let $AP(\mathbb{E})$ denote the space of almost periodic functions.

Denote by $PAP_0(\mathbb{E})$ the space of ergodic perturbations defined by

$$PAP_0(\mathbb{E}) := \left\{ f \in BC(\mathbb{R}, \mathbb{E}) : \lim_{R\to\infty} \frac{1}{2R} \int_{-R}^{R} \|f(t)\|\, dt = 0 \right\}$$

**Definition 2 ([9]).** A function $f : \mathbb{R} \longrightarrow \mathbb{E}$ is called pseudo almost periodic if $f = g + \phi$, where $g \in AP(\mathbb{E})$ and $\phi \in PAP_0(\mathbb{E})$.

The collection of all pseudo almost periodic functions from $\mathbb{R}$ into $\mathbb{E}$ is denoted by $PAP(\mathbb{E})$.

**Definition 3 ([10]).** Let $\rho \in \mathbb{U}_\infty$. We define the weighted ergodic space by

$$PAP_0(\mathbb{E}, \rho) := \left\{ f \in BC(\mathbb{R}, \mathbb{E}) : \lim_{R\to\infty} \frac{1}{m(R, \rho)} \int_{-R}^{R} \|f(t)\|\rho(t)\, dt = 0 \right\}$$

**Definition 4 ([10]).** Let $\rho \in \mathbb{U}_\infty$. A function $f : \mathbb{R} \longrightarrow \mathbb{E}$ is called weighted pseudo almost periodic (or $\rho$-pseudo almost periodic) if it is expressed as follows:

$$f = g + \phi,$$

where $g \in AP(\mathbb{E})$ and $\phi \in PAP_0(\mathbb{E}, \rho)$.

The collection of all weighted-pseudo almost periodic functions from $\mathbb{R}$ into $\mathbb{E}$ is denoted by $PAP(\mathbb{E}, \rho)$.

*Remark 1 ([10]).* The functions $g$ and $\phi$ appearing in definition 4 are, respectively, called the almost periodic and the weighted ergodic components of $f$.

*Remark 2 ([10]).* The decomposition of a $\rho$-pseudo almost periodic function $f = g + \phi$, where $g \in AP(\mathbb{E})$ and $\phi \in PAP_0(\mathbb{E}, \rho)$, is unique. This is mainly based upon the fact that $g(\mathbb{R}) \subset \overline{f(\mathbb{R})}$. Hence, $PAP(\mathbb{E}, \rho) = AP(\mathbb{E}) \oplus PAP_0(\mathbb{E}, \rho)$.

**Definition 5 ([10]).** Let $(\mathbb{F}, \| \ \|)$ be a Banach space. A function $F : \mathbb{R} \times \mathbb{F} \to \mathbb{E}$ is called almost periodic in $t \in \mathbb{R}$ uniformly in $y \in \mathbb{F}$ if for each $\varepsilon > 0$ and any compact $K \subset \mathbb{F}$ there exists $l(\varepsilon)$ such that every interval of length $l(\varepsilon)$ contains a number $\tau$ with the property that

$$\|F(t+\tau, y) - F(t, y)\| < \varepsilon \quad \text{for each } t \in \mathbb{R} \text{ and } y \in K.$$

The collection of those functions is denoted by $AP(\mathbb{F}, \mathbb{E})$.

In the same way, we define $PAP_0(\mathbb{F}, \mathbb{E}, \rho)$ as the collection of jointly continuous functions $F : \mathbb{R} \times \mathbb{F} \to \mathbb{E}$ such that $F(\cdot, y)$ is bounded and

$$\lim_{R\to\infty} \frac{1}{m(R,\rho)} \int_{-R}^{R} \|F(s,y)\|\rho(s)\ ds = 0 \qquad \text{for all } \ y \in \mathbb{F}.$$

**Definition 6 ([10]).** A function $F : \mathbb{R} \times \mathbb{F} \to \mathbb{E}$ is called weighted pseudo almost periodic if

$$F = G + H,$$

where $G \in AP(\mathbb{F}, \mathbb{E})$ and $H \in PAP_0(\mathbb{F}, \mathbb{E}, \rho)$.

The class of such functions is denoted by $PAP(\mathbb{F}, \mathbb{E}, \rho)$.

We give now some properties of a weighted pseudo almost periodic functions.

**Definition 7 ([10]).** Let $\rho_1, \rho_2 \in \mathbb{U}_\infty$. One says that $\rho_1$ is equivalent to $\rho_2$ or $\rho_1 \sim \rho_2$ whenever $\frac{\rho_1}{\rho_2} \in \mathbb{U}_B$.

**Theorem 1 ([10]).** *Let $\rho_1, \rho_2 \in \mathbb{U}_\infty$. If $\rho_1$ is equivalent to $\rho_2$, then $PAP(\mathbb{E}, \rho_1) = PAP(\mathbb{E}, \rho_2)$.*

An immediate consequence of Theorem 1 is the next corollary, which enables us to connect the Zhang's space $PAP(\mathbb{E}) = AP(\mathbb{E}) \oplus PAP_0(\mathbb{E})$ with a weighted pseudo almost periodic class $PAP(\mathbb{E}, \rho)$.

**Corollary 1 ([10]).** *If $\rho \in \mathbb{U}_\mathbb{B}$, then $PAP(\mathbb{E}, \rho) = PAP(\mathbb{E})$.*

**Theorem 2 ([7]).** *The space $PAP(\mathbb{E}, \rho)$ is a closed subspace of $BC((\mathbb{R}, \mathbb{E}), \|\ \|_\infty)$ provided with the uniform norm topology. This yields $PAP(\mathbb{E}, \rho)$ is a Banach space.*

*Example 1.* Let $\rho(t) = e^t \qquad$ for each t $\in \mathbb{R}$.

It is easy to see that $m(R, \rho) = e^R - e^{-R}$ and hence $\rho \in \mathbb{U}_\infty$.

Set $f(t) = sin\ t + sin\ \sqrt{2}\ t + e^{-t}$. It is clear that $f$ belongs to $PAP(\mathbb{R}, \rho)$.

Namely, $sin\ t + sin\ \sqrt{2}\ t$ is its almost periodic component and the weighted ergodic component of $f$ verify

$$\lim_{R\to\infty} \frac{1}{e^R - e^{-R}} \int_{-R}^{R} e^{-t}\ e^t\ dt = 0,$$

while $\qquad \displaystyle\lim_{R\to\infty} \frac{1}{2R} \int_{-R}^{R} e^{-t}\ dt = +\infty.$

Hence, $f$ does not belong to $PAP(\mathbb{R})$.

## 3 Partial Functional Differential Equations

Here and hereafter, we suppose that

**($\mathbf{H_0}$)** $A$ satisfies the Hille-Yoshida condition: there exists $M \geq 0,\ \omega \in \mathbb{R}$ such that

$$(\omega, +\infty) \subset \rho(A) \text{ and } |(\lambda I - A)^{-n}| \leq \frac{M}{(\lambda-\omega)^n} \quad \text{for } n \in \mathbb{N} \text{ and } \lambda > \omega,$$

where $\rho(A)$ is the resolvent set of $A$.

To Eq. (1), we associate the following initial value problem

$$\begin{cases} \dfrac{d}{dt}x(t) = Ax(t) + L(x_t) + f(t) & \text{for } t \geq 0, \\ x_0 = \varphi \in C := C([-r, 0]; \mathbb{E}). \end{cases} \tag{2}$$

**Definition 8 ([1]).** We say that a continuous function $x$ from $[-r, \infty)$ into $\mathbb{E}$ is an integral solution of Eq. (2), if the following conditions hold:

(i) $\int_0^t x(s)\, ds \in D(A) \quad \text{for } t \geq 0,$

(ii) $x(t) = \varphi(0) + A\int_0^t x(s)\; ds + \int_0^t (L(x_s) + f(s))\; ds \quad \text{for } t \geq 0,$

(iii) $x_0 = \varphi.$

If $\overline{D(A)} = \mathbb{E}$, the integral solutions coincide with the known mild solutions. From the closed-ness property of $A$, we can see that if $x$ is an integral solution of Eq. (2), then $x(t) \in \overline{D(A)}$ for all $t \geq 0$, in particular $\varphi(0) \in \overline{D(A)}$.

Let us introduce the part $A_0$ of the operator $A$ in $\overline{D(A)}$ which is defined by

$$\begin{cases} D(A_0) = \left\{ y \in D(A) :\ Ay \in \overline{D(A)} \right\}, \\ A_0 y = Ay \quad \text{for } y \in D(A_0). \end{cases}$$

**Lemma 1 ([14]).** *$A_0$ generates a strongly continuous semigroup $(T_0(t))_{t\geq 0}$ on $\overline{D(A)}$.*

For the existence of integral solutions, we have the following result.

**Theorem 3 ([1]).** *Assume that ($\boldsymbol{H_0}$) holds. Then for all $\varphi \in C$ such that $\varphi(0) \in \overline{D(A)}$, Eq. (2) has a unique integral solution $x$ on $[-r, +\infty)$. Moreover, $x$ is given by*

$$x(t) = T_0(t)\;\varphi(0) + \lim_{\lambda\to\infty} \int_0^t T_0(t-s)B_\lambda(L(x_s) + f(s))\; ds \quad \textit{for } \; t \geq 0,$$

*where $B_\lambda = \lambda(\lambda I - A)^{-1} \quad$ for $\lambda > \omega$.*

The phase space $C_0$ of Eq. (2) is defined by

$$C_0 = \left\{ \varphi \in C : \varphi(0) \in \overline{D(A)} \right\}.$$

For each $t \geq 0$, we define the linear operator $T(t)$ on $C_0$ by

$$T(t)\varphi = x_t(0,\varphi) \quad \text{for } t \geq 0,$$

where $x(0,\varphi)$ is the solution of the following linear equation

$$\begin{cases} \dfrac{d}{dt}x(t) = Ax(t) + L(x_t) \quad \text{for } t \geq 0, \\ \\ x_0 = \varphi \in C := C([-r,0];\mathbb{E}). \end{cases}$$

**Proposition 1 ([1]).** *$(T(t))_{t\geq 0}$ is a strongly continuous semigroup of linear operators on $C_0$ :*

*(i) for all $t \geq 0$, T(t) is a bounded linear operator on $C_0$;*
*(ii) T(0)=I;*
*(iii) $T(t+s) = T(t)T(s)$ for all $t,s \geq 0$;*
*(iv) for all $\varphi \in C_0$, $T(t)\varphi$ is a continuous function of $t \geq 0$ with values in $C_0$.*

**Theorem 4 ([2]).** *Let $\mathscr{A}_T$ be defined on $C_0$ by*

$$\begin{cases} D(\mathscr{A}_T) = \Big\{\varphi \in C^1([-r,0];X) : \ \varphi(0) \in D(A), \ \varphi'(0) \in \overline{D(A)} \text{ and } \varphi'(0) = A\varphi(0) + L(\varphi)\Big\}, \\ \mathscr{A}_T\varphi = \varphi' \quad \text{for } \varphi \in D(\mathscr{A}_T). \end{cases}$$

*Then, $\mathscr{A}_T$ is the infinitesimal generator of the semigroup $(T(t))_{t\geq 0}$ on $C_0$.*

In order to give the variation of constants formula, we need to recall some notations and results which are taken from [2].

Let $\langle X_0 \rangle$ be the space defined by

$$\langle X_0 \rangle = \Big\{X_0 c : c \in X\Big\},$$

where the function $X_0 c$ is defined by

$$(X_0 c)(\theta) = \begin{cases} 0 & \text{if } \theta \in [-r,0), \\ c & \text{if } \theta = 0. \end{cases}$$

The space $C_0 \oplus \langle X_0 \rangle$ is equipped with the norm

$$|\phi + X_0 c| = |\phi|_C + |c| \quad \text{for } (\phi, c) \in C_0 \times X,$$

is a Banach space and consider the extension $\tilde{\mathscr{A}_T}$ defined on $C_0 \oplus \langle X_0 \rangle$ by

$$\begin{cases} D(\tilde{\mathscr{A}_T}) = \Big\{\varphi \in C^1([-r,0];\mathbb{E}) : \ \varphi(0) \in D(A) \ \text{ and } \varphi'(0) \in \overline{D(A)}\Big\}, \\ \mathscr{A}_T\varphi = \varphi' + X_0(A\varphi(0) + L(\varphi) - \varphi'(0)). \end{cases}$$

**Lemma 2 ([2]).** *Assume that ($\boldsymbol{H_0}$) holds. Then, $\tilde{\mathscr{A}_T}$ satisfies the Hille-Yoshida condition on $C_0 \oplus \langle X_0 \rangle$: there exist $\tilde{M} \geq 0$ and $\tilde{\omega} \in \mathbb{R}$ such that*

$$(\tilde{\omega}, +\infty) \subset \rho(\tilde{\mathscr{A}_T}) \ \text{ and } \ |(\lambda I - \tilde{\mathscr{A}_T})^{-n}| \leq \frac{\tilde{M}}{(\lambda - \tilde{\omega})^n} \quad \text{for } n \in \mathbb{N} \ \text{ and } \ \lambda > \tilde{\omega}.$$

*Moreover, the part of $\tilde{\mathscr{A}_T}$ on $\overline{D(\tilde{A}_T)} = C_0$ is exactly the operator $\mathscr{A}_T$.*

**Theorem 5 ([2]).** *Assume that ($\boldsymbol{H_0}$) holds. Then for all $\varphi \in C_0$, the solution $x$ of Eq. (2) is given by the following variation of constants formula*

$$x_t = T(t)\,\varphi + \lim_{\lambda \to +\infty} \int_0^t T(t-s)\tilde{B}_\lambda(X_0 f(s))\, ds \quad \text{for } \ t \geq 0,$$

*where $\tilde{B}_\lambda = \lambda\, (\lambda I - \tilde{\mathscr{A}_T})^{-1}$ for $\lambda > \tilde{\omega}$.*

In the following, we assume that ($\boldsymbol{H_1}$) The operator $T_0(t)$ is compact on $\overline{D(A)}$ for every $t > 0$.

**Theorem 6 ([2]).** *Assume that ($\boldsymbol{H_0}$) and ($\boldsymbol{H_1}$) hold, then T(t) is compact for $t > r$.*

As a consequence from the compactness property of the operator $T(t)$, we have that the spectrum $\sigma(\mathscr{A}_T)$ is the point spectrum

$$\sigma(\mathscr{A}_T) = \big\{\lambda \in \mathbb{C} : \ker \Delta(\lambda) \neq \{0\}\big\}$$

where the linear operator $\Delta(\lambda) : D(A) \to \mathbb{E}$ is given by

$$\Delta(\lambda) = \lambda I - A - L(e^{\lambda} I)$$

and $e^{\lambda} I : \mathbb{E} \to C$ is defined by

$$(e^{\lambda} x)(\theta) = e^{\lambda\theta} x \quad \text{for } \ x \in \mathbb{E} \ \text{ and } \theta \in [-r, 0].$$

## 4 Weighted Pseudo Almost Periodic Solution

In this section we study the existence and uniqueness of a weighted pseudo almost periodic solution to Eq. (1).

**Definition 9 ([16]).** We say that the semigroup $(T(t))_{t\geq 0}$ is hyperbolic if $\sigma(\mathscr{A}_T) \cap i\mathbb{R} = \varnothing$.

**Theorem 7 ([14]).** *Assume that ($\boldsymbol{H_0}$) and ($\boldsymbol{H_1}$) hold. If the semigroup $(T(t))_{t\geq 0}$ is hyperbolic, then the space $C_0$ is decomposed as the direct sum of the stable and unstable subspaces*

$$C_0 = S \oplus U$$

*and there exist positive constants K and c such that*

$$\|T(t)\varphi\| \leq Ke^{-ct}\|\varphi\| \quad \textit{for } \ t \geq 0 \textit{ and } \ \varphi \in S,$$

$$\|T(t)\varphi\| \leq Ke^{ct}\|\varphi\| \quad \textit{for } \ t \leq 0 \textit{ and } \ \varphi \in U.$$

**Theorem 8 ([14]).** *Assume that ($\boldsymbol{H_0}$) and ($\boldsymbol{H_1}$) hold and the semigroup $(T(t))_{t\geq 0}$ is hyperbolic. If f is bounded on $\mathbb{R}$, then Eq. (1) has a unique bounded solution on $\mathbb{R}$ which is given by the following formula*

$$x_t = \lim_{n\to+\infty} \int_{-\infty}^{t} T^s(t-\tau)\Pi^s(\tilde{B}_n X_0 f(\tau))\, d\tau + \lim_{n\to+\infty} \int_{+\infty}^{t} T^u(t-\tau)\Pi^u(\tilde{B}_n X_0 f(\tau))\, d\tau$$

*where $\Pi^s$ and $\Pi^u$ are the projections of C onto the stable and unstable subspaces, respectively.*

We give the main result of this work, which shows the existence and uniqueness of $\rho$-pseudo almost periodic solution if the input function $f$ is $\rho$-pseudo almost periodic.

**Theorem 9.** *Fix $\rho \in \mathbb{U}_\infty$. Assume that ($\boldsymbol{H_0}$) and ($\boldsymbol{H_1}$) hold and the semigroup $(T(t))_{t\geq 0}$ is hyperbolic. If f is $\rho$-pseudo almost periodic in $t \in \mathbb{R}$ and $\rho$ is decreasing with*

$$P(c) := \sup_{R>0} \left( \int_{-R}^{R} e^{-c(t+R)} \rho(t)\, dt \right) < \infty, \tag{3}$$

*then, Eq. (1) has one and only one bounded solution which is also $\rho$-pseudo almost periodic.*

*Proof.* Eq. (1) has one and only one bounded solution on $\mathbb{R}$ which is given by

$$\begin{aligned} x_t = &\lim_{n\to+\infty} \int_{-\infty}^{t} T^s(t-\tau)\ \Pi^s(\tilde{B}_n X_0 f(\tau))\, d\tau \\ &+ \lim_{n\to+\infty} \int_{+\infty}^{t} T^u(t-\tau)\ \Pi^u(\tilde{B}_n X_0 f(\tau))\, d\tau \end{aligned}$$

We will show that both functions

$$\lim_{n\to+\infty} \int_{-\infty}^{t} T^s(t-\tau)\ \Pi^s(\tilde{B}_n X_0 f(\tau))\, d\tau \ \text{ and}$$

$$\lim_{n\to+\infty} \int_{+\infty}^{t} T^u(t-\tau)\ \Pi^u(\tilde{B}_n X_0 f(\tau))\, d\tau$$

are weighted pseudo almost periodic.

Since $f$ is $\rho$-pseudo almost periodic then, $f = g + \phi$ where $g$ is almost periodic and $\phi \in PAP_0(\mathbb{E}, \rho)$.

Recall that $\phi \in PAP_0$ $(\mathbb{E}, \rho)$ if and only if

$$\phi \in BC(\mathbb{R}, \mathbb{E}) \text{ and } \lim_{R\to\infty} \frac{1}{m(R,\rho)} \int_{-R}^{R} \|\phi(t)\| \, \rho(t)dt = 0.$$

$$\begin{aligned}
&\lim_{n\to+\infty} \int_{-\infty}^{t} T^s(t-\tau)\, \Pi^s(\tilde{B}_n X_0 f(\tau))\, d\tau \\
&\quad = \lim_{n\to+\infty} \int_{-\infty}^{t} T^s(t-\tau)\, \Pi^s(\tilde{B}_n X_0 g(\tau))\, d\tau \\
&\qquad + \lim_{n\to+\infty} \int_{-\infty}^{t} T^s(t-\tau)\, \Pi^s(\tilde{B}_n X_0 \phi(\tau))\, d\tau
\end{aligned}$$

$$\begin{aligned}
&\lim_{n\to+\infty} \int_{+\infty}^{t} T^u(t-\tau)\, \Pi^u(\tilde{B}_n X_0 f(\tau))\, d\tau \\
&\quad = \lim_{n\to+\infty} \int_{+\infty}^{t} T^u(t-\tau)\, \Pi^u(\tilde{B}_n X_0 g(\tau))\, d\tau \\
&\qquad + \lim_{n\to+\infty} \int_{+\infty}^{t} T^u(t-\tau)\, \Pi^u(\tilde{B}_n X_0 \phi(\tau))\, d\tau
\end{aligned}$$

We will show that both

$$\lim_{n\to+\infty} \int_{-\infty}^{t} T^s(t-\tau)\, \Pi^s(\tilde{B}_n X_0 g(\tau))\, d\tau \text{ and } \lim_{n\to+\infty} \int_{+\infty}^{t} T^u(t-\tau)\, \Pi^u(\tilde{B}_n X_0 g(\tau))\, d\tau$$

are almost periodic.

Let $y = (y_m)$ be a real sequence. By almost periodicity of $g$, there exists a subsequence of $y_m$ noted by $y_m^{'}$ and a continuous function $h(t)$ such that $h(t) = \lim_{m\to\infty} g(t + y_m^{'})$ uniformly in $\mathbb{R}$.

So,

$$\begin{aligned}
&\lim_{m\to\infty}\Big[\lim_{n\to\infty}\int_{-\infty}^{t} T^s(t-\tau)\Pi^s(\tilde{B}_nX_0g(\tau+y_m^{'}))d\tau\Big]\\
&=\lim_{n\to\infty}\int_{-\infty}^{t} T^s(t-\tau)\Pi^s(\tilde{B}_nX_0\lim_{m\to\infty}g(\tau+y_m^{'}))d\tau\\
&=\lim_{n\to+\infty}\int_{-\infty}^{t} T^s(t-\tau)\ \Pi^s(\tilde{B}_nX_0h(\tau))\ d\tau
\end{aligned}$$

uniformly in $\mathbb{R}$.

Similarly,

$$\begin{aligned}
&\lim_{m\to\infty}\Big[\lim_{n\to\infty}\int_{+\infty}^{t} T^u(t-\tau)\Pi^u(\tilde{B}_nX_0g(\tau+y_m^{'}))d\tau\Big]\\
&=\lim_{n\to\infty}\int_{+\infty}^{t} T^u(t-\tau)\Pi^u(\tilde{B}_nX_0\lim_{m\to\infty}g(\tau+y_m^{'}))d\tau\\
&=\lim_{n\to+\infty}\int_{+\infty}^{t} T^u(t-\tau)\Pi^u(\tilde{B}_nX_0h(\tau))d\tau
\end{aligned}$$

uniformly in $\mathbb{R}$.

Thus,

$$\lim_{n\to+\infty}\int_{-\infty}^{t} T^s(t-\tau)\Pi^s(\tilde{B}_nX_0g(\tau))\ d\tau \ \text{ and } \ \lim_{n\to+\infty}\int_{+\infty}^{t} T^u(t-\tau)\Pi^u(\tilde{B}_nX_0g(\tau))\ d\tau$$

are almost periodic.

It remains to show that

$$\lim_{R\to\infty}\frac{1}{m(R,\rho)}\int_{-R}^{R}\Big\|\lim_{n\to+\infty}\int_{-\infty}^{t} T^s(t-\tau)\Pi^s(\tilde{B}_nX_0\phi(\tau))\ d\tau\Big\|\rho(t)\ dt=0$$

and

$$\lim_{R\to\infty}\frac{1}{m(R,\rho)}\int_{-R}^{R}\Big\|\lim_{n\to+\infty}\int_{+\infty}^{t} T^u(t-\tau)\Pi^u(\tilde{B}_nX_0\phi(\tau))\ d\tau\Big\|\rho(t)\ dt=0$$

Let us put

$$\begin{aligned}
I(t)&=\lim_{n\to+\infty}\int_{-\infty}^{t} T^s(t-\tau)\Pi^s(\tilde{B}_nX_0\phi(\tau))\ d\tau\\
J(t)&=\lim_{n\to+\infty}\int_{+\infty}^{t} T^u(t-\tau)\Pi^u(\tilde{B}_nX_0\phi(\tau))\ d\tau
\end{aligned}$$

Using the relation (3) and (2) there exists a positive constants $K$ and $\tilde{M}$ such that

$$\begin{aligned}
&\lim_{R\to\infty} \frac{1}{m(R,\rho)} \int_{-R}^{R} \|I(t)\|\, \rho(t)\, dt \\
&\leq \lim_{R\to\infty} \frac{K\,\tilde{M}}{m(R,\rho)} \int_{-R}^{R} \Big[ \int_{-\infty}^{t} e^{-c(t-\tau)}\, \|\phi(\tau)\|\, d\tau \Big] \rho(t)\, dt \\
&= \lim_{R\to\infty} \frac{K\,\tilde{M}}{m(R,\rho)} \int_{-R}^{R} \Big[ \int_{-R}^{t} e^{-c(t-\tau)}\, \|\phi(\tau)\|\, d\tau \Big] \rho(t)\, dt \\
&\quad + \lim_{R\to\infty} \frac{K\,\tilde{M}}{m(R,\rho)} \int_{-R}^{R} \Big[ \int_{-\infty}^{-R} e^{-c(t-\tau)}\, \|\phi(\tau)\|\, d\tau \Big] \rho(t)\, dt \\
&= I_1(\rho) + I_2(\rho)
\end{aligned}$$

By using the Fubini's theorem, one has

$$\begin{aligned}
I_1(\rho) &= \lim_{R\to\infty} \frac{K\,\tilde{M}}{m(R,\rho)} \int_{-R}^{R} \Big[ \int_{-R}^{t} e^{-c(t-\tau)}\, \|\phi(\tau)\|\, d\tau \Big] \rho(t)\, dt \\
&= \lim_{R\to\infty} \frac{K\,\tilde{M}}{m(R,\rho)} \int_{-R}^{R} \|\phi(\tau)\| \Big[ \int_{\tau}^{R} e^{-c(t-\tau)}\, \rho(t)\, dt \Big]\, d\tau
\end{aligned}$$

Since $\rho$ is a decreasing function then we have

$$I_1(\rho) \leq \lim_{R\to\infty} \frac{K\,\tilde{M}}{m(R,\rho)} \int_{-R}^{R} \|\phi(\tau)\|\ \ \rho(\tau) \Big[ \frac{1}{c} \Big(1 - e^{-c(R-\tau)}\Big) \Big]\, d\tau.$$

Furthermore $-R \leq t \leq R$ and $c > 0$ then $\frac{1}{c}\left(1 - e^{-c(R-\tau)}\right)$ is bounded uniformly in $\tau$.

$$I_1(\rho) \leq \lim_{R\to\infty} \frac{K\,\tilde{M}}{c\, m(R,\rho)} \int_{-R}^{R} \|\phi(\tau)\|\, \rho(\tau)\, d\tau = 0.$$

By (3) we have

$$\begin{aligned}
I_2(\rho) &= \lim_{R\to\infty} \frac{K\,\tilde{M}}{m(R,\rho)} \int_{-R}^{R} \Big[ \int_{-\infty}^{-R} e^{-c(t-\tau)}\, \|\phi(\tau)\|\, d\tau \Big] \rho(t)\, dt \\
&= \lim_{R\to\infty} \frac{K\,\tilde{M}}{m(R,\rho)} \int_{-R}^{R} e^{-ct} \Big[ \int_{-\infty}^{-R} e^{c\tau}\, \|\phi(\tau)\|\, d\tau \Big] \rho(t)\, dt \\
&= \lim_{R\to\infty} \frac{K\,\tilde{M}}{m(R,\rho)} \int_{-R}^{R} e^{-ct}\ \ \rho(t)\, dt \Big[ \int_{-\infty}^{-R} e^{c\tau}\, \|\phi(\tau)\|\, d\tau \Big]
\end{aligned}$$

$$= \lim_{R\to\infty} \frac{K\,\tilde{M}}{c\ m(R,\rho)\ e^{cR}}\ \sup_{\tau\in\mathbb{R}} \|\phi(\tau)\| \int_{-R}^{R} e^{-ct}\ \rho(t)\ dt$$

$$= 0.$$

By the similar argument we show that

$$\lim_{R\to\infty} \frac{1}{m(R,\rho)} \int_{-R}^{R} \|J(t)\|\ \rho(t)\ dt.$$

This completes the proof of the theorem.

## 5 Nonlinear Partial Functional Differential Equation

In this section, we consider the nonlinear partial differential equation

$$\frac{d}{dt}\,x(t) = Ax(t) + L(x_t) + h(t, x(t-r)) \qquad \text{for } t \in \mathbb{R}. \tag{4}$$

**($\mathbf{H_2}$)** $h : \mathbb{R}\times\mathbb{E} \longrightarrow \mathbb{E}$ is continuous and Lipschitzian with respect to the second argument:
there exists $\tilde{K} > 0$ such that

$$\|h(t,x) - h(t,y)\| \le \tilde{K}\ \|x-y\| \quad \text{for all } x,\ y \in \mathbb{F} \quad \text{and } t \in \mathbb{R}.$$

**Theorem 10 ([10]).** *Assume that ($\boldsymbol{H_0}$), ($\boldsymbol{H_1}$), and ($\boldsymbol{H_2}$) hold. Then, if $u \in PAP(\mathbb{E},\rho)$, then $h(\cdot, u(\cdot - r)) \in PAP(\mathbb{E},\rho)$.*

**Theorem 11.** *Assume that ($\boldsymbol{H_0}$), ($\boldsymbol{H_1}$), and ($\boldsymbol{H_2}$) hold, the semigroup $(T(t))_{t\ge 0}$ is hyperbolic and $\rho$ is decreasing with*

$$P(c) := \sup_{R>0} \left( \int_{-R}^{R} e^{-c(t+R)} \rho(t)\ dt \right) < \infty.$$

*Then, (4) has one and only one bounded solution which is also $\rho$-pseudo almost periodic whenever $\tilde{K}$ is small enough.*

*Proof.* Let $v \in PAP(\mathbb{E},\rho)$ consider the following equation

$$\frac{d}{dt}\,u(t) = Au(t) + L(u_t) + h(t, v(t-r)) \qquad \text{for } t \in \mathbb{R}. \tag{5}$$

By theorem 10, $PAP(\mathbb{E},\rho)$ is translation invariant we have that

$$h(t, v(t-r)) \in PAP(\mathbb{E},\rho).$$

Then Eq. (5) has one and only one solution $u$ in $PAP(\mathbb{E},\rho)$ which is given by

$$u_t = \lim_{n\to+\infty}\int_{-\infty}^{t} T^s(t-\tau)\Pi^s(\tilde{B}_n X_0 h(\tau, v(\tau-r))d\tau$$
$$+ \lim_{n\to+\infty}\int_{+\infty}^{t} T^u(t-\tau)\Pi^u(\tilde{B}_n X_0 h(\tau, v(\tau-r)))d\tau$$

Let the operator $\mathscr{H}$ be defined by

$$\mathscr{H} : PAP(\mathbb{E},\rho) \longrightarrow PAP(\mathbb{E},\rho)$$
$$v \longmapsto \mathscr{H}(v) = u$$

Due to the hyperbolicity, we can see that for some positive constant $\tilde{N}$

$$\|\mathscr{H}(v) - \mathscr{H}(w)\| \le \tilde{N}\tilde{K}\,\|v-w\|$$

If $\tilde{N}\tilde{K} < 1$, then $\mathscr{H}$ has a unique fixed point which is the unique $\rho$-pseudo almost periodic solution of Eq. (4).

## 6 Example

To illustrate the result in Theorem 9, we consider the following partial functional differential equation with diffusion which describes the evolution of a single diffusive animal species with population density $v$. For more details about this model, we refer to [17].

$$\begin{cases} \dfrac{\partial}{\partial t}\, y(t,x) = \dfrac{\partial^2}{\partial x^2}\, y(t,x) + \displaystyle\int_{-r}^{0} q(\theta)\, y(t+\theta,x)\, d\theta + z(t,x) \quad \text{for } t\in\mathbb{R} \text{ and } x\in[0,\pi], \\ y(t,0) = y(t,\pi) = 0 \quad \text{for } t\in\mathbb{R}, \end{cases} \tag{6}$$

where $q : [-r,0] \longrightarrow \mathbb{R}$ is continuous, $z : \mathbb{R}\times[0,\pi] \longrightarrow \mathbb{R}$ is continuous and defined by

$$z(t,x) = \psi(t)\ \ \eta(x), \quad \eta : [0,\pi] \longrightarrow \mathbb{R} \text{ is a continuous function,}$$

where $\psi(t) = \sin\, t + \sin\, \sqrt{2}\, t + e^{\alpha t}$ for each $t\in\mathbb{R}$ and $\alpha > 0$.

Let $\mathbb{E} = C\,([0,\pi];\mathbb{R})$ be the space of continuous functions from $[0,\pi]$ to $\mathbb{R}$ equipped with the uniform norm topology.

Define the operator $A : D(A) \subset \mathbb{E} \longrightarrow \mathbb{E}$ by

$$\begin{cases} D(A) = \Big\{y \in C^2([0,\pi];\mathbb{R}) : y(0) = y(\pi) = 0\Big\}, \\ \\ Ay = y''. \end{cases}$$

**Lemma 3 ([14]).** $(0, +\infty) \subset \rho(A)$ *and* $|(\lambda I - A)^{-1}| \leq \dfrac{1}{\lambda}$ *for* $\lambda > 0$.

Moreover, $\overline{D(A)} = \Big\{y \in \mathbb{E} : y(0) = y(\pi) = 0\Big\}$. Consequently, the condition $(\mathbf{H_0})$ is satisfied.

In order to rewrite Eq. (6) in the abstract form (1), we introduce the operator $L : C \longrightarrow \mathbb{E}$ defined by

$$L(\phi)(x) = \int_{-r}^{0} q(\theta)\, \phi(\theta)(x)\, d\theta \quad \text{for } x \in [0,\pi] \text{ and } \phi \in C,$$

and the function $f : \mathbb{R} \to \mathbb{E}$ defined by

$$f(t)(x) = z(t,x) = \psi(t) \;\; \eta(x) \quad \text{for } t \in \mathbb{R} \text{ and } \; x \in [0,\pi].$$

Then, $L$ is a bounded linear operator from $C$ to $\mathbb{E}$ and from continuity of $\psi$ we get that $f$ is a continuous function from $\mathbb{R}$ to $\mathbb{E}$.

Then, the equation (6) takes the abstract form (1).

Let $A_0$ be the part of A in $\overline{D(A)}$. Then, $A_0$ is given by

$$\begin{cases} D(A_0) = \Big\{y \in C^2([0,\pi];\mathbb{R}) : y(0) = y(\pi) = y''(0) = y''(\pi) = 0\Big\}, \\ \\ A_0 y = Ay \quad \text{for } y \in D(A_0). \end{cases}$$

$A_0$ generates a strongly continuous compact semigroup $(T_0(t))_{t\geq 0}$ on $\overline{D(A)}$.

In order to study the existence and uniqueness of a bounded solution of Eq. (1), we suppose that

$(\mathbf{H_3})$ $\displaystyle\int_{-r}^{0} |q(\theta)|\, d\theta < 1.$

**Proposition 2.** *Assume that ($\boldsymbol{H_3}$) holds. Then, the semigroup $(T(t))_{t\geq 0}$ is exponentially stable: there exist constants $M \geq 1$ and $\omega > 0$ such that $\|T(t)\| \leq M\, e^{-\omega t}$ for all $t \geq 0$.*

*Proof.* Let $\lambda \in \sigma(\mathscr{A}_T)$, then there exists $x \in D(A))$, $x \neq 0$ such that $\Delta(\lambda)\, x = 0$, which implies that

$$\lambda x - Ax - \Big(\int_{-r}^{0} q(\theta)\, e^{\lambda\theta}\, d\theta\Big)x = 0,$$

and

$$\lambda - \int_{-r}^{0} q(\theta)\, e^{\lambda\theta}\, d\theta \in \sigma_p\, (A),$$

where $\sigma_p\ (A)$ is the point spectrum $\sigma_p\ (A)$ of $A$ and is

$$\sigma_p\, (A) = \Big\{ -n^2 : n \in \mathbb{N}^* \Big\}.$$

Consequently, $\lambda \in \sigma(\mathscr{A}_T)$ if and only if

$$\lambda - \int_{-r}^{0} q(\theta)\, e^{\lambda\theta}\, d\theta = -n^2 \quad \text{for some } n \geq 1. \tag{7}$$

Taking the real part in formula (7), we obtain that

$$Re(\lambda) = \int_{-r}^{0} q(\theta)\, e^{Re(\lambda\theta)}\, cos(Im\lambda\ \theta)\, d\theta - n^2 \qquad \text{for } \ n \geq 1.$$

Assume that $Re(\lambda) \geq 0$, then

$$Re(\lambda) \leq \int_{-r}^{0} |q(\theta)|\, d\theta - 1 < 0.$$

This gives a contradiction. Consequently, $\sigma(\mathscr{A}_T) \subset \{\lambda \in \mathbb{C} : Re\ (\lambda) < 0\}$, which implies that the semigroup $(T(t))_{t\geq 0}$ is exponentially stable.

Set

$$\rho(t) = \begin{cases} 1 & \text{if } \ t < 0, \\ e^{-\beta t} & \text{if } \ t \geq 0 \ \text{ and } \beta > 0. \end{cases}$$

Then, $\lim_{R\to+\infty} m(R, \rho) = +\infty$ and hence $\rho \in \mathbb{U}_\infty$.

**Proposition 3.** *Assume that* $0 < \alpha < \beta$. *Then, Eq.* (6) *has a unique bounded and weighted pseudo almost periodic solution.*

*Proof.* If $\alpha < \beta$, the condition (3) is satisfied as

$$P(\omega) := \sup_{R>0} \Big( \frac{1}{e^{\omega R}} \int_{-R}^{R} e^{-\omega t} \rho(t)\, dt \Big) < \infty.$$

It is easy to check that $\psi$ doesn't belong to $PAP(\mathbb{R})$ since

$$\lim_{R\to\infty} \frac{1}{2R} \int_{-R}^{R} e^{\alpha t}\, dt = \infty.$$

and $\psi \in PAP(\mathbb{R}, \rho)$ with $sin\ t + sin(\sqrt{2}\ t)$ as its almost periodic component and $e^{\alpha t}$ as its weighted ergodic component which verify

$$\lim_{R\to\infty} \frac{1}{m(R,\rho)} \int_{-R}^{R} e^{\alpha t} \ \ \rho(t)\, dt = 0.$$

By Theorem 9, we deduce that Eq. (6) has a unique bounded and weighted pseudo almost periodic solution.

## References


1. M. Adimy, K. Ezzinbi, Existence and linearized stability for partial neutral functional differential equations, Differential Equations and Dynamical Systems, Volume 7, (1999), 371–417.
2. M. Adimy, K. Ezzinbi, M. Laklach, Spectral decomposition for partial neutral functional differential equations, Can. Appl. Math. Q., Volume 9, Issue 1, (2001), 1–34.
3. E. Ait Dads and O. Arino, Exponential dichotomy and existence of pseudo almost periodic solutions of some differential equations, Nonlinear Analysis, Volume 27, Issue 4, (1996), 369–386.
4. E. Ait Dads, K. Ezzinbi and O. Arino, Pseudo almost periodic solutions for some differential equations in Banach space, Nonlinear Analysis, Volume 28, Issue 7, (1997), 1141–1155.
5. E. Ait Dads and K. Ezzinbi, Pseudo almost periodic solutions of some delay differential equations, Journal of Mathematical Analysis and Applications, Volume 201, Issue 287, (1996), 840–850.
6. C. Corduneanu, Almost Periodic Functions, Wiley, New York, 1968. Reprinted, Chelsea, New York, 1989.
7. T. Diagana, Weighted pseudo-almost periodic solutions to some differential equations, Nonlinear Analysis, Volume 68, (2008), Issue 8, 2250–2260.
8. T. Diagana, Existence and uniqueness of pseudo almost periodic solutions to some classes of partial evolution equations, Nonlinear Analysis, Volume 66, Issue 2, (2007), 384–395.
9. T. Diagana, Pseudo Almost Periodic Functions in Banach Spaces, Nova Science Publishers, Inc., New York, 2007.
10. T. Diagana, Weighted pseudo almost periodic functions and applications, C.R.A.S, Volume 343, Issue 10, (2006), 643–646.
11. T. Diagana, C.M. Mahop, G.M. N'Guérékata, Pseudo-almost-periodic solutions to some semilinear differential equations, Mathematical and Computer Modelling, Volume 43, Issues 1-2, (2006), 89–96.
12. T. Diagana, C.M. Mahop, G.M. N'Guérékata, B. Toni, Existence and uniqueness of pseudo almost periodic solutions to some classes of semilinear differential equations and applications, Nonlinear Analysis, Volume 64, Issue 11, (2006), 2442–2453.
13. T. Diagana, Pseudo almost periodic solutions to some differential equations, Nonlinear Analysis, Volume 60, Issue 7, (2005), 1277–1286.
14. K. Ezzinbi, G.M. N'Guérékata, Almost automorphic solutions for some partial functional differential equations, Journal of Mathematical Analysis and Applications, Volume 328, (2007), 344–358.
15. A.M. Fink, Almost Periodic Differential Equations, Lecture Notes in Mathematics, Volume 377, Springer-Verlag, New York, 1974.
16. J.K. Hale and V. Lunel, Introduction to Functional Differential Equations, Springer-verlag, New-York/Berlin, 1993.

17. J. Wu, Theory and Applications of Partial Functional Differential Equations, Springer-verlag, New-York/Berlin, 1996.
18. C.Y. Zhang, Integration of vector-valued pseudo almost periodic functions, Proceedings of the American Mathematical Society, Volume 121, Issue 1, (1994), 167–174.
19. C.Y. Zhang, pseudo almost-periodic solutions of some Differential Equations, Journal of Mathematical Analysis and Applications, Volume 181, Issue 1, (1994), 62–76.
20. C.Y. Zhang, Pseudo almost periodic solutions of some differential equations II, Journal of Mathematical Analysis and Applications, Volume 192, Issue 2, (1995), 543–561.
21. L. Zhang and Y. Xu, Weighted pseudo almost periodic solutions for functional differential equations, Electronic Journal of Differential Equations, Volume 2007, Issue 146, (2007), 1–7.

# Using B-splines functions and EM algorithm for Hidden Markov Model-based Unsupervised Image Segmentation

**Atizez Hadrich, Mourad Zribi, and Afif Masmoudi**

**Abstract** Hidden Markov models have been used in image processing, especially in image segmentation. In this paper, we propose a new approach for the unsupervised image segmentation, based on hidden Markov models and B-splines functions. The estimation of the new parameters for the hidden Markov model by using B-splines functions is performed from the expectation maximization (EM) algorithm. Then, we introduce a new algorithm (EMMB) based on EM Markov B-spline. Experimental results on synthetic and color images show that the new approach can provide a more homogeneous segmentation than the classical one.

**Keywords** Hidden Markov models • Unsupervised image segmentation • EM algorithm • Color image • B-splines function

## 1 Introduction

Hidden Markov random field (HMRF) models revealed themselves as a powerful tool for image segmentation [9]. They are very applied in accounting for spatial dependencies between the different pixels of an image but these spatial dependencies are also responsible for a typically large amount of computation. In practice,

A. Hadrich (✉)
Faculté des Sciences de Sfax, Laboratoire de Probabilités et Statistique, Université de Sfax, B. P. 1171 Sfax, Tunisie

Laboratoire d'Informatique, Signal et Image de la Côte, d'Opale (LISIC-EA 4491), ULCO, 50 Rue Ferdinand Buisson, BP 719, 62228 Calais Cedex, France
e-mail: hadrich_atizez@lisic.univ-littoral.fr

M. Zribi
Laboratoire d'Informatique, Signal et Image de la Côte, d'Opale (LISIC-EA 4491), ULCO, 50 Rue Ferdinand, Buisson BP 719, 62228 Calais Cedex, France
e-mail: Mourad.Zribi@lisic.univ-littoral.fr

A. Masmoudi
Faculté des Sciences de Sfax, Laboratoire de Probabilités et Statistique, Université de Sfax, B. P. 1171 Sfax, Tunisie
e-mail: Afif.Masmoudi@fss.rnu.tn



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_11

Markov model-based segmentation requires the estimation of the model parameters. A common approach consists of alternatively restoring the unknown segmentation based on a maximum a posteriori (MAP) rule and then estimating the model parameters using the observations and the restored data. In our work, we introduce a new hidden Markov model by using B-splines functions. In order to estimate the model parameters, we apply the EM algorithm called EM Markov B-spline algorithm (EMMB). This procedure is widely used in the context of incomplete data and in particular to estimate independent mixture models due to its simplicity. In this paper, we propose a generalization of the mean field principle of statistical physics to make the EM algorithm tractable. More specifically, we consider approximations of Markov models, with complex dependencies, by systems of independent variables. These approximations are obtained by fixing the neighborhood of each pixel to arbitrary constants. They lead to valid probability models, with factorization properties, much simpler to deal with. We then use these approximations to carry out the EM algorithm and derive a class of algorithms. They can be interpreted as the EM algorithm for independent mixture models with the difference that the mixture model adaptatively changes at each iteration depending on the current choice for the pixels' neighbors labels. It follows algorithms that have the advantage to take the spatial information into account while keeping each iteration as simple as in the independent mixture case. The paper is organized as follows. Section II specifies the context of nonparametric density estimation methods and B-splines functions. Hidden Markov models for image segmentation in Section III. Parameters estimation using the EMMB algorithm in Section IV. Performance comparison in section V and the conclusion in section VI.

## 2 Nonparametric density estimation methods and B-splines functions.

In the mathematical subfield of numerical analysis, a B-splines [1–4, 11] is a spline function that has a minimum support with respect to a given degree, smoothness, and domain partition. It's well known that every spline density function can be represented as a finite linear combination of B-splines. The term B-splines stands for basis splines according to Isaac Jacob Schoenberg [3]. A B-splines nonparametric density estimator with uniformly spaced knots convenient for large data sets was discussed by Gehringer and Redner [6].

Curry and Schoenberg (1966) [3] have proved that every spline function $S$ of degree $d$ ($d = 1, 2, \ldots$) with $m$ knots ($m = 1, 2, \ldots$) has a unique expansion

$$S(y) = \sum_{l=1}^{m+d} \frac{b_l}{h_l} B_l^d(y), \text{ for } a < y < b \text{ and } h_l = \int B_l^d(y)dy, \tag{1}$$

where $a$, $b \in \mathbb{R}$, and $b_l$'s are unknown parameters which need to be estimated. Note that, $b_l \geq 0$, and $\sum_{l=1}^{m+d} b_l = 1$ is a special requirement when using B-splines to estimate probability density functions. For general splines, this special requirement is not true. The case $d = 1$ corresponds to a piecewise approximation which is attractively simple but produces a visible roughness, unless the knots are close to each other.

The B-splines of $d$ degree are defined recursively by

$$B_l^d(y) = \frac{y - y_l}{y_{l+d} - y_l} B_l^{d-1}(y) + \frac{y_{l+d+1} - y}{y_{l+d+1} - y_{l+1}} B_{l+1}^{d-1}(y) \tag{2}$$

where

$$B_l^0(y) = \begin{cases} 1, \text{ if } y \in [y_l, y_{l+1}) \\ 0, \text{ elsewhere.} \end{cases}$$

In more technical terms, a spline function $S$ of degree $d$ with $m$ knots, $a = y_1 < y_2 < \ldots < y_{m+2} = b$ is $(d-1)$ continuous derivative function such that $S \in P^d$, ($P^d$ is the class of polynomials of a maximum degree $d$ in each of intervals $(a, y_2), (y_2, y_3), \ldots, (y_{m+1}, b)$.

In our work, we use the second order B-splines functions $(B_l^2)_{l=1\ldots m+2}$ defined by

$$B_l^2(y) = \begin{cases} \dfrac{(y - y_l)^2}{(y_{l+1} - y_l)(y_{l+2} - y_l)}, & y \in [y_l, y_{l+1}) \\ \dfrac{1}{y_{l+2} - y_{l+1}} [\dfrac{(y - y_l)(y_{l+2} - y)}{y_{l+2} - y_l} + \\ \dfrac{(y - y_{l+1})(y_{l+3} - y)}{y_{l+3} - y_{l+1}}], & y \in [y_{l+1}, y_{l+2}) \\ \dfrac{(y_{l+3} - y)^2}{(y_{l+3} - y_{l+1})(y_{l+3} - y_{l+2})}, & y \in [y_{l+2}, y_{l+3}) \\ 0, & \text{elsewhere} \end{cases}$$

where $l$ is integer as usual. We notice that $0 \leq B_l^2(y) \leq 1$ is always verified.

Non-zero parts of three $B^2 = (B_l^2)_{l=1\ldots m+2}$ splines are plotted in Fig. 1. The support of each spline covers three intervals. It is a quadratic polynomial on each support interval. Note that the peak value of $B_l^2(y)$ is less than 1. If all the $B_l^2$ splines could be plotted, in any interval, there would be contributions from three splines. But we only see this in $[y_{l+1}, y_{l+2}]$ in Fig. 1. It's interesting to notice that these three contributions sum to unity.

Let $Y_1, Y_2, \ldots, Y_n$ be $n$ random variables with an unknown common pdf $f$. By using the second order B-splines, we can approximate $f$ by the following 'mixture' of B-splines:

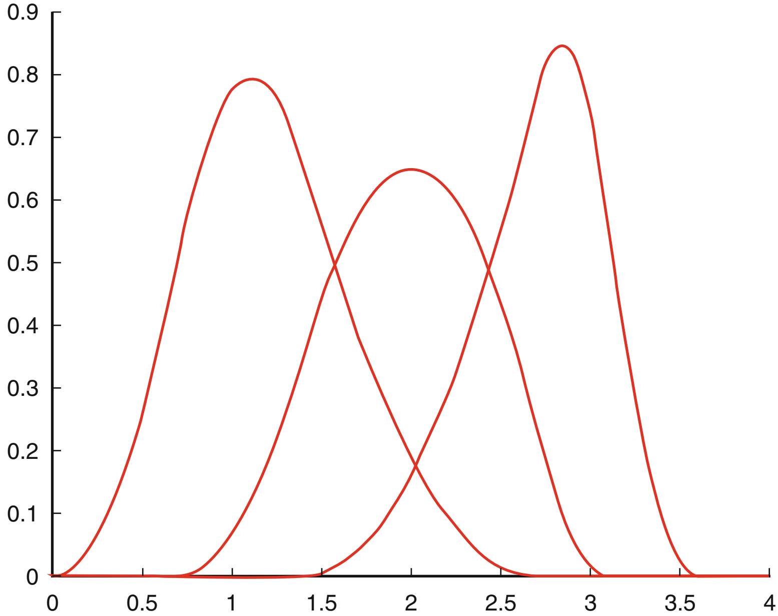


**Fig. 1** Non-zero parts of B-splines of degree 2. Note that the peak values are less than 1.

$$f(y,b) = \sum_{l=1}^{m+2} b_l B^2_{l,h_l}(y)$$

with $b_l \geq 0$, $\sum_{l=1}^{m+2} b_l = 1$, $B^2_{l,h_l}(y) = \dfrac{B^2_l(y)}{h_l}$ and $h_l = \int B^2_l(y)dy$.

Although the mixture model is widely used [14], it is not considered to be a complete model in practice because it only describes the data statistically and no spatial information about the data is utilized. Under certain intensity distributions, we want the model to be "adaptive" to structural information or spatially dependent in order to fit the actual image better. This leads to the consideration of Hidden Markov model (HMM).

## 3 Hidden Markov model

The spatial property can be modeled through different aspects, amongst which the contextual constraint is a general and powerful one. HMM theory provides a convenient and consistent way to model context-dependent entities such as image pixels and correlated features.

Let $\mathbb{K} = \{1,2,\ldots,K\}$, $\mathbb{D} = \{1,2,\ldots,D\}$.

Let $S = \{1, 2, \ldots, N\}$ be the set of indexes and $R = \{r_i, i \in S\}$ denotes any family of random variables indexed by $S$, in which each random variable $R_i$ takes a value $r_i$ in its state space. Such a family $r$ is called a random field. The joint event $(R_i = r_i, \ldots, R_N = r_N)$ is simplified to $R = r$ where $r = \{r_1, \ldots, r_N\}$ is a configuration of $R$, corresponding to a realization of this random field. Let $X$ and $Y$ be two such random fields whose state spaces are $K$ and $D$, respectively, so that for $\forall i \in S$ we have $X_i \in \mathbb{K}$ and $Y_i \in \mathbb{D}$. Let $x$ denote a configuration of $X$ and $\mathscr{X}$ be the set of all possible configurations so that

$$\mathscr{X} = \{x = (x_1, \ldots, x_N) \mid x_i \in \mathbb{K}, i \in S\}.$$

The state of $\mathscr{X}$ is unobservable.

Similarly, let $y$ be a configuration of $Y$ and $\mathscr{Y}$ be the set of all possible configurations so that

$$\mathscr{Y} = \{y = (y_1, \ldots, y_N) \mid y_i \in \mathbb{D}, i \in S\}.$$

Let $b = \{b_{.,k}/k \in \mathbb{K}\}$ and let $b_{.,k} = (b_{1,k}, \ldots, b_{m+2,k}),\ k \in \mathbb{K}$.

Given $X_i = k$, $Y_i$ follows a conditional probability distribution

$$p(y_i \mid k) = f(y_i; b_{.,k}) = \sum_{l=1}^{m+2} b_{l,k} B_{l,h_l}^2(y_i), \forall k \in K;\ l = 1, \ldots, m+2 \tag{3}$$

where $b_{.,k}$ is the set of parameters. For all $l$, the function family $f(.; b_{.,k})$ has the same known analytic form. We also have that

$$\begin{aligned} p(x \mid y) \propto &\quad p(y \mid x)p(x) \\ &\propto p(x) \prod_{i \in S} p(y_i \mid x_i) \end{aligned} \tag{4}$$

In an HMM, the sites in $S$ are related to one another via a neighborhood system, which is defined as $N = \{N_i, i \in S\}$, where $N_i$ is the set of sites neighboring $i$, $i \notin N_i$ and $i \in N_j \iff j \in N_i$. A random field $X$ is said to be an HMM on $S$ with respect to a neighborhood system $N$ if and only if

$$p(x) > 0, \forall x \in \mathscr{X}$$

$$p(x_i \mid x_{S|\{i\}}) = p(x_i \mid x_{N_i}).$$

Note that the neighborhood system can be multidimensional. According to the Hammersley-Clifford theorem [8], an HMM can equivalently be characterized by a Gibbs distribution. Thus

$$p(x) = Z^{-1} \exp(-U(x)) \tag{5}$$

where $Z$ is a normalizing constant called the partition function, and $U(x)$ is an energy function of the form

$$U(x) = \sum_{c \in C} V_c(x) \tag{6}$$

which is a sum of clique potentials $V_c(x)$ over all possible cliques $C$. A clique is defined as a subset of sites in $S$ in which all the pairs of distinct sites are neighbors, except the single-site cliques. The value of $V_c(x)$ depends on the local configuration of clique $c$. For more details on HMM and Gibbs distribution, see [13].

## 4 Parameters estimation using the EMMB algorithm

For any $x \in \mathscr{X}$, the realizations $y_1, \ldots, y_N$ are conditionally independent:

$$p(y \mid x) = \prod_{i \in S} p(y_i \mid x_i). \tag{7}$$

The image classification problem we consider involves assigning to each pixel a class label taking a value from the set $\mathbb{K}$. Each pixel is characterized by an intensity value $y_i$ from the set $\mathbb{D}$. A labeling of $S$ is denoted by $x$, where $x_i$, $i \in S$ is the corresponding class label of pixel $i$. We write $x^*$ for the true but unknown labeling configuration and $\hat{x}$ for an estimate of $x^*$, both of which are interpreted as particular realizations of a random field $X$, which is an HMM with a specified distribution $p(x)$. The observable image itself is denoted by $y$. The problem of classification is the problem of recovering $x^*$, given the observed image $y$.

We seek a labeling $\hat{x}$ of an image, which is an estimate of the true labeling $x^*$, according to the MAP criterion

$$\hat{x} = \arg\max_{x \in \mathscr{X}} \{p(y \mid x) p(x)\}. \tag{8}$$

From (8), we need to compute the prior probability of the class and the likelihood probability of the observation. Since $x$ is considered as a realization of an HMM, its prior probability can be derived from

$$p(x) = Z^{-1} \exp(-U(x)) \tag{9}$$

It is also assumed that the pixel intensity $y_i$ follows a B-spline distribution with parameters $b_{l,x_i}$, given the class label $x_i$

$$p(y_i \mid x_i) = \sum_{l=1}^{m+2} b_{l,x_i} B^2_{l,h_l}(y_i) \tag{10}$$

Based on the conditional independence assumption of $y$ and according to (7), the likelihood function is given by

$$p(y \mid x) = \prod_{i \in S} p(y_i \mid x_i)$$
$$= \prod_{i \in S} (\sum_{l=1}^{m+2} b_{l,x_i} B^2_{l,h_l}(y_i))$$

which can be written as

$$p(y \mid x) = \tilde{Z}^{-1} \exp(-U(y \mid x)) = \tilde{Z}^{-1} \exp(-\sum_{i \in S} U(y_i \mid x_i)) \tag{11}$$

where $U(y_i \mid x_i) = -\log(\sum_{l=1}^{m+2} b_{l,x_i} B^2_{l,h_l}(y_i))$ and

$$U(y \mid x) = \sum_{i \in S} U(y_i \mid x_i)$$
$$= -\sum_{i \in S} \log(\sum_{l=1}^{m+2} b_{l,x_i} B^2_{l,h_l}(y_i)) \tag{12}$$

and the constant normalization term $\tilde{Z} = 1$. It is easy to show that

$$\log p(x \mid y) \propto -U(x \mid y), \tag{13}$$

where

$$U(x \mid y) = U(y \mid x) + U(x) + \text{const} \tag{14}$$

is the posterior energy. The MAP estimation is equivalent to minimizing the posterior energy function

$$\hat{x} = \arg\min_{x \in \mathscr{X}} \{U(y \mid x) + U(x)\} \tag{15}$$

The strategy underlying the EM algorithm [5, 7] consists of the following: estimate the missing part as $\hat{x}$, given the current $b$ estimate and then use it to form the complete data set $\{\hat{x}, y\}$; new $b$ can be estimated by maximizing the expectation of the complete-data Log likelihood. Mathematically, the EM algorithm can be described by the following.

Start an initial estimate $b^{(o)}$.

The E-step calculates the conditional expectation

$$Q = Q(b \mid\mid b^{(t)}) = \sum_{x \in \mathscr{X}} p^{(t)}(x \mid y, b^{(t)}) \log p(x, y \mid b). \tag{16}$$

where

$$log(p(x,y) = \sum_{i \in S}\{\log(\sum_{l=1}^{m+2} b^{(t)}_{l,x_i} B^2_{l,h_l}(y_i)) + \log(p^{(t)}(x_i))\},$$

$$p^{(t)}(x \mid y, b^{(t)}) = \prod_{i \in S} p^{(t)}(x_i \mid y_i, b^{(t)})$$

and

$$p^{(t)}(x_i \mid y_i, b^{(t)}) = \frac{p^{(t)}(y_i \mid x_i, b^{(t)})p^{(t)}(x_i)}{p(y_i)}$$
$$\propto \sum_{l=1}^{m+2}(b^{(t)}_{l,x_i} B^2_{l,h_l}(y_i))p^{(t)}(x_i).$$

The M-step maximizes $Q(b \mid\mid b^{(t)})$ to obtain the next estimate

$$b^{(t+1)} = \arg\max_b Q(b \mid\mid b^{(t)}). \tag{17}$$

Let $b^{(t+1)} \longrightarrow b^{(t)}$ and repeat from the E-step.

Under certain reasonable conditions, EM estimates converge locally to the maximum likelihood estimates. We denote by $p^{(t)}(k \mid y_i, b^{(t)})$ the locally dependent probability of $x_i = k$ and the parameters $b = \{b_{.,k}/k \in \mathbb{K}\}$.

By taking first and second derivative of $Q$ with respect to $b_{l,k}$ and $b^2_{l,k}$ we have

$$\frac{\partial Q}{\partial b_{l,k}} = \sum_{i \in S} p^{(t)}(k \mid y_i, b^{(t)})\left(\frac{B^2_{l,h_l}(y_i) - B^2_{k,h_l}(y_i)}{p(k, y_i \mid b)}\right) = 0$$

and

$$\frac{\partial^2 Q}{\partial b^2_{l,k}} = -\sum_{i \in S} p^{(t)}(k \mid y_i, b^{(t)})\left(\frac{B^2_{l,h_l}(y_i) - B^2_{k,h_l}(y_i)}{p(k, y_i \mid b)}\right)^2;$$

$l = 1, \ldots, m+2; k \in \mathbb{K}$.

Let

$$\frac{\partial Q}{\partial b_{l,k}} = F_l(b) \text{ and } \frac{\partial^2 Q}{\partial b^2_{l,k}} = F'_l(b);$$

$l = 1, \ldots, m+2; k \in \mathbb{K}$.

In order to solve $F_l(b) = 0$, we use the Newton Raphson method or the steepest descent method

$$F'_l(b^{(t)})(b^{(t+1)} - b^{(t)}) = -F_l(b^{(t)});$$

$l = 1, \ldots, m+2$.

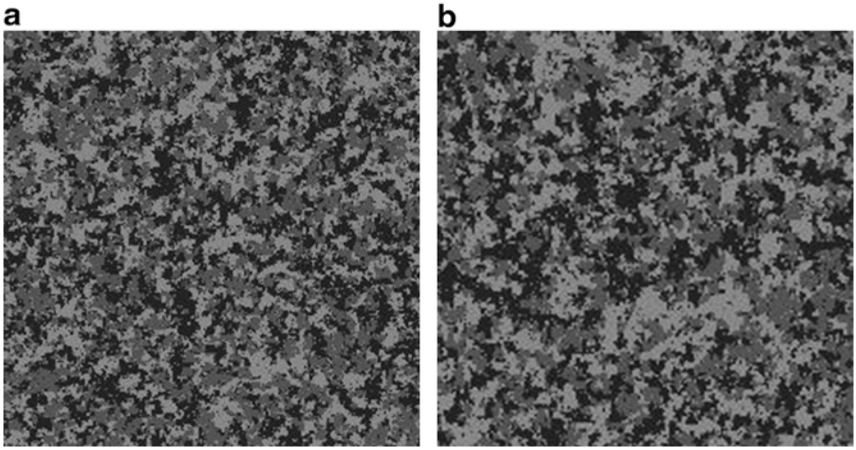


**Fig. 2** Image simulation by the Standard EM model and the EMMB model with 3-class case and 4-connected using Potts model.

Hence, we obtain

$$b_{l,k}^{(t+1)} = b_{l,k}^{(t)} + \frac{\displaystyle\sum_{i \in S} p^{(t)}(k \mid y_i, b^{(t)})\left(\frac{B_{l,h_l}^2(y_i) - B_{k,h_l}^2(y_i)}{p(k, y_i \mid b)}\right)}{\displaystyle\sum_{i \in S} p^{(t)}(k \mid y_i, b^{(t)})\left(\frac{B_{l,h_l}^2(y_i) - B_{k,h_l}^2(y_i)}{p(k, y_i \mid b)}\right)^2} \tag{18}$$

for $l = 1, \ldots, m+2, k \in \mathbb{K}$.

The calculation of the conditional $p^{(t)}(k \mid y_i, b^{(t)})$ involves the estimation of the class labels $x_i$, which are obtained through the estimation (8).

Our algorithm converges to the estimated density $\hat{f}(y_i) = \sum_{l=1}^{m+2} b_{l,\hat{x}_i} B_{l,h_l}^2(y_i)$ where $\hat{x} = (\hat{x}_1, \ldots, \hat{x}_N)$ denotes the set of class label of pixel $i$ (see Figs. 2 and 3).

## 5 Performance comparison

The EMMB algorithm presented in Section V not only provides an effective method for parameters estimation, but also it gives a complete framework for unsupervised classification using iterative updating.

Without prior information, histogram analysis is widely used out initial estimation using a discriminant measure based thresholding method proposed by [12]. The basic idea is to find thresholds maximizing the interclass variances. According to theories of discriminant analysis, such thresholds are optimal solutions. Once

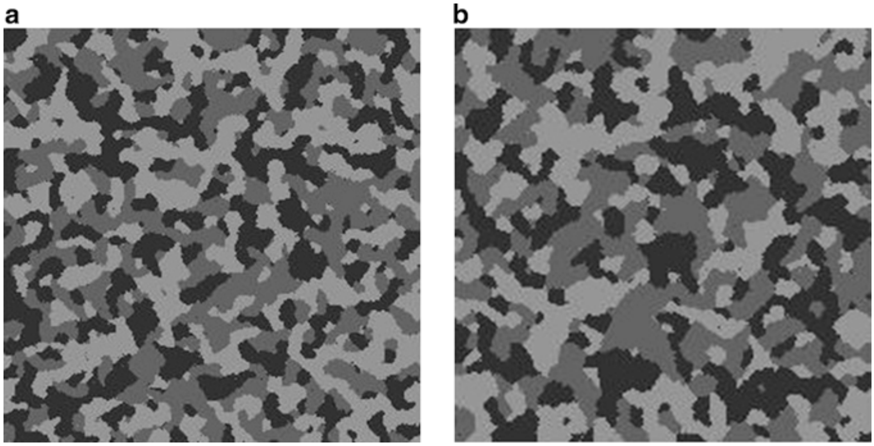

**Fig. 3** Image simulation by the Standard EM model and the EMMB model with 3-class case and 8-connected using Potts model.

the optimal thresholds have been determined, the initial classification can also be obtained either directly through the thresholding, or through an ML estimation with those known parameters.

We illustrate the performance of EMMB segmentation with different examples. First, we show a comparison between the standard EM method and our EMMB method for segmenting and parameter estimating. Second, we calculate the Mean Squared Errors (MSE) which is defined by $\text{MSE} = \frac{1}{n}\sum_{j=1}^{n}(f(y_j) - \hat{f}(y_j))^2$ and the Kullback-Leibler divergence (KL) which is defined by $\text{KL} = \frac{1}{n}\sum_{j=1}^{n} f(y_j)\log(\frac{f(y_j)}{\hat{f}(y_j)})$ of the estimated mixture density $\hat{f}$ from the true density $f$ for each method.

In what follows, we consider the following clique potential function defined by $V_c(x) = \begin{cases} -\delta, & \text{if } x_i = x_j \\ \delta, & \text{if } x_i \neq x_j \end{cases}$ where $\delta > 0$.

In Table 1, we have first computed the MSE and KL between the empirical distribution and the estimated mixture density by using the standard EM. Second,

**Table 1** MSE and KL of mixture distribution obtained by EM and EMMB.

| MSE | Image vegetable | img R | img G | img B |
|---|---|---|---|---|
| Standard EM | 0.2981 | 0.3124 | 0.4236 | 0.1344 |
| EMMB | 0.0406 | 0.0068 | 0.0497 | 0.0232 |
| KL | Image vegetable | img R | img G | img B |
| Standard EM | 0.329 | 0.125 | 0.286 | 0.147 |
| EMMB | 0.0258 | 0.0012 | 0.097 | 0.095 |

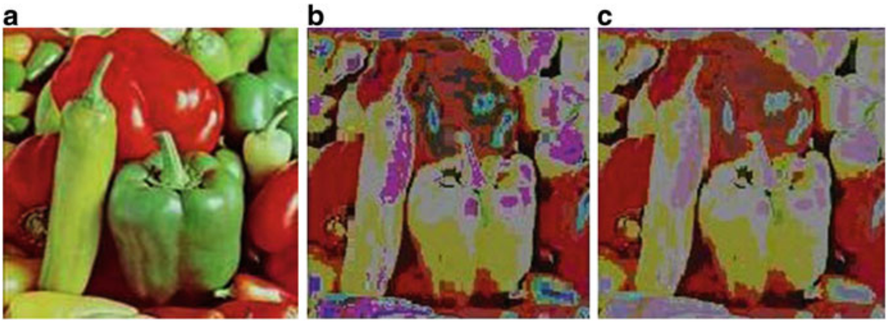


**Fig. 4** Image segmented with EMMB and standard EM.

we have computed the MSE and KL between the empirical distribution and the estimated mixture density by using the EMMB algorithm. This work was done for an image (vegetable) which will be described later. We notice that the EMMB method gives a much lower MSE and KL for both images.

In Fig 4, we consider a real image (vegetable) which is a photograph of $512 \times 512 \times 3$ pixels resolution and is represented by 256 color levels segmented with EMMB model and standard EM.

## 6 Conclusions

In this paper, we have introduced a new nonparametric B-splines using Hidden Markov model. Many results presented prove that the estimation of density by using the proposed estimator is better than those of other methods.

## References

1. Adell J. A, Sanguesa C., 2005. Approximation by B-splines Convolution operators. A probabilistic approach. *Computational and applied Mathematics*, Vol. 174, n$^o$1, pp. 79–99.
2. Bartels R, Beautty J, Barsky B., 1987. An Introduction to Splines for use in Computer Graphics and Geometric Modeling. *Morgan Kaufmann Publishers*, Lost Altos, CA.
3. Curry H.B, Schoenberg I.J., 1966. On poly freqency functions IV: The fundamental spline functions and their limits. *J. Anal. Math*, Vol. 17, pp. 71–107.
4. De Boor C., 1978. A practical guide to splines. *Springer*, New York.
5. Dempester A.P., Laird N.M., 1977. Rubin D.B. Maximum Likelihood from Incomplete Data via the E.M. Algorithm, *Journal of the Royal Statistical Society* , Ser. B, Vol. 39, pp. 1–38.
6. Gehringer, K.R. and Redner, R.A., 1992, Nonparametric density estimation using tensor product splines. *Communications in Statistics-Simulation and Computation*, Vol. 21, pp. 849–878.
7. Homer Walker and Richard Redner,1984. Mixture Densities, Maximum Likelihood and the EM Algorithm, *SIAM Rev*, Vol. 26, pp. 195–239.

8. J. Besag, "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *J. Roy. Stat. Soc.*, ser. B, vol. 36, no. 2, pp. 192–326, 1974.
9. Kelly P.A, Derin H, Hartt K.D., 1988. Adaptive segmentation of speckled images using a hierarchical random field model. *IEEE Tran. on Accoust, Speech and Signal Processing* Vol. 36, pp. 1628–1641.
10. Kim J. and al., 2005. A Nonparametric Statistical Method for image segmentation using information theory and curve evolution. *IEEE transactions on image processing*, Vol. 14, pp. 1486–1502.
11. Masri R, Redner R.A., 2005. Convergence rates for uniform B-splines density estimators on bounded and semi-infinite domains. *Nonparametric Statistics*, Vol. 17, pp. 555–582.
12. N. Otsu, "A threshold selection method from gray-level histogram," *IEEE Trans. Syst. Man Cybern.*, vol. 9, pp. 62–66, Jan. 1979.
13. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, no. PAMIŰ6, pp. 721–741, June 1984.
14. T. Lei and W. Sewchand, "Statistical approach to X-ray CT imaging and its applications in image analysis—Part II: A new stochastic modelbased image segmentation technique for X-ray CT image," *IEEE Trans. Med. Imag.*, vol. 11, no. 1, pp. 62–69, 1992.

# Iris Localization Using Mixture of Gamma Distributions in the Segmentation Process

**Fatma Mallouli, Atef Masmoudi, Afif Masmoudi, and Mohamed Abid**

**Abstract** This paper contributes to more accurate iris segmentation. We propose a new approach for iris image segmentation based on mixture of Gamma distributions modeling and an extended Expectation Maximization (EM) algorithm. We apply our approach to segment iris images from the CASIA (Chinese Academy of Sciences Institute of Automation)-Iris-Twins testing database. The accuracy of our algorithm is proved based on Kullback Leibler distance computation.

**Keywords** Iris • Segmentation • Mixture of Gamma distributions • EM algorithm • CASIA Iris images database

## 1 Introduction

Biometric identification is a process based on measurable biological and behavioral characteristic information that can be used for automated recognition. This process is preferred on traditional methods, while its information constitutes an important emphasis on security and cannot be stolen virtually. Iris recognition is a biometric authentication approach that applies pattern recognition procedure based on high resolution of iris images of a person's eye [5]. An iris recognition system is composed of various sub-systems: Picture taking process, segmentation of the picture, normalization, encoding and matching processes [3]. Every one of these steps is important to better identify subjects, but in an early step, iris segmentation plays the important role. In other words, a well-segmented picture would be an

F. Mallouli (✉) • M. Abid
Laboratory CES-Lab, National School of Engineering of Sfax, Sfax, Tunisia

A. Masmoudi
Sfax Preparatory Engineering Institute, University of Sfax, Route Menzel Chaker, BP 1172-3018 Sfax, Tunisia
e-mail: afif.masmoudi@fss.rnu.tn

A. Masmoudi
Laboratory of Probability and Statistics, Faculty of Sciences of Sfax University of Sfax. B. P. 1171, Sfax, Tunisia



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_12

excellent input to the steps that follow. In real life, images are captured at a distance and on the move which makes iris segmentation much more difficult due to the effects of: eye position and size variation, eyebrows, eyelashes, glasses and contact lenses, and hair; all along with changing illumination and focusing condition. Image segmentation is among the most important frequently addressed fundamental problems in image analysis and pattern recognition. It is the process of division of the image into homogenous regions. In the past decades, many segmentation methods for analyzing various images have been studied [1, 8, 10, 13]. Statistical modeling for analyzing various images has been reported [11]. The process of image segmentation, particularly, iris segmentation [15], is a complex step in irisrecognition systems. The level of complexity is well multiplied when trying to segment non-cooperative and noisy iris images [7, 11]. The segmentation process leads to: Localization of the iris, eye lid boundaries (lower and upper), skin region detection, eyebrow and eye lashes detection, and pupil identification [2]. Our paper deals with finding the region where the iris is located. In many studies [7, 11] authors eliminate the skin regions, the eyebrows, and most of miscellaneous regions. Then, the output is a binary image that could be a much better input to locating the iris than the raw image. This approach is very crucial to noisy images taken on the move or/and at a distance. After that, they apply segmentation algorithms to segment it into three regions where one of them contains the iris [7]. Many of these algorithms utilize probability models mainly finite mixture Gaussian distributions [6]. In the present paper we propose to model different regions of iris image by mixture of gamma distributions. The parameters of the proposed mixture model are estimated by using the Expectation-Maximization algorithm through maximum likelihood. Note that, we extend the EM algorithm by adding an update step to the M step, in order to better estimate the parameters of our model. To confirm the efficiency and robustness of the proposed method, we apply our gamma modeling to the CASIA Iris images database. We compare our results to classical Gaussian approach. The outline of this paper is as follows: In the second section, we present the mixture of gamma model and the EM algorithm. In the third section, we apply our extended EM algorithm to segment the iris image into three regions. Finally we discuss results and we conclude.

## 2 Mixture of Gamma Distributions and EM Algorithm

### *2.1 Mixture of Gamma distributions*

The finite mixture of gamma distributions has been used as the statistical modeling of a continuous feature space. The feature space can be modeled as a finite mixture of gamma distributions with a known number $K$ of components. Let $X = x_1, \ldots, x_N$ be a set of observable sample drawn independently according to the density mixture

$f(x)$ and $x_i$ denotes the observation at the $i^{th}$ pixel of an image modeled as a mixture of Gamma distributions. The mixture model with $K$ components is given by:

$$f(x) = \sum_{k=1}^{K} \pi_k \gamma(x/a_k, b_k) \tag{1}$$

where $\pi_k$ is the mixing proportions such that $0 < \pi_k < 1$ and $\sum_{k=1}^{K} \pi_k = 1$

$$\gamma(x/a_k, b_k) = \frac{b_k^{a_k}}{\Gamma(a_k)} e^{-b_k(x)} x^{a_k - 1} I_{\mathbb{R}^+}(x)$$

denotes the density of Gamma distribution with shape parameter $b_k$ and with scale parameter $a_k$. The log likelihood function of the mixture of Gamma distributions for a parameter $\theta = (a_1, \ldots, a_k; b_1, \ldots, b_k; \pi_1, \ldots, \pi_k)$ is given by:

$$\begin{aligned} L(\theta) &= \sum_{i=1}^{N} Log[f(x_i)] \\ &= \sum_{i=1}^{N} Log[\sum_{k=1}^{K} \pi_k \gamma(x_i/a_k, b_k)] \end{aligned} \tag{2}$$

In this section, we will present our approach for estimating the parameters $(a_1, \ldots, a_k; b_1, \ldots, b_k; \pi_1, \ldots, \pi_k)$ by using our extension of the EM algorithm.

## 2.2 *Extended EM algorithm*

The EM algorithm is a general iterative technique for computing maximum-likelihood (ML) estimates when the observed data can be regarded as incomplete [4]. In maximum-likelihood estimation [9, 12], the unknown parameter $\theta$ is estimated by maximizing the log-likelihood function which is given by equation (2). The initial condition of the EM algorithm can be chosen using K-means method. The EM algorithm consists of an E-step and an M-step is proposed to estimate this problem. Suppose that $\theta^n$ denotes the estimation of obtained after the $n^{th}$ iteration of the algorithm. Then at the $(n+1)^{th}$ iteration, the E-step computes the expected complete data log-likelihood function:

$$Q(\theta, \theta^n) = \sum_{i=1}^{N} \sum_{k=1}^{K} \tau_k^n(x_i) Log[\pi_k \gamma(x_i/a_k, b_k)] \tag{3}$$

where $\tau_k^n(x_i)$ is a posterior probability and is computed as:

$$\tau_k^n(x_i) = \frac{\pi_k^n \gamma(x_i/a_k^n, b_k^n)}{\sum_{k=1}^{k} \pi_k^n \gamma(x_i/a_k^n, b_k^n)} \quad (4)$$

The M-step maximizes $Q(\theta, \theta^n)$ function with respect to $\theta$ in order to obtain the new parameter value $\theta^{n+1}$ using the following equation:

$$\theta^{n+1} = \arg\max_{\theta} Q(\theta, \theta^n) \quad (5)$$

After some calculations, the estimations of $a_k$, $b_k$, and $\pi_k$ are, respectively, given by $a_k^{n+1} = \psi_0^{-1}[\frac{\sum_{i=1}^{N} \tau_k^n(x_i) Log(x_i b_k^n)}{\sum_{i=1}^{N} \tau_k^n(x_i)}]$, $b_k^{n+1} = \frac{\sum_{i=1}^{N} a_k^n \tau_k^n(x_i)}{\sum_{i=1}^{N} x_i \tau_k^n(x_i)}$ and

$$\pi_k^{n+1} = \frac{1}{N} \sum_{i=1}^{N} \tau_k^n(x_i) \quad (6)$$

where $\psi_0^{-1}$ denotes the reciprocal function of the digamma function $\psi_0(x) = \frac{\Gamma(x)}{..?}$ and $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$ for all $x > 0$. Note that $\mu_k^{n+1}$ and $S_k^{n+1}$ represent, respectively, the estimation of the mean and the standard deviation parameters in the mixture of Gaussian distributions. The problem of using the EM algorithm in image segmentation lies in the difficulty to estimate the number of components in mixture. Many approaches, for example, the method used in [14] assumes that the number is known in advance, this means that, the number of segmentation region is determined in advance by the user. In our case the number of segmentation region is fixed to three: iris, papillary, and spectra (K=3). The segmentation is carried out by assigning each pixel into a proper class according to the Bayesian rule. After the mixture identification, the Bayesian rule is applied in order to classify the pixels according to their gray level $x$. Let $j(x)$ be the label of the class of the pixel $x$ and $\widehat{a_k}$, $\widehat{b_k}$, and $\widehat{\pi_k}$ are the estimated mixture parameters:

$$j(x) = \arg\max_{1 \le k \le 3} \pi_k \gamma(x_i/a_k^n, b_k^n) \quad (7)$$

# 3 Experimentations and Results

## 3.1 Data

CASIA Iris Image Database (CASIA-Iris) [2] has been released to the international biometrics community and updated from CASIA-IrisV1 to CASIA-IrisV4. Great progress of iris recognition has been achieved since 1990s. However, iris images captured at a distance are more challenging than traditional close-up iris images. Most current iris recognition methods have been typically evaluated on medium sized iris image databases with a few hundreds of subjects. However, more and more large-scale iris recognition systems are deployed in real-world applications. Many new problems are met in classification and indexing of large-scale iris image databases. CASIA-Iris-Twins contains iris images of pairs of twins, which were collected using OKI's IRISPASS-h camera [2].

We took 100 iris images of the CASIA database, in order to segment them. We used the classical model and the Mixture of Gamma distributions for each eye image. Figure 1 shows a sample of the results composed of the original image, segmented image by usual approach and segmented image by our approach.

We notice from this figure that the segmented images using our approach are better than those segmented with the classical approach. In fact, the first are more close to the original eye images.

Based on Kullback distance, the performance of our method is evaluated using two criteria, the first one is the evaluation of the estimation method and the second is the evaluation of the segmentation results.

## 3.2 Segmentation method evaluation

As a criterion of similarity to evaluate the estimation methods, the Kullback Leibler divergence is computed between the usual empirical distribution and the estimated Gamma or Normal distribution for each eye image (Figure 2). The fact that one model is better than the other can be observed by the minimal Kullback Leibler distance or divergence. From Figure 2 we notice that for large number of eyes, the divergence between the empirical and the Gamma distributions is lower than the divergence between the empirical and the Normal distributions, which implies that our method performs better than the usual normal method. On the other hand, and in order to evaluate the segmentation process, we calculated Kullback Leibler distance between each two estimated density regions, after that we sum the three different distances for each segmented eye image. This sum is calculated for the classical method and for our approach. The better is the segmentation, the higher is the distance between the classes, and therefore, the best method gives the larger distance value. The graphical representation of the sum of three class

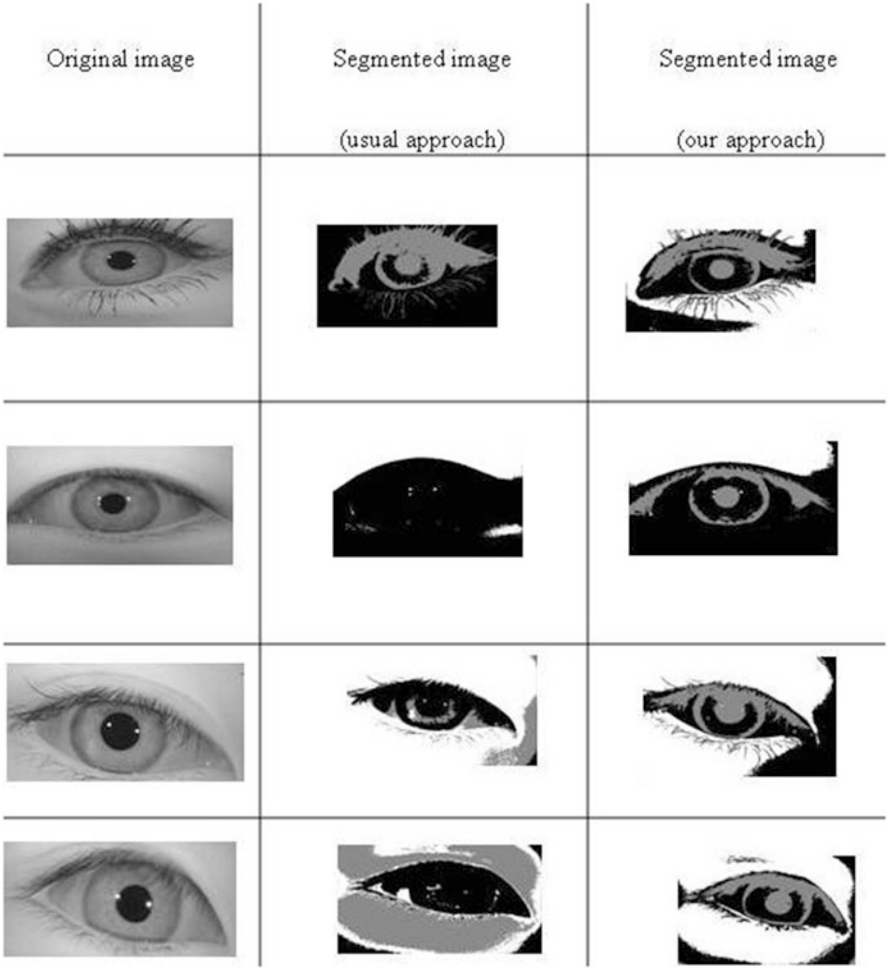


**Fig. 1** Original and segmented eye images by usual and our approaches

distances for every individual eye, estimated by Gamma and Normal approaches is shown in Figure 3.

This figure shows a good performance of our algorithm in iris image processing, since our approach gives higher sum of distances for most of the eyes tested, which implies a better segmentation given by the adopted mixture of the Gamma model.

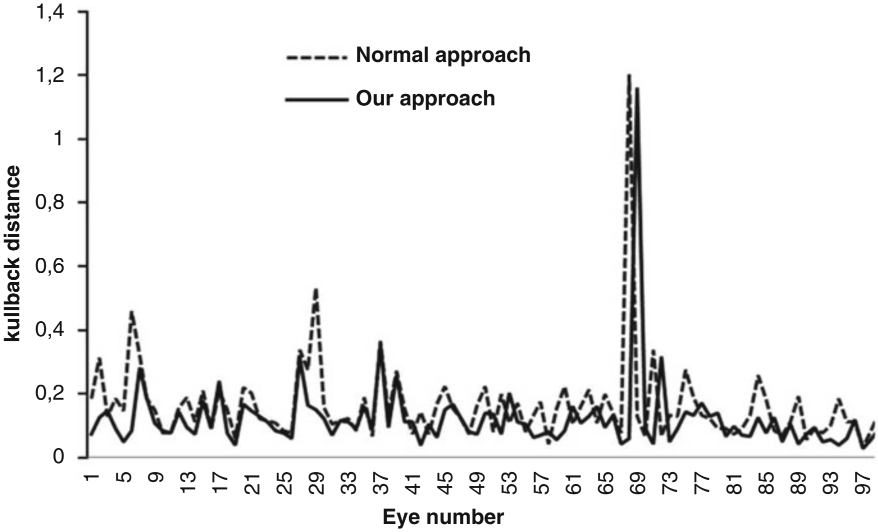


**Fig. 2** Kullback distances between Gamma and empirical Gaussian estimated distributions

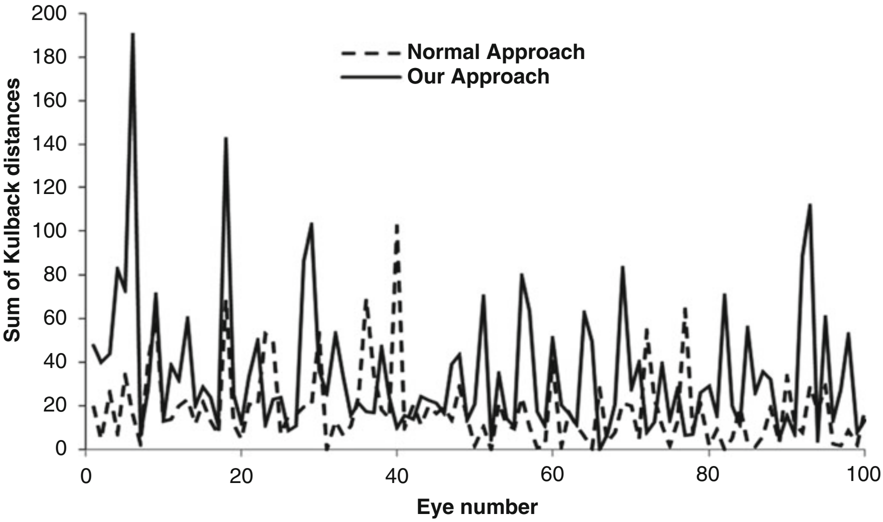


**Fig. 3** Graphical representation of class distances between Gamma and Gaussian approaches

## 4 Conclusion

Mixture of distributions has provided a mathematical based approach to the statistical modeling of a wide variety of random phenomena. These models are required in many applications in particular in automated recognition. In this paper, we considered eye image segmentation process. We apply mixture of Gamma distributions and we extend the EM algorithm to improve the segmentation process. Iris images from CASIA database were segmented based on two methods (usual and our method). The efficiency of our approach is proven by comparison to the reference method segmentation.

## References


1. K.W.Bowyer, and K.Hollingsworthand Flynn P.J., *Image understanding for iris biometrics: a survey, COMPUT. VIS. IMAGE UND.* 110 (2) 281–307, 2008. University of Notre Dame Notre Dame, Indiana 46556
2. CASIA-MS PalmprintV1, http://biometrics.idealtest.org/
3. Daugman, J. G. High confidence visual recognition of persons by a test of statistical independence, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25 (11) 1148–1161, 1993.
4. Dempster A., Laird, N., and Rubin, D. Maximum Likelihood From Incomplete Data via the EM Algorithm, *J. Royal Statistical Soc.* 1–38, 1977.
5. Gaganpreet K., Akshay, G. and Manvjeet, K. Enhanced Iris Recognition System - an Integrated Approach to Person Identification. *Int. J. Comput. Appl.* 8(1) 1–5, 2010.
6. Kim, J. Cho, S. and Choi, J. Iris recognition using wavelet features. *J. VLSI Signal Processing.* 38, 147–256, 2004.
7. Peihua L., Xiaomin L., Lijuan X. and Qi, S. Robust and accurate iris segmentation in very noisy iris images. *Image Vision Comput.* 28 (2) 246–253, 2010.
8. Proença, H. and Alexandre, L.A. Iris segmentation methodology for noncooperative recognition, *IEEE Proceedings of the Vision, Image and Signal Processing* 153 (2) 199–205, 2006.
9. Redner, R. A., Walker, H. F. Mixture densities, maximum likelihood and the EM algorithm, *Soc. Ind. Appl. Math.* 26, 195–239, 1984.
10. Singh, S. and Singh, K. Segmentation Techniques for Iris Recognition. *Int. J. Sci. Eng. Res.*, 2 (4) 163–170, 2011.
11. Tieniu T., Zhaofeng, H., and Zhenan, S. Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. *Image Vision Comput.* 28, 223–230, 2010.
12. Ueda, N. and Nakano, R. Mixture density estimation via EM algorithm with deterministic annealing. *Neural Networks for Signal Processing IV*, 69–77, 1994.
13. Wildes, R. Iris recognition: an emerging biometric technology. Proceedings of the IEEE (85) 1348–1363, 1997.
14. Yamazaki, T. Introduction of EM Algorithm into Color Image Segmentation. *Proceedings of ICIRS'98*, 368–371, 1998.
15. Zhaofeng, H. Tieniu, T. Zhenan, S. Qiu, X.C. Towards accurate and fast iris segmentation for iris biometrics, IEEE Trans. Pattern Anal. *Mach. Intell.* 31(9) 1670–1684, 2009.

# Gamma stopping and drifted stable processes

**Mahdi Louati, Afif Masmoudi, and Farouk Mselmi**

**Abstract** Let $Y = (Y(t))_{t\geq 0}$ be a Lévy process on the real line and $T$ be a Gamma random variable independent from $Y$. We proved that, for all $p > 1$, $Y(T)$ and $Y(T)/T^p$ are independent if, and only if, $Y$ is stable with parameter $1/p$. This represents an extension of the result given by Letac and Seshadri [4] which represents the case where $T$ is an exponential random variable.

**Keywords** Cumulant function • infinitely divisible process • Lévy process

## 1 Introduction

In probability theory, a stable process is a type of stochastic process. It includes stochastic processes whose associated probability distributions are stable distributions. This is why, many research have been devoted to this class of stochastic processes. References [7] and [5] studied this class by characterizing it by its Fourier transform and Laplace transform. Combining this class with the class of Lévy processes which represent a very important class of stochastic processes, we get many important results by using the formula of Lévy-Khinchine. A major emphasis was put on Lévy processes in many applied fields such as the theory of financial mathematics. Hence, several papers and books have been devoted to these processes in the past few years. Some of these works have dealt with the relationship between Lévy processes and infinitely divisible distributions (see [6] and [1]). First, we consider a Lévy process $Y = (Y(t))_{t\geq 0}$ defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with a value on the real line. Let $\mu_t$ be the distribution of $Y(t)$ and let $k(s)$ be the cumulant function of $\mu_1$, *i.e.,*

$$k(s) = \log \mathbb{E}\left(e^{sY(1)}\right)$$

M. Louati (✉) • A. Masmoudi • F. Mselmi
Faculty of Science, Department of Mathematics, University of Sfax, Sfax, Tunisia
e-mail: Mahdi.Louati@ipeis.rnu.tn



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_13

defined on $\Theta = int\{s \in \mathbb{R}; \mathbb{E}\left(e^{sY(1)}\right) < +\infty\} \neq \emptyset$. So, the cumulant function of $\mu_t$ is $k_t(s) = tk(s)$. An interesting paper of [4] has characterized a Lévy process $Y = (Y(t))_{t\geq 0}$ by using a mixture of $Y$ and an exponential standard random variable $T$ independent from $Y$ in order to get the cumulant function and the stability of $Y(t)$.

The idea of this work is based on the results given by [4] in order to extend these results by replacing the exponential distribution by a Gamma distribution and find a new characterization and find the stability of the class of Lévy processes on the real line.

The remaining part of this paper is structured as follows. After recalling some preliminary results in section 2, we establish our main results in section 3. In section 4, we give a conclusion.

To make clear the results of this paper, we need to introduce some notations that will be used in this paper.

## 2 Preliminary

### 2.1 Natural exponential family

To make clear our introduction, we need to introduce some basic notations. As a reference of these notations, we have [3]. Let $\mu$ be a positive random measure on the real line, we denote by

$$L_\mu(s) = \int_{\mathbb{R}} \exp(sx)\mu(dx) < \infty$$

its Laplace transform and

$$\Theta(\mu) = int\{s \in \mathbb{R};\ L_\mu(s) < \infty\}.$$

Let $\mu$ be a probability measure such that $\Theta(\mu) \neq \emptyset$. We define the cumulant function of the measure $\mu$ by

$$k_\mu(s) = \ln\left(L_\mu(s)\right).$$

The set

$$\Lambda(\mu) = \{\lambda > 0;\ \exists\ \mu_\lambda \text{ such that } k_{\mu_\lambda}(s) = \lambda k_\mu(s), \forall s \in \Theta(\mu)\}$$

is called the Jørgensen set of $\mu$ and the measure $\mu_\lambda$ is its $\lambda-$ power of convolution. The set $\Lambda(\mu)$ is equal to $]0, +\infty[$ if and only if $\mu$ is infinitely divisible (see [7], p. 155). For $\mu$ a positive measure on the real line, the probability set

$$F = F(\mu) = \{P(s,\mu)(dx) = e^{sx-k_\mu(s)}\mu(dx); s \in \Theta(\mu)\}$$

is called the natural exponential family (NEF) generated by $\mu$.

## *2.2 Lévy process*

Let $(Y(t))_{t\geq 0}$ be a Lévy process on the real line. An important result which characterizes Lévy processes is the Lévy-Khinchine formula expressed by the cumulant function, for all $s \leq 0$,

$$k_{Y(t)}(s) = t\left[\frac{1}{2}\sigma^2 s^2 + bs + \int_{\mathbb{R}\backslash\{0\}} (e^{sx} - 1 + s\tau(x))\nu(dx)\right],$$

where $b, \sigma \in \mathbb{R}$, the measure $\nu$ satisfies the condition $\int_{\mathbb{R}\backslash\{0\}} \min(1, x^2)\nu(dx) < \infty$ and $\tau$ is a some fixed bounded continuous function such that $\lim_{x\to 0} \frac{\tau(x)-x}{x^2} < \infty$ and nonzero. The measure $\nu$ is called Lévy measure. If $\nu(]0, +\infty[) = 0$, then we said that $\nu$ is spectrally negative Lévy measure and it is said spectrally positive, if $\nu(]-\infty, 0[) = 0$. If $\nu$ is bounded, the semigroup $(\mu_t)_{t\geq 0}$ is said to be of type 0. If $\nu$ is unbounded but such that $\int_{\mathbb{R}\backslash\{0\}} \min(1, |x|)\nu(dx) < +\infty$, the semigroup $(\mu_t)_{t\geq 0}$ is said to be of type 1. It is said to be of type 2 if $\int_{\mathbb{R}\backslash\{0\}} \min(1, |x|)\nu(dx)$ diverges. The function $\tau$ is not useful for types 0 or 1 and we can write in this case

$$k_{Y(t)}(s) = t\left[\frac{1}{2}\sigma^2 s^2 + bs + \int_{\mathbb{R}\backslash\{0\}} (e^{sx} - 1)\nu(dx)\right]. \tag{1}$$

## *2.3 Stable process*

Let $\alpha \in ]0, 2]$, a random variable $X$ on $\mathbb{R}$ is $\alpha$-stable in the broad sense if for each $n \geq 2$, there exist $f_n \in \mathbb{R}$ and $n$ random variables $X_1, X_2, \ldots, X_n$ i.i.d such that

$$X_1 + X_2 + \ldots + X_n = n^{1/\alpha}X + f_n. \tag{2}$$

Furthermore $X$ is strictly stable if (2) holds with $f_n = 0$ for all $n \geq 2$.

Let $p > 1$. If a random variable $X$ is $\frac{1}{p}$-stable on the real line, then its Laplace transform is given as follows, for all $s \leq 0$

$$L_X(s) = e^{-(-s)^{1/p}}.$$

Using a result given by [2], we get that there exist $t, \beta > 0$ such that a stable process $(X(t))_{t\geq 0}$ has the following Laplace transform, for all $s \leq \frac{1}{\beta}$

$$L_{X(t)}(s) = e^{t(1-(1-\beta s)^{1/p})}.$$

## 3 Main Results

In this section, we study a generalization of [4] on the real line of a Lévy process $Y = (Y(t))_{t\geq 0}$ on $\mathbb{R}$. For this reason, we denote by $\mu_t$ the distribution of $Y(t)$. Without loss of generality, we suppose that $0 \in \Theta(\mu_1)$ and $\mu_1$ is non-Dirac distribution. Then, we get these following theorems.

**Theorem 3.1.** *Let $Y = (Y(t))_{t\geq 0}$ be a non-negative, non-Dirac real Lévy process. Let $T$ be a random variable with Gamma $\gamma(n, b)$ distribution independent from $Y$. Assume that for $p > 1$, $Y(T)/T^p$ and $Y(T)$ are independent. Then*

*i)* $\mathbb{P}(Y(t) = 0) = 0$.

*ii)* *For all* $s < \frac{1}{\beta}$

$$\mathbb{E}(e^{sY(t)}) = \exp\left(\frac{t}{b}\left(1 - (1-\beta s)^{1/p}\right)\right), \tag{3}$$

*where* $\frac{1}{\beta} = \frac{n\Gamma(n)}{pb^p\Gamma(n+p)}\mathbb{E}\left(\frac{T^p}{Y(T)}\right)$.

*iii)* *For all* $s < \frac{1}{\beta}$

$$\mathbb{E}(e^{sY(T)}) = (1-\beta s)^{-\frac{n}{p}}. \tag{4}$$

*iv)* *For all* $a > 0$

$$\mathbb{E}\left(\left(\frac{T^p}{Y(T)}\right)^a\right) = \frac{\Gamma(n/p)\Gamma(pa+n)b^{pa}}{\Gamma(n/p+a)\Gamma(n)\beta^a}. \tag{5}$$

*Proof.* *i)* Since $Y$ is a Lévy process, then as a consequence we get for all $s \leq 0$,

$$L_{\mu_t}(s) = \int_0^{+\infty} e^{sy}\mu_t(dy) = e^{tk_{\mu_1}(s)}.$$

For all $t > 0$, denote $p_t = \mathbb{P}(Y(t) = 0)$, then

$$p_t = \lim_{s\to-\infty}\int_0^{+\infty} e^{sy}\mu_t(dy) = \mu_t(0) = p_1^t.$$

Assume that $p_t > 0$, for $t > 0$, then

$$\begin{aligned} q &= \mathbb{P}(Y(T) = 0) \\ &= \int_0^{+\infty} \frac{e^{-\frac{t}{b}} t^{n-1}}{\Gamma(n) b^n} \mathbb{P}(Y(t) = 0) dt \\ &= \int_0^{+\infty} \frac{e^{-\frac{t}{b}} t^{n-1}}{\Gamma(n) b^n} p_1^t dt \\ &= \int_0^{+\infty} \frac{e^{-\frac{t}{b}(1-b\ln p_1)} t^{n-1}}{\Gamma(n) b^n} p_1^t dt = (1 - b \ln p_1)^{-n} > 0. \end{aligned}$$

Since $Y(T)/T^p$ and $Y(T)$ are independent, hence

$$q = \mathbb{P}\left(Y(T) = 0; Y(T)/T^p = 0\right) = \mathbb{P}\left(Y(T) = 0\right)^2 = q^2.$$

Hence $q = 1$, this contradicts that $Y$ is non-Dirac. Then for all $t \geq 0$, $\mathbb{P}(Y(t) = 0) = 0$.

*ii)* There exists $\lambda > 0$ such that

$$\mathbb{E}\left(\frac{T^p}{Y(T)}\right) = \mathbb{E}\left(\frac{T^p}{Y(T)}\Big| Y(T)\right) = \lambda.$$

This implies that

$$\mathbb{E}\left(T^p \Big| Y(T)\right) = \lambda Y(T).$$

Using this, then we have

$$\int_0^{+\infty} t^p \frac{e^{-\frac{t}{b}} t^{n-1}}{\Gamma(n) b^n} \mu_t(y) dt = \int_0^{+\infty} \lambda y \frac{e^{-\frac{t}{b}} t^{n-1}}{\Gamma(n) b^n} \mu_t(y) dt \tag{6}$$

Then, for $s < 0$, we multiply both sides of (1) by $e^{sy}$ and integrate with respect to $y$, we get

$$\int_0^{+\infty} \int_0^{+\infty} e^{sy} t^p \frac{e^{-\frac{t}{b}} t^{n-1}}{\Gamma(n) b^n} \mu_t(dy) dt = \lambda \int_0^{+\infty} \int_0^{+\infty} e^{sy} y \frac{e^{-\frac{t}{b}} t^{n-1}}{\Gamma(n) b^n} \mu_t(dy) dt.$$

Using the property of infinite divisibility of $Y$, we get

$$\int_0^{+\infty} e^{tk_{\mu_1}(s)} \frac{e^{-\frac{t}{b}} t^{p+n-1}}{\Gamma(n) b^n} dt = \lambda \frac{\partial}{\partial s} \left( \int_0^{+\infty} e^{tk_{\mu_1}(s)} \frac{e^{-\frac{t}{b}} t^{n-1}}{\Gamma(n) b^n} dt \right).$$

Thus

$$\int_0^{+\infty} \frac{e^{-\frac{t}{b}(1-bk_{\mu_1}(s))}t^{p+n-1}}{\Gamma(n)b^n}dt = \lambda\frac{\partial}{\partial s}\left(\int_0^{+\infty} \frac{e^{-\frac{t}{b}(1-bk_{\mu_1}(s))}t^{n-1}}{\Gamma(n)b^n}dt\right).$$

It follows that

$$(1-bk_{\mu_1}(s))^{-p-n}\frac{b^p\Gamma(n+p)}{\Gamma(n)} = \lambda\frac{\partial}{\partial s}\left((1-bk_{\mu_1}(s))^{-n}\right).$$

This implies that

$$(1-bk_{\mu_1}(s))^{p-1}bk'_{\mu_1}(s) = -\frac{b^p\Gamma(n+p)}{\lambda n\Gamma(n)}.$$

Then we get

$$k_{\mu_1}(s) = \frac{1}{b}(1-(1-\beta s)^{1/p}),$$

where $\frac{1}{\beta} = \frac{n\Gamma(n)}{pb^p\Gamma(n+p)}\mathbb{E}\left(\frac{T^p}{Y(T)}\right)$.

*iii)* Using the result (3), we have

$$\begin{aligned}\mathbb{E}\left(e^{sY(T)}\right) &= \int_0^{+\infty}\frac{e^{-\frac{t}{b}}t^{n-1}}{\Gamma(n)b^n}\mathbb{E}\left(e^{sY(t)}\right)dt\\ &= \int_0^{+\infty}\frac{e^{-\frac{t}{b}}t^{n-1}}{\Gamma(n)b^n}e^{\frac{t}{b}(1-(1-\beta s)^{1/p})}dt\\ &= \int_0^{+\infty}\frac{t^{n-1}}{\Gamma(n)b^n}e^{-\frac{t}{b}(1-\beta s)^{1/p}}dt\\ &= (1-\beta s)^{-n/p}.\end{aligned}$$

Then the distribution of $Y(T)$ is Gamma $\gamma(n/p, \beta)$.

*iv)* Since $Y(T)/T^p$ and $Y(T)$ are independent, then

$$\mathbb{E}\left(\left(\frac{T^p}{Y(T)}\right)^a\right) = \mathbb{E}\left(\frac{T^{pa}}{Y(T)^a}\right)\frac{\mathbb{E}\left(Y(T)^a\right)}{\mathbb{E}\left(Y(T)^a\right)} = \frac{\mathbb{E}\left(T^{pa}\right)}{\mathbb{E}\left(Y(T)^a\right)}.$$

Using the fact that $T \sim \gamma(n, b)$ and $Y(T) \sim \gamma(n/p, \beta)$, then we get

$$\mathbb{E}\left(\left(\frac{T^p}{Y(T)}\right)^a\right) = \frac{\mathbb{E}\left(T^{pa}\right)}{\mathbb{E}\left(Y(T)^a\right)} = \frac{\Gamma(n/p)\Gamma(pa+n)b^{pa}}{\Gamma(n/p+a)\Gamma(n)\beta^a}.$$

**Theorem 3.2.** *Let $Y = (Y(t))_{t\geq 0}$ be a non-negative, non-Dirac real Lévy process. Let $T$ be a random variable with Gamma $\gamma(n, b)$ distribution independent from $Y$. Then for $p > 1$, $Y(T)/T^p$ and $Y(T)$ are independent if, and only if, $Y$ is stable and has the following Laplace transform, for all $s \leq \frac{1}{\beta}$*

$$L_{Y(t)}(s) = e^{\frac{t}{b}(1-(1-\beta s)^{1/p})}.$$

*Proof.* Assume that $Y$ is stable and has the following Laplace transform, for all $s \leq \frac{1}{\beta}$

$$L_{Y(t)}(s) = e^{\frac{t}{b}(1-(1-\beta s)^{1/p})}.$$

As a consequence, we get (4) and (5). On the other hand,

$$\mathbb{E}\left(\left(\frac{T^p}{Y(T)}\right)^a e^{sY(T)}\right) = \int_0^{+\infty}\int_0^{+\infty} \frac{t^{pa}}{y^a} e^{sy} \frac{e^{-\frac{t}{b}} t^{n-1}}{b^n \Gamma(n)} \mu_t(dy)dt$$

Using the fact that $y^{-a} = \frac{1}{\Gamma(a)}\int_0^{+\infty} e^{-yv} v^{a-1} dv$, then we get

$$\begin{aligned}
\mathbb{E}\left(\left(\frac{T^p}{Y(T)}\right)^a e^{sY(T)}\right) &= \frac{1}{\Gamma(a)}\int_0^{+\infty}\int_0^{+\infty} v^{a-1}\frac{e^{-\frac{t}{b}} t^{pa+n-1}}{b^n\Gamma(n)}\left(\int_0^{+\infty} e^{(s-v)y}\mu_t(dy)\right)dtdv\\
&= \frac{1}{\Gamma(a)}\int_0^{+\infty}\int_0^{+\infty} v^{a-1}\frac{e^{-\frac{t}{b}(1-bk_{\mu_1}(s-v))} t^{pa+n-1}}{b^n\Gamma(n)}dtdv\\
&= \frac{1}{\Gamma(a)}\int_0^{+\infty} v^{a-1}\left(\int_0^{+\infty}\frac{e^{-\frac{t}{b}(1-bk_{\mu_1}(s-v))} t^{pa+n-1}}{b^n\Gamma(n)}dt\right)dv\\
&= \frac{\Gamma(pa+n)}{\Gamma(a)\Gamma(n)}\int_0^{+\infty} v^{a-1}(1-bk_{\mu_1}(s-v))^{-pa-n} b^{pa} dv\\
&= \frac{\Gamma(pa+n)b^{pa}}{\Gamma(a)\Gamma(n)}\int_0^{+\infty} v^{a-1}(1-\beta(s-v))^{-a-\frac{n}{p}} dv.
\end{aligned}$$

Assume that $u = \frac{\beta}{1-\beta s}v$, thus

$$\begin{aligned}
\mathbb{E}\left(\left(\frac{T^p}{Y(T)}\right)^a e^{sY(T)}\right) &= \frac{\Gamma(pa+n)b^{pa}}{\Gamma(a)\Gamma(n)}\frac{1}{(1-\beta s)^{\frac{n}{p}}\beta^a}\int_0^{+\infty}\frac{u^{a-1}}{(1+u)^{a+\frac{n}{p}}}du\\
&= \frac{\Gamma(n/p)\Gamma(pa+n)b^{pa}}{\Gamma(a+n/p)\Gamma(n)\beta^a}(1-\beta s)^{-\frac{n}{p}}\\
&= \mathbb{E}\left(\left(\frac{T^p}{Y(T)}\right)^a\right)\mathbb{E}\left(e^{sY(T)}\right).
\end{aligned}$$

This implies that $Y(T)/T^p$ and $Y(T)$ are independent.

*Remark 3.1.* Using the Lévy-Khinchine formula given in (1) with $\sigma = b = 0$ and spectrally positive Lévy measure $\nu$, we get that the Laplace transform $L_{Y(t)}(s) = e^{\frac{t}{b}(1-(1-\beta s)^{1/p})}$ has the following Lévy measure of type 1:

$$\nu(dx) = -\frac{1}{b}e^{-\frac{x}{\beta}}\frac{x^{-\frac{1}{p}-1}\beta^{1/p}}{\Gamma(1-\frac{1}{p})}1_{]0,+\infty[}(x)dx.$$

Consider now a Lévy process $X = (X_t)_{t\geq 0}$ governed by a convolution semigroup $(\mu_t)_{t\geq 0}$ which is spectrally negative, let us fix a level $x_0 > 0$ and consider the hitting time

$$Y(x_0) = \inf\{t \geq 0; X(t) = x_0\}$$

with the convention that $Y(x_0) = +\infty$ if this set of $t$ is empty. Because $X$ is spectrally negative, this is a consequence of the general theory of Lévy processes (see [1]) that $Y(x_0) < +\infty$ if and only if there exists $t$ such that $X(t) > x_0$. This occurs $\limsup_{t\to+\infty} X(t) = +\infty$ when almost surely.

**Theorem 3.3.** *Let $X = (X_t)_{t\geq 0}$ be a spectrally negative Lévy process such that $\limsup_{t\to+\infty} X(t) = +\infty$ a.s. Let $x_0$ be a realization of Gamma $\gamma(n,b)$ random variable $T$ independent from X. Define $Y(x_0) = \inf\{t \geq 0; X(t) = x_0\}$. Then there exists $p \in ]1,2]$ such that $Y(T)/T^p$ and $Y(T)$ are independent if and only if there exists $\beta > 0$ such that for $z \geq 0$, one has*

$$E(e^{zX(t)}) = e^{\frac{t}{\beta}((1+bz)^p-1)}.$$

*Proof.* Consider the Lévy process $Y = (Y(x_0))_{x_0>0}$. Then $T$ and $Y$ are independent. From Theorem 3.1, there exists a number $\beta > 0$ such that, for all $s \leq 0$,

$$L_{Y(x_0)}(s) = e^{\frac{x_0}{b}(1-(1-\beta s)^{1/p})}.$$

Since $X$ is a Lévy process, then, for all $z \geq 0$

$$L_{X(t)}(z) = e^{tk_{X(1)}(z)}.$$

Using the fact that $X$ is spectrally negative, then we get that for suitable $\sigma^2$, $b$, $\nu$ one has the Lévy-Khinchine formula

$$k_{X(1)}(z) = \frac{1}{2}\sigma^2 z^2 + bz + \int_{-\infty}^{0}(e^{zx} - 1 - z\tau(x))\nu(dx).$$

Note that $k'_{X(1)}(0) = b + \int_{-\infty}^{0} (x - \tau(x))\nu(dx) \in [-\infty, +\infty[$. Since $\limsup_{t\to+\infty} X(t) = +\infty$ a.s, then $\mathbb{E}(X(t)) = tk'_{X(1)}(0) \geq 0$. Using the fact that $k_{X(1)}(0) = 0$ and $k_{X(1)}$ is convex, thus $k_{X(1)}(z) > 0$, for all $z > 0$. For $z > 0$, consider the process $M_z = (M_z(t))_{t\geq 0}$ defined by $M_z(t) = e^{zX(t)-tk_{X(1)}(z)}$. With respect to the natural filtration of $X$, we get that $M_z$ is a martingale. For fixed $d > 0$, consider the regular stopping time $Y_d = \min(Y(x_0), d)$. Hence

$$\begin{aligned} 1 &= \mathbb{E}(M_z(Y_d)) = \mathbb{E}(M_z(Y_d)1_{\{Y(x_0)>d\}}) + \mathbb{E}(M_z(Y_d)1_{\{Y(x_0)\leq d\}}) \\ &= \mathbb{E}(e^{zX(d)-dk_{X(1)}(z)}1_{\{Y(x_0)>d\}}) + \mathbb{E}(e^{zx_0-x_0k_{X(1)}(z)}1_{\{Y(x_0)\leq d\}}). \end{aligned}$$

Since

$$0 \leq \mathbb{E}(e^{zX(d)-dk_{X(1)}(z)}1_{\{Y(x_0)>d\}}) \leq e^{zx_0-dk_{X(1)}(z)} \longrightarrow_{d\to+\infty} 0.$$

This implies that

$$1 = e^{zx_0}\mathbb{E}(e^{-Y(x_0)k_{X(1)}(z)}) = e^{zx_0}e^{\frac{x_0}{b}(1-(1+\beta k_{X(1)}(z))^{1/p})}.$$

Then, we get

$$k_{X(1)}(z) = \frac{1}{\beta}((1+bz)^p - 1).$$

## 4 Conclusion

The aim of this work was to explain the importance of the role of a Gamma random variable in terms of finding a characterization of a very important class of stochastic processes which is the class of Lévy processes. This was achieved by mixing a Lévy process with a Gamma random variable and using some independence conditions between this random matrix and the Lévy process. A fundamental property which has been used many times was the property of infinite divisibility Lévy processes which allowed us to get these results. Hence, the main result was to give the cumulant function of the Lévy process.

As a motivation of this work is to extend this on the multivariate case on the cone of symmetric matrices. This work has many important results because the natural extension of the Gamma distribution on the cone of symmetric matrices is the Wishart distribution. Moreover, it isn't an easy task to find a characterization and the stability of Lévy processes on the cone of symmetric matrices.

## References


1. Bertoin, J, (1996). Lévy Processes. *Cambridge University Press, Cambridge*.
2. Hougaard, P. (1986). Survival models for homogenous populations derived from stable distributions. *Biometrika*, 73, 387–396.
3. Letac, G. (1992). Lecture on natural exponential families and their variance functions. *Conselho Nacional de Desenvolvimento Cientifico e Tecnológico, Instituto de Matemàtica Pura e Aplicada*
4. Letac, G. and Seshadri, V. (2005). Exponential stopping and drifted stable processes. *Statist. Probab. Lett.*, 72, 137–143.
5. Samorodnitsky, G. and Taqqu, M., (1994). Stable non-Gaussian random processes.
6. Sato, K.I, (1999). Lévy Processes and infinitely divisible distributions. *Cambridge University Press, Cambridge*.
7. Seshadri, V. (1994). The Inverse Gaussian Distribution. A Case Study in Exponential Families. Oxford: Oxford University Press.

# On the Born-Oppenheimer asymptotic expansions

**Abderrahmane Senoussaoui**

**Abstract** We study the discrete spectrum of a general class of Born-Oppenheimer Hamiltonians of the type:

$$H = -h^2 \Delta_x + P\left(x, y, D_y\right) \text{ on } L^2\left(\mathbb{R}_x^n \times \mathbb{R}_y^p\right), n, p \in \mathbb{N}^*$$

when $h$ tends to $0^+$, here $P\left(x, y, D_y\right)$ is a pseudodifferential operator on $L^2\left(\mathbb{R}_y^p\right)$. In the case where the first eigenvalue $\lambda_1(x)$ of $P\left(x, y, D_y\right)$ on $L^2\left(\mathbb{R}_y^p\right)$ admits one nondegenerate point well, we obtain WKB-type expansions for all order in $h^{1/2}$ of eigenvalues (in the interval $[0, C_0 h]$, $C_0 > 0$) and associated normalized eigenfunctions of $H$, and this for all orders in $h^{1/2}$.

## 1 Introduction

The Born-Oppenheimer approximation is a method introduced in [2] to analyze the spectrum of molecules. It consists in studying the behavior of the associate Hamiltonian when the nuclear mass tends to infinity. This Hamiltonian can be written in the form:

$$P = -h^2 \Delta_x - \Delta_y + V(x, y)$$

where $x \in \mathbb{R}^n$ represents the position of the nuclei, $y \in \mathbb{R}^p$ is the position of the electrons, $h$ is proportional to the inverse of the square-root of the nuclear mass, and $V(x, y)$ is the interaction potential.

In the last decade, many efforts have been made in order to study in the semiclassical limit the spectrum of $P$ (see, e.g., [4, 6–10],...). These authors have shown that in many situations it is still possible to perform, by Grushin's method, semiclassical constructions related to the existence of some hidden effective semiclassical operator.

A. Senoussaoui (✉)
Faculté des Sciences, Département de Mathématiques, Université d'Oran,
B.P. 1524 El-Mnaouer, Oran, Algeria
e-mail: senoussaoui_abdou@yahoo.fr



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings
in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_14

It has been proved, both for smooth potentials [7] and for the physically interesting case of Coulomb interaction potentials (see, [6, 9]), the existence for the operator $P$ of asymptotic expansions for eigenvalues and associated eigenfunctions of the types:

$$\sum_{j\geq 0}\alpha_j h^{j/2} \text{ and } e^{-\psi(x)/h}\left(\sum_{j\geq 0}a_j\left(x,y\right)h^{j/2}\right),$$

where $\psi\left(x\right)$ is the Agmon distance between $x$ and the potential well.

Here we plan to give a unified version of the two results in [7] and [6], which can be applied to the general class of operators of the type $H=-h^2\Delta_x+P\left(x,y,D_y\right)$, where $P\left(x,y,D_y\right)$ is a pseudodifferential operator on $\mathscr{H}_2=L^2\left(\mathbb{R}_y^p\right)$ (the so-called electronic Hamiltonian and its eigenvalues are the so-called electronic levels).

By using the $h$-pseudodifferential operators with operator-valued symbol and the general Feshbach reduction scheme (see [1, 11]), the spectral study of $H$ on $L^2\left(\mathbb{R}_x^n\times\mathbb{R}_y^p\right)$ is reduced to that of a matrix of $h$-pseudodifferential operators $F\left(\lambda\right)$ on $\left(L^2\left(\mathbb{R}_x^n\right)\right)^{\oplus m}$ (the so-called effective Hamiltonian) with principal symbol the diagonal matrix $diag\left(\xi^2+\lambda_j\left(x\right)\right)_{1\leq j\leq m}$ where $m>0$ depends on the energy level and $\left(\lambda_j\left(x\right)\right)_{1\leq j\leq m}$ are the electronic levels. In particular, we obtain the following equivalence:

$$\lambda\in Sp\left(H\right)\Longleftrightarrow\lambda\in Sp\left(F\left(\lambda\right)\right)$$

(here $Sp$ stands for the spectrum).

The general theory of Helffer and Sjöstrand in [5] can be applied to the operator $F\left(\lambda\right)$ and shows the existence of formal WKB-type expansions for eigenfunctions of this operator. This finally gives the formal WKB-type expansions for the operator $H$ itself.

The argument of Martinez in [7] gives a justification to the formal WKB-constructions by showing that, modulo an error of size $\mathscr{O}\left(h^\infty\right)$, the formal eigenfunctions approximate correctly the true eigenfunctions of $H$.

The plan of this paper is the following. In the first section we introduce our assumptions and give preliminaries. The spectral reduction of the problem is given in the second section. The third section is devoted to state our main result and apply the reduction theorem obtained in the section 2 to establish the proof.

## 2 Assumptions and preliminaries

On the pseudodifferential operator $Q\left(x\right)=P\left(x,y,D_y\right)$, we assume (H1), (H2), and (H3) below.

(H1) For every $x\in\mathbb{R}^n$, $Q\left(x\right)$ is self-adjoint and bounded from below on $\mathscr{H}_2$ with $x$-independent domain $\mathscr{H}_1$.

(H2) The spectrum of the pseudodifferential operator $Q(x)$ has two disjoint components for every $x \in \mathbb{R}^n$:

$$Sp(Q(x)) = \{\lambda_1(x), \lambda_2(x), \ldots, \lambda_m(x)\} \cup \Gamma(x)$$

where $\lambda_1(x), \lambda_2(x), \ldots, \lambda_m(x)$ depend continuously on $x$, there is a gap between them and the rest of $Sp(Q(x))$ more precisely:

$$\exists\delta > 0, \quad \inf \Gamma(x) > \max\{\lambda_1(x), \lambda_2(x), \ldots, \lambda_m(x)\} + \delta, \ \forall x \in \mathbb{R}^n$$

and remain uniformly separated outside some compact subset of $\mathbb{R}^n$:

$$\exists \tilde{C} > 0, \ \inf_{\substack{j \neq k \\ |x| \geq C}} \left|\lambda_j(x) - \lambda_k(x)\right| \geq \tilde{C}, \ C > 0.$$

(H3) $Q(x) \in C_b^\infty(\mathbb{R}^n, \mathscr{L}(\mathscr{H}_1, \mathscr{H}_2))$, $Q(x)$ depends smoothly on $x$ and is uniformly bounded together with all its derivatives as an operator from $\mathscr{H}_1$ to $\mathscr{H}_2$.

*Example 1.* • The operator $Q(x) = -\frac{d^2}{dy^2} + \left(1 + x^2\right)^{2l} y^2$, $x \in \mathbb{R}, l \in \mathbb{R}$ satisfies the assumptions (H1) to (H3) with domain

$$\mathscr{H}_1 = H^2\left(\mathbb{R}_y\right) \cap \left\{\varphi \in L^2\left(\mathbb{R}_y\right);\ y^2\varphi \in L^2\left(\mathbb{R}_y\right)\ \right\},$$

$$\lambda_j(x) = (2j+1)\left(1+x^2\right)^l;\ j = 1, \ldots, m \text{ and}$$

$$\Gamma(x) = \left\{(2j+1)\left(1+x^2\right)^l;\ j \geq m+1\right\}.$$

• A second example is given in [9, 10] for the differential operator $\hat{Q}(x) = U(x)\left(-\Delta_y + V(x,y)\right)U^{-1}(x)$ where $U(x)$ is a diffeomorphism regularizing the physical case of the Coulomb interaction potential $V(x,y)$.

We denote by $S^m\left(\mathbb{R}^{2n}, \mathscr{L}(A,B)\right)$ the space of operator-valued symbols of order $m \in \mathbb{R}$:

$$\left\{a : \mathbb{R}^{2n} \mapsto \mathscr{L}(A,B) \in C^\infty;\ \forall(\alpha,\beta) \in \mathbb{R}^{2n},\ \left\|\partial_x^\alpha \partial_\xi^\beta a(x,\xi)\right\|_{\mathscr{L}(A,B)} = \mathscr{O}(< \xi >^m)\right\}$$

with $< \xi >= \left(1 + |\xi|^2\right)^{1/2}$ and $\mathscr{L}(A,B)$ the space of bounded linear operators from the Hilbert space $A$ to the Hilbert space $B$.

For $\varphi \in \mathscr{S}(\mathbb{R}^n, A)$ (the Schwartz space), $x \in \mathbb{R}^n$ and $a \in S^m\left(\mathbb{R}^{2n}, \mathscr{L}(A,B)\right)$, the $h$-pseudodifferential operator (the Weyl quantification of the symbol $a$) is defined by:

$$Op_h^w(a)\varphi(x) = (2\pi h)^{-1} \int_{\mathbb{R}^{2n}} e^{\frac{i}{h}<x-y,\xi>} a\left(\frac{x+y}{2}, \xi\right)\varphi(y)\,dy d\xi.$$

Note that $Op_h^w(a)$ maps continuously $\mathscr{S}(\mathbb{R}^n, A)$ into $\mathscr{S}(\mathbb{R}^n, B)$. In particular, due to a slight generalization of the Caldéron-Vaillancourt theorem (see [3, 11]), if $m \leq 0$ then $Op_h^w(a) \in \mathscr{L}\left(L^2(\mathbb{R}^n, A), L^2(\mathbb{R}^n, B)\right)$.

Using the constructions made in [6] lemma 1.1, we have the following lemma:

**Lemma 1.** *Under (H1) to (H3), there exist an orthonormal family* $\{u_1(x), u_2(x), \ldots, u_m(x)\}$ *in* $\mathscr{H}_2$ *such that:*

1. $\forall j \in \{1, \ldots, m\}, u_j(x) \in C_b^\infty(\mathbb{R}^n, \mathscr{H}_2)$,
2. $\{u_1(x), u_2(x), \ldots, u_m(x)\}$ *generates the space* $\bigoplus_{j=1}^{m} \ker\left(Q(x) - \lambda_j(x)\right)$.

## 3 Feshbach reduction

If $\overset{m}{\underset{j=1}{\oplus}} \psi_j = (\psi_1, \ldots, \psi_m) \in \left(L^2(\mathbb{R}^n)\right)^{\oplus m}$ and $\varphi \in L^2(\mathbb{R}^n, \mathscr{H}_1)$, then we define:

$$\overset{m}{\underset{j=1}{\oplus}} u_j(x)\left(\overset{m}{\underset{j=1}{\oplus}} \psi_j\right) = \sum_{j=1}^{m} u_j(x)\,\psi_j,$$

$$< \varphi, \overset{m}{\underset{j=1}{\oplus}} u_j(x) >_{\mathscr{H}_2} = \overset{m}{\underset{j=1}{\oplus}} (< \varphi, u_j(x) >_{\mathscr{H}_2}).$$

For $\lambda \in \mathbb{R}$, we consider the following matrix operator (the so-called Grushin operator):

$$\mathscr{P}(\lambda) = \begin{pmatrix} H - \lambda & \overset{m}{\underset{j=1}{\oplus}} u_j(x) \\ < ., \overset{m}{\underset{j=1}{\oplus}} u_j(x) >_{\mathscr{H}_2} & 0 \end{pmatrix} \text{ on } L^2(\mathbb{R}^n, \mathscr{H}_1) \oplus \left(L^2(\mathbb{R}^n)\right)^{\oplus m}.$$

Denote by $\lambda_+ = \inf_{x \in \mathbb{R}^n} \{Sp(Q(x)) \setminus \{\lambda_1(x), \lambda_2(x), \ldots, \lambda_m(x)\}\}$. Then we have:

**Theorem 1.** *Assume (H1)–(H3). Then for any* $\lambda < \lambda_+$, *the Grushin operator* $\mathscr{P}(\lambda) : H^2(\mathbb{R}^n, \mathscr{H}_1) \oplus \left(L^2(\mathbb{R}^n)\right)^{\oplus m} \longrightarrow L^2(\mathbb{R}^n, \mathscr{H}_2) \oplus \left(L^2(\mathbb{R}^n)\right)^{\oplus m}$ *is invertible and its inverse can be written as:*

$$\mathscr{P}(\lambda)^{-1} = \begin{pmatrix} E(\lambda) & E_+(\lambda) \\ E_-(\lambda) & E_{-+}(\lambda) \end{pmatrix}$$

*where* $E(\lambda), E_\pm(\lambda)$ *and* $E_{-+}(\lambda)$ *are h-pseudodifferential operators.*

*Moreover, we have the following equivalence:*

$$\lambda \in Sp(H) \Longleftrightarrow \lambda \in Sp(F(\lambda)) \tag{1}$$

*where* $F(\lambda) = \lambda - E_{-+}(\lambda)$ *is an* $m \times m$ *matrix of h-pseudodifferential operators on* $\left(L^2(\mathbb{R}^n)\right)^{\oplus m}$ *with the diagonal matrix diag* $\left(\xi^2 + \lambda_j(x)\right)_{1 \le j \le m}$ *as principal symbol.*

*Proof.* We can consider the Grushin operator $\mathscr{P}(\lambda)$ as an $h$-pseudodifferential operator with operator-valued symbol $p_\lambda(x,\xi)$ given by:

$$p_\lambda(x,\xi) = \begin{pmatrix} \xi^2 + Q(x) - \lambda & \overset{m}{\underset{j=1}{\oplus}} u_j(x) \\ < ., \overset{m}{\underset{j=1}{\oplus}} u_j(x) >_{\mathscr{H}_2} & 0 \end{pmatrix} : \mathscr{H}_1 \oplus \mathbb{C}^m \longrightarrow \mathscr{H}_2 \oplus \mathbb{C}^m.$$

Using the fact that for any $\lambda < \lambda_+$ and $x \in \mathbb{R}^n$,

$$\hat{\pi}(x)\, Q(x)\, \hat{\pi}(x) - \lambda > 0 \tag{2}$$

the symbol $p_\lambda(x,\xi)$ is invertible and its inverse $q_\lambda(x,\xi)$ is given by:

$$q_\lambda(x,\xi) = \begin{pmatrix} r(x,\xi,\lambda) & \overset{m}{\underset{j=1}{\oplus}} u_j(x) \\ < ., \overset{m}{\underset{j=1}{\oplus}} u_j(x) >_{\mathscr{H}_2} & \left(\lambda - \xi^2 - \lambda_j(x)\right)_{1 \le j \le m} \end{pmatrix}$$

where $\hat{\pi}(x) = 1 - \pi(x)$, $\pi(x)$ denotes the orthogonal projection on the space $\bigoplus_{j=1}^{m} \ker\left(Q(x) - \lambda_j(x)\right)$ and $r(x,\xi,\lambda) = \hat{\pi}(x)\left(\xi^2 + \hat{\pi}(x)\, Q(x)\, \hat{\pi}(x) - \lambda\right)^{-1} \hat{\pi}(x)$.

Due to (H3) and (2), we can consider the Weyl quantification $Q(\lambda) = Op_h^w(q_\lambda)$ : $L^2(\mathbb{R}^n, \mathscr{H}_2) \oplus \left(L^2(\mathbb{R}^n)\right)^{\oplus m} \longrightarrow H^2(\mathbb{R}^n, \mathscr{H}_1) \oplus \left(L^2(\mathbb{R}^n)\right)^{\oplus m}$.

The symbolic calculus and especially the composition theorem of $h$-pseudodifferential operators allow us to obtain

$$\begin{cases} \mathscr{P}(\lambda)\, Q(\lambda) = I + hR_1; \ \|R_1\|_{\mathscr{L}\left(L^2(\mathbb{R}^n,\mathscr{H}_2) \oplus \left(L^2(\mathbb{R}^n)\right)^{\oplus m}\right)} = \mathscr{O}(1) \\ Q(\lambda)\, \mathscr{P}(\lambda) = I + hR_2; \ \|R_2\|_{\mathscr{L}\left(H^2(\mathbb{R}^n,\mathscr{H}_1) \oplus \left(L^2(\mathbb{R}^n)\right)^{\oplus m}\right)} = \mathscr{O}(1) \end{cases}.$$

Here, the estimates of $\|R_1\|$ and $\|R_2\|$ are uniform with respect to $h$. As a consequence, for $h$ small enough, $\mathscr{P}(\lambda)$ is invertible and its inverse is given by the Neumann series:

$$\mathscr{P}(\lambda)^{-1} = Q(\lambda)\left(I + \sum_{k=1}^{+\infty} h^k R_1^k\right) = \left(I + \sum_{k=1}^{+\infty} h^k R_2^k\right) Q(\lambda). \tag{3}$$

In view of (4) and the expression of the symbol $q_\lambda(x,\xi)$ it remains to prove the equivalence (1). This comes from the two following algebraic identities:

$$((H-\lambda)\,u=v)\Leftrightarrow \mathscr{P}(\lambda)\,(u\oplus 0)=v\oplus(<u,\bigoplus_{j=1}^{m}u_j(x)>_{\mathscr{H}_2})$$

$$\Leftrightarrow (u\oplus 0)=\mathscr{P}(\lambda)^{-1}\,(v\oplus(<u,\bigoplus_{j=1}^{m}u_j(x)>_{\mathscr{H}_2}))$$

$$((H-\lambda)\,u=v)\Leftrightarrow \begin{cases} u=E(\lambda)\,v+E_{+}(\lambda)\,(<u,\bigoplus_{j=1}^{m}u_j(x)>_{\mathscr{H}_2}) \\ 0=E_{-}(\lambda)\,v+E_{-+}(\lambda)\,(<u,\bigoplus_{j=1}^{m}u_j(x)>_{\mathscr{H}_2}) \end{cases} \tag{4}$$

and

$$(E_{-+}(\lambda)\,\alpha=\beta)\Leftrightarrow \mathscr{P}(\lambda)^{-1}\,(0\oplus\alpha)=(E_{+}(\lambda)\,\alpha)\oplus\beta$$

$$\Leftrightarrow 0\oplus\alpha=\mathscr{P}(\lambda)\,((E_{+}(\lambda)\,\alpha)\oplus\beta)$$

$$(E_{-+}(\lambda)\,\alpha=\beta)\Leftrightarrow \begin{cases} 0=(H-\lambda)\,(E_{+}(\lambda)\,\alpha)+\bigoplus_{j=1}^{m}u_j(x)\,\beta \\ \alpha=<E_{+}(\lambda)\,\alpha,\bigoplus_{j=1}^{m}u_j(x)>_{\mathscr{H}_2} \end{cases}. \tag{5}$$

If $\lambda\notin Sp\,(H)$, then from (7) we deduce:

$$E_{-+}(\lambda)\,\alpha=\beta\Leftrightarrow \begin{cases} E_{+}(\lambda)\,\alpha=-(H-\lambda)^{-1}\,(\bigoplus_{j=1}^{m}u_j(x)\,\beta) \\ \alpha=<-(H-\lambda)^{-1}\,(\bigoplus_{j=1}^{m}u_j(x)\,.),\bigoplus_{j=1}^{m}u_j(x)>_{\mathscr{H}_2}\beta \end{cases}.$$

In particular,

$$0\notin Sp\,(E_{-+}(\lambda))\ \text{ and}$$

$$E_{-+}(\lambda)^{-1}=-<(H-\lambda)^{-1}\,(\bigoplus_{j=1}^{m}u_j(x)\,.),\bigoplus_{j=1}^{m}u_j(x)>_{\mathscr{H}_2}.$$

Conversely, if $0\notin Sp\,(E_{-+}(\lambda))$, then (6) gives:

$$(H-\lambda)\,u=v\Leftrightarrow \begin{cases} <u,\bigoplus_{j=1}^{m}u_j(x)>_{\mathscr{H}_2}=-E_{-+}(\lambda)^{-1}\,(E_{-}(\lambda)\,v) \\ u=E(\lambda)\,v-E_{+}(\lambda)\,E_{-+}(\lambda)^{-1}\,E_{-}(\lambda)\,v \end{cases}.$$

As a consequence,

$$\lambda\notin Sp\,(H)\ \text{ and }\ (H-\lambda)^{-1}=E(\lambda)-E_{+}(\lambda)\,E_{-+}(\lambda)^{-1}\,E_{-}(\lambda)$$

## 4 WKB-Constructions

In this section we will give asymptotic expansions in powers of $h^{1/2}$ for eigenvalues and eigenfunctions of $H$ near the potential well formed by the first electronic level $\lambda_1(x)$ of $Q(x)$. Assume (H2) with $m = 1$. By adding a constant to $Q(x)$ and a translation in the variable $x$, we assume that $\lambda_1(x)$ has a minimum strict at zero (a one point well not degenerate):

$$0 = \inf_{x\in\mathbb{R}^n} \lambda_1(x), \quad \underline{\lim}_{|x|\mapsto\infty} \lambda_1(x) > 0, \ \lambda_1^{-1}(0) = \{0\}, \ \lambda_1''(0) > 0.$$

Denoting by $\psi(x)$ the distance between $x \in \mathbb{R}^n$ and 0 in the Agmon metric $\lambda_1(x)\, dx^2$, it is known (see [5]) that there is a neighborhood $\Omega$ of 0 such that:

$$\psi \in C^\infty(\Omega, \mathbb{R}), \ (\nabla\psi)^2(x) = \lambda_1(x), \ \forall x \in \Omega.$$

We fix some (arbitrarily large) constant $C_0 > 0$ outside the spectrum of the harmonic oscillator $H_0 = -\Delta_x + \frac{1}{2} < \lambda_1''(0)\, x, x >_{\mathbb{R}^n}$. Denote by $e_1, \ldots, e_{N_0}$ the eigenvalues of $H_0$ in $[0, C_0]$,

$$Sp(H_0) = \left\{ \sum_{i=1}^{n} (2\alpha_i + 1)\sqrt{\mu_i};\ \alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n \right\}$$

where $\mu_1, \ldots, \mu_n$ are the eigenvalues of the matrix $\lambda_1''(0)$. The main result is as follows:

**Theorem 2.** *Assume* $(H1)$–$(H3)$. *$H$ has $N_0$ eigenvalues $E_1(h), E_2(h), \ldots, E_{N_0}(h)$ in $[0, C_0 h]$ such that, for every $j \in \{1, \ldots, N_0\}$ and for $h$ sufficiently small, $E_j(h)$ admit the following asymptotic expansion:*

$$E_j(h) = e_j h + \sum_{k\geq 1} \alpha_{j,k} h^{1+k/2} \ \ \textit{modulo}\ \mathcal{O}(h^\infty) \tag{6}$$

$\alpha_{j,k} \in \mathbb{R}$. *If $E_j(h)$ is asymptotically simple (in the sense that the expansion* (6) *determines $E_j(h)$ in a unique way), then the associated normalized eigenfunction $\varphi_j(x, y; h)$ satisfies:*

$$e^{\psi(x)/h}\varphi_j(x, y; h) = h^{-m_j} \sum_{k\geq 0} a_{j,k}(x, y)\, h^{k/2} \ \ \textit{modulo}\ \mathcal{O}(h^\infty)$$

*in $C^\infty(\Omega, \mathcal{H}_1)$, where $a_{0,0} = \tilde{a}_0(x)\, u_1(x, y)$, $\tilde{a}_0(x) \neq 0$, $\forall x \in \Omega$, $m_j \in \mathbb{R}$, $m_1 = n/4$.*

Before turning to the proof of the theorem 2, let us recall some basic facts on formal $h$-pseudodifferential operators with operator valued symbol. If $\Omega$ is an open set in $\mathbb{R}^n$, $H$ is a Hilbert space and $m \in \mathbb{R}$, we introduce the space of formal power series:

$$S^m(\Omega, H) = \left\{ \sum_{k \geq 0} h^{-m+k/2} s_k(x)\, ;\; s_k \in C^\infty(\Omega, H) \right\}.$$

For $\psi \in C^\infty(\Omega, \mathbb{R})$ and $\mathscr{V}$ a neighborhood of 0 in $\mathbb{R}^n$ we set

$$\Omega^* = \{(x, \xi) \in \Omega \times \mathbb{C}^n;\; \xi - i\nabla\psi(x) \in \mathscr{V}\} \tag{7}$$

and the space of formal symbol

$$S^0\left(\Omega^*, \mathscr{L}(H, K)\right) = \left\{ \sum_{k \geq 0} h^k p_k(x, \xi)\, ;\; p_k \in C^\infty\left(\Omega^*, \mathscr{L}(H, K)\right) \right\}$$

where $H$, $K$ denote Hilbert spaces. For any symbol $b(x, \xi, h) = b = \sum\limits_{k \geq 0} h^k b_k \in S^0(\Omega^*, \mathscr{L}(H, K))$, one can define the action of the operator of symbol $b$ on $e^{-\psi(x)/h} S^m(\Omega, H)$ by setting for $s \in S^m(\Omega, H)$:

$$e^{\psi(x)/h} Op_h^w(b)\left(e^{-\psi(x)/h} s\right)$$
$$= \sum_{\alpha \in \mathbb{N}^n} \frac{h^{|\alpha|}}{i^{|\alpha|}\alpha!} \partial_z^\alpha \left[ \partial_\xi^\alpha b\left(\frac{x+z}{2}, i\nabla\psi(x), h\right) s(z, h)\, e^{-\mathscr{K}(x,z)/h} \right]_{z=x} \tag{8}$$

where

$$\mathscr{K}(x, z) = \psi(z) - \psi(x) - (z - x)\nabla\psi(x) = \mathscr{O}\left(|z - x|^2\right).$$

$Op_h^w(b)$ is called a formal $h$-pseudodifferential operator, this definition coincides with a formal stationary phase expansion.

One can verify as in [11] that

$$e^{\psi(x)/h} Op_h^w(b)\left(e^{-\psi(x)/h} s\right) \in S^m(\Omega, K).$$

Furthermore, if $G$ is an Hilbert space and $c = \sum\limits_{j \geq 0} h^j c_j \in S^0(\Omega^*, \mathscr{L}(K, G))$, then $Op_h^w(c) \circ Op_h^w(b) = Op_h^w(d)$ is a formal $h$-pseudodifferential operator with symbol $d(x, \xi, h) = \sum\limits_{j \geq 0} h^j d_j(x, \xi) \in S^0(\Omega^*, \mathscr{L}(H, G))$ given by:

$$d_j(x, \xi) = \sum_{|\alpha|+|\beta|+k+l=j} \frac{(-1)^{|\beta|}}{\alpha!\beta!(2i)^{|\alpha|+|\beta|}} \left(\partial_\xi^\alpha \partial_x^\beta b_k\right)\left(\partial_\xi^\alpha \partial_x^\alpha c_l\right). \tag{9}$$

*Proof.* The main idea of the proof is to use a simplified formal version of the theorem 1. For $\lambda < \lambda_+$ we set

$$\tilde{\mathscr{P}}(\lambda) = \begin{pmatrix} H - \lambda & u_1(x) \\ < ., u_1(x) >_{\mathscr{H}_2} & 0 \end{pmatrix} \text{ on } e^{-\psi(x)/h} S^m(\Omega, \mathscr{H}_1 \oplus \mathbb{C}),$$

$\tilde{\mathscr{P}}(\lambda)$ is a formal $h$-pseudodifferential operator with symbol

$$\tilde{p}(x, \xi; \lambda) = \begin{pmatrix} \xi^2 + Q(x) - \lambda & u_1(x) \\ < ., u_1(x) >_{\mathscr{H}_2} & 0 \end{pmatrix} \in S^0\left(\Omega^*, \mathscr{L}(\mathscr{H}_1 \oplus \mathbb{C}, \mathscr{H}_2 \oplus \mathbb{C})\right)$$

where $\Omega^*$ is defined as in (7) with $\psi(x)$ being the Agmon distance. As in the proof of the theorem 1, the inverse of the symbol $\tilde{p}(x, \xi; \lambda)$ is given by:

$$q_0(x, \xi; \lambda) = \begin{pmatrix} \hat{\pi}(x)\left(\xi^2 + \hat{\pi}(x) Q(x) \hat{\pi}(x) - \lambda\right)^{-1} \hat{\pi}(x) & u_1(x) \\ < ., u_1(x) >_{\mathscr{H}_2} & \lambda - \xi^2 - \lambda_1(x) \end{pmatrix}.$$

This permits us, using the composition formula (9), to construct a formal $h$-pseudodifferential operator $\tilde{\mathscr{Q}}(\lambda)$ of symbol (see [7, 11])

$$q_\lambda(x, \xi, h) = q_0(x, \xi; \lambda) + \sum_{k \geq 1} h^k q_k(x, \xi; \lambda) \in S^0\left(\Omega^*, \mathscr{L}(\mathscr{H}_2 \oplus \mathbb{C}, \mathscr{H}_1 \oplus \mathbb{C})\right)$$

such that

$$\tilde{\mathscr{P}}(\lambda) \tilde{\mathscr{Q}}(\lambda) = Id \quad \text{on } e^{-\psi(x)/h} S^m(\Omega, \mathscr{H}_2 \oplus \mathbb{C}). \tag{10}$$

Writing

$$\tilde{\mathscr{Q}}(\lambda) = \begin{pmatrix} \tilde{E}(\lambda) & \tilde{E}_+(\lambda) \\ \tilde{E}_-(\lambda) & \tilde{E}_{-+}(\lambda) \end{pmatrix}$$

and setting $\tilde{F}(\lambda) = \lambda - \tilde{E}_{-+}(\lambda)$, by construction, the operator $\tilde{F}(\lambda)$ is a nice formal $h$-pseudodifferential operator with scalar symbol

$$e_\lambda(x, \xi, h) = \xi^2 + \lambda_1(x) + \sum_{k \geq 1} h^k e_k(x, \xi; \lambda) \in S^0\left(\Omega^*, \mathbb{C}\right).$$

Since $\tilde{F}(\lambda)$ is formally self-adjoint and $\lambda_1(x)$ admits a nondegenerate minimum at 0, the construction of Helffer and Sjöstrand in [5] Sect. 3 gives $N_0$ formal series $\tilde{E}_j(h; \lambda)$ for $j \in \{1, .., N_0\}$ of the form:

$$\tilde{E}_j(h; \lambda) = e_j h + \sum_{k \geq 1} h^{1+k/2} e_{j,k}(\lambda) \quad \text{and} \quad a_j(x, h; \lambda) \in S^{m_j}(\Omega; \mathbb{C})$$

such that

$$\left(\tilde{F}(\lambda) - \tilde{E}_j(h;\lambda)\right)\left(e^{-\psi(x)/h}a_j(x,h;\lambda)\right) = 0 \text{ in } e^{-\psi(x)/h}S^{m_j}(\Omega;\mathbb{C}).$$

Fix $j \in \{1,..,N_0\}$. Using the analytical dependence in $\lambda$ of the symbol $e_\lambda(x,\xi,h)$ of $\tilde{F}(\lambda)$ and applying again the construction of Helffer and Sjöstrand for $\tilde{F}\left(e_j h + \lambda' h^{3/2}\right)$ ($\lambda'$ indicates a new parameter), this gives formal series of the form:

$$\tilde{E}_j'\left(h;\lambda'\right) = e_j h + e_{j,1}h^{3/2} + \sum_{k\geq 2} h^{1+k/2}\tilde{e}_{j,k}\left(\lambda'\right).$$

Setting $\lambda' = e_{j,1} + \lambda'' h^{1/2}$ (where $\lambda''$ indicates a new parameter) and reiterating this process, we obtain finally formal series (independent of $\lambda$):

$$\tilde{E}_j(h) = e_j h + \sum_{k\geq 1} h^{1+k/2}e_{j,k} \text{ and } a_j(x,h) \in S^{m_j}(\Omega;\mathbb{C}).$$

Furthermore, we have:

$$\left(\tilde{F}\left(\tilde{E}_j(h)\right) - \tilde{E}_j(h)\right)\left(e^{-\psi(x)/h}a_j(x,h)\right) = 0 \text{ in } e^{-\psi(x)/h}S^{m_j}(\Omega;\mathbb{C}) \tag{11}$$

$$< a_j(.,h), a_{j'}(.,h) >_\psi = \delta_{jj'} + \mathcal{O}(h) \tag{12}$$

where (12) holds in the sense of formal power series in $h$ with complex coefficients. The inner product $< .,. >_\psi$ is defined by a formal stationary phase expansion at 0:

$$< u(x,h), v(x,h) >_\psi = \int_\Omega e^{-2\psi(x)/h}u(x,h)\overline{v(x,h)}dx.$$

Using (10) and the definitions of $\tilde{\mathscr{P}}(\lambda)$ and $\tilde{\mathscr{Q}}(\lambda)$, the equation (11) yields that the formal symbol:

$$b_j(x,y;h) = e^{\psi(x)/h}\tilde{E}_+\left(\tilde{E}_j(h)\right)\left(e^{-\psi(x)/h}a_j(x,h)\right) \in S^{m_j}(\Omega;\mathscr{H}_1)$$

solves

$$\left(H - \tilde{E}_j(h)\right)\left(e^{-\psi(x)/h}b_j(x,y;h)\right) = 0 \text{ in } e^{-\psi(x)/h}S^{m_j}(\Omega;\mathscr{H}_2).$$

Since

$$\tilde{E}_+(\lambda) = u_1(x,.) + \mathcal{O}(h) \text{ and } < u_1(x,.), u_1(x,.) >_{\mathscr{H}_2} = 1,$$

we get

$$< e^{-\psi(x)/h} b_j(.,.;h), e^{-\psi(x)/h} b_{j'}(.,.;h) >_{L^2(\Omega\times\mathbb{R}^p)} = \delta_{jj'} + \mathscr{O}(h).$$

By a standard argument in [7] and [5] Sec. 5, one can show that the eigenvalues $E_j(h)$ of $H$ in $[0, C_0 h]$ ($C_0 > 0$ arbitrarily large) admit for asymptotic expansions $\tilde{E}_j(h)$ found above. Moreover, if $E_j(h)$ is asymptotically simple, the formal series $e^{-\psi(x)/h} b_j(x,y;h)$ are the asymptotic expansions for the associated normalized eigenfunctions $\varphi_j(x,y;h)$ ($j = 1, \ldots, N_0$). This is the case for the first eigenfunction $\varphi_1(x,y;h)$ (since $E_1(h)$ is simple) and $m_1 = n/4$ is chosen for the normalization.

## References


1. Balazard-Konlein, A.: *Calcul fonctionnel pour des opérateurs h-admissibles à symboles opérateurs et applications. Thèse de 3ème cycle, University of Nantes, France, (1985).. Calcul fonctionnel pour des opérateurs h-admissibles à symboles opérateurs et applications. Thèse de 3ème cycle, University of Nantes, France, (1985).*
2. Born, M., Oppenheimer, R.: *Zur Quantentheorie der Molekeln. Annal. Phys. vol 84, 457, 1927.*
3. Calderon, A. P., Vaillancourt, R.: *On the boundness of pseudo-differential operators. J. Math. Soc. Japan 23, 1972, p374–378.*
4. Gérard, C., Martinez, A., Sjöstrand, J.: *A mathematical approach to the effective hamiltonian in perturbed periodic problems. Comm. Math. Physics vol 142 (4), 1991.*
5. Helffer, B., Sjöstrand, J.: *Multiple wells in the semiclassical limit I. Comm. P.D.E. vol 9 (4), 1984, p337–408.*
6. Klein, M., Martinez, A., Seiler, R., Wang, X.P.: *On the Born-Oppenheimer expansion for polyatomic molecules. Comm. Maths. Phys. 1992, p607–639.*
7. Martinez, A.: *Développements asymptotiques et effet tunnel dans l'approximation de Born-Oppenheimer. Ann. I.H.P. vol 49, (3), 1989, p239–257.*
8. Martinez, A., Messirdi, B.: *Resonances of diatomic molecules in the Born-Oppenheimer approximation. Comm. P.D.E. vol 19 (7/8), 1994, p1139–1162.*
9. Messirdi, B., Senoussaoui, A.: *Méthode BKW formelle et spectre des molécules polyatomiques dans l'approximation de Born-Oppenheimer. Canadian Journal of Physics, vol 79, (4), 2001, p757–771.*
10. Messirdi, B., Senoussaoui, A., Djelouli, G.: *Resonances of polyatomic molecules in the Born-Oppenheimer approximation. To appear in Journal of Mathematical Physics, 2005.*
11. Senoussaoui, A.: *Opérateurs h-admissibles matriciels à symboles opérateurs. African Diaspora Journal of Mathematics, Vol 4, no.1, (2007), 7–27.*

# Finite Kibble's Bivariate Gamma Mixtures for Color Image Segmentation

**Taher Ben Arab, Mourad Zribi, and Afif Masmoudi**

**Abstract** The segmentation of a color image is an important research field of image processing. A color image could be considered as the result of a finite mixture model. Although the most well known used distribution when considering mixture models is the Gaussian one, it is certainly not the best approximation for image segmentation. It is well known that the statistics of natural images are not Gaussian at all. In this paper, an efficient method of image segmentation is proposed. The method uses Kibble's bivariate Gamma mixture model and K-Means algorithm. Using the K-Means algorithm, the number of image regions is identified and the model parameters inside the image regions are estimated by using the EM algorithm. Experimental results that demonstrate the performance of the proposed model for image segmentation are presented.

**Keywords** Segmentation • Mixture model • Kibble's Bivariate Gamma • K-Means algorithm • EM algorithm

T. Ben Arab (✉)
Laboratoire d'Informatique: Signal et Image de la Côte d'Opale (LISIC-EA 4491), ULCO, 50 Rue Ferdinand Buisson, BP 719, 62228 Calais Cedex, France

Faculté des Sciences de Sfax, Laboratoire de Probabilités et Statistique, Université de Sfax, B. P. 1171 Sfax, Tunisie
e-mail: taher.ben_arab@yahoo.fr

M. Zribi
Laboratoire d'Informatique: Signal et Image de la Côte d'Opale (LISIC-EA 4491), ULCO, 50 Rue Ferdinand, Buisson BP 719, 62228 Calais Cedex, France
e-mail: Mourad.Zribi@lisic.univ-littoral.fr

A. Masmoudi
Faculté des Sciences de Sfax, Laboratoire de Probabilités et Statistique, Université de Sfax, B. P. 1171 Sfax, Tunisie
e-mail: Afif.Masmoudi@fss.rnu.tn



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_15

# 1 Introduction

In the last few decades, the researchers have shown a major focus on satellite image segmentation in terms of image processing. Several papers have been published, focusing mainly on gray scale images, hence giving less attention to color image segmentation [6], which is able to give us much more information about the object or images [1, 3, 7, 12, 16]. Image Segmentation is typically used in order to locate the objects and the boundaries of images [8]. Image Segmentation is a process based on selected image features in order to ensure the partition of image pixels. The pixels belonging to the same region must be spatially connected and must have similar image features. If the selected segmentation feature is color, then the image segmentation process will separate pixels having distinct color features into different regions and, simultaneously, will gather the pixels which are spatially connected and have a similar color into the same region [9]. In color imagery [15], image pixels can be represented by a number of different color spaces, e.g. Red Green Blue (RGB), Hue Saturation Value (HSV) or Hue Saturation Intensity (HSI) [11].

In the literature, several works have been dealing with color image segmentation under HSI color space, since the Hue and Saturation Values of the pixels in the image are not negative [10, 14, 18].

In this paper, we introduce a Kibble's Gamma bivariate Mixture model in order to ensure the segmentation of a color image. The Expectation Maximization (EM) algorithm [4] is used to estimate the model parameters. The number of image segments can be initialized by using the $K$-means algorithm [16].

This paper is organized as follows. In section II, we describe briefly the bivariate Kibble's Gamma distribution. Section III is dealing with the estimation of unknown parameters of Kibble's Gamma distribution using two methods, namely the Maximum Likelihood Estimation (ML) and the Method of the Moments (MOM). Section IV describes the bivariate Kibble's Gamma Mixture model. This description shows how to initialize the model parameters and also presents the segmentation algorithm. The experimental results are given in section V before concluding in section VI.

# 2 Kibble's Bivariate Gamma

Let $Y = (Y_1, Y_2)$ be a bivariate random variable representing the feature vector of pixels of a satellite image region. Since these two variables are highly influenced by various random factors such as vision, lighting, moisture, environmental conditions, etc., this feature vector can be viewed as a bivariate random vector. To model the bivariate features of the image, it is very common to assume that the feature vector of the image follows a Kibble's bivariate Gamma distribution (KBGD) [2, 13].

**Definition 1.** A random vector $Y = (Y_1, Y_2)$ is distributed according to a Kibble's Bivariate Gamma distribution with positive shape parameter $s$ and the scale vector parameter $\Sigma = (\Sigma_{11}, \Sigma_{22}, \Sigma_{12}^2)$, where $0 < \Sigma_{12}^2 < \Sigma_{11}\Sigma_{22}$, if its probability density function (pdf) is defined as follows:

$$f_Y(y_1, y_2) = \exp(-\frac{\Sigma_{22}y_1 + \Sigma_{11}y_2}{\Sigma_{11}\Sigma_{22} - \Sigma_{12}^2}) \times \frac{(y_1y_2)^{s-1}}{(\Sigma_{11}\Sigma_{22} - \Sigma_{12}^2)^s\Gamma(s)} \times f_s(\delta y_1 y_2)\mathbf{1}_{[0,+\infty[^2}(y_1, y_2), \quad (1)$$

where $\mathbf{1}_{[0,+\infty[^2}(y_1, y_2)$ is the indicator function defined on $[0, +\infty[^2$, $\delta = \frac{\Sigma_{12}^2}{(\Sigma_{11}\Sigma_{22} - \Sigma_{12}^2)^2}$, and $f_s(x)$ is given by $f_s(x) = \sum_{k=0}^{\infty} \frac{x^k}{\Gamma(s+k)k!}$, $\forall\ x \in \mathbb{R}$.

In Fig. 1 and Fig. 2, we present a KBGD for different values of parameters.

The marginal distribution $Y_i$, $i = 1, 2$, is distributed according to a univariate Gamma distribution with pdf defined by

$$f_{Y_i}(y_i) = (\frac{y_i}{\Sigma_{ii}})^{s-1}\frac{\exp(-\frac{y_i}{\Sigma_{ii}})}{\Sigma_{ii}\Gamma(s)}\mathbf{1}_{[0,+\infty[}(y_i). \quad (2)$$

where $s > 0$ represents the shape parameter and $\Sigma_{ii} > 0$, $i = 1, 2$, is called the scale parameter.

To define the parameters of the random variable $Y$, we introduce the following proposition.

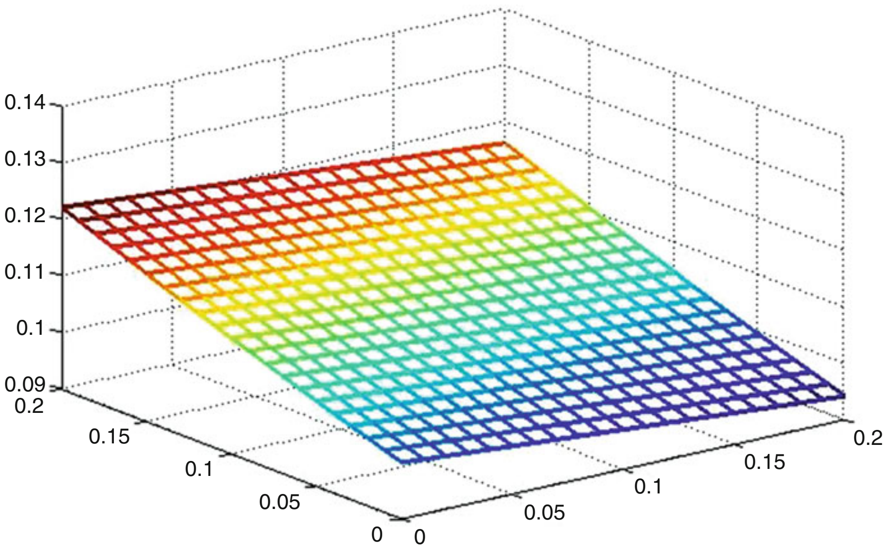


**Fig. 1** Pdf of KBGD with parameters $s = 1$, $\Sigma_{11} = 10$, $\Sigma_{22} = 3$, and $\Sigma_{12}^2 = 2$

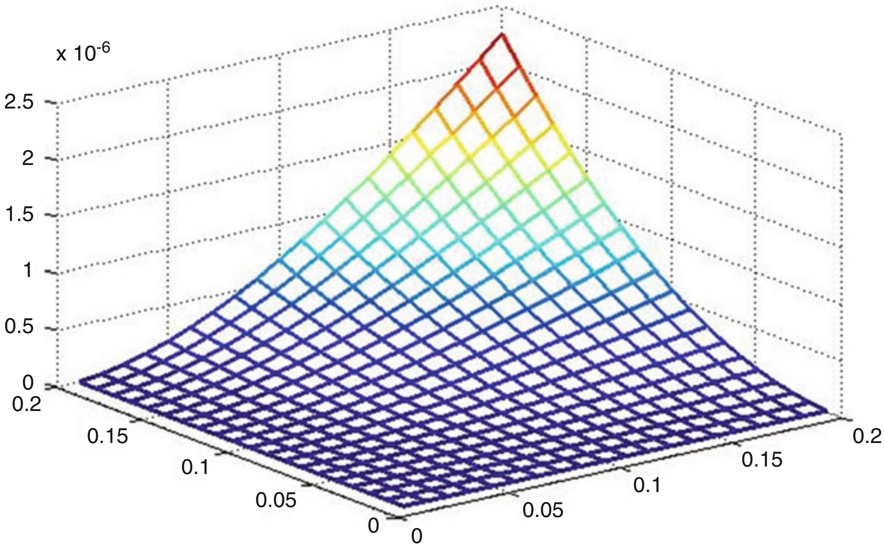


**Fig. 2** Pdf of KBGD with parameters $s = 3$, $\Sigma_{11} = 1$, $\Sigma_{22} = 10$, and $\Sigma_{12}^2 = 5$

**Proposition 1.** *Let $Y = (Y_1, Y_2)$ have a Kibble's bivariate Gamma distribution with positive shape parameter $s$ and the scale parameter $\Sigma = (\Sigma_{11}, \Sigma_{22}, \Sigma_{12}^2)$, where $0 < \Sigma_{12}^2 < \Sigma_{11}\Sigma_{22}$. Then, for all $\theta_1, \theta_2 \le 0$, the Laplace transform of $Y$ is given by*

$$L((\theta_1, \theta_2)) = \Big[1 - \theta_{11}\Sigma_{11} - \theta_{22}\Sigma_{22} + (\Sigma_{11}\Sigma_{22} - \Sigma_{12}^2)\theta_{11}\theta_{22}\Big]^{-s}$$

*with*

$$\theta_{11}\Sigma_{11} + \theta_{22}\Sigma_{22} - (\Sigma_{11}\Sigma_{22} - \Sigma_{12}^2)\theta_{11}\theta_{22} < 1.$$

For a proof, see [13].

**Proposition 2.** *The moments and the covariance of the random variable $Y$, with pdf defined in Eq. 4, are given by*

$$E[Y_1] = s\Sigma_{11},\ E[Y_2] = s\Sigma_{22}$$

$$var(Y_1) = s\Sigma_{11}^2,\ var(Y_2) = s\Sigma_{22}^2$$

$$cov(Y_2, Y_2) = s\Sigma_{12}^2.$$

These different moments cited in Proposition 2 can be obtained by differentiating this expression of the Laplace transform defined in Proposition 1 with respect to $\theta_{i,j},\ i,j = 1, 2$.

## 3 Parameter Estimation

This section deals with the problem of estimating the unknown parameter vector $\Sigma = (\Sigma_{11}, \Sigma_{22}, \Sigma_{12}^2)$, for a known parameter $s$, from $n$ independent vectors $Y^1, \ldots, Y^n$, where $Y^i = (Y_1^i, Y_2^i)$ is distributed according to a Kibble's bivariate Gamma distribution with a parameter vector $\Sigma$.

### *3.1 Maximum Likelihood Estimation (ML)*

The log-likelihood function of a sample bivariate observations $Y^1, Y^2, \ldots, Y^n$, with same density defined in Eq. 1 is given by

$$l(\Sigma) = -ns \ \ln(\Sigma_{1.2}) - \frac{n\Sigma_{22}\overline{Y}_1}{\Sigma_{1.2}} + (s-1)\sum_{i=1}^{n}\log(Y_1^iY_2^i) - \log\Gamma(s) - \frac{n\Sigma_{11}\overline{Y}_2}{\Sigma_{1.2}} + \sum_{i=1}^{n}\ln(f_s(\delta Y_1^iY_2^i)), \tag{3}$$

where $\overline{Y}_j = \frac{1}{n}\sum_{i=1}^{n} Y_j^i$, $j = 1, 2$, $\Sigma_{1.2} = \Sigma_{11}\Sigma_{22} - \Sigma_{12}^2$ and $\delta = \dfrac{\Sigma_{12}^2}{\Sigma_{1.2}^2}$.

By taking the differential, with respect to $\Sigma_{11}$, $\Sigma_{22}$, and $\Sigma_{12}^2$, one has

$$-ns + \frac{n\Sigma_{22}\overline{Y}_1}{\Sigma_{1.2}} - \frac{n\overline{Y}_2}{\Sigma_{22}} + \frac{n\Sigma_{11}\overline{Y}_2}{\Sigma_{1.2}} - 2\frac{\Sigma_{12}^2}{\Sigma_{1.2}^2}H = 0, \tag{4}$$

$$-ns + \frac{n\Sigma_{11}\overline{Y}_2}{\Sigma_{1.2}} - \frac{n\overline{Y}_1}{\Sigma_{11}} + \frac{n\Sigma_{22}\overline{Y}_1}{\Sigma_{1.2}} - 2\frac{\Sigma_{12}^2}{\Sigma_{1.2}^2}H = 0, \tag{5}$$

$$ns - \frac{n\Sigma_{22}\overline{Y}_1}{\Sigma_{1.2}} - \frac{n\Sigma_{11}\overline{Y}_2}{\Sigma_{1.2}} + \frac{\Sigma_{12}^2 + \Sigma_{11}\Sigma_{22}}{\Sigma_{1.2}^2}H = 0, \tag{6}$$

with $H = \sum_{i=1}^{n} y_1^i y_2^i \dfrac{f_{s+1}(\delta y_1^i y_2^i)}{f_s(\delta y_1^i y_2^i)}$ and $f_{s+1} = f_s'$.

From Eq. 4, Eq. 5, and Eq. 6, the ML estimators of $\hat{\Sigma}_{11}$ and $\hat{\Sigma}_{22}$ are defined by

$$\hat{\Sigma}_{11} = \frac{\overline{Y}_1}{s}, \ \hat{\Sigma}_{22} = \frac{\overline{Y}_2}{s} \tag{7}$$

By replacing $\Sigma_{11}$ and $\Sigma_{22}$ by their estimators in Eq. 6, then the estimator of $\Sigma_{12}^2$ is the root of the following function

$$\phi(\Sigma_{12}^2) = n - \frac{s}{\overline{Y}_1\overline{Y}_2 - s^2\Sigma_{12}^2}\sum_{i=1}^{n} y_1^i y_2^i \frac{f_{s+1}(\hat{\delta} y_1^i y_2^i)}{f_s(\hat{\delta} y_1^i y_2^i)} = 0, \tag{8}$$

where $\hat{\delta} = \dfrac{s_1^4\Sigma_{12}^2}{(\overline{Y}_1\overline{Y}_2 - s^2\Sigma_{12}^2)^2}$.

Closed-form solutions of Eq. 8 do not exist for the parameter $\hat{\Sigma}_{12}^2$. We can get a solution by using a Newton-Raphson procedure [19]. We generally get the convergence of the Newton-Raphson procedure after few iterations.

## *3.2 Method of Moments (MOM)*

The estimator of the vector parameter $\Sigma = (\Sigma_{11}, \Sigma_{22}, \Sigma_{12}^2)$ by the MOM is the solution of the following system

$$\begin{cases} \overline{Y}_1 = E[Y_1],\ \overline{Y}_2 = E[Y_2] \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(Y_1^{(i)} - \overline{Y}_1)^t(Y_2^{(i)} - \overline{Y}_2) = cov(Y_2, Y_2) \end{cases}$$

Consequently,

$$\hat{\Sigma}_{11} = \frac{\overline{Y}_1}{s},\ \hat{\Sigma}_{22} = \frac{\overline{Y}_2}{s} \tag{9}$$

$$\hat{\Sigma}_{12}^2 = \frac{1}{ns}\sum_{i=1}^{n}(Y_1^{(i)} - \overline{Y}_1)^t(Y_2^{(i)} - \overline{Y}_2). \tag{10}$$

In this case, we conclude that the parameters $\Sigma_{11}$ and $\Sigma_{22}$ have the same estimators obtained by the ML and the MOM.

## *3.3 Simulations*

In order to compare the performance of the MOM estimator and the ML estimator, we propose some simulations.

We generate a random vector $Y$ according to a KBGD with different parameters

*i)* $s = 3,\ \Sigma_{11} = 0.3,\ \Sigma_{22} = 0.5,$
*ii)* $s = 3,\ \Sigma_{11} = 10,\ \Sigma_{22} = 5.$

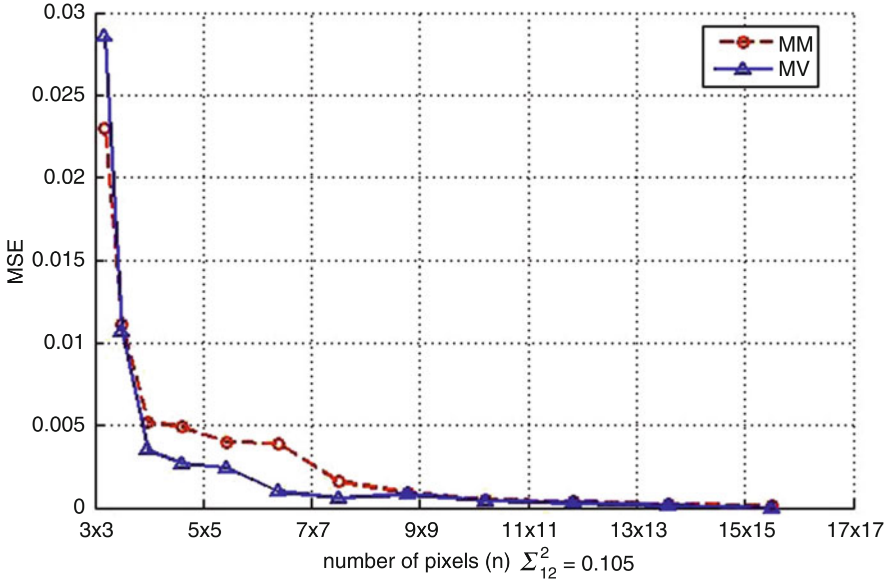


**Fig. 3** MSE versus n for parameter $\Sigma_{12}^2$ ($s = 3$, $\Sigma_{11} = 0.3$, $\Sigma_{22} = 0.5$)

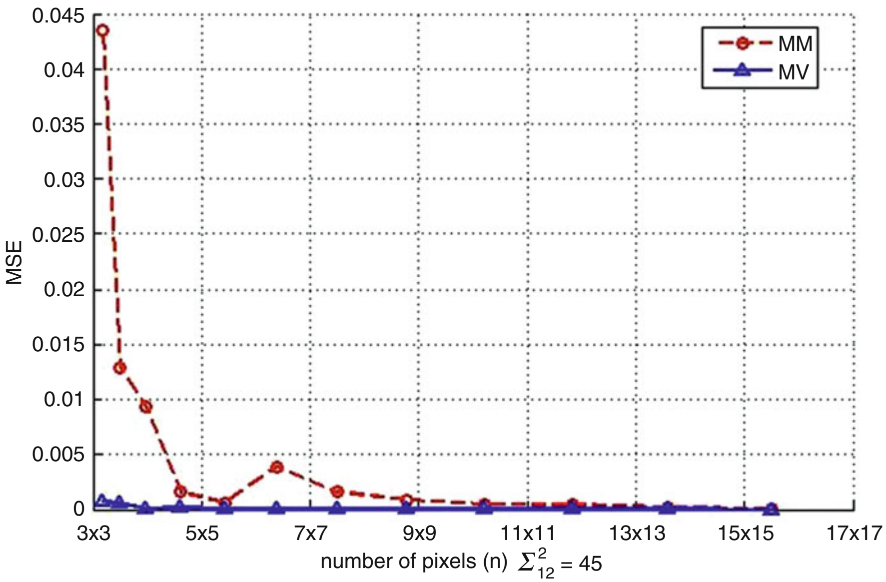


**Fig. 4** MSE versus n for parameter $\Sigma_{12}^2$ ($s = 3$, $\Sigma_{11} = 10$, $\Sigma_{22} = 5$)

The comparative study of the MOM and ML is characterized by the MSE as a function of $n$, where $n$ is the size of the sample. The number of resampling is $N = 1000$. We present, respectively, in Fig. 3 and Fig. 4 the MSE of the estimator of the parameter $\Sigma_{12}^2$ in two cases ($\Sigma_{12}^2 = 0.105$ and $\Sigma_{12}^2 = 45$). The circle curves correspond to the estimator of MOM whereas the triangle curves correspond to the estimator of ML. We observe from the two figures that the ML method is more efficient than the MOM method.

# 4 Kibble's Bivariate Gamma Segmentation Algorithm

## *4.1 Kibble's bivariate Gamma mixture model*

A crucial problem here is the choice of the mixture pdf. Generally the Gaussian distribution is considered [17], yet recent works have shown that other distributions may provide better modelling capabilities. Among these distributions that we shall consider in this work, we have the Kibble's bivariate Gamma KBG for a known value of $s$ which represents the KBGD defined in Definition II.1. It is important to note that for a known value of $s$, a KBG is fully characterized by $\Sigma = (\Sigma_{11}, \Sigma_{22}, \Sigma_{12}^2)$, where $0 < \Sigma_{12}^2 < \Sigma_{11}\Sigma_{22}$.

Since the entire image is a collection of regions, which are characterized by KBG presented in Definition II.1, it can be characterized through a K-component finite KBG and its pdf is of the form

$$\begin{aligned} f((y_1, y_2), \Phi) &= \sum_{k=1}^{K} \pi_k f((y_{1,k}, y_{2,k}); \Sigma_k) \\ &= \sum_{k=1}^{K} \pi_k \exp(-\frac{\Sigma_{22,k} y_{1,k} + \Sigma_{11,k} y_{2,k}}{\Sigma_{11,k}\Sigma_{22,k} - \Sigma_{12,k}^2}) \\ &\quad \frac{(y_{1,k} y_{2,k})^{s-1}}{(\Sigma_{11,k}\Sigma_{22,k} - \Sigma_{12,k}^2)^s \Gamma(s)} \\ &\quad \times f_s(\delta_k y_{1,k} y_{2,k}) \end{aligned}$$

where $\Phi = (\pi_1, \ldots, \pi_K, \Sigma_{11,1}, \ldots, \Sigma_{11,K}, \Sigma_{22,1}, \ldots, \Sigma_{22,K}, \Sigma_{12,1}^2, \ldots, \Sigma_{12,K}^2)$, $\delta_k = \dfrac{\Sigma_{12,k}}{\Sigma_{1.2,k}^2}$, $0 < \alpha_k < 1$ and $\sum_{k=1}^{K} \alpha_k = 1$.

## *4.2 Estimation of the parameters mixture by EM-Algorithm*

The problem of estimating the parameters which determine a mixture has been the subject of diverse studies. During the last two decades, the MML has become the most common approach to deal with this problem. Of the variety of iterative methods which have been suggested as alternatives to optimize the parameters of a mixture [3], the one most widely used is expectation maximization (EM). EM was originally proposed by [4] for estimating the ML of stochastic models. The algorithm employs an iterative procedure and the practical form is usually simple.

To obtain the estimation of the model parameters, we utilized the EM-algorithm by maximizing the expected likelihood function for carrying out the EM-algorithm. The log-likelihood function of bivariate observations $(y^1_{1,.}, y^1_{2,.}) \dots, (y^n_{1,.}, y^n_{2,.})$ drawn from an image is

$$
\begin{aligned}
l(\Phi) &= \sum_{j=1}^{n} \log\Big(\sum_{k=1}^{K} \pi_k f((y^j_{1,k}, y^j_{2,k}); \Sigma_k)\Big) \\
&= \sum_{j=1}^{n} \log\Big(\sum_{k=1}^{K} \pi_k \exp(-\frac{\Sigma_{22,k} y^j_{1,k} + \Sigma_{11,k} y^j_{2,k}}{\Sigma_{11,k}\Sigma_{22,k} - \Sigma^2_{12,k}}) \\
&\quad \frac{(y^j_{1,k} y^j_{2,k})^{s-1}}{(\Sigma_{11,k}\Sigma_{22,k} - \Sigma^2_{12,k})^s \Gamma(s)} \times f_s(\delta_k y^j_{1,k} y^j_{2,k})\Big)
\end{aligned}
$$

where $\delta_k = \dfrac{\Sigma^2_{12,k}}{(\Sigma_{11,k}\Sigma_{22,k} - \Sigma^2_{12,k})^2}$.

The model parameters

$$
\Phi = (\pi_1, \dots, \pi_K, \Sigma_{11,1}, \Sigma_{22,1}, \Sigma^2_{12,1}, \dots, \Sigma_{11,K}, \Sigma_{22,K}, \Sigma^2_{12,K})
$$

is estimated by using the EM algorithm.

The EM algorithm is decomposed in the following two steps:

**Step E:** The updated equation of the parameter $\pi_k$ is

$$
\pi_k^{(l+1)} = \frac{1}{n} \sum_{j=1}^{n} \tau_{k,j}(y^j_{1,k}, y^j_{2,k}; \Sigma^{(l)}) \tag{11}
$$

$$
= \frac{1}{n} \sum_{j=1}^{n} \frac{\pi_k^{(l)} f_k(y^j_{1,k}, y^j_{2,k}; \Sigma^{(l)})}{\sum_{k'=1}^{K} \pi_{k'}^{(l)} f_{k'}(y^j_{1,k'}, y^j_{2,k'}; \Sigma^{(l)})} \tag{12}
$$

where $f_k(y_{1,j}, y_{2,j}; \Sigma^{(l)})$ is given by Eq. 4.

**Step M:** The updated equations of $\Sigma_{11,k}$, $\Sigma_{22,k}$, and $\Sigma^2_{12,k}$ at $(l+1)^{th}$ iteration is

$$\sum_{j=1}^{n} \tau_{k,j}^{(l)}(y_{1,k}^j, y_{2,k}^j; \Sigma^{(l)})\Big[-s + \frac{\Sigma_{22,k} y_{1,k}^j}{\Sigma_{1.2,k}} - \frac{y_{2,k}^j}{\Sigma_{22,k}} + \frac{\Sigma_{11,k} y_{2,k}^j}{\Sigma_{1.2,k}} - 2y_{1,k}^j y_{2,k}^j \frac{\Sigma_{22,k}\Sigma^2_{12,k}}{\Sigma^2_{1.2,k}} \frac{f_{s+1}(\delta_k Y_{1,k}^j y_{2,k}^j)}{f_s(\delta_k y_{1,k}^j Y_{2,k}^j)}\Big] = 0, \quad (13)$$

$$\sum_{j=1}^{n} \tau_{k,j}^{(l)}(y_{1,k}^j, y_{2,k}^j; \Sigma^{(l)})\Big[-s + \frac{\Sigma_{11,k} y_{2,k}^j}{\Sigma_{1.2,k}} - \frac{y_{1,k}^j}{\Sigma_{11,k}} + \frac{\Sigma_{22,k} y_{1,k}^j}{\Sigma_{1.2,k}} - 2y_{1,k}^j Y_{2,k}^j \frac{\Sigma_{11,k}\Sigma^2_{12,k}}{\Sigma^2_{1.2,k}} \frac{f_{s+1}(\delta_k y_{1,k}^j y_{2,k}^j)}{f_s(\delta_k y_{1,k}^j y_{2,k}^j)}\Big] = 0, \quad (14)$$

$$\sum_{j=1}^{n} \tau_{k,j}^{(l)}(y_{1,k}^j, y_{2,k}^j; \Sigma^{(l)})\Big[s - \frac{\Sigma_{22,k} y_{1,k}^j + \Sigma_{11,k} y_{2,k}^j}{\Sigma_{1.2,k}} + y_{1,k}^j y_{2,k}^j \times \frac{\Sigma^2_{12,k} + \Sigma_{11,k}\Sigma_{22,k}}{\Sigma^2_{1.2,k}} \frac{f_{s+1}(\delta_k y_{1,k}^j y_{2,k}^j)}{f_s(\delta_k y_{1,k}^j y_{2,k}^j)}\Big] = 0, \quad (15)$$

where $f_{s_1}(x) = \sum_{l=0}^{\infty} \frac{x^l}{\Gamma(s_1+l)l!}$, $\forall\, x \in \mathbb{R}, f_{s+1} = f_s'$, et $\Sigma_{1.2,k} = \Sigma_{11,k}\Sigma_{22,k} - \Sigma^2_{12,k}$. A closed-form solution does not exist for the $\Sigma_{11,k}$, $\Sigma_{22,k}$ and $\Sigma^2_{12,k}$, $k = 1, 2, \ldots, K$, parameters. We can get a solution by using a Newton-Raphson procedure initialized by estimator $(\Sigma_{11,k}, \Sigma_{22,k}, \Sigma^2_{12,k})$ defined in Eq. 9 and Eq. 10. The convergence of the Newton-Raphson procedure is generally obtained after few iterations.

The efficiency of the EM-algorithm in estimating the parameters is heavily dependent on the number of regions in the image. The number of mixture components initially taken for K-means algorithm is by plotting the histogram of the pixel intensities of the two images. The number of peaks in the histogram can be taken as the initial value of the number of regions K. Usually the mixing parameter $\pi_k$ and the region parameters $(\Sigma_{11,k}, \Sigma_{22,k}, \Sigma^2_{12,k})$ are unknown. A commonly used method in initialization is by drawing a random sample in the entire image data. This method performs well only when the sample size is large, and the computation time is heavily increased. When the sample size is small it is likely that some small regions may not be sampled. To overcome this problem, we use K-means algorithm [16] to divide the whole images into various homogeneous regions. We obtain the initial estimates of the parameters $(\Sigma_{11,k}, \Sigma_{22,k}, \Sigma^2_{12,k})$ for each image region using the method of moment estimators for KBGD and for the parameters $\pi_k$ as $\pi_k = \frac{1}{K}$. Therefore the initial estimates of $(\Sigma_{11,k}, \Sigma_{22,k}, \Sigma^2_{12,k})$ can be obtained by the MOM presented in section III.B.

### 4.3 Application of the KBG mixture in segmentation

After estimating the parameters of the KBG mixture the prime step in image segmentation is allocating the pixels to the segments of the image. The image segmentation steps are the following:

**Step 1)** Plot the histogram of the pixel intensities of the two images.
**Step 2)** Obtain the initial estimates of the model parameters using K-means algorithm and moment estimators as discussed in section III.
**Step 3)** Obtain the refined estimates of the model parameters by using the EM-algorithm with the updated equations given in section IV.B.
**Step 4)** Assign each pixel into the corresponding $k^{th}$ region (segment). That is,

$$j(y) = \arg \max_{1 \le k \le K} (\pi_k f_k(y)).$$

where $j(y)$ represents the label of the class of the pixel $y = (y_{1,.}, y_{2,.})$.

## 5 Experimentation

In order to demonstrate the utility of the image segmentation algorithm developed in this paper, an experiment is conducted with two colors satellites images (sat 1 and sat 2) which the size is $256 \times 256 \times 3$. A random sample of this images is taken the feature vector consisting of HS for each pixel of the each image is computed utilizing HSI color space. In HSI color space the HS values are computed from the values of RGB for each pixel in the image using the formula for

$$H = \cos^{-1}\left(\frac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}}\right),\ B < G,$$

$$S = 1 - \frac{\min(R, G, B)}{I},\ \text{where } I = \frac{R+G+B}{3}.$$

With the feature vector $(H, S)$ each image is modelled by using the two-component mixture KBG. The number of segments in each of the four colors images considered for experimentation is determined by the histogram of pixel intensities. The histograms of the two images are shown in Fig. 5.

The initial estimates of the number of regions $K$ in each image are obtained and given in Table 1.

**Table 1** Initial Estimates of $K$

| Image | sat 1 | sat 2 |
|---|---|---|
| Estimate of $K$ | 3 | 3 |

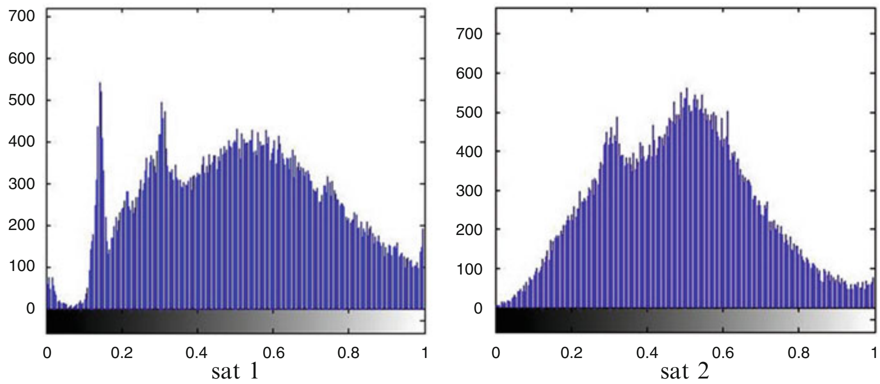


**Fig. 5** Histograms of the pixel intensities of the two images

**Table 2** Estimated of Mixture parameters for image sat 1, after 10 iterations, with known parameter value $s$

| | Initial by K-means | | | Final by EM Algorithm | | |
|---|---|---|---|---|---|---|
| Parameters | $k_1$ | $k_2$ | $k_3$ | $k_1$ | $k_2$ | $k_3$ |
| KBGM | | | | | | |
| s | 9 | 5 | 2 | | | |
| $\pi_i$ | 1/3 | 1/3 | 1/3 | 0.0819 | 0.5768 | 0.3413 |
| $\Sigma_{11}$ | 0.17 | 0.0528 | 0.0939 | 1.1879 | 2.3531 | 1.4263 |
| $\Sigma_{22}$ | 0.19 | 0.0171 | 0.0239 | 0.0031 | 0.0354 | 0.0332 |
| $\Sigma_{12}^2$ | 0.003 | 0.0003 | 0.0014 | 0.0027359 | 0.0010 | 0.0290 |
| BGAUM | | | | | | |
| $\pi_i$ | 1/3 | 1/3 | 1/3 | 0.9363 | 0.0038 | 0.0017 |
| $\mu_1$ | 0.0152 | 0.2642 | 0.1878 | 0.0091 | 0.3657 | 0.1106 |
| $\mu_2$ | 0.0175 | 0.0857 | 0.0477 | 0.0162 | 0.0607 | 0.0001 |
| $\sigma_1$ | 0.0013 | 0.0596 | 0.1178 | 0.6613 | 0.0482 | 0.0675 |
| $\sigma_2$ | 0.0005 | 0.0083 | 0.0217 | 0.4483 | 0.0018 | $10^{-4}$ |
| $\rho$ | 0.0003 | 0.0016 | 0.0278 | 0.1807 | 0.0001 | $10^{-5}$ |

After assigning these initial values of $K$ to each image data set, the $K$-means algorithm is performed. Using these initial parameters estimates $(\pi_k, \mu_{1,k}, \mu_{2,k}, \sigma_{1,k}, \sigma_{2,k}, \rho_k)$ for a bivariate Gaussian mixture (BGAUM) and $(\pi_k, \Sigma_{11,k}, \Sigma_{22,k}, \Sigma_{1.2,k})$ for the Kibble's bivariate Gamma mixture (KBGM), where $k = 1, \ldots, K$, by the $K$-means algorithm to the EM algorithm. The computed values of the initial estimates and the final estimates of the two models parameters for each image are shown in Tables 2 and 3.

Substituting the final estimates of the model parameters, the pdfs of the feature vector of each image are estimated. Using the estimated pdfs and the image segmentation algorithm given in sub-section IV.C, the image segmentation is achieved for each of the two color images under consideration. After developing

**Table 3** Estimated of Mixture parameters for image sat 2, after 10 iterations, with known parameter value *s*

| | Initial by K-means | | | Final by EM Algorithm | | |
|---|---|---|---|---|---|---|
| Parameters | $k_1$ | $k_2$ | $k_3$ | $k_1$ | $k_2$ | $k_3$ |
| KBGM | | | | | | |
| s | 9 | 4 | 2 | | | |
| $\pi_i$ | 1/3 | 1/3 | 1/3 | $10^{-5}$ | 0.6895 | 0.3105 |
| $\Sigma_{11}$ | 0.015 | 0.0452 | 0.0425 | 6.4627 | 1.1649 | 1.6137 |
| $\Sigma_{22}$ | 0.29 | 0.0442 | 0.0498 | 0.0034 | 0.0105 | 0.0521 |
| $\Sigma_{12}^2$ | 0.0005 | 0.0018 | 0.0017 | 0.0021 | 0.0013 | 0.0487 |
| BGAUM | | | | | | |
| $\pi_i$ | 1/3 | 1/3 | 1/3 | 0.9295 | 0.0314 | 0.0392 |
| $\mu_1$ | 0.0131 | 0.1808 | 0.085 | 0.011 | 0.0998 | 0.0001 |
| $\mu_2$ | 0.026 | 0.1941 | 0.0996 | 0.0225 | 0.2174 | 0.78 |
| $\sigma_1$ | 0.004 | 0.0443 | 0.075 | 0.0003 | 0.0147 | 0.001 |
| $\sigma_2$ | 0.0025 | 0.0185 | 0.0437 | 0.0021 | 0.0092 | 0.3367 |
| $\rho$ | 0.0005 | 0.007 | 0.0349 | 0.0005 | 0.0003 | $10^{-5}$ |

**Table 4** Comparative study of Image Quality Metrics from MBGM and MBGAUM

| Images | Quality Metrics | KBGM | BGAUM |
|---|---|---|---|
| sat 1 | MSE | 0.4101 | 0.8402 |
| | PSNR | 8.9124 | 1.7417 |
| | M D | 0.5499 | 0.8915 |
| sat 2 | MSE | 0.5928 | 0.9516 |
| | PSNR | 5.2288 | 0.4964 |
| | M D | 0.6194 | 0.9313 |

the image segmentation method, it is necessary to verify the utility of segmentation in model building of the image for image retrieval. Using the estimated pdfs of the images under consideration, we are able to get the retrieved images. The original, segmented, and retrieved images are shown in Fig. A.1 and Fig. A.2 (see Appendix).

The performance evaluation of the retrieved images is made by Subjective Image Quality (SIQ) testing or by Objective Image Quality (OIQ) testing. The OIQ testing methods are often used since the numerical results of an objective measure are readily computed and allow a significant comparison between the different algorithms. There are SIQ measures available for performance evaluation of the image segmentation method. An extensive survey of quality measures is given by [5]. For the performance evaluation of the developed segmentation algorithm, we consider the Image Quality Measures [5], namely MSE, Peak Signal to Noise Ratio (PSNR) and Maximum Distance (MD), which are computed for the two images with respect to the developed method and earlier methods and are presented in Table 4.

From the Table 4, we notice that all the image quality metrics for the two images are satisfying the standard criteria. This implies that using the proposed algorithm allows the images to be retrieved accurately. A comparative study of the proposed algorithm, with the ones based on the KBGM and BGAUM models with $K$-means, reveals that the MSE of the proposed model KBGM is lower than the one associated with BGAUM. Concerning all other quality metrics, it is also observed that the performance of the proposed model in retrieving the images is again better when compared with the BGAUM.

## 6 Conclusion

In this paper, we have proposed a segmentation algorithm adapted to color image by the use of the Kibble's bivariate Gamma distribution. Here, it is assumed that the color image is characterized by the HSI color space, in which the HS values are nonnegative. The model parameters are estimated using the EM-algorithm. The initialization and the number of image segments are both determined through the $K$-means algorithm and the Moment Method of estimation. Experimental results have demonstrated the efficiency of the proposed method KBGM.

## Appendix

See Figs. A.1 and A.2.

## References

1. Belongie S., et. al., *Color and texture-based image segmentation using EM and its application to content-based image retrieval*, Proc. of ICCV, pp. 675–82, 1998.
2. Chao-Wei Chou, Wen-Jang Huangb, *A note on characterizations of the bivariate gamma distribution*, Journal of Statistical Planning and Inference, 2003.
3. Delignon Y., et. al., *Estimation of generalized mixtures and its application in image segmentation*, IEEE Trans. on Image Processing, vol. 6, no. 10, pp. 1364–76, 1997.
4. Dempster A., Laird N., and Rubin D., *Maximum Likelihood From Incomplete Data via the EM Algorithm*, J. Royal Statistical Soc., Series B, pp. 1–38, 1977.
5. Eskicioglu A.M. and Fisher P.S, *Image Quality Measures and their Performance*, IEEE Transactions On comm., vol.43, No.12, pp.2959–2965, 1995.
6. Fadilah, Norasyikin, Junita Mohamad-Saleh, Zaini Abdul Halim, Haidi Ibrahim, and Syed Syed Ali. *Intelligent Color Vision System for Ripeness Classification of Oil Palm Fresh Fruit Bunch*, Sensors, 2012.
7. Griffin L.D, Colchester A.C.F, Robinson G.P, *Scale and segmentation of grey-level images using maximum gradient paths*, Image and Vision Computing, vol 10, pp. 389–402, 1992.
8. Hadrich, Atizez, Mourad Zribi, and Afif Masmoudi. *A proposed normalized B-spline density estimator and it application in unsupervised statistical image segmentation*, 2012 IEEE International Conference on Systems Man and Cybernetics (SMC), 2012.
9. Haritha, D., Rao, K. Srinivasa and Satyanaraya, Ch.*Face Recognition Algorithm Based on Doubly Truncated Gaussian Mixture Model Using Hierarchical Clustering Algorithm*, International Journal of Computer Science Issues (IJCSI), 2012.

10. B.N.Jagadesh and al, A Robust Skin Colour Segmentation Using Bivariate Pearson Type II (Bivariate Beta) Mixture Model, *I.J. Image, Graphics and Signal Processing*, 2012.
11. Rajkumar, G. V. S.*Image Segmentation Method Based On Finite Doubly Truncated Bivariate Gaussian Mixture Model with Hierarchical Clustering*, International Journal of Computer Science Issues (IJCSI)/16940784, 20110701.
12. Seshashayee, M., Rao, K. Srinivas; Satyanarayana, Ch. and Rao, P. Srinivasa. *Image Segmentation Based on a Finite Generalized New Symmetric Mixture Model with K-Means*, International Journal of Computer Science Issues (IJCSI), 2011.
13. Taher Ben Arab, Mourad Zribi and Afif Masmoudi, Parameter Estimation Of The Diagonal Of The Modified Riesz Distribution, *IEEE 17th International Conference on Intelligent Engineering Systems*, June 19–21, 2013 in Costa Rica 2013.
14. Taher Ben Arab, Mourad Zribi and Afif Masmoudi. Unsupervised Satellite Image Segmentation using a Bivariate Beta type-II Mixture Model. *IEEE 9th International Conference on Computational Cybernetics*, Page(s): 95–100, 2013.
15. Tie Qi Chen. *Color Image Segmentation in Color and Spatial Domain*, Lecture Notes in Computer Science, 2003
16. Vlassis, N., *The global k-means clustering algorithm*, Pattern Recognition, vol 36, pp.451–461,2003.
17. Vlassis A. M, and Verbeek J., and Klose B., *Greedy gaussian mixture model learning for tetxture image segmentation*, Proceedings of theWorkshop on kernel and Subspace Methods for Computer Vision, pp. 37–46, August 2001.
18. G.V.S. Rajkumar, K.Srinivasa Rao, P.Srinivasa Rao, Studies on Colour Image Segmentation T echniqu e Based on Finite Left Truncated Bivariate Gaussian Mixture Model with K-Means, *Global Journal of Computer Science and Technology* , Volume 11 Issue 18 Version 1.0 October 2011.
19. T.J Ypma, Historical developement of the Newton-Raphson method. *SIAM Review*, 37, 531–551, 1995.

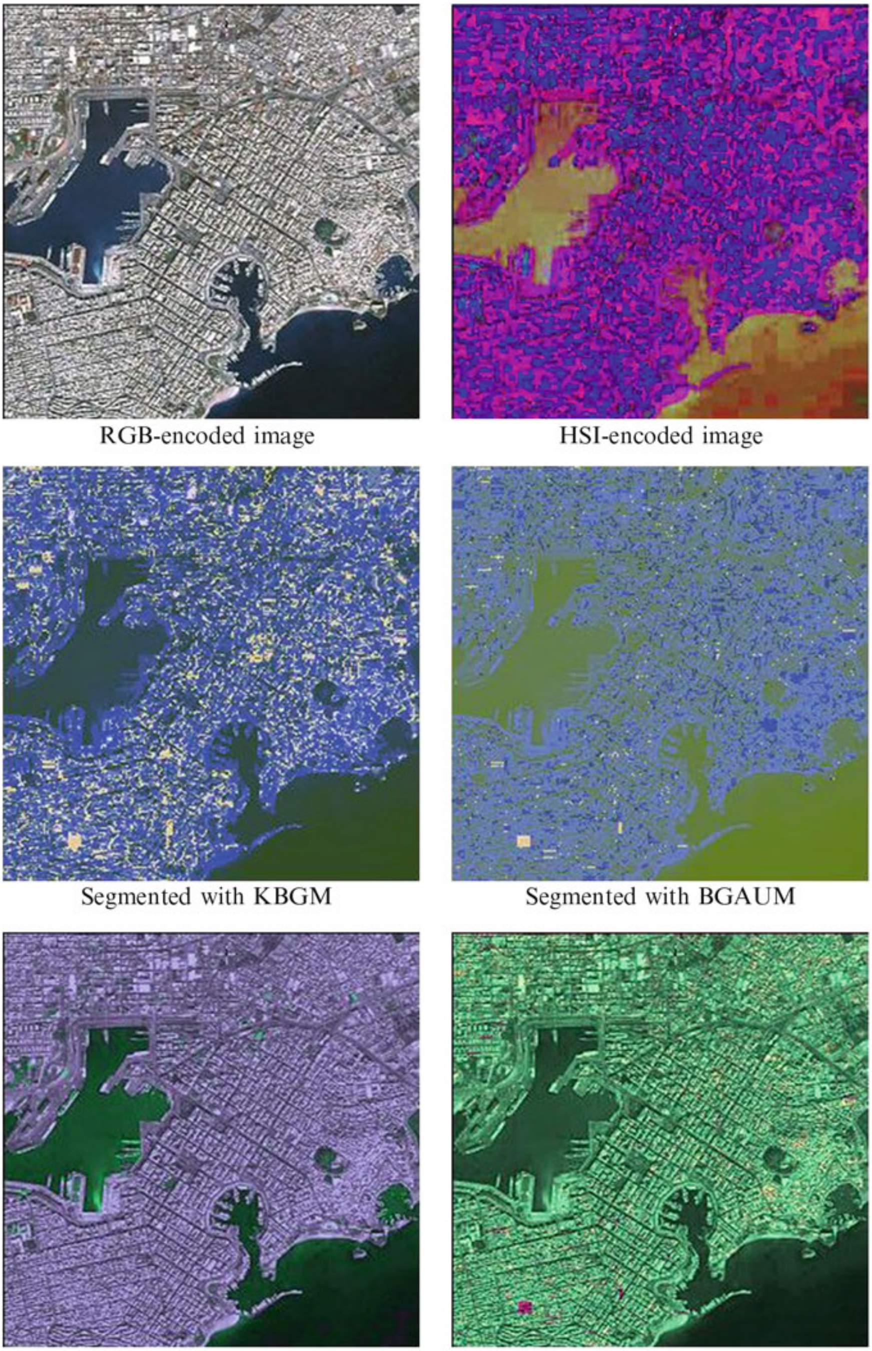


**Fig. A.1** Sat 1 segmented with KBGM and BGAUM

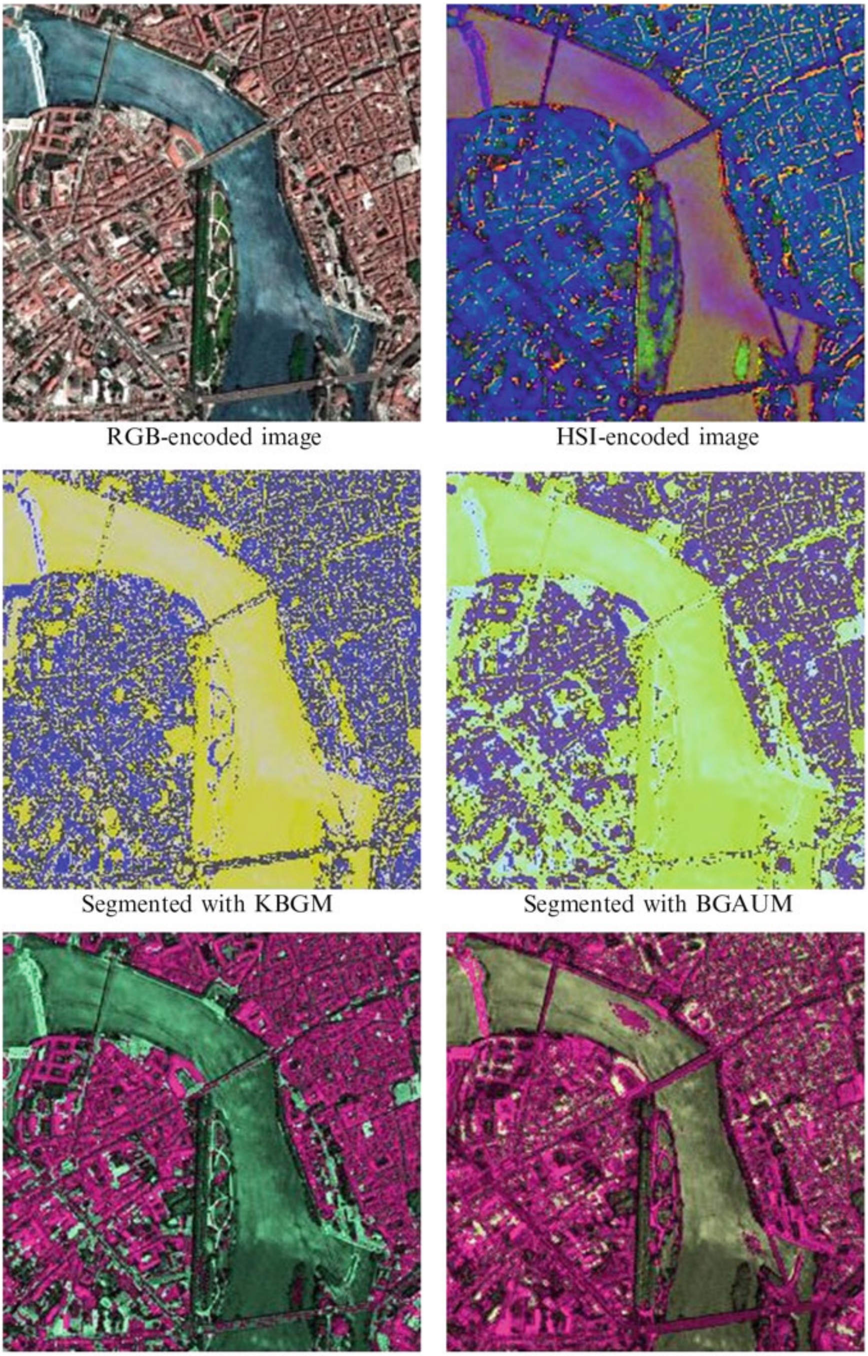


**Fig. A.2** Sat 2 segmented with KBGM and BGAUM

# Stabilization of a class of time-varying systems in Hilbert spaces

**Hanen Damak and Mohamed Ali Hammami**

**Abstract** This paper studies the problem of stabilization of the infinite-dimension time-varying systems in Hilbert space where the associated nominal system is a certain class of linear time-varying systems and the perturbation term satisfies some certain conditions. In contrast to the previous results, the stabilizability conditions are obtained by solving a Ricatti differential equation and do not involve any stability property of the evolution operator. Our goal is to prove the sufficient conditions for the case of uniform exponential stability of the origin. The obtained result extends existing results in the literature to infinite-dimensional and time-varying control systems.

## 1 Introduction

The problem of controllability and stabilizability for linear control systems has received a considerable amount of interest in the past decades [7–11] and the references therein. This problem regarding as an extension of the classical Kalman result [3] on controllability and stability of linear control systems is to find an admissible control $u(t)$, such that the corresponding solution $x(t)$ of the system has desired properties. Depending on the properties involved one defines various quantitative problems. For time-varying control systems in finite-dimensional spaces, using Kalman's decomposition method, in [2] we prove that the system is completely stabilizable if it is uniformly globally null-controllable. In [6], we develop the relationship between the exact controllability and complete stabilizability for linear time-varying control systems in Hilbert spaces.

In this paper, we discuss the problem of global uniform stabilizability for a class of nonlinear dynamical systems in Hilbert spaces, it means that the solutions converge exponentially to the origin.

---

H. Damak (✉)
University of Monastir Tunisia, Monastir, Tunisia
e-mail: hanen.damak@yahoo.fr

M.A. Hammami
University of Sfax Tunisia, Sfax, Tunisia
e-mail: MohamedAli.Hammami@fss.rnu.tn

 
A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_16

## 2 Stabilization of a class of time-varying systems in Hilbert spaces

We will use the following notation throughout this paper: $X$ denotes a Hilbert space with the norm $\|.\|_X$ and inner product $< .,. >_X$ . $L(X)$ (respectively, $L(X,Y)$) denotes the Banach space of all linear bounded operators $S$ mapping $X$ into $X$ (respectively, $X$ into $Y$) endowed with the norm

$$\|S\| = \sup\{\|Sx\| : x \in X, \|x\| \leq 1\}.$$

$X^*$ denotes the dual space of $X$; $L_2([t,s],X)$ denotes the set of all strongly measurable $L_2$-integrable and $X$-valued functions on $[t,s]$; $D(A)$, $A^{-1}$ and $A^*$ denote the domain, the inverse and the adjoint of the operator $A$, respectively; $clM$ denotes the closure of a set $M$; $I$ the identity operator; $LO([t,+\infty[,X^+)$ denotes the set of all linear bounded self-adjoint non-negative definite operator-valued function on $[t,+\infty[$. Let $X$, $U$ be real Hilbert spaces.

We consider the control dynamical system

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) + F(t,x(t)) \\ x(t_0) = x_0 \end{cases} \tag{1}$$

where $x(t) \in X$ is the system state, $u(t) \in U$ is the control input, $A(t) : X \longrightarrow X$ and $B(t) \in L(U,X)$. We assume that $F(t,x) : [0,+\infty[\times X \longrightarrow X$ is a nonlinear operator, continuous and satisfying the following inequality

$$\|F(t,x)\| \leq \gamma(t)\|x\|, \quad \forall t \geq 0, \quad \forall x \in X$$

where $\gamma : [0,+\infty[\longrightarrow \mathbb{R}$ is continuous non-negative function with

$$\int_0^{+\infty} \gamma(s)ds \leq M_\gamma < +\infty.$$

The corresponding nominal system is described by

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) \\ x(t_0) = x_0 \end{cases} \tag{2}$$

In the sequel, we say that control $u(t)$ is admissible if $u(t) \in L_2([t_0,+\infty[,U)$. As in [1] we will assume the following conditions that guarantee the existence and uniqueness of the solution of linear control system (2).

$\mathscr{H}1)$ Operator functions $A(.)x$ and $B(.)u$ are continuous and bounded in $t \geq t_0 \geq 0$ for all $x \in X$, $u \in U$;

$\mathscr{H}2)$ $clD(A(t)) = X$, $t \geq 0$ and $A(.)x$ is a continuous function on $[t_0, +\infty[$ for every $x \in D(A(.))$.

$\mathscr{H}3)$ For each $t \geq t_0 \geq 0$, $A(t)$ generates a $C_0$-semigroup on $X$ and there is an evolution operator $U(t, s) : \{(t, s) : t \geq s \geq t_0 \geq 0\} \longrightarrow L(X)$, such that $U^*(t, s)$ is continuous in $t, s$ and for each $x \in D(A(t))$, $U(t, s)x \in D(A(t))$ the following conditions holds:

*i)* $\dfrac{\partial U(t,s)x}{\partial t} = A(t)U(t,s)x, \quad U(s,s) = I,$

$$\lim_{n\to+\infty} U_n(t,s)x = U(t,s)x,$$

where $U_n(t, s)$ is the evolution operator generated by the Yosida approximation [5]

$$A_n(t) = n^2[nI - A(t)]^{-1} - nI$$

of $A(t)$.

*ii)* $U(t, s) = U(t, r)U(r, s)$, for all $t \geq r \geq s \geq t_0 \geq 0$.

In this case, we say that $A(t)$ generates a strongly continuous evolution operator $U(t, s)$ and then for every initial state $x_0 \in X$, for every admissible control $u(t)$, the linear control system (2) has a mild solution given by

$$x(t) = U(t, t_0)x_0 + \int_{t_0}^{t} U(t, s)B(s)u(s)ds.$$

*Remark 1.* It is well known that if the operator $A(t) \in L(X), t \geq 0$, which is bounded in $[0, +\infty[$, then the semigroup evolution operator $U(t, s)$ satisfying the above conditions always exists. However, if $A(t), \quad t \in [0, +\infty[$ is unbounded, the evolution operator $U(t, s)$ exists provided additional assumptions, see [1, 5] for the details.

**Definition 1.** An Operator $Q(t) \in L(X)$ is called definite positive if there exist a positive constant $m > 0$, such that

$$< Q(t)x, x >\geq m\|x\|^2, \quad \forall x \in X, \quad \forall t \geq 0.$$

**Definition 2.** The system $[A(t), B(t)]$ is called globally null-controllable (GNC) in finite time if for every $x_0 \in X$, there exist a number $T > 0$ and an admissible control $u(t)$, such that

$$U(T, t_0)x_0 + \int_{t_0}^{t} U(T, s)B(s)u(s)ds = 0.$$

We state the following well-known controllability criterion for infinite-dimensional control system that be used later.

**Proposition 1 ([1, 2]).** *The system* $[A(t), B(t)]$ *is globally null-controllable in time* $T > 0$ *if and only if one of the following conditions hold:*

($i$) *There is a number* $c > 0$, *such that*

$$\int_{t_0}^{t} \|B^*(s)U^*(t,s)x\|^2 ds \geq c\|U^*(T,t_0)x\|^2, \quad \forall x \in X.$$

($ii$) $\int_{t_0}^{t} U(T,s)B(s)B^*(s)E^*(T,s)ds$ *is positive definite.*

**Definition 3.** The system (2) is called completely stabilizable (CSz) if for every number $\delta > 0$, there exists a feedback control $u(t) = K(t)x(t)$, where $K(t) \in L(X,U)$ is bounded on $[t_0, +\infty[$, such that the solution $x(t, x_0)$ of the closed-loop system $\dot{x}(t) = [A(t) + B(t)K(t)]x(t)$, $x(t_0) = x_0$ satisfies

$$\exists N > 0 : \ \|x(t)\| \leq k\|x(t_0)\|e^{(-\delta(t - t_0))}, \quad \forall t \geq t_0.$$

The solution to the stabilizability problem involves a Riccati operator equation (ROE) of the form

$$\dot{P}(t) + A^*(t)P(t) + P(t)A(t) - P(t)B(t)B^*(t)P(t) + Q(t) = 0 \tag{3}$$

where $Q(t) \geq 0$ is a given self-adjoint operator function. Since $A(t), t \geq t_0 \geq 0$ is an unbounded operator, it is not clear a priori what a solution of (ROE) means. We will define, as in [1], the solution of (ROE) as follows.

**Definition 4.** The solution of (ROE) (3) is a linear operator function $P(t) \in L(X)$ satisfying the following two conditions:

$i$) The scalar function $< P(.)x, y >$ is differentiable on $[t_0, +\infty[$ for every $x, y \in D(A(.))$.

$ii$) For all $x, y \in D(A(t)), t \geq t_0 \geq 0$ :

$$\frac{d}{dt} < P(t)x, y > + < P(t)x, A(t)y > + < P(t)A(t)x, y >$$
$$- < P(t)B(t)B(t)P(t)x, y > + < Q(t)x, y >= 0.$$

**Definition 5.** Let $Q(t) \in LO([t_0, +\infty[, X^+)$. The control system (2) is called $Q(t)-$stabilizable if for every initial state $x_0$, there is a control $u(t) \in L_2([t_0, +\infty[, U)$ such that the cost function

$$J(u) = \int_{t_0}^{\infty} [\|u(t)\|^2 + < Q(t)x(t, x_0), x(t, x_0) >]dt, \tag{4}$$

exists and is finite.

**Proposition 2 ([1]).** *If linear control system (2) is $Q(t)$-stabilizable, then the Riccati differential equation (ROE) (3) has the solution $P(t) \in LO([t_0, +\infty[, X^+)$ bounded on $[t_0, +\infty[$.*

**Definition 6.** The system (1) is called uniformly exponentially stable if there exists an operator function $K(t) \in L(X, U)$, such that the solution of the closed-loop system $\dot{x}(t) = [A(t) + B(t)K(t)]x(t) + F(t, x)$ satisfies

$$\|x(t)\| \leq k\|x(t_0)\|e^{(-\gamma(t - t_0))},$$

where $\gamma > 0$, $k \geq 0$, $t_0 \in [0, +\infty[$.

**Proposition 3 ([4]).** *If the system $[A(t), B(t)]$ is globally null-controllable (GNC) in finite time, then the linear time-varying control system (2) is completely stabilizable (CSz).*

**Theorem 1.** *Under assumptions $\mathscr{H}1), \mathscr{H}2)$ and $\mathscr{H}3)$, if the system $[A(t), B(t)]$ is globally null-controllable (GNC) in finite time, then the system (1) is globally uniformly exponentially stable.*

*Proof.* We choose an operator function $Q(t) \in LO([t_0, +\infty[, X^+)$ bounded on $[0, +\infty[$, such that

$$Q(t) > 2A(t) + B(t)B^*(t), \quad t \geq 0.$$

Then, there exists $c_1 > 0$, such that

$$< Q(t) - 2A(t) + B(t)B^*(t)x, x > \geq c_1\|x\|^2, \quad t \geq 0.$$

The system $[A(t), B(t)]$ is globally null-controllable (GNC) in finite time. Then, for every initial state $x_0 \in X$ there is an admissible control $u(t) \in L_2([t_0, T], U)$, such that the solution $x(t)$ of the system (2), according to the control $u(t)$ satisfies

$$x(t_0) = x_0, \quad x(T) = 0.$$

Let $u_x(t)$ denote an admissible control according to the solution $x(t)$ of the system. Define

$$\tilde{u}_x(t) = \begin{cases} u_x(t) & if \ \ t \in [t_0, T], \\ 0 \ \ if & t > T\cdot \end{cases}$$

Since $Q(t) \in LO([t_0, +\infty[, X^+)$ and $\tilde{u}_x \in L_2([t_0, T], U)$, we have

$$J(\tilde{u}_x) = \int_{t_0}^{\infty} [\|\tilde{u}_x(t)\|^2 + < Q(t)x(t, x_0), x(t, x_0) >]dt$$

$$= \int_{t_0}^{T} [\|u_x(t)\|^2 + < Q(t)x(t,x_0), x(t,x_0) >]dt < +\infty,$$

which means that the linear control system $[A(t), B(t)]$ is $Q(t)$-stabilizable. Applying Proposition 2, we can find an operator function $P(t) \in LO([t_0, +\infty[, X^+)$, which is a solution of the following (ROE)

$$\dot{P}(t) + A^*(t)P(t) + P(t)A(t) - P(t)B(t)B^*(t)P(t) + Q(t) = 0. \tag{5}$$

We now consider a Lyapunov-like function

$$V(t,x) =< P(t)x, x > + \|x\|^2,$$

and construct a feedback control of the form

$$u(t) = -\frac{1}{2}B^*(t)[P(t) - I]x(t), \quad t \geq t_0.$$

Since $P(t) \in LO([t_0, +\infty[, X^+)$, then there exists $M > 0$, such that

$$\|P(t)\| \leq M, \quad \forall t \geq t_0.$$

It is easy to verify that

$$\|x(t)\|^2 \leq V(t,x(t)) \leq (M+1)\|x(t)\|^2, \quad \forall t \geq t_0.$$

Taking the derivative of $V$ in $t$ along the solution of $x(t)$ of the system (1) and using the chosen feedback control and the (ROE) (5), we have

$$\dot{V}(t,x(t)) =< \dot{P}(t)x(t), x(t) > +2 < P(t)\dot{x}(t), x(t) > +2 < \dot{x}(t), x(t) > .$$

We obtain

$$\dot{V}(t,x(t)) \leq -\left(\frac{c_1}{M+1} - 2(M+1)\gamma(t)\right) V(t,x(t)).$$

Then, $V(t)$ satisfies the following estimation

$$V(t) \leq V(t_0)e^{-\int_{t_0}^{t} \alpha(s)\, ds}$$

with

$$\alpha(t) = \frac{c_1}{(M+1)} - 2(M+1)\delta(t).$$

Moreover,

$$e^{-\int_{t_0}^{t} \alpha(s)\, ds} \leq e^{2(M+1)M_\gamma} e^{-\frac{c_1}{M+1}(t-t_0)}.$$

Thus, we obtain

$$V(t) \leq V(t_0) e^{2(M+1)M_\gamma} e^{-\frac{c_1}{M+1}(t-t_0)}.$$

It follows that

$$\|x(t)\| \leq \sqrt{(M+1)} e^{(M+1)M_\gamma} \|x(t_0)\| e^{-\frac{c_1}{2(M+1)}(t-t_0)}.$$

Hence, the system (1) in closed-loop with the linear feedback $u(t) = K(t)x(t)$ is globally uniformly exponentially stable. □

**Acknowledgements** The authors thank the referees for their careful reading of the manuscript and insightful comments.

## References

1. Bensoussan, A., Da Prato, G., Delfour, M.C, Mitter, S.K.: Representation and Control of Infinite Dimensional Systems. Birkhauser, (1992)
2. Ikeda, M., Maed, H., Kodama, S.: Stabilization of linear systems. SIAM J. on control. **10**, 716–729 (1972)
3. Kalman, R.E., Ho, Y.C., Narenda, K.S.: Controllability of linear dynamical systems. Differential Equations. **1**, 189–213 (1963)
4. Niamsup, P., Phat, V.N.: Linear time varying systems in Hilbert spaces: Exact controllability implies complete stabilizability. Thai J. Math. **7**, 189–199 (2009)
5. Pazy, A., Semigroup of Linear Operators and Applications to Partial Differential Equations. Springer, New York (1983)
6. Phat, V.N., Ha, Q.P.: New characterization of stabilizability via Riccati equations for LTV systems. IMA J. Math. Contr. Inform. **25**, 419–429 (2008)
7. Phat, V.N.: Stabilization of linear continuous time-varying systems with state delays in Hilbert spaces. Electro J. Differ. Equ. **2001**, 1–13 (2001)
8. Phat, V.N., Constrained Control Problems of Discrete Processes. World Scientific, Singapore (1998)
9. Phat, V.N.: New stabilization criteria for linear time-varying systems with state delay and normed bounded uncertainties. IEEE Trans. Aut. Control. **12**, 2095–2098 (2002)
10. Wonham, W.M., Linear Multivariable Control: A Geometric Approach. Springer-Verlag, Berlin (1979)
11. Zabzyk, J., Mathematical Control Theory: An Introduction. Birkhauzer, Boston (1992)

# Weighted Sobolev Spaces for the Laplace Equation in an Exterior Domain

**Hela Louati**

**Abstract** This work solves Dirichlet problem for the Laplace operator in an exterior domain of $\mathbb{R}^3$. We are interested in the existence and the uniqueness of weak and strong solutions in $L^p$-theory which makes our work more difficult. Our analysis is based on the principle that linear exterior problems can be solved by combining their properties in the whole space $\mathbb{R}^3$ and the properties in bounded domains. Our approach rests on the use of weighted Sobolev spaces.

## 1 Introduction

Let $\Omega'$ be a bounded connected open set in $\mathbb{R}^3$ with boundary $\partial\Omega' = \Gamma$ of class $C^{1,1}$ representing an obstacle and let $\Omega$ its complement which means that $\Omega = \mathbb{R}^3 \setminus \overline{\Omega'}$. Let $\mathbf{x} = (x_1, x_2, x_3)$ be a typical point in $\mathbb{R}^3$ and let $r = |\mathbf{x}| = (x_1^2 + x_2^2 + x_3^2)^{1/2}$ denote its distance to the origin. In order to control the behavior at infinity of our functions and distributions we use for basic weights the quantity $\rho(\mathbf{x}) = (1 + r^2)^{1/2}$ which is equivalent to $r$ at infinity, and to one on any bounded subset of $\mathbb{R}^3$ and the quantity $\ln(2 + r^2)$. We define $\mathscr{D}(\Omega)$ to be the linear space of infinite differentiable functions with compact support on $\Omega$. Now, let $\mathscr{D}'(\Omega)$ denote the dual space of $\mathscr{D}(\Omega)$, often called the space of distributions on $\Omega$. We denote by $< .,. >$ the duality pairing between $\mathscr{D}'(\Omega)$ and $\mathscr{D}(\Omega)$. For each $p \in \mathbb{R}$ and $1 < p < \infty$, the conjugate exponent $p'$ is given by the relation $\frac{1}{p} + \frac{1}{p'} = 1$. Then, for any nonnegative integers $m$ and real numbers $p > 1$ and $\alpha$, setting

$$k = k(m, p, \alpha) = \begin{cases} -1, & \text{if } \frac{3}{p} + \alpha \notin \{1, \ldots, m\}, \\ m - \frac{3}{p} - \alpha, & \text{if } \frac{3}{p} + \alpha \in \{1, \ldots, m\}, \end{cases}$$

H. Louati (✉)
University of Sfax, Sfax, Tunisia
e-mail: hela.louati@gmail.com



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_17

we define the following space:

$$\begin{aligned} W_{\alpha}^{m,p}(\Omega) = \{u \in \mathscr{D}'(\Omega);\\ &\forall\, \lambda \in \mathbb{N}^3 :\ 0 \leq |\lambda| \leq k,\ \rho^{\alpha-m+|\lambda|}(\ln(2+r^2))^{-1}D^{\lambda}u \in L^p(\Omega);\\ &\forall\, \lambda \in \mathbb{N}^3 :\ k+1 \leq |\lambda| \leq m, \rho^{\alpha-m+|\lambda|}D^{\lambda}u \in L^p(\Omega)\}. \end{aligned}$$

It is a reflexive Banach space equipped with its natural norm:

$$\begin{aligned} ||u||_{W_{\alpha}^{m,p}(\Omega)} = \Bigg( &\sum_{0\leq|\lambda|\leq k} ||\rho^{\alpha-m+|\lambda|}(\ln(2+r^2))^{-1}D^{\lambda}u||^p_{L^p(\Omega)}\\ &+ \sum_{k+1\leq|\lambda|\leq m} ||\rho^{\alpha-m+|\lambda|}D^{\lambda}u||^p_{L^p(\Omega)}\Bigg)^{1/p}. \end{aligned}$$

For $m = 0$, we set

$$W_{\alpha}^{0,p}(\Omega) = \{u \in \mathscr{D}'(\Omega);\ \ \rho^{\alpha}\, u \in L^p(\Omega)\}.$$

We note that the logarithmic weight only appears if $\frac{3}{p} + \alpha \in \{1, \ldots, m\}$ and all the local properties of $W_{\alpha}^{m,p}(\Omega)$ coincide with those of the corresponding classical Sobolev spaces $W^{m,p}(\Omega)$. We set $\mathring{W}_{\alpha}^{m,p}(\Omega)$ as the adherence of $\mathscr{D}(\Omega)$ for the norm $\| \,.\, \|_{W_{\alpha}^{m,p}(\Omega)}$. Then, the dual space of $\mathring{W}_{\alpha}^{m,p}(\Omega)$, denoting by $W_{-\alpha}^{-m,p'}(\Omega)$, is a space of distributions. When $\Omega = \mathbb{R}^3$, we have $W_{\alpha}^{m,p}(\mathbb{R}^3) = \mathring{W}_{\alpha}^{m,p}(\mathbb{R}^3)$. If $\Omega$ is a Lipschitz exterior domain, then we have

$$\mathring{W}_{\alpha}^{1,p}(\Omega) = \left\{v \in W_{\alpha}^{1,p}(\Omega),\ v = 0 \quad \text{on} \quad \partial\Omega\right\},$$

and

$$\mathring{W}_{\alpha}^{2,p}(\Omega) = \left\{v \in W_{\alpha}^{2,p}(\Omega),\ v = \frac{\partial v}{\partial \boldsymbol{n}} = 0 \quad \text{on} \quad \partial\Omega\right\},$$

where $\frac{\partial v}{\partial \boldsymbol{n}}$ is the normal derivate of $v$.

$$W_{\alpha}^{1,p}(\Omega) \hookrightarrow W_{\alpha}^{0,p*}(\Omega) \quad \text{where} \quad p* = \frac{3p}{3-p} \tag{1}$$

and, by duality, we have

$$W_{-\alpha}^{0,q}(\Omega) \hookrightarrow W_{-\alpha}^{-1,p'}(\Omega) \quad \text{where} \quad q = \frac{3p'}{3+p'}.$$

On the other hand, if $m > 0$ and $\frac{3}{p} + \alpha \neq 1$ or $m \leq 0$ and $3/p' - \alpha \neq 0$, we have the following continuous embedding:

$$W^{m,p}_{\alpha}(\Omega) \hookrightarrow W^{m-1,p}_{\alpha-1}(\Omega). \tag{2}$$

Moreover the space $W^{m,p}_{\alpha}(\Omega)$ sometimes contains some polynomials functions. For any $q \in \mathbb{N}$, $\mathscr{P}_q$ (respectively, $\mathscr{P}^{\Delta}_q$) stands for the space of polynomials (respectively, harmonic polynomials) of degree $\leq q$. If $q$ is a strictly negative integer, we set by convention $\mathscr{P}_q = \{0\}$. For $m \geq 0$ and if $\frac{3}{p} + \alpha$ does not belong to $\{i \in \mathbb{Z}\,;\, i \leq m\}$, then $\mathscr{P}_{[m-\alpha-\frac{3}{p}]}$ is the space of all polynomials included in $W^{m,p}_{\alpha}(\Omega)$. The norm of the quotient space $W^{m,p}_{\alpha}(\Omega)/\mathscr{P}_{[m-\alpha-\frac{3}{p}]}$ is:

$$||u||_{W^{m,p}_{\alpha}(\Omega)/\mathscr{P}_{[m-\alpha-\frac{3}{p}]}} = \inf_{Q \in \mathscr{P}_{[m-\alpha-\frac{3}{p}]}} ||u + Q||_{W^{m,p}_{\alpha}(\Omega)}. \tag{3}$$

In addition, the Hardy inequality holds, for $1 < p < \infty$,

$$\forall\ u \in \mathring{W}^{m,p}_{\alpha}(\Omega), \quad ||u||_{W^{m,p}_{\alpha}(\Omega)} \leq C||\nabla\, u||_{\mathbf{W}^{m-1,p}_{\alpha}(\Omega)},$$

where $C = C(p,\alpha,\Omega) > 0$ and when $\Omega = \mathbb{R}^3$, we have

$$\forall u \in W^{1,p}_{\alpha}(\mathbb{R}^3), \quad \begin{cases} ||u||_{W^{1,p}_{\alpha}(\mathbb{R}^3)} \leq ||\nabla\, u||_{\mathbf{W}^{0,p}_{\alpha}(\mathbb{R}^3)}, & \text{if} \quad 3/p + \alpha > 1, \\ ||u||_{W^{1,p}_{\alpha}(\mathbb{R}^3)/\mathscr{P}_0} \leq ||\nabla\, u||_{\mathbf{W}^{0,p}_{\alpha}(\mathbb{R}^3)}, & \text{otherwise,} \end{cases}$$

where $\mathscr{P}_0$ stands for the space of constant functions in $W^{1,p}_{\alpha}(\mathbb{R}^3)$ when $3/p + \alpha \leq 1$ and $C$ satisfies $C = C(p,\alpha) > 0$.

Finally, given a Banach space $B$, with dual space $B'$ and a closed subspace $X$ of $B$, we denote by $B' \perp X$ the subspace of $B'$ orthogonal to $X$, i.e.

$$B' \perp X = \{f \in B'; < f, v >= 0\ \forall\, v \in X\} = (B/X)'.$$

The space $B' \perp X$ is also called the polar space of $X$ in $B'$.

## 2 Preliminary results

Now, we give some results related to solving the Dirichlet problem and Neumann problem which are essential to ensure the existence and the uniqueness of some vectors potentials and one usually forces either the normal component to vanish or the tangential components to vanish. We start by giving the definition of the kernel of the Laplace operator for any integer $\alpha \in \mathbb{Z}$:

$$\mathscr{A}^{\Delta}_{\alpha,p}(\Omega) = \left\{\chi \in W^{1,p}_{\alpha}(\Omega); \quad \Delta\,\chi = 0 \quad \text{in} \quad \Omega \quad \text{and} \quad \chi = 0 \quad \text{on}\ \Gamma\right\}.$$

In contrast to a bounded domain, the Dirichlet problem for the Laplace operator with zero data can have nontrivial solutions in an exterior domain; it depends upon the exponent of the weight. The result that we state below is established by Giroire when $p = 2$; see [5] for more details.

**Proposition 1.** *For any integer $\alpha < 0$, the space $\mathscr{A}^{\Delta}_{\alpha,2}(\Omega)$ is a subspace of all functions in $W^{1,2}_{\alpha}(\Omega)$ of the form $v(p) - p$, where $p$ runs over all polynomials of $\mathscr{P}^{\Delta}_{-\alpha-1}$ and $v(p)$ is the unique solution in $W^{1,2}_{0}(\Omega)$ of the Dirichlet problem*

$$\Delta\, v(p) = 0 \quad \text{in} \quad \Omega \quad \text{and} \quad v(p) = p \quad \text{on } \Gamma. \tag{4}$$

*$\mathscr{A}^{\Delta}_{\alpha,2}(\Omega)$ is a finite-dimensional space of the same dimension as $\mathscr{P}^{\Delta}_{-\alpha-1}$ and $\mathscr{A}^{\Delta}_{\alpha,2}(\Omega) = \{0\}$ when $\alpha \geq 0$.*

Our analysis is based on the principle that linear exterior problems can be solved by combining their properties in the whole space $\mathbb{R}^3$ and the properties in bounded domains. Let us begin by recalling some results in $\mathbb{R}^3$ previously established by Amrouche, Girault, and Giroire in [3].

**Theorem 1.** *Let $\alpha \geq 0$ be an integer. The following Laplace operators are isomorphisms:*

$$\Delta : W^{1,p}_{\alpha}(\mathbb{R}^3)/\mathscr{P}_{[1-\alpha-\frac{3}{p}]} \mapsto W^{-1,p}_{\alpha}(\mathbb{R}^3) \perp \mathscr{P}^{\Delta}_{[1+\alpha-\frac{3}{p'}]} \qquad \textit{if } \frac{3}{p'} \notin \{1, \dots, \alpha\}$$

$$\Delta : W^{1,p}_{-\alpha}(\mathbb{R}^3)/\mathscr{P}_{[1+\alpha-\frac{3}{p}]} \mapsto W^{-1,p}_{-\alpha}(\mathbb{R}^3) \qquad \textit{if } \alpha > 0 \textit{ and } \frac{3}{p} \notin \{1, \dots, \alpha\}.$$

## 3 The Dirichlet problem for the Laplace operator

In this section, we propose to solve the Laplace equation with Dirichlet boundary condition:
Let $p > 1$ and $f$ given in $W^{-1,p}_{\alpha}(\Omega)$ and $g$ given in $W^{1/p',p}(\Gamma)$, find $u$ in $W^{1,p}_{\alpha}(\Omega)$ solution of:

$$-\Delta\, u = f \quad \text{in} \quad \Omega, \quad u = g \quad \text{on} \quad \Gamma. \tag{5}$$

We are really interested by the case where $\alpha$ in $H_1$ or in $H_2$ with

$$H_1 = \{\alpha \in \mathbb{Z} \text{ such that } \alpha < 0 \text{ and } 3/p \notin \{1, \dots, -\alpha\}\}$$

and

$$H_2 = \left\{\alpha \in \mathbb{Z} \text{ such that } \alpha > 0 \text{ and } 3/p' \notin \{1, \dots, \alpha\}\right\},$$

with the convention that $\{1, \dots, -\alpha\}$ or $\{1, \dots, \alpha\}$ are empty if $\alpha = 0$.

Note that when $\alpha = 0$, problem (5) has been studied in [2]. We start our study by solving the Laplace equation in $\mathbb{R}^3$:

$$-\Delta u = f \quad \text{in} \quad \mathbb{R}^3, \tag{6}$$

with a right-hand side that has a bounded support.

**Lemma 1.** *Assume that $p > 2$ and $\alpha \in \mathbb{Z}$. Let $f$ in $W_\alpha^{-1,p}(\mathbb{R}^3)$ with compact support. Then*

*i) if $\alpha \in H_1$, problem (6) has a solution $u$ in $W_\alpha^{1,2}(\mathbb{R}^3) \cap W_\alpha^{1,p}(\mathbb{R}^3)$ unique up to an element of $\mathscr{P}_{[-\alpha-\frac{1}{2}]}$.*

*ii) If $\alpha \in H_2$, problem (6) has a unique solution $u$ in $W_\alpha^{1,2}(\mathbb{R}^3) \cap W_\alpha^{1,p}(\mathbb{R}^3)$ if and only if*

$$\forall \lambda \in \mathscr{P}^{\Delta}_{[\alpha-\frac{1}{2}]}, \quad \langle f, \lambda \rangle_{W_\alpha^{-1,2}(\mathbb{R}^3) \times W_{-\alpha}^{1,2}(\mathbb{R}^3)} = 0. \tag{7}$$

*Proof.* At first, note that the duality pairing $\langle f, \lambda \rangle_{W_\alpha^{-1,2}(\mathbb{R}^3) \times \mathring{W}_{-\alpha}^{1,2}(\mathbb{R}^3)}$ is well defined for all $f \in W_\alpha^{-1,p}(\mathbb{R}^3)$ with compact support and $p > 2$.

The proof of point *i*) and *ii*) is very similar, so we do only the proof of the first result ($\alpha < 0$ and $3/p \notin \{1, \ldots, -\alpha\}$). The proof follows the idea of [2]. Since $p > 2$ and support of $f$ is compact, we have $f$ in $W_\alpha^{-1,2}(\mathbb{R}^3)$. Using Theorem 1 we deduce that problem (6) has exactly one solution $u$ in $W_\alpha^{1,2}(\mathbb{R}^3)/\mathscr{P}_{[-\alpha-\frac{1}{2}]}$. The remainder of the proof is devoted to establish that $u$ in $W_\alpha^{1,p}(\mathbb{R}^3)$. For any positive real number $R_0$, let $B_{R_0}$ denote the open ball centered at the origin, with radius $R_0$; and assuming that $R_0$ is sufficiently large for $\overline{\Omega'} \subset B_{R_0}$, we denote by $\Omega_{R_0}$ the intersection $\Omega \cap B_R$. Take $R_0$ sufficiently large so that the support of $f$ is contained in $B_{R_0}$. Let $\lambda$ and $\mu$ be two scalar, nonnegative functions in $C^\infty(\mathbb{R}^3)$ that satisfy

$$\forall \ x \in B_{R_0}, \quad \lambda(x) = 1, \quad \operatorname{supp} \lambda \subset B_{R_0+1}, \quad \forall \ x \in \mathbb{R}^3, \quad \lambda(x) + \mu(x) = 1.$$

Then, we can write

$$u = \lambda\, u + \mu\, u.$$

As $\mu$ is very smooth and vanishes on $B_{R_0}$, then $\mu f = 0$. After an easy calculation, we obtain that $\mu\, u$ satisfies the following equations in $\mathbb{R}^3$:

$$-\Delta(\mu\, u) := f_1 \quad \text{in} \quad \mathbb{R}^3,$$

with $f_1 = -(\Delta\, \mu)u - 2\nabla\, \mu \cdot \nabla\, u$. Moreover, owing to the support of $\mu$, $f_1$ belongs to $L^2(\mathbb{R}^3)$. In addition, if $\mathscr{O}$ is a Lipschitzian bounded domain, we have $L^2(\mathscr{O}) \hookrightarrow W^{-1,q}(\mathscr{O})$ for any $2 \leq q \leq 6$. Hence, we shall assume for the time being that $2 < p \leq 6$ and afterward, we shall use a bootstrap argument. Then $f_1$

belongs to $W_\alpha^{-1,p}(\mathbb{R}^3)$. Since $3/p \notin \{1,\dots,-\alpha\}$, it follows from Theorem 1 that there exists a unique $v$ in $W_\alpha^{1,p}(\mathbb{R}^3)/\mathscr{P}_{[1-\alpha-3/p]}$ such that $-\Delta\mu\, u = \Delta\, v$. Hence $\mu u - v$ is a harmonic tempered distribution and therefore a polynomial belonging to $W_\alpha^{1,p}(\mathbb{R}^3) + W_\alpha^{1,2}(\mathbb{R}^3)$. Thus there exists a polynomial $K \in \mathscr{P}_{[1-\alpha-3/p]} \subset W_\alpha^{1,p}(\mathbb{R}^3)$ such that $\mu u = v + K$. Consequently, $\mu u$ belongs to $W_\alpha^{1,p}(\mathbb{R}^3)$. In particular, we have $\mu\, u = u$ outside $B_{R_0+1}$, so the restriction of $u$ to $\partial B_{R_0+1}$ belongs to $W^{1/p',p}(\partial B_{R_0+1})$. Therefore, $u$ satisfies:

$$-\,\Delta\, u = f \quad \text{in} \quad B_{R_0+1}, \quad u|_{\partial B_{R_0+1}} = \mu\, u. \tag{8}$$

This problem has a unique solution $u \in W^{1,p}(B_{R_0+1})$. This implies that $u \in W_\alpha^{1,p}(\mathbb{R}^3)$ if $2 < p \leq 6$. Now, suppose that $p > 6$. The above argument shows that $u$ belongs to $W_\alpha^{1,6}(\Omega)$ and we can repeat the same argument with $p = 6$ instead of $p = 2$ using the fact that if $\mathscr{O}$ is a Lipschitzian bounded domain, we have $L^6(\mathscr{O}) \hookrightarrow W^{-1,t}(\mathscr{O})$ for any real number $t$. This establishes the existence of solution $u$ in $W_\alpha^{1,2}(\mathbb{R}^3) \cap W_\alpha^{1,p}(\mathbb{R}^3)$ of Problem (6).

The next lemma solves problem (5) with homogeneous boundary conditions and a right-hand side $f$ with bounded support:

$$-\,\Delta\, u = f \quad \text{in} \quad \Omega, \quad u = 0 \quad \text{on} \quad \Gamma. \tag{9}$$

**Lemma 2.** *Let $p > 2$ and $\alpha \in \mathbb{Z}$. Suppose that $\Gamma$ is of class $C^{1,1}$ and let $f$ in $W_\alpha^{-1,p}(\Omega)$ with compact support. Then*

*i) if $\alpha \in H_1$, problem (9) has a solution $u$ in $W_\alpha^{1,2}(\Omega) \cap W_\alpha^{1,p}(\Omega)$ unique up to an element of $\mathscr{A}_{\alpha,2}^{\Delta}(\Omega)$.*

*ii) If $\alpha \in H_2$, problem (9) has a unique solution $u$ in $W_\alpha^{1,2}(\Omega) \cap W_\alpha^{1,p}(\Omega)$ if and only if*

$$\forall \lambda \in \mathscr{A}_{-\alpha,2}^{\Delta}(\Omega), \quad \langle f, \lambda \rangle_{W_\alpha^{-1,2}(\Omega)\times \mathring{W}_{-\alpha}^{1,2}(\Omega)} = 0. \tag{10}$$

*Proof.* The proof of point *i*) and *ii*) is very similar, so we do only the proof of the first result. Proceed as in Lemma 1. Since $p > 2$ and support of $f$ is compact, we have $f$ which belongs to $W_\alpha^{-1,2}(\Omega)$. Due to Theorem 3.5 of [4], problem (9) has exactly one solution $u$ in $W_\alpha^{1,2}(\Omega)/\mathscr{A}_{\alpha,2}^{\Delta}(\Omega)$. The remainder of the proof is devoted to establish that $u$ in $W_\alpha^{1,p}(\Omega)$. Take $R_0$ sufficiently large so that both the support of $f$ and $\overline{\Omega'}$ are contained in $B_{R_0}$. Let us extend $u$ by zero in $\Omega'$. Then, the extended distribution denoted by $\tilde{u}$ belongs to $W_\alpha^{1,2}(\mathbb{R}^3)$. With the above partition of unity, we split $u$ into $\lambda\, u + \mu\, u$ and after an easy calculation, we obtain that $\mu\tilde{u}$ satisfies the following equations in $\mathbb{R}^3$:

$$-\,\Delta(\mu\,\tilde{u}) := f_1 \quad \text{in} \quad \mathbb{R}^3 \quad \text{with} \quad f_1 = -(\Delta\,\mu)\tilde{u} - 2\nabla\,\mu\cdot\nabla\,\tilde{u}. \tag{11}$$

Suppose at first that $2 < p \leq 6$, then $f_1 \in W_\alpha^{-1,p}(\mathbb{R}^3)$. Since $3/p \notin \{1, \ldots, -\alpha\}$, it follows from Theorem 1 that (11) has a unique solution in $W_\alpha^{1,p}(\mathbb{R}^3)/\mathscr{P}_{[1-\alpha-3/p]}$ and hence $\mu\,\tilde{u}$ belongs to $W_\alpha^{1,p}(\mathbb{R}^3)$. Thus, as $\mu\,\tilde{u} = u$ outside $B_{R_0+1}$, the restriction of $u$ to $\partial B_{R_0+1}$ belongs to $W^{1/p',p}(\partial B_{R_0+1})$. Therefore, $u$ satisfies:

$$-\Delta\, u = f \quad \text{in} \quad \Omega_{R_0+1}, \quad u|_{\partial B_{R_0+1}} = \tilde{u} \quad \text{and} \quad u|_\Gamma = 0,$$

where $\Omega_{R_0+1}$ denotes the intersection $\Omega \cap B_{R_0+1}$. Since the boundary of $\Omega_{R_0+1}$ is of class $C^{1,1}$, this problem has a unique solution $u$ in $W^{1,p}(\Omega_{R_0+1})$. This implies that $u$ belongs to $W_\alpha^{1,p}(\Omega)$ if $2 < p \leq 6$ and the same bootstrap argument extends this result to any real value of $p > 2$.

As a consequence of Lemma 2, we can solve the following problem:

$$-\,\Delta\, u = 0 \quad \text{in} \quad \Omega, \quad u = g \quad \text{on} \quad \Gamma. \tag{12}$$

**Corollary 1.** *Let $p > 2$ and $\alpha \in \mathbb{Z}$. Suppose that $\Gamma$ is of class $C^{1,1}$ and let $g$ in $W^{1/p',p}(\Gamma)$. Then*

*i) if $\alpha \in H_1$, problem (12) has a solution $u$ in $W_\alpha^{1,2}(\Omega) \cap W_\alpha^{1,p}(\Omega)$ unique up to an element of $\mathscr{A}_{\alpha,2}^{\Delta}(\Omega)$.*

*ii) If $\alpha \in H_2$, problem (12) has a unique solution $u$ in $W_\alpha^{1,2}(\Omega) \cap W_\alpha^{1,p}(\Omega)$ if and only if*

$$\forall \lambda \in \mathscr{A}_{-\alpha,2}^{\Delta}(\Omega), \quad \left\langle g, \frac{\partial \lambda}{\partial \boldsymbol{n}} \right\rangle_\Gamma = 0, \tag{13}$$

*where the duality on $\Gamma$ is defined by:*

$$< .,. >_\Gamma = < .,. >_{H^{1/2}(\Gamma)\times H^{-1/2}(\Gamma)}\ .$$

*Proof.* The proof of point *i*) and *ii*) is very similar, so we do only the proof of the second result. Let $R$ be chosen so that $\overline{\Omega'}$ is contained in $B_R$ and let $v$ be the lifting function of $g$ satisfying:

$$-\Delta\, v = 0 \quad \text{in} \quad \Omega_R, \quad v|_{\partial B_R} = 0 \quad \text{and} \quad v|_\Gamma = g.$$

This set of equations defines a unique function $v$ in $W^{1,p}(\Omega_R)$ and when we extend it by zero outside $B_R$, the extended function, still denoted by $v$, belongs to $W_\alpha^{1,2}(\Omega) \cap W_\alpha^{1,p}(\Omega)$. Then problem (12) is equivalent to

$$-\,\Delta\, z = \Delta\, v \quad \text{in} \quad \Omega, \quad z = 0 \quad \text{on} \quad \Gamma, \tag{14}$$

where $\Delta\, v$ belongs to $W_\alpha^{-1,p}(\Omega)$ and has a bounded support and thus belongs to $W_\alpha^{-1,2}(\Omega)$. Thanks to the density of $\mathscr{D}(\overline{\Omega})$ in $W_\alpha^{1,2}(\Omega)$, for all $\lambda \in \mathscr{A}_{-\alpha,2}^{\Delta}(\Omega)$

$$\langle \Delta\, v, \lambda \rangle_{W_\alpha^{-1,2}(\Omega)\times \mathring{W}_{-\alpha}^{1,2}(\Omega)} = -\left\langle \frac{\partial \lambda}{\partial \boldsymbol{n}}, v \right\rangle_\Gamma = -\left\langle \frac{\partial \lambda}{\partial \boldsymbol{n}}, g \right\rangle_\Gamma .$$

Applying (13) and Lemma 2, we deduce that problem (14) has a unique solution $z$ in $W_\alpha^{1,2}(\Omega) \cap W_\alpha^{1,p}(\Omega)$. Thus $u = z + v$ is the unique solution of problem (12).

The next theorem characterizes the kernel $\mathscr{A}_{\alpha,p}^{\Delta}(\Omega)$ with $\alpha \in H_2$ of the Laplace operator with Dirichlet boundary conditions. We will start by $p > 2$ and we shall see at the end of this section the characterization of the kernel $\mathscr{A}_{\alpha,p}^{\Delta}(\Omega)$ when $p \leq 2$.

**Theorem 2.** *Assume that $p > 2$ and $\alpha \in H_2$ and suppose that $\Gamma$ is of class $C^{1,1}$. Then*

$$\mathscr{A}_{\alpha,p}^{\Delta}(\Omega) = \{0\} .$$

*Proof.* The proof follows the idea of [5]. Let $p > 2$ and $\alpha \in H_2$. Let $z$ be an element of $\mathscr{A}_{\alpha,p}^{\Delta}(\Omega)$ and let us extend it by zero in $\Omega'$. The extended function, denoted by $\tilde{z}$ belongs to $W_\alpha^{1,p}(\mathbb{R}^3)$ and thus $\Delta\,\tilde{z}$ belongs to $W_\alpha^{-1,p}(\mathbb{R}^3)$ with compact support. Since $3/p' \notin \{1, \ldots, \alpha\}$ and $\langle \Delta\,\tilde{z}, \lambda \rangle = 0$ for all $\lambda \in \mathscr{P}_{[\alpha-\frac{1}{2}]}^{\Delta}$, it follows from Lemma 1 that there exists a solution $w$ in $W_\alpha^{1,2}(\mathbb{R}^3) \cap W_\alpha^{1,p}(\mathbb{R}^3)$ such that $\Delta\,\tilde{z} = \Delta\, w$. Hence $w - \tilde{z}$ is a harmonic tempered distribution belonging to $W_\alpha^{1,p}(\mathbb{R}^3)$ and thus $w - \tilde{z}$ belongs to $\mathscr{P}_{[1-\alpha-3/p]}^{\Delta}$. Since $\alpha > 0$, we deduce that $w - \tilde{z} = 0$. Therefore, $z$ belongs to $\mathscr{A}_{\alpha,2}^{\Delta}(\Omega)$ which is reduced to $\{0\}$ (see Proposition 1).

*Remark 1.* 1. Note that if $\alpha = 0$, $\mathscr{A}_{0,p}^{\Delta}(\Omega)$ has been studied in Theorem 2.7 of [2]:

i) If $1 < p < 3$, $\mathscr{A}_{0,p}^{\Delta}(\Omega) = \{0\}$.
ii) If $p \geq 3$,

$$\mathscr{A}_{0,p}^{\Delta}(\Omega) = \{w(\lambda) - \lambda \quad \text{with} \ \ \lambda \in \mathbb{R}\} ,$$

where $w(\lambda)$ denotes the unique solution in $W_\alpha^{1,2}(\Omega) \cap W_\alpha^{1,p}(\Omega)$ of the following equations:

$$-\,\Delta\, w(\lambda) = 0 \quad \text{in} \quad \Omega, \quad w(\lambda) = \lambda \quad \text{on} \quad \Gamma. \tag{15}$$

2. We shall see at the end of this section that in fact $\mathscr{A}_{\alpha,p}^{\Delta}(\Omega) = \{0\}$ for $1 < p < 2$ and $\alpha > 0$ such that $3/p' \notin \{1, \ldots, \alpha\}$.

The proof of the following theorem is very similar to that of Theorem 2.

**Theorem 3.** *Assume that $1 < p < \infty$ and $\alpha \leq 0$ such that $3/p \notin \{1, \ldots, -\alpha\}$, with the convention that this set is empty if $\alpha = 0$ and suppose that $\Gamma$ is of class $C^{1,1}$. Then*

$$\mathscr{A}_{\alpha,p}^{\Delta}(\Omega) = \left\{ w(\lambda) - \lambda \quad \text{with} \ \ \lambda \in \mathscr{P}_{[1-\alpha-3/p]} \right\} ,$$

*where* $w(\lambda)$ *denotes the unique solution in* $W^{1,2}_{\alpha}(\Omega) \cap W^{1,p}_{\alpha}(\Omega)$ *of problem* (15) *if* $p > 2$ *and* $w(\lambda)$ *denotes the unique solution in* $W^{1,p}_{\alpha}(\Omega)$ *of problem* (15) *if* $p \leq 2$.

We are now in a position to solve problem (5) for $p \geq 2$.

**Theorem 4.** *Let* $p \geq 2$ *and* $\alpha \in \mathbb{Z}$. *Suppose that* $\Gamma$ *is of class* $C^{1,1}$ *and let* $f$ *in* $W^{-1,p}_{\alpha}(\Omega)$ *and* $g$ *in* $W^{1/p',p}(\Gamma)$. *Then*

*i) if* $\alpha \in H_1$, *problem* (5) *has a unique solution* $u$ *in* $W^{1,p}_{\alpha}(\Omega)/\mathscr{A}^{\Delta}_{\alpha,p}(\Omega)$.

*ii) If* $\alpha \in H_2$, *problem* (5) *has a unique solution* $u$ *in* $W^{1,p}_{\alpha}(\Omega)$ *if and only if*

$$\forall \lambda \in \mathscr{A}^{\Delta}_{-\alpha,p'}(\Omega), \quad \langle f, \lambda \rangle_{W^{-1,p}_{\alpha}(\Omega)\times \mathring{W}^{1,p'}_{-\alpha}(\Omega)} = \left\langle g, \frac{\partial \lambda}{\partial \boldsymbol{n}} \right\rangle_{\Gamma}, \tag{16}$$

*where the duality on* $\Gamma$ *is defined by:*

$$< .,. >_{\Gamma} = < .,. >_{W^{1/p',p}(\Gamma)\times W^{-1/p',p'}(\Gamma)} .$$

*Proof.* Note that the case $p = 2$ has been studied in Theorem 3.5 of [4]. So we can suppose that $p > 2$. The proof of point *i*) and *ii*) is very similar, so we do only the proof of the second result. The proof can be done in two steps:

**First case: we suppose that** $g = 0$. We want to extend the data $f$ by zero in $\Omega'$. As $f$ belongs to $W^{-1,p}_{\alpha}(\Omega)$, it follows from Corollary 1.3 of [1] that there exists a function $\boldsymbol{F}$ in $\boldsymbol{W}^{0,p}_{\alpha}(\Omega)$ such that $f = \operatorname{div} \boldsymbol{F}$ in $\Omega$. Let $\tilde{\boldsymbol{F}}$ denote the extension by zero of $\boldsymbol{F}$ in $\Omega'$ and set $\tilde{f} = \operatorname{div} \tilde{\boldsymbol{F}}$. Then $\tilde{f}$ belongs to $W^{-1,p}_{\alpha}(\mathbb{R}^3)$. It follows from Theorem 1 that there exists a unique $\tilde{w}$ in $W^{1,p}_{\alpha}(\mathbb{R}^3)$ such that $-\Delta \tilde{w} = \tilde{f}$ in $\mathbb{R}^3$. Denoting by $w$ the restriction of $\tilde{w}$ to $\Omega$ and by $\gamma w \in W^{1/p',p}(\Gamma)$ the trace of $w$ on $\Gamma$. Thanks to the density of $\mathscr{D}(\overline{\Omega})$ in $W^{1,p}_{\alpha}(\Omega)$, we have for all $\lambda \in \mathscr{A}^{\Delta}_{-\alpha,p'}(\Omega)$

$$\langle f, \lambda \rangle_{W^{-1,p}_{\alpha}(\Omega)\times \mathring{W}^{1,p'}_{-\alpha}(\Omega)} = \langle \Delta w, \lambda \rangle_{W^{-1,p}_{\alpha}(\Omega)\times \mathring{W}^{1,p'}_{-\alpha}(\Omega)} = \left\langle \frac{\partial \lambda}{\partial \boldsymbol{n}}, -\gamma w \right\rangle_{\Gamma} .$$

Applying (16) and Corollary 1, we have the existence of a unique solution $\xi$ in $W^{1,2}_{\alpha}(\Omega) \cap W^{1,p}_{\alpha}(\Omega)$ such that

$$-\Delta \xi = \boldsymbol{0} \quad \text{in} \quad \Omega, \quad \xi = -\gamma w \quad \text{on} \quad \Gamma. \tag{17}$$

Hence, $u = w + \xi$ belongs to $W^{1,p}_{\alpha}(\Omega)$ and satisfies problem (5) with $g = 0$.

**Second case: nonhomogeneous boundary data**. Each $g$ in $W^{1/p',p}(\Gamma)$ has a lifting function $v$ introduced in the proof of Corollary 1. Then problem (5) is equivalent to

$$-\Delta z = f + \Delta v \quad \text{in} \quad \Omega, \quad v = 0 \quad \text{on} \quad \Gamma. \tag{18}$$

As $f + \Delta v$ belongs to $W^{-1,p}_{\alpha}(\Omega)$ and for all $\lambda \in \mathscr{A}^{\Delta}_{-\alpha,p'}(\Omega)$ we have

$$\langle f + \Delta v, \lambda \rangle_{W_\alpha^{-1,p}(\Omega) \times \mathring{W}_{-\alpha}^{1,p'}(\Omega)} = -\left\langle \frac{\partial \lambda}{\partial \boldsymbol{n}}, g \right\rangle_\Gamma + \langle f, \lambda \rangle_{W_\alpha^{-1,p}(\Omega) \times \mathring{W}_{-\alpha}^{1,p'}(\Omega)} = 0,$$

it follows from the first case ($g = 0$) that problem (18) has a solution in $W_\alpha^{1,p}(\Omega)$. Hence $u = v + z$ belongs to $W_\alpha^{1,p}(\Omega)$ and satisfies problem (5). Uniqueness follows from the definition of the kernel $\mathscr{A}_{\alpha,p}^{\Delta}(\Omega)$.

The following existence result can be stated via a dual argument.

**Theorem 5.** *Suppose that $1 < p < 2$ and $\alpha \in H_2$ and assume that $\Gamma$ is of class $C^{1,1}$. Then for any $f$ in $W_\alpha^{-1,p}(\Omega)$ and $g$ in $W^{1/p',p}(\Gamma)$ such that (16) is satisfied, problem (5) has a unique solution $u$ in $W_\alpha^{1,p}(\Omega)$.*

*Proof. i)* **First case**: we suppose that $g = 0$. Then problem (5) has the following equivalent variational formulation: Find $u$ in $\mathring{W}_\alpha^{1,p}(\Omega)$ such that for any $\varphi$ in $\mathring{W}_{-\alpha}^{1,p'}(\Omega)$,

$$\langle u, -\Delta \varphi \rangle_{\mathring{W}_\alpha^{1,p}(\Omega) \times W_{-\alpha}^{-1,p'}(\Omega)} = \langle f, \varphi \rangle_{W_\alpha^{-1,p}(\Omega) \times \mathring{W}_{-\alpha}^{1,p'}(\Omega)}.$$

According to Theorem 4, for any $f'$ in $W_{-\alpha}^{-1,p'}(\Omega)$ there exists a unique solution $\varphi$ in $\mathring{W}_{-\alpha}^{1,p'}(\Omega)/\mathscr{A}_{-\alpha,p'}^{\Delta}(\Omega)$ such that $-\Delta \varphi = f'$ in $\Omega$ and we have

$$\inf_{\lambda \in \mathscr{A}_{-\alpha,p'}^{\Delta}(\Omega)} ||\varphi + \lambda||_{W_{-\alpha}^{1,p'}(\Omega)} \leq C||f'||_{W_{-\alpha}^{-1,p'}(\Omega)}. \tag{19}$$

Let $T$ be a linear form defined from $W_{-\alpha}^{-1,p'}(\Omega)$ onto $\mathbb{R}$ by:

$$T: \ (f') \mapsto \langle f, \varphi \rangle_{W_\alpha^{-1,p}(\Omega) \times \mathring{W}_{-\alpha}^{1,p'}(\Omega)}.$$

Observe that for any $f' \in W_{-\alpha}^{-1,p'}(\Omega)$ and $\lambda \in \mathscr{A}_{-\alpha,p'}^{\Delta}(\Omega)$, we have

$$\begin{aligned} |T(f')| &= | \langle f, \varphi + \lambda \rangle_{W_\alpha^{-1,p}(\Omega) \times \mathring{W}_{-\alpha}^{1,p'}(\Omega)} | \\ &\leq ||f||_{W_\alpha^{-1,p}(\Omega)} ||\varphi + \lambda||_{W_{-\alpha}^{1,p'}(\Omega)} \leq C||f||_{W_\alpha^{-1,p}(\Omega)} ||f'||_{W_{-\alpha}^{-1,p'}(\Omega)}. \end{aligned}$$

Thus the linear form $T$ is continuous on the following space $W_{-\alpha}^{-1,p'}(\Omega)$ and we deduce that there exists a unique $u$ in $\mathring{W}_\alpha^{1,p}(\Omega)$ such that

$$T(f') = \langle u, f' \rangle_{\mathring{W}_\alpha^{1,p}(\Omega) \times W_{-\alpha}^{-1,p'}(\Omega)},$$

with

$$||u||_{W_\alpha^{1,p}(\Omega)} \leq C||f||_{W_\alpha^{-1,p}(\Omega)}.$$

By definition of $T$, it follows

$$\langle f,\varphi\rangle_{W_\alpha^{-1,p}(\Omega)\times \mathring{W}_{-\alpha}^{1,p'}(\Omega)} = \langle u,f'\rangle_{\mathring{W}_\alpha^{1,p}(\Omega)\times W_{-\alpha}^{-1,p'}(\Omega)},$$

which is the variational formulation of problem (5).

*ii)* **Second case**: the nonhomogeneous boundary data, i.e. $u = g$ on $\Gamma$. Each $g$ in $W^{1/p',p}(\Gamma)$ has a lifting function $v$ in $W_\alpha^{1,p}(\Omega)$. Then problem (5) is equivalent to (18) and to conclude, we can use the first case.

In the same way, we can prove the following theorem:

**Theorem 6.** *Suppose that $1 < p < 2$ and $\alpha \in H_1$ and assume that $\Gamma$ is of class $C^{1,1}$. Then for any $f$ in $W_\alpha^{-1,p}(\Omega)$ and $g$ in $W^{1/p',p}(\Gamma)$, problem (5) has a unique solution $u$ in $W_\alpha^{1,p}(\Omega)/\mathscr{A}_{\alpha,p}^{\Delta}(\Omega)$.*

*Proof.* *i)* **First case**: we suppose that $g = 0$. Then problem (5) has the following equivalent variational formulation: Find $u$ in $\mathring{W}_\alpha^{1,p}(\Omega)$ such that for any $\varphi$ in $\mathring{W}_{-\alpha}^{1,p'}(\Omega)$,

$$\langle u,-\Delta\,\varphi\rangle_{\mathring{W}_\alpha^{1,p}(\Omega)\times W_{-\alpha}^{-1,p'}(\Omega)} = \langle f,\varphi\rangle_{W_\alpha^{-1,p}(\Omega)\times \mathring{W}_{-\alpha}^{1,p'}(\Omega)}.$$

According to Theorem 4, for any $f'$ in $W_{-\alpha}^{-1,p'}(\Omega) \perp \mathscr{A}_{\alpha,p}^{\Delta}(\Omega)$ there exists a unique solution $\varphi$ in $\mathring{W}_{-\alpha}^{1,p'}(\Omega)$ such that $-\Delta\,\varphi = f'$ in $\Omega$ and we have

$$||\varphi||_{W_{-\alpha}^{1,p'}(\Omega)} \leq C||f'||_{W_{-\alpha}^{-1,p'}(\Omega)}. \tag{20}$$

Let $T$ be a linear form defined from $W_{-\alpha}^{-1,p'}(\Omega) \perp \mathscr{A}_{\alpha,p}^{\Delta}(\Omega)$ onto $\mathbb{R}$ by:

$$T:\ (f') \mapsto \langle f,\varphi\rangle_{W_\alpha^{-1,p}(\Omega)\times \mathring{W}_{-\alpha}^{1,p'}(\Omega)}.$$

Observe that for any $f' \in W_{-\alpha}^{-1,p'}(\Omega)$, we have

$$\begin{aligned}|T(f')| &= |\,\langle f,\varphi\rangle_{W_\alpha^{-1,p}(\Omega)\times \mathring{W}_{-\alpha}^{1,p'}(\Omega)}\,|\\ &\leq ||f||_{W_\alpha^{-1,p}(\Omega)}||\varphi||_{W_{-\alpha}^{1,p'}(\Omega)} \leq C||f||_{W_\alpha^{-1,p}(\Omega)}||f'||_{W_{-\alpha}^{-1,p'}(\Omega)}.\end{aligned}$$

Thus the linear form $T$ is continuous on the following space $W_{-\alpha}^{-1,p'}(\Omega)$ and we deduce that there exists a unique $u$ in $\mathring{W}_\alpha^{1,p}(\Omega)$ such that

$$T(f') = \langle u,f'\rangle_{\mathring{W}_\alpha^{1,p}(\Omega)\times W_{-\alpha}^{-1,p'}(\Omega)},$$

with

$$||u||_{W_{\alpha}^{1,p}(\Omega)} \leq C||f||_{W_{\alpha}^{-1,p}(\Omega)}.$$

By definition of $T$, it follows

$$\langle f, \varphi \rangle_{W_{\alpha}^{-1,p}(\Omega)\times \mathring{W}_{-\alpha}^{1,p'}(\Omega)} = \langle u, f' \rangle_{\mathring{W}_{\alpha}^{1,p}(\Omega)\times W_{-\alpha}^{-1,p'}(\Omega)},$$

which is the variational formulation of problem (5).

*ii)* **Second case**: the nonhomogeneous boundary data, i.e. $u = g$ on $\Gamma$. Each $g$ in $W^{1/p',p}(\Gamma)$ has a lifting function $v$ in $W_{\alpha}^{1,p}(\Omega)$. Then problem (5) is equivalent to (18) and to conclude, we can use the first case.

The next theorem summarizes the result of this section.

**Theorem 7.** *Let $\Gamma$ be of class $C^{1,1}$ if $p \neq 2$ or Lipschitz-continuous if $p = 2$. Let $f$ in $W_{\alpha}^{-1,p}(\Omega)$ and $g$ in $W^{1/p',p}(\Gamma)$ with $\alpha \in \mathbb{Z}$. Then*

1. *if $\alpha \in H_1$, problem (5) has a unique solution $u$ in $W_{\alpha}^{1,p}(\Omega)/\mathscr{A}_{\alpha,p}^{\Delta}(\Omega)$.*
2. *If $\alpha \in H_2$, problem (5) has a unique solution $u$ in $W_{\alpha}^{1,p}(\Omega)$ if and only if $f$ and $g$ satisfy the compatibility condition (16).*
3. *If $\alpha = 0$, problem (5) has a unique solution $u$ in $W_{\alpha}^{1,p}(\Omega)/\mathscr{A}_{0,p}^{\Delta}(\Omega)$ if $p \geq 2$ and if $p < 2$, problem (5) has a unique solution $u$ in $W_{\alpha}^{1,p}(\Omega)$ if and only if $f$ and $g$ satisfy the compatibility condition (16).*

*In addition, there exists a constant $C$, independent of $u$, $f$, and $g$, such that*

$$||u||_{W_{\alpha}^{1,p}(\Omega)/\mathscr{A}_{\alpha,p}^{\Delta}(\Omega)} \leq C\left(||f||_{W_{\alpha}^{-1,p}(\Omega)} + ||g||_{W^{1/p',p}(\Gamma)}\right). \tag{21}$$

We will prove now a regularity result, when the external forces belong to $W_{\alpha}^{0,p}(\Omega)$ with $\alpha \in \mathbb{Z}$.

**Theorem 8.** *Let $\Gamma$ be of class $C^{1,1}$ if $p \neq 2$ or Lipschitz-continuous if $p = 2$. Let $f$ in $W_{\alpha}^{0,p}(\Omega)$ and $g$ in $W^{1+1/p',p}(\Gamma)$ with $\alpha \in \mathbb{Z}$. Then*

1. *if $\alpha \in H_1$ and $3/p + \alpha \neq 1$, problem (5) has a unique solution $u$ in $W_{\alpha}^{2,p}(\Omega)/\mathscr{A}_{\alpha-1,p}^{\Delta}(\Omega)$.*
2. *If $\alpha \in H_2$, problem (5) has a unique solution $u$ in $W_{\alpha}^{2,p}(\Omega)/\mathscr{A}_{\alpha-1,p}^{\Delta}(\Omega)$ if and only if $f$ and $g$ satisfy the compatibility condition (16).*
3. *If $\alpha = 0$ and $p \neq 3$, problem (5) has a unique solution $u$ in $W_{0}^{2,p}(\Omega)/\mathscr{A}_{-1,p}^{\Delta}(\Omega)$.*

*In addition, there exists a constant $C$, independent of $u$, $f$, and $g$, such that*

$$||u||_{W_{\alpha}^{2,p}(\Omega)/\mathscr{A}_{\alpha,p}^{\Delta}(\Omega)} \leq C\left(||f||_{W_{\alpha}^{0,p}(\Omega)} + ||g||_{W^{1+1/p',p}(\Gamma)}\right). \tag{22}$$

*Proof.* The proof of point 1), 2), and 3) is very similar, so we do only the proof of the second result. First observe that $W^{2-1/p,p}(\Gamma) \hookrightarrow W^{1/p',p}(\Gamma)$ and since $3/p' \neq \alpha$ we have $W_{\alpha}^{0,p}(\Omega) \hookrightarrow W_{\alpha-1}^{-1,p}(\Omega)$, then we deduce thanks to Theorem 7 that problem (5) has a solution $u \in W^{1,p}_{\alpha-1}(\Omega)$ unique if $\alpha \geq 2$ and unique up to an element of $\mathscr{A}^{\Delta}_{0,p}(\Omega)$ if $\alpha = 1$. The rest of the proof is similar to that of Lemma 2, we introduce the same partition of unity as in Lemma 2. With the same notation, we can write

$$u = \lambda\, u + \mu\, u.$$

Let us extend $\mu\, u$ by zero in $\Omega'$. Then, the extended distributions denoted by $\widetilde{\mu\, u}$ belongs to $W^{1,p}_{\alpha-1}(\mathbb{R}^3)$. A quick computation in $\mathscr{D}'(\mathbb{R}^3)$ shows that $\widetilde{\mu\, u}$ satisfies the following equations:

$$-\Delta(\widetilde{\mu\, u}) := f_1$$

with

$$f_1 = \mu \tilde{f} - (\Delta\, \mu)\tilde{u} - 2\nabla\, \mu \cdot \nabla\, \tilde{u}.$$

Moreover, owing to the support of $\mu$, $f_1$ belongs to $W^{0,p}_{\alpha}(\mathbb{R}^3) \perp \mathscr{P}^{\Delta}_{[\alpha-3/p']}$. According to Theorem 9.9 see [3], there exists $z \in W^{2,p}_{\alpha}(\mathbb{R}^3)$ such that $\Delta z = \Delta(\widetilde{\mu\, u})$ in $\mathbb{R}^3$ ($z$ is unique up to an element of $\mathscr{P}_{[2-\alpha-3/p]}$). Hence $z - \widetilde{\mu\, u}$ is a harmonic tempered distribution belonging to $W^{2,p}_{\alpha}(\mathbb{R}^3) + W^{1,p}_{\alpha-1}(\mathbb{R}^3)$, therefore a polynomial $k \in \mathscr{P}_{[2-\alpha-3/p]} \subset W^{2,p}_{\alpha}(\mathbb{R}^3)$. Then, we deduce that $\widetilde{\mu\, u} = z + k$ belongs to $W^{2,p}_{\alpha}(\mathbb{R}^3)$. In particular, we have $\widetilde{\mu\, u} = u$ outside $B_{R_0+1}$, so the restriction of $u$ to $\partial B_{R_0+1}$ belongs to $W^{2-1/p,p}(\partial\, B_{R_0+1})$. Therefore, $u$ satisfies:

$$-\Delta\, u = f \quad \text{in} \quad \Omega_{R_0+1} \quad u|_{\partial B_{R_0+1}} = \widetilde{\mu\, u} \quad \text{and} \quad u|_{\Gamma} = g.$$

Since the boundary of $\Omega_{R_0+1}$ is of class $C^{1,1}$, this problem has a unique solution $u$ in $W^{2,p}(\Omega_{R_0+1})$. This implies that $u$ belongs to $W^{2,p}_{\alpha}(\Omega)$. The uniqueness of the solution $u$ follows from this inclusion $W^{2,p}_{\alpha}(\Omega) \subset W^{1,p}_{\alpha-1}(\Omega)$ which is valid if $3/p + \alpha \neq 1$ and since $\alpha > 0$ this inclusion holds.

*Remark 2.* To prove the uniqueness of the solution $u$ in Theorem 8 we need to have the following condition:

$$3/p + \alpha \neq 1. \tag{23}$$

because the inclusion $W^{2,p}_{\alpha}(\Omega) \subset W^{1,p}_{\alpha-1}(\Omega)$ holds if (23) is satisfied.

**Acknowledgements** The authors thank the referees for their careful reading of the manuscript and insightful comments.

## References

1. F. ALLIOT, C. AMROUCHE. Weak solutions for exterior Stokes problem in weighted Sobolev spaces. *Mathematical Methods in the Applied Sciences* **23**, 575–600, 2000.
2. C. AMROUCHE, V. GIRAULT AND J. GIROIRE. Dirichlet and Neumann exterior problems for the n-dimensional laplace operator, an approach in weighted Sobolev spaces. *J. Math. Pures Appl.* **76**(1), 55–81, 1997.
3. C. AMROUCHE, V. GIRAULT AND J. GIROIRE. Weighted Sobolev spaces for the laplace equation in $\mathbb{R}^n$. *J. Math. Pures et Appl.* **20**, 579–606, 1994.
4. V. GIRAULT. The Stokes problem and vector potential operator in three-dimensional exterior domains: An approch in weighted Sobolev spaces. *Differential and Integral Equations* **7**(2), 535–570, 1994.
5. J. GIROIRE. Etude de quelques problèmes aux Limites Extérieurs et Résolution par Equations Intégrales. *Thèse,* **Université Paris VI**, 1987.

# About the quotient of two bounded operators

**Abdellah Gherbi and Bekkai Messirdi**

**Abstract** The quotient operators are introduced in order to extend the class of all closed operators acting on a Hilbert space $H$. In fact, Kaufman proved in Kaufman (Proc Am Math Soc 72:531–534, 1978) using $A$ and $B$ such that $R(A^*) + R(B^*) = \{A^*x + B^*y : x; y \in H\}$ is closed in $H$, that a linear operator $T$ on $H$ is closed if and only if $T$ is represented as a quotient $B/A$. So that every closed operator is included in the class of quotients. Moreover, he proved that if $T$ is a closed densely defined operator, then $T$ is represented as $T = B/(I - B^*B)^{\frac{1}{2}}$ using a unique pure contraction $B$, i.e., an operator such that $\|Bx\| < \|x\|$ for all nonzero $x$ in $H$. In this paper we attempt to study some algebraic and topological properties of quotient operators acting on Hilbert space, such that the boundedness, compactness and invertibility, other results such that the powers of quotient and the quotient character of limit of converging sequence of quotient operators are also established.

**Keywords** Quotient operator • closed range • generalized inverse • nilpotent • idempotent and compact operator

**2010 Mathematics Subject Classifications:** 47 A05

## 1 Introduction and preliminaries

Throughout this paper, let $\mathscr{B}(H)$ denote the algebra of all bounded operators acting on a complex Hilbert space $H$ equipped with the inner product $\langle .; . \rangle$ and the associated norm $\|.\|$. For $T$ closed densely defined linear operator on $H$, we denote by $N(T)$ and $R(T)$ the null space and range of $T$. For two bounded operators $A$ and

A. Gherbi (✉)
EPST d'Oran, 31000 Oran, Algeria
e-mail: gherbi1@live.fr

B. Messirdi
University of Oran, 31000 Oran, Algeria
e-mail: bmessirdi@yahoo.com



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_18

$B$ in $\mathscr{B}(H)$ with the kernel condition $N(A) \subset N(B)$, we define the quotient operator $B/A$ to be the mapping $Au \longrightarrow Bu$ for all $u \in H$. It's clear that $R(A)$ and $R(B)$ are, respectively, the domain and the range of $B/A$. Note that the quotient of two bounded operators is not necessarily bounded. It was shown in [9, 10] that the sum and the product of two quotients are again represented as quotients, and that if a quotient operator is densely defined, its adjoint is also represented as quotient. In [12] W. E. Kaufman showed, using $A$ and $B$ such that $R(A^*) + R(B^*) = \{A^*x + B^*y : x; y \in H\}$ is closed in $H$, that a linear operator $T$ on $H$ is closed if and only if $T$ is represented as a quotient $B/A$. So that every closed operator is included in the class of quotients. Moreover, he proved that if $T$ is a closed densely defined operator, then $T$ is represented as $T = B/(I - B^*B)^{\frac{1}{2}}$ using a unique pure contraction $B$, i.e., an operator such that $\|Bx\| < \|x\|$ for all nonzero $x$ in $H$. Let the function $\Gamma$ defined by $\Gamma(B) = B/(I - B^*B)^{\frac{1}{2}}$, then W. Kaufman proved that there is a one-to-one correspondence between $\mathscr{C}_0(H)$ and $\mathscr{C}(H)$ via $\Gamma$. This function is also used to reformulate questions about unbounded operators in terms of bounded ones

- In [12, 14], Kaufman proved that the map $\Gamma$ preserves many properties of operators: self-adjointness, nonnegative conditions, normality, and quasinormality.
- In [8] Hirasawa showed that a pure contraction $B$ is hyponormal if and only if $T = B/(I - B^*B)^{\frac{1}{2}}$ is formally hyponormal , and if $B$ is quasinormal then $T^n = B^n/(I - B^*B)^{\frac{n}{2}}$ is quasinormal for all integers $n \geq 2$.

The aim of this paper is to characterize some algebraic, topological properties of quotient operator. In fact, our work is organized as follows:

In the second section we study some algebraic and topological properties of quotient operator such as boundedness, compactness, invertibility. The powers of quotient (idempotent, nilpotent, and quasi-nilpotent quotients) and the limit of converging sequence of quotient operators are also established.

Throughout this paper, $B/A$ is quotient operator of two bounded operators $A, B \in \mathscr{B}(H)$.

## 2 Mains results

### *2.1 Bounded, compact quotient operator*

First, we recall the Douglas majorization lemma.

**Lemma 1 ([5, 6]).** *Let $A, B \in \mathscr{B}(H)$. Then the following conditions are equivalent:*

1. *$R(B) \subset R(A)$.*
2. *$BB^* \leq \lambda AA^*$.*
3. *There exists a bounded operator $X \in \mathscr{B}(H)$; such that $B = AX$.*

*If one of these conditions holds, then there exists a unique operator* $D \in \mathscr{B}(H)$ *such that* $AD = B$ *and* $R(D) \subseteq \overline{R(A^*)}$. $D$ *is called Douglas solution of the equation* $AX = B$

For a quotient operator $B/A$, we can easily deduce from this lemma the following:

**Corollary 1.** *If* $R(B^*) \subset R(A^*)$*, then* $B/A$ *is bounded.*

*Proof.* Since $R(B^*) \subset R(A^*)$, there exists $X^* \in \mathscr{B}(H)$ such that $A^*X^* = B^*$, in other words, $B = XA$ where $(X^*)^* = X$. Hence $X$ is a bounded extension of $B/A$, so, $B/A$ is bounded on $H$.

Obviously, if $R(B^*) \subset R(A^*)$, then $N(A) \subset N(B)$. Thus, for what conditions on $A$ and $B$ we have the converse implication?

As answer of this question, we have the following result due to Barnes [2].

**Proposition 1.** *For two bounded operator* $A, B \in \mathscr{B}(H)$ *such that* $A$ *has closed range and* $N(A) \subset N(B)$ *we have* $R(B^*) \subset R(A^*)$.

This proposition imply immediately the following:

**Theorem 1.** *If* $B/A$ *is quotient operator with closed domain, that is,* $R(A)$ *is closed in* $(H)$*, then* $B/A$ *is bounded.*

Recall from [1, 4] and [15] that if $R(A)$ is closed, then there exists a unique bounded operator $A^\dagger$ called the Moore penrose generalized inverse of $A$ such that

$$AA^\dagger A = A;\ A^\dagger AA^\dagger = A^\dagger;\ (A^\dagger A)^* = A^\dagger A;$$
$$(AA^\dagger)^* = AA^\dagger;\ A^\dagger A = P_{R(A^\dagger)};\ AA^\dagger = P_{R(A)}.$$

This yields a question about the expression of the quotient operator $B/A$ using $A^\dagger$. For this, we have the following corollary.

**Corollary 2.** *If* $B/A$ *is quotient operator with closed domain, then* $B/A = BA^\dagger$. *(Respectively, if* $A$ *is invertible, then* $B/A = BA^{-1}$*).*

*Proof.* It follows immediately from the properties of $A^\dagger$. It my be very important to note that Kaufman proved in [13] that the quotient operator is only what was called semi closed operator and we obtain from [3] and [17] the following result:

**Corollary 3.** *If* $B/A$ *is quotient operator, then it is bounded from* $(R(A), \langle .,. \rangle_{B/A})$ *onto* $H$*, where* $\langle .,. \rangle_{B/A}$ *is the inner product define for all* $x, y \in R(A)$ *by*

$$\langle x, y \rangle_{B/A} = \langle x, y \rangle + \langle (B/A)_x, (B/A)_y \rangle$$

Recall from [7, 11] that an operator $T \in \mathscr{B}(H)$ is called compact if for any bounded sequence $(f_n)_{n\in\mathbb{N}}$ in $H$, the sequence $(Tf_n)_{n\in\mathbb{N}}$ has a Cauchy subsequence.

According to this definition, we characterize the compact quotient operator as follows:

**Proposition 2.** *$B/A$ is also compact if only if $B$ is compact on $H$.*

*Proof.* Let $(f_n)_{n\in\mathbb{N}}$ be a bounded sequence in $H$. Since $A$ is bounded in $H$, the sequence $(Af_n)_{n\in\mathbb{N}}$ is also in $H$, and by the compactness of $B$, the sequence $((B/A)Af_n)_{n\in\mathbb{N}} = (Bf_n)_{n\in\mathbb{N}}$ has a Cauchy subsequence in $H$. Hence $B/A$ is compact on $H$. With similar manner we prove the converse implication.

Note that if $B/A$ is compact, then $B/A$ is bounded. In fact, we have

$$\sup_{\|f\|\leq 1} \|(B/A)f\| < \infty$$

Otherwise we have a sequence $(f_n)_{n\in\mathbb{N}}$ with $\|f_n\| \leq 1$ for all $n \in \mathbb{N}$ such that $\lim_{n\longrightarrow\infty} \|(B/A)f_n\| = \infty$, which excludes the existence of a Cauchy subsequence of $((B/A)f_n)_{n\in\mathbb{N}}$.

**Corollary 4.** *The quotient of two compact operators is compact.*

## *2.2 About the inverse of quotient operator*

Our intention in this paragraph is to prove the following theorem concerning the invertibility of quotient operator.

**Theorem 2.** *Let $B/A$ be a quotient operator on $H$. Then*

1. *If $N(A) = N(B)$, then $B/A$ is invertible and $(B/A)^{-1} = A/B$.*
2. *If $B/A$ is a bounded quotient with a closed range $R(B)$ in $H$, then $B/A$ has a generalized inverse $(B/A)^{\dagger} = AB^{\dagger}$*

*$(B/A)^{\dagger}$ is then called the Moore-Penrose generalized inverse of $B/A$.*

*Proof.* 1. The operator $A/B$ is well defined from the condition $N(A) = N(B)$.
Since the domain of $A/B$ is $R(B)$, we notice that the compositions

$$(A/B)(B/A):\ Au \longrightarrow Bu \longrightarrow Au \ \text{ for all } u \in H$$

$$(B/A)(A/B):\ Bv \longrightarrow Av \longrightarrow Bv \ \text{ for all } v \in H$$

give the desired equality.

2. If $B/A$ is bounded, then $R(A)$ is closed in $H$ and $B/A = BA^{\dagger}$. Thus,

$$(B/A)^{\dagger}(B/A) = AB^{\dagger}BA^{\dagger} = A(B^{\dagger}B)A^{\dagger} = P_{R(A)}$$

$$(B/A)(B/A)^{\dagger} = BA^{\dagger}AB^{\dagger} = B(A^{\dagger}A)B^{\dagger} = P_{R(B)}$$

**Corollary 5.** *The quotient operator $B/A$ has an everywhere defined and bounded inverse if and only if the operator $B$ is invertible and*

$$T^{-1} = A/B = AB^{-1}$$

For a closed densely defined operator $T \in \mathscr{C}(H)$ we have the following theorem.

**Theorem 3.** *Let $T \in \mathscr{C}(H)$ represented by $T = \Gamma(B) = B/A = BA^{-1}$ with $B \in \mathscr{C}_0(H)$ and $A = (I - B^*B)^{\frac{1}{2}}$. If $B^\dagger$ exists, then $T^\dagger$ exists and $T^\dagger = AB^\dagger$.*

*Proof.* Define $S = AB^\dagger$. By the definition of $\Gamma(B)$, we have $D(T) = R(A)$ and $R(T) = R(B)$. Since $B^*B$ commutes with $B^\dagger B$, and $AA^{-1}x = x$ for any $x \in R(A)$, we have

$$STx = AB^\dagger BA^{-1}x = B^\dagger BAA^{-1}x = B^\dagger Bx$$

for any $x \in D(T)$. On the other hand,

$$TSy = BA^{-1}AB^\dagger y = BB^\dagger y$$

for any $y \in R(T)$, and the uniqueness of $T^\dagger$ implies that $S = T^\dagger$.

## 2.3 Powers of quotient operator

First, we note that it is necessary to assume that $R(B) \subset R(A)$, so that we can discuss about the powers of the quotient operator $B/A$.

**Theorem 4.** *Let $B/A$ be quotient of two commuting bounded operators $A$ and $B$ such that $R(B) \subset R(A)$. Then*

$$(B/A)^n = B^n/A^n \text{ for all } n \in \mathbb{N}.$$

*Proof.* We proceed by induction on the values of $n$ using the definition of product of quotient operators.

The quotient $B/A$ is nilpotent (resp. idempotent) if $R(B) \subset R(A)$ and $(B/A)^2 = 0$ (resp. $(B/A)^2 = B/A$). This implies immediately the following.

**Theorem 5.** *The quotient operator $B/A$ is nilpotent (resp. idempotent) if the solution $X$ of the Douglas equation $AX = B$ is nilpotent (resp. idempotent).*

It follows from this theorem that the quotient of two idempotent operators is idempotent, and that if $B$ is nilpotent then $B/A$ is again nilpotent.

In general, for $n \in \mathbb{N}$, $B/A$ is $n$-nilpotent (resp. $n$-idempotent) if $AX = B$ and $X^n = 0$ (resp. $X^n = X$).

The quotient operator $B/A$ is said to be quasi-nilpotent if it is a null spectral radius, that is

$$\lim_{n\longrightarrow\infty} \|(B/A)^n Ax\|^{\frac{1}{n}} = 0$$

**Theorem 6.** *The quotient operator $B/A$ is quasi-nilpotent if the solution $X$ of the Douglas equation $AX = B$ is quasi-nilpotent.*

*Proof.* Since

$$\|(B/A)^n Ax\|^{\frac{1}{n}} = \|AX^n x\|^{\frac{1}{n}} \text{ for all } x \in H.$$

Then, if $X$ is quasi-nilpotent, we have

$$\lim_{n\longrightarrow\infty} \|(B/A)^n Ax\|^{\frac{1}{n}} = \lim_{n\longrightarrow\infty} \|A\|^{\frac{1}{n}} \|X^n x\|^{\frac{1}{n}} = 0$$

## 2.4 Limit of a sequence of quotient operators

In the following theorem, we prove that the limit of a converging quotient operators sequence is also quotient operator.

**Theorem 7.** *Let $(B_n/A_n)_{n\in\mathbb{N}}$ be a sequence of quotient operators converging to an operator $C$ with domain $\mathscr{D}(C) = \bigcap_{n\in\mathbb{N}} R(A_n) \cap K$, where $K$ the Hilbert space of all $x$ in $H$ such that $\lim_{n\longrightarrow\infty} (B_n/A_n)_n x$ exists. Then $C$ is quotient of two bounded operators.*

*Proof.* First let $Q_n = B_n/A_n$ for all $n \in \mathbb{N}$. As we have done above (Corollary 3), we consider for all $(x,y) \in (\mathscr{D}(C))^2$ the inner product:

$$\langle x,y\rangle_C = \langle x,y\rangle_K + \langle Cx, Cy\rangle = \langle x,y\rangle_K + \lim_{n\longrightarrow\infty} \langle Q_n x, Q_n y\rangle.$$

We now show that $(\mathscr{D}(C), \|.\|_C)$ is complete.

Let $(x_m)_{m\in\mathbb{N}}$ be a Cauchy sequence in $(\mathscr{D}(C), \|.\|_C)$. Clearly, $(x_m)_{m\in\mathbb{N}}$ is Cauchy in $K, H$ and $(R(A_n), \|.\|_{Q_n})$. Hence, $(x_m)_{m\in\mathbb{N}}$ converges to $x$ in $\mathscr{D}(C)$. We have from Corollary 3

$$\|Q_n x_m - Q_n x\| \underset{m\longrightarrow\infty}{\longrightarrow} 0.$$

Therefore

$$\|x_m - x\|_C = \|x_m - x\|_K + \lim_{n\longrightarrow\infty} \|Q_n x_m - Q_n x\| \underset{m\longrightarrow\infty}{\longrightarrow} 0$$

So, $(\mathscr{D}(C), \|.\|_C)$ is complete.

It follows from a result of Mac-Nerney [16],Theorem 3 that there exists an operator $A \in \mathscr{B}(H)$ such that $R(A) = \mathscr{D}(C)$ and for all $(x, y) \in (\mathscr{D}(C))^2$

$$\langle x, y\rangle_C = \langle A^{-1}x, A^{-1}y\rangle$$

Set $B = CA$. Then we have for all $x \in H$

$$\|Bx\|^2 \leq \langle Ax, Ax\rangle_C \leq \|x\|^2$$

Hence, $B$ is bounded on $H$ and $C$ is the quotient $B/A$.

## References

1. A. Ben-Israel and T.N.E. Greville, *Generalized Inverses: Theory and Applications*. Second edition, Springer, New York, 2003.
2. Bruce, A. Barnes, *Majorization, range inclusion, and factorisation for bounded linear operators*, Proc. Amer. Math. Soc, 2005, 133, 155–162
3. G. Djellouli, S. Messirdi and B. Messirdi, *Some Stronger Topologies for Closed Operators in Hilbert Space*.Int. J. Contemp. Math. Sciences, 2010, 5, no. 25, 1223–1232.
4. D. S. Djordjević, *Characterizations of normal, hyponormal and EP operators*. J. Math. Anal. Appl. 2007, 329, 1181–1190.
5. R. G. Douglas, *O*n majorization and range inclusion of operators in Hilbert space, Proc. Amer. Math. Soc. 1966, 17, 413–416.
6. P.A. Fillmore, J.P. Williams, *On operator ranges*, Adv. Math. 1971, 7, 254–281.
7. S. Goldberg, *Unbounded Linear Operators.* McGraw-Hill, New-York, 1966.
8. G. Hirasawa, *quotient of bounded operators and Kaufman's theorem*. Math. J. Toyama Univ. 1995, 18, 215–224.
9. S. Izumino, *Quotients of bounded operators*, Proc. Amer. Math. Soc. 1989, 106, 427–435.
10. S. Izumino, *Quotients of bounded operators and their weak adjoints*, J. Operator theory. 1993, 29,83–96
11. T. Kato, *Perturbation theory for linear operators*, Springer-Verlag, New York, 1966.
12. W. E. Kaufman, *Representing a closed operator as a quotient of continuous operators*, Proc. Amer. Math. Soc. 1978, 72, 531–534.
13. W. E. Kaufman, *Semiclosed operators in Hilbert space*, Proc. Amer. Math. Soc. 1979, 76, 67–73.
14. W. E. Kaufman, *Closed operators and pure contractions in Hilbert space*, Proc. Amer. Math. Soc. 1983, 87, 83–87.
15. J-Ph. Labrousse, *Inverses généralisés d'opérateurs non bornés*, Proc. Amer. Math. Soc. 1992, 115, 125–129.
16. J. S. Mac-Nerney, *Investigation concerning positive definite continued fractions*. Duke Math.J. 1959, 26,663–678. Amer. Math. Soc. 1978, 72, 531–534.
17. S. Messirdi, M. Djaa and B. Messirdi, *Stability of Almost Closed Operators on a Hilbert Space*. Sarajevo J. of Math. 2009, 5(17), 133–141.

# Exact Controllability For Korteweg-De Vries Equation and its Cost in the Zero-Dispersion Limit

**Hajer Dbebria and Ali Salem**

**Abstract** In this paper, we consider the problem of exact boundary controllability of a linear Korteweg-de Vries (KdV) equation in a bounded domain when the condition for the control is the difference between the derivative of the solution in the left and right endpoint. We prove the existence of a countable set of critical lengths out of which we have the exact controllability. In the second part of this paper, we study the behavior of the optimal control and how the cost of controllability evolves as the dispersive term brought to zero.

**Keywords** Exact Controllability • Korteweg-de Vries equation • Hilbert Uniqueness Method (HUM) • Ingham's inequality • Cost of null controllability

## 1 Introduction and Main Results

The Korteweg-de Vries (KdV) equation:

$$y_t + y_x + y_{xxx} + yy_x = 0, \tag{1}$$

may serve as a model for propagation of small amplitude long water waves in a uniform channel. In this context, $t$ is the time variable, $x$ is the space variable, and $y$ the state, stands for the deviation of the liquid surface from the equilibrium position.

Let us be more specific on the problem under view. Let $T > 0$ (final time), $L > 0$ (the length of the domain), and $y^0$, $y^T \in H^{-1}(0, L)$, does it exist a control function

H. Dbebria (✉)
Laboratory of Engineering Mathematics, Polytechnic School of Tunisia, University of Carthage, Carthage, Tunisia
e-mail: hajer.dbebria@gmail.com

A. Salem
College of Arts and Sciences in Noairyya, University of Dammam, King of Saudi Arabia, Dammam, Saudi Arabia

Laboratory of Engineering Mathematics, Polytechnic School of Tunisia, University of Carthage, Carthage, Tunisia
e-mail: a3li.salem@gmail.com



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_19

$v(t) \in L^2(0,T)$ such that the solution of the following Cauchy problem:

$$\begin{cases} y_t + y_x + \varepsilon y_{xxx} = 0 & in\ (0,T)\times(0,L), \\ y(t,0) = y(t,L) = 0 & in\ (0,T), \\ y_x(t,L) - y_x(t,0) = v(t) & in\ (0,T), \\ y(0,x) = y^0(x) & in\ (0,L) \end{cases} \tag{2}$$

satisfies $y(T,x) = y_T(x)$?

Here, $\varepsilon$ is a positive dispersion cœfficient and $v(t) \in \mathbb{R}$ constitute the control of our system.

In this paper, we are interested in two types of controllability results concerning this system. These two types are the following.

- First, we consider the problem of exact controllability for system (2), when the dispersion coefficient is fixed (Theorems 3).
- Next, we are interested in how the cost of null controllability evolves as the dispersive term tends to 0 (Theorem 4).

Many results of controllability have been studied in recent years for KdV equation (1). In particular several different cases have been considered: the case where all three boundary conditions are used as controls (see [4]). If we act on the left Dirichlet boundary condition and homogeneous data is considered at the right, then the system behaves like a heat equation and only null controllability can be proven [6, 11]. On the other hand, if we act on the two right data and homogeneous boundary condition is considered at the left, then the system behaves like a wave equation with an infinite speed of propagation (see [10]). When we put only one control input at the right endpoint and keep homogeneous the other two boundary conditions: there exist some spatial domains (intervals of some given lengths) for which the corresponding linearized KdV equation is not any more controllable [10, 11]. In spite of that, in these critical cases the nonlinearity gives the exact controllability of the nonlinear KdV equation [2, 3, 5].

This paper is organized as follows. In Section 2, we recall the well-posedness results for the linear KdV control system and prove the exact controllability. It is done by using spectral analysis and the HUM method; see [9]. In Section 3 we establish that the cost of null controllability will dramatically increase as the dispersive term brought to zero.

## 2 Exact Controllability

In this section we aim to apply the HUM [9], so we have to study the backward problem:

$$\begin{cases} z_t + z_x + \varepsilon z_{xxx} = 0 \\ z(t,0) = z(t,L) = 0, \\ z_x(t,L) - z_x(t,0) = 0, \\ z(T,x) = z^T(x). \end{cases} \tag{3}$$

Performing the change of variables $\tau = T - t$ and $\xi = -x$ in (3) and replacing $(\tau, \xi)$ by $(t, x)$, we get the following homogeneous problem:

$$\begin{cases} z_t + z_x + \varepsilon z_{xxx} = 0, \\ z(t,0) = z(t,L) = 0, \\ z_x(t,L) - z_x(t,0) = 0, \\ z(0,x) = z^0(x). \end{cases} \tag{4}$$

We begin by showing the well-posedness of homogeneous problem (4).

## 2.1 *Well-posedness of Cauchy Problem*

Let $A$ denote the operator

$$A : z \mapsto -z_x - \varepsilon z_{xxx} \tag{5}$$

on the (dense) domain $\mathscr{D}(A) \subset \mathrm{L}^2(0,L)$ defined by

$$\mathscr{D}(A) := \Big\{ z \in H^3(0,L); z(0) = z(L) = 0; z_x(0) = z_x(L) \Big\}. \tag{6}$$

Hence, from the classical semigroup results, one sees that the operator $A$ is an infinitesimal generator of a continuous group. Also it's not difficult to see that the skew-adjoint operator $A$ has a compact resolvent. Hence the spectrum $\sigma(A)$ of $A$ consists only of eigenvalues. Furthermore the spectrum of A is a discrete subset of $iR$ and the eigenfunctions form an orthonormal basis of $H^1(0,L)$. We denote by $(i\lambda_k)_{k\in Z}$ the eigenvalues of $A$ and by $(\phi_k)_{k\in\mathbb{Z}}$ its eigenfunctions. We can then state the following well-posedness result:

**Theorem 2.1.** *For any $z^0 \in H^1(0,L)$, there exists a unique solution of the homogeneous problem (4), which belongs to $H^1(0,L)$ and is given by:*

$$\begin{cases} z(t,x) = \sum_{k\in Z} e^{i\lambda_k t} z_0^k \phi_k(x), \\ z^0 = \sum_{k \in Z} z_0^k \phi_k. \end{cases} \tag{7}$$

*Moreover,*

$$\forall t \in \mathbb{R} : \|z(t,.)\|_{H^1(0,L)} = \|z_0\|_{H^1(0,L)}.$$

Let us now define what we mean by a solution of our control system (2).

**Definition 1.** Let $\varepsilon > 0,\ T > 0,\ y_0 \in H^{-1}(0,L)$ and $v \in L^2(0,T)$. A solution of the nonhomogeneous problem (2) is a function $y \in C([0,T], H^{-1}(0,L))$ satisfying $y(0) = y^0$ and $\forall\ \tau \in [0,T],\ \forall\ z^0 \in H^1(0,L)$,

$$< y(\tau), z(\tau) >_{-1,1} = < y^0, z^0 >_{-1,1} + \int_0^T \varepsilon z_x(t,L) v(t) dt.$$

With this definition, we obtain the following result

**Proposition 1.** *Let $\varepsilon > 0,\ T > 0$. Let $y^0 \in H^{-1}(0,L)$ and $v \in L^2(0,T)$. Then the problem (2) has a unique weak solution (defined by transposition).*

## 2.2 *Observability of the Homogeneous Problem*

In this subsection, we apply the duality between controllability and observability which reduces the control problem to an observability problem to the adjoint problem (3). There are many techniques that are useful to address the problem of observability such as the Ingham's inequalities.

### 2.2.1 Ingham's Inequality

**Lemma 1 ([7]).** *Let $T > 0$. Let $(\beta_k)_{k\in\mathbb{Z}} \in \mathbb{R}$ be a sequence of pairwise distinct real numbers such that*

$$\lim_{|k|\to\infty} \beta_{k+1} - \beta_k = +\infty.$$

*Then there exist two strictly positive constants $c_1(T)$ and $c_2(T)$ such that for any sequence $(\gamma_k)_{k\in\mathbb{Z}} \in \mathbb{R}$ satisfying $\sum_{k\in\mathbb{Z}} \gamma_k^2 < +\infty$, the series $f(t) = \sum_{k\in\mathbb{Z}} \gamma_k e^{i\beta_k t}$ converges in $L^2(0,T)$ and satisfies*

$$c_1(T) \sum_{k\in\mathbb{Z}} \gamma_k^2 \leq \int_0^T |\, f(t)\,|^2\, dt \leq c_2(T) \sum_{k\in\mathbb{Z}} \gamma_k^2.$$

In order to apply this lemma we first study the sequences of eigenvalues and eigenfunctions of $A$.

**Proposition 2.** *The real numbers $(\lambda_k)_{k\in\mathbb{Z}}$ have the asymptotic form*

$$\lambda_k = 8\varepsilon \frac{k^3\pi^3}{L^3} + o(k^2) \text{ as } k \to \pm\infty. \tag{8}$$

*Proof.* The eigenvalue problem to be considered is

$$\begin{cases} -\phi' - \varepsilon\phi''' = i\lambda\phi, \\ \phi(0) = \phi(L) = 0, \\ \phi'(0) = \phi'(L). \end{cases} \tag{9}$$

The characteristic equation of (9) is

$$\varepsilon z^3 + z + i\lambda = 0 \tag{10}$$

Let us denote a solution of this equation $z = 2ia$, where $a \in \mathbb{R}$. It follows that the eigenvalues are

$$\lambda = 2a(4\varepsilon a^2 - 1).$$

Thus the three solutions of (10) are

$$z_0 = \sqrt{| 3a^2 - \frac{1}{\varepsilon} |} - ia, \; z_1 = -\sqrt{| 3a^2 - \frac{1}{\varepsilon} |} - ia, \; z_2 = 2ia.$$

We distinguish 3 cases.

1. $3a^2 - \dfrac{1}{\varepsilon} < 0.$

   In this case, it is easy to see that the eigenfunction $\phi$ of $A$ associated with the eigenvalue $\lambda = 2a(4\varepsilon a^2 - 1)$ may be written

$$\phi(x) = e^{-iax}\alpha\cos(x\sqrt{-(3a^2 - \frac{1}{\varepsilon})}) + e^{-iax}\beta\sin(x\sqrt{-(3a^2 - \frac{1}{\varepsilon})}) + \gamma e^{2iax}$$

   where $\alpha, \beta$, and $\gamma$ are some constants such that:

$$\phi(0) = \phi(L) = 0 \quad \text{and} \quad \phi'(0) = \phi'(L).$$

   That means, such that

$$\alpha + \gamma = 0.$$

$$e^{-iaL}\alpha\cos(L\sqrt{-(3a^2 - \frac{1}{\varepsilon})}) + e^{-iaL}\beta\sin(L\sqrt{-(3a^2 - \frac{1}{\varepsilon})}) - \alpha e^{-2iaL} = 0. \tag{11}$$

$$
\begin{aligned}
&- 3ia\alpha + \beta\sqrt{-3a^2 + \frac{1}{\varepsilon}} \\
&\quad = -iae^{-iaL}\left(\alpha\cos(L\sqrt{-3a^2 + \frac{1}{\varepsilon}}) + \beta\sin(L\sqrt{-(3a^2 - \frac{1}{\varepsilon})})\right) \\
&\quad\quad - e^{-iaL}\alpha\sqrt{-3a^2 + \frac{1}{\varepsilon}}\sin(L\sqrt{-3a^2 + \frac{1}{\varepsilon}}) \\
&\quad\quad + e^{-iaL}\beta\sqrt{-3a^2 + \frac{1}{\varepsilon}}\cos(L\sqrt{-3a^2 + \frac{1}{\varepsilon}})) - 2ia\alpha e^{2iaL}. \qquad (12)
\end{aligned}
$$

From (11), one obtains

$$
\beta = \alpha\frac{e^{3ial} - \cos(L\sqrt{-(3a^2 - \frac{1}{\varepsilon})})}{\sin(L\sqrt{-(3a^2 - \frac{1}{\varepsilon})})}.
$$

Taking the real part of equation (12), one obtains that a must satisfy

$$
\begin{aligned}
\sqrt{-(3a^2 - \frac{1}{\varepsilon})}\cos(2aL) &= 3a\sin(aL)\sin(L\sqrt{-(3a^2 - \frac{1}{\varepsilon})}) + \\
&\sqrt{-(3a^2 - \frac{1}{\varepsilon})}\cos(aL)\cos(L\sqrt{-(3a^2 - \frac{1}{\varepsilon})}) \qquad (13)
\end{aligned}
$$

The number of parameters $a$ satisfying (13) is finite and depends on $L$ and $\varepsilon$. As if $a$ satisfies (13), then $(-a)$ so, we find in this case $2N_{L,\varepsilon}$ eigenvalues $\{\lambda_{-N_{L,\varepsilon}}, \ldots, \lambda_{-1}, \lambda_1, \ldots, \lambda_{N_{L,\varepsilon}}\}$.

2. $3a^2 - \frac{1}{\boldsymbol{\varepsilon}} = 0$

   In this case, we don't find any eigenfunction satisfy the boundary conditions.

3. $3a^2 - \frac{1}{\boldsymbol{\varepsilon}} > 0$

   In this case, it is not difficult to see that the eigenfunction $\phi$ of $A$ associated with the eigenvalue $\lambda = 2a(4\varepsilon a^2 - 1)$ may be written as

$$
\phi(x) = e^{-iax}\alpha\cosh(x\sqrt{3a^2 - \frac{1}{\varepsilon}}) + e^{-iax}\beta\sinh(x\sqrt{3a^2 - \frac{1}{\varepsilon}}) + \gamma e^{2iax} \qquad (14)
$$

where $\alpha$, $\beta$, and $\gamma$ are some constants such that $\phi(0) = \phi(L) = 0$ and $\phi'(0) = \phi'(L)$. That means, such that

$$
\alpha + \gamma = 0,
$$

$$e^{-iaL}\left(\alpha\cosh(L\sqrt{3a^2-\frac{1}{\varepsilon}})+\beta\sinh(L\sqrt{3a^2-\frac{1}{\varepsilon}})\right)-\alpha e^{2iaL} \;=\; 0 \tag{15}$$

$$\begin{aligned}&-iae^{-iaL}\left(\alpha\cosh(L\sqrt{3a^2-\frac{1}{\varepsilon}})+\beta\sinh(L\sqrt{3a^2-\frac{1}{\varepsilon}})\right)+\\&\quad e^{-iaL}\sqrt{3a^2-\frac{1}{\varepsilon}}\left(\alpha\sinh(L\sqrt{3a^2-\frac{1}{\varepsilon}})+\beta\cosh(L\sqrt{3a^2-\frac{1}{\varepsilon}})\right)\\&\quad -2ia\alpha e^{2iaL}=-3ia\alpha+\beta\sqrt{3a^2-\frac{1}{\varepsilon}}L.\end{aligned} \tag{16}$$

We deduce from (15)-(16) that

$$\beta=\alpha\frac{e^{3iaL}-\cosh(L\sqrt{3a^2-\frac{1}{\varepsilon}})}{\sinh(L\sqrt{3a^2-\frac{1}{\varepsilon}})},$$

and

$$\begin{aligned}&-3a\alpha+\mathscr{I}(\beta)\sqrt{3a^2-\frac{1}{\varepsilon}}=-3a\alpha\cos(2aL)+\alpha\sqrt{3a^2-\frac{1}{\varepsilon}}\sin(-aL)\\&\left(\sinh(L\sqrt{3a^2-\frac{1}{\varepsilon}})+\mathscr{R}(\beta)\cosh(L\sqrt{3a^2-\frac{1}{\varepsilon}})\right)\\&+\sqrt{3a^2-\frac{1}{\varepsilon}}\cos(aL)\mathscr{I}(\beta)\cosh(L\sqrt{3a^2-\frac{1}{\varepsilon}}).\end{aligned}$$

From these equations, one obtains that $a$ satisfies the following one

$$-3a\sinh(L\sqrt{3a^2-\frac{1}{\varepsilon}})\sin(aL)+\sqrt{3a^2-\frac{1}{\varepsilon}}\cos(2aL)-\cosh(\sqrt{3a^2-\frac{1}{\varepsilon}})\cos(aL)=0. \tag{17}$$

If one neglects the terms $e^{-L\sqrt{3a^2-\frac{1}{\varepsilon}}}$ as $a\to\pm\infty$, one gets

$$e^{L\sqrt{3a^2-\frac{1}{\varepsilon}}}=\frac{\cos(2aL)}{\frac{\sqrt{3}}{2}\sin(aL)+\frac{1}{2}\cos(aL)}=\frac{\cos(2aL)}{\cos(aL-\frac{\pi}{3})}$$

and hence there exists a unique solution $a_{k+N_{L,\varepsilon}}$ defined by equation (17) and given asymptotically by

$$a_k = \frac{5\pi}{6L} + \frac{k\pi}{L} + o(\frac{1}{k}) \quad \text{(respectively} \quad a_{-k} = \frac{-5\pi}{6L} - \frac{k\pi}{L} + o(\frac{1}{k})). \tag{18}$$

The associated eigenfunction $\phi_k$ is

$$\phi_k(x) = \alpha_k \left[ e^{-ia_k x} \left( cosh(x\sqrt{3a^2 - \frac{1}{\varepsilon}}) + \frac{e^{3ia_k L} - \cosh(L\sqrt{3a^2 - \frac{1}{\varepsilon}})}{\sinh(L\sqrt{3a^2 - \frac{1}{\varepsilon}})} \sinh(x\sqrt{3a^2 - \frac{1}{\varepsilon}}) \right) \right] - \alpha_k e^{2ia_k x}. \tag{19}$$

where $\alpha_k$ is chosen in such a way that $\| \ \phi_k \ \|_{H^1(0,L)} = 1$. Thus, from (18), one deduces the asymptotic behavior of the eigenvalues and therefore the proof of this proposition is complete. □

From the proof of the last proposition, we deduce the following lemma.

**Lemma 2.** *There exists a constant* $C > 0$ *such that*

$$\lim_{k\to\infty} \frac{|\ \phi_k'(L)\ |}{|\ k\ |} = C. \tag{20}$$

*Proof.* By using the formula of the eigenfunctions, we get

$$|\ \phi_k'(L)\ | = |\ \alpha_k\ | \left| -3ia_k + \sqrt{3a_k^2 - \frac{1}{\varepsilon}} \left( \frac{\cosh(L\sqrt{3a_k^2 - \frac{1}{\varepsilon}})}{\sinh(L\sqrt{3a_k^2 - \frac{1}{\varepsilon}})} - \frac{e^{-3ia_k L}}{\sinh(L\sqrt{3a_k^2 - \frac{1}{\varepsilon}})} \right) \right|.$$

With (18), we find that

$$\begin{aligned} \lim_{k\to\pm\infty} \frac{|\ \phi_k'(L)\ |}{|\ k\ |} &= |\ \alpha\ | \left| -3i\frac{\pi}{L} + \frac{\pi}{L}\sqrt{3} \right| \\ &= |\ \alpha\ | \ \frac{2\pi\sqrt{3}}{L} > 0. \end{aligned}$$

□

We apply the Ingham's inequality lemma to the function:

$$f(t) = \sum_{k\in\mathbb{Z}} \gamma_k e^{i\beta_k t} = \varepsilon z_x(t, L)$$

where

$$\forall k \in \mathbb{Z}, \quad \begin{cases} \gamma_k = \varepsilon z_0^k \phi_k'(L), \\ z^0 = \sum_{k\in\mathbb{Z}} z_0^k \phi_k \in H^1(0,L), \\ \beta_k = \lambda_k. \end{cases}$$

Then, we get the existence of two constants $c_1(T),\ c_2(T) > 0$ (see [8]) such that for any $z^0 \in H^1(0,L)$,

$$c_1 \sum_{k\in\mathbb{Z}} \varepsilon^2 \mid z_0^k \mid^2 \mid \phi_k'(L) \mid^2 \leq \int_0^T \mid \varepsilon z_x(t,L) \mid^2 dt \leq c_2 \sum_{k\in\mathbb{Z}} \varepsilon^2 \mid z_0^k \mid^2 \mid \phi_k'(L) \mid^2 . \tag{21}$$

*Remark 2.1.* The left-hand inequality in (21) is called an observability inequality and the right-hand one is called an admissibility inequality.

We can estimate by above this two inequalities in terms of the $H^1$-norm of $z^0$. In order to do that the following condition

$$\forall k \in \mathbb{Z},\ \phi_k'(L) \neq 0 \tag{22}$$

is required.

In the following lemma we focus on the length of the domain $L$ such that $\phi_k'(L) \neq 0$.

**Lemma 3.** *Let* $L > 0$. *Consider the following assertion:*
$(\mathscr{A})$ : $\exists \lambda \in \mathbb{C},\ \exists \varphi \in H^3(0,L)\backslash\{0\}$ *such that:*

$$\begin{cases} \lambda\varphi + \varphi' + \varepsilon\varphi''' = 0, \\ \varphi(0) = \varphi(L) = \varphi'(0) = \varphi'(L) = 0. \end{cases}$$

*Then* $(\mathscr{A}) \Leftrightarrow L \in \mathscr{N}_\varepsilon = \left\{ 2\pi\sqrt{\varepsilon\frac{k^2+kl+l^2}{3}};\ k,\ l \in \mathbb{N}^* \right\}$.

*Remark 2.2.* 1. $\sum_{k\in\mathbb{Z}} k^2 \mid z_0^k \mid^2$ is the square norm of $\|z_0\|^2_{H^1(0,L)}$. In fact, we have

$$\|z_0\|^2_{H^1(0,L)} = \left\| \sum_{k\in\mathbb{Z}} z_0^k \phi_k \right\|^2_{H^1(0,L)} = \sum_{k\in\mathbb{Z}} (1+ \mid \lambda_k \mid)^{2/3} \mid z_0^k \mid^2 . \tag{23}$$

From the asymptotic form of the eigenvalues, we can deduce

$$\begin{aligned} \|z_0\|^2_{H^1(0,L)} &= \sum_{k\in\mathbb{Z}} (1+ \mid \lambda_k \mid)^{2/3} \mid z_0^k \mid^2 \\ &= \sum_{k\in\mathbb{Z}} (C(\varepsilon))^{2/3} k^2 \mid z_0^k \mid^2 + o(1/k)^{4/3}. \end{aligned}$$

2. Given the asymptotic behavior of $\phi'_k(L)$ as $k \to \pm\infty$, there exist two constants $m(L)$ and $M(L)$ such that

$$| m(L) | k \leq | \phi'_k(L) | \leq | M(L) | k$$

With those remarks we obtain the following result:

**Theorem 2.2.** *Let $z^0 \in H^1(0,L)$, there exist two positive constants $c_{T,L,\varepsilon}$ and $C_{T,L,\varepsilon}$ such that*

$$c_{T,L,\varepsilon} \|z_0\|^2_{H^1(0,L)} \leq \int_0^T | \varepsilon z_x(t,L) |^2 \, dt \leq C_{T,L,\varepsilon} \|z_0\|^2_{H^1(0,L)} \tag{24}$$

*where $c_{T,L,\varepsilon} = \dfrac{L^2}{4\pi^2} c_1 m(L)^2 \varepsilon^{4/3}$ and $C_{T,L,\varepsilon} = \dfrac{L^2}{4\pi^2} c_2 M(L)^2 \varepsilon^{4/3}$ with $c_1$, $c_2$ are the constants of Ingham's lemma.*

### 2.2.2 Application of HUM

Thanks to the observability inequality, we can apply the Hilbert Uniqueness Method. We consider the map:

$$\Lambda : \ H^1(0,L) \to H^{-1}(0,L)$$
$$z^0 \mapsto \ y(T,.).$$

where $y$ is the solution of nonhomogenous problem (2) and $z$ the solution of homogeneous problem (4). Thanks to the time reversibility of (2) and Proposition 1, $\Lambda$ is a well-defined continuous map from $H^1(0,L)$ into its dual $H^{-1}(0,L)$. On the other hand,

$$< \Lambda(z_0), z_0 >_{H^{-1},H^1} = \int_0^T | \varepsilon z_x(t,L) |^2 \, dt.$$

Thanks to the following observability inequality

$$c_{T,L,\varepsilon} \|z_0\|^2_{H^1(0,L)} \leq \int_0^T | \varepsilon z_x(t,L) |^2 \, dt,$$

we conclude that

$$< \Lambda(z_0), z_0 >_{-1,1} \geq c_{T,L,\varepsilon} \|z_0\|^2_{H^1(0,L)} \text{ where } c_{T,L,\varepsilon} = \frac{L^2}{4\pi^2} c_1 \, | \, m(L) \, |^2 \, \varepsilon^{4/3}.$$

We deduce then from the Lax-Milgram's theorem [1] that $\Lambda$ is an isomorphism from $H^1(0, L)$ into $H^{-1}(0, L)$. We now prove the exact controllability of the linear KdV equation and we have the following theorem of controllability.

**Theorem 2.3.** *Let $T > 0$ and $L > 0$ be such that $L \in \mathbb{R}_+ \backslash \mathcal{N}$ where $\mathcal{N}_\varepsilon := \{2\pi\sqrt{\varepsilon\frac{k^2+kl+l^2}{3}},\ k,\ l \in \mathbb{N}^*\}$. Let $y_0,\ y_T \in H^{-1}(0, L)$. Then there exists a control $v \in L^2(0, T)$ such that the solution of (4) satisfies $y(T, .) = y_T$.*

*Remark 2.3.* Thanks to the HUM, one can choose a control $v \in L^2(0, T)$ of minimal $L^2$-norm among all the controls driving the system from $y^0$ at $t = 0$ to $y^T$ at $t = T$.

*Remark 2.4.* Thanks to the linearity and the reversibility in time of the KdV equation, we have equivalence between the exact controllability and the null controllability.

# 3 Cost of Null Controllability

In this section our goal is to define the quantity which measures the cost of the null controllability for system (2) and to give an estimate to this cost as $\varepsilon$ is brought to $0^+$. We begin by introducing some results. Then in the second paragraph we study the behavior of the cost of the null controllability.

## *3.1 Main Results*

The cost of null controllability for linear KdV equation has been studied by O. Glass and S. Guerrero (see [6]). The control problem is the following:

$$\begin{cases} y_t + (My)_x + \varepsilon y_{xxx} = 0, & \text{in } (0, T) \times (0, 1), \\ y_{|x=0} = u_1, y_{|x=1} = u_2, y_{x|x=1} = u_3 & \text{in } (0, T), \\ y_{|t=0} = y_0 & \text{in } (0, 1). \end{cases} \tag{25}$$

when $\varepsilon$ is a positive dispersion coefficient, $M$ is a transport coefficient, $u_i$ $(i = 1,\ 2,\ 3)$ are time-dependent functions which constitute the controls of the system. In this section we use only the Neumann boundary control on the right $(u_2 = u_3 = 0)$. Our system is the following one:

$$\begin{cases} y_t + y_x + \varepsilon y_{xxx} = 0, \\ y(t,0) = 0, \quad y(t,1) = 0, \quad y_x(t,1) = u(t) \\ y(0,x) = y_0(x). \end{cases} \tag{26}$$

Performing the change of variables

$$u(t) = y_x(t,0) + v(t)$$

our system becomes

$$\begin{cases} y_t + y_x + \varepsilon y_{xxx} = 0, \\ y(t,0) = y(t,L) = 0, \\ y_x(t,L) - y_x(t,0) = v(t), \\ y(0,x) = y_0(x). \end{cases} \tag{27}$$

Now we focus on the behavior of the cost of null controllability when $\varepsilon$ vanishes.

## *3.2 Behavior of Cost of Controllability*

For $y^0 \in H^{-1}(0,L)$, we denote by $U(\varepsilon,T,L,y^0)$ the set of controls $v \in L^2(0,T)$ such that the corresponding solution of (2) satisfies $y(T,.) = 0$. It's easy to see that the set $U(\varepsilon,T,L,y^0)$ is a closed affine subspace of $L^2(0,T)$. Let us denote by $\mathscr{U}^T(y^0)$ the projection of 0 on this closed affine subspace, i.e., the element of $U(\varepsilon,T,L,y^0)$ of the smallest $L^2(0,T)$-norm. Then it is not hard to see that the map

$$\begin{aligned} \mathscr{U}^T \ : H^{-1}(0,L) \ &\rightarrow \ L^2(0,T) \\ y^0 \ &\mapsto \ \mathscr{U}^T(y^0) \end{aligned}$$

is a linear continuous map. Let us now define the quantity which measures the cost of the null controllability for system (2):

$$K(\varepsilon,T,L) \ = \ \sup_{\|y^0\|_{H^{-1}(0,L)=1}} \{\min\{\|v\|_{L^2(0,T)} \ : \ v \ \in \ U(\varepsilon,T,L,y^0)\}\}, \tag{28}$$

i.e.,

$$K(\varepsilon,T,L) = \| \ \mathscr{U}^T \ \|_{\mathscr{L}(H^{-1}(0,L),L^2(0,T))} \ .$$

Our result is the following:

**Proposition 3.** *Let $T > 0$. Let $z^0$ be the initial data and $z(t,x)$ the solution of the homogeneous problem (4). The control system (2) is exactly controllable in time T* *if and only if there exists* $c_{T,L,\varepsilon} > 0$ *such that*

$$c_{T,L,\varepsilon}\|z^0\|^2_{H^1(0,L)} \leq \int_0^T \mid \varepsilon z_x(t,L) \mid^2 dt \tag{29}$$

*Moreover, if such a* $c_{T,L,\varepsilon} > 0$ *exists and if* $c^T$ *is the maximum of the set of* $c_{T,L,\varepsilon} > 0$ *such that (29) holds, one has*

$$K(\varepsilon, T, L) = \frac{1}{\sqrt{c^T}}.$$

Then from the observability inequality

$$c_{T,L,\varepsilon}\|z^0\|^2_{H^1(0,L)} \leq \int_0^T \mid \varepsilon z_x(t,L) \mid^2 dt$$

we have

$$\|z_0\|^2_{H^1} \leq \frac{1}{c_{T,L,\varepsilon}} \int_0^T \mid \varepsilon z_x(t,L) \mid^2 dt.$$

The constant $\dfrac{1}{\sqrt{c_{T,L,\varepsilon}}}$ is called constant of observability $C_{obs}$.

On the other hand, from the admissibility inequality:

$$\int_0^T \mid \varepsilon z_x(t,L) \mid^2 dt \leq C_{T,L,\varepsilon}\|z_0\|^2_{H_1}$$

we have

$$\frac{1}{C_{T,L,\varepsilon}} \int_0^T \mid \varepsilon z_x(t,L) \mid^2 dt \leq \|z_0\|^2_{H^1}.$$

The constant $\dfrac{1}{\sqrt{C_{T,L,\varepsilon}}}$ is called admissibility constant $C_{ad}$.

Given the definition of the cost of controllability, we have

$$C_{ad} \ < \ K(\varepsilon, T, L) \ < \ C_{obs}.$$

In order to study the behavior of $K(\varepsilon, T, L)$ as $\varepsilon$ leads to 0, it is natural to look at the limits of the observability's constant and the admissibility one as $\varepsilon$ leads to 0.

$$\lim_{\varepsilon \to 0+} C_{obs} = \lim_{\varepsilon \to 0+} \frac{2\pi}{L \mid m(L) \mid \varepsilon^{2/3}\sqrt{c_1(T)}} = +\infty$$

and

$$\lim_{\varepsilon \to 0^+} C_{ad} = \lim_{\varepsilon \to 0^+} \frac{2\pi}{L \mid M(L) \mid \varepsilon^{2/3} \sqrt{c_2(T)}} = +\infty.$$

It follows directly that

**Theorem 3.1.**

$$\lim_{\varepsilon \to 0^+} K(\varepsilon, T, L) = +\infty.$$

## References


1. Brezis H., *Analyse Fonctionnelle: Théorie et Application*, 5[ième] édition, Masson.
2. Cerpa E. *Exact controllability of a nonlinear Korteweg-de Vries equation on a critical spatial domain*, SIAM J. Control Optim., Vol. 46, No. 3, (2007), pp. 877–899
3. Cerpa E., Crépeau E. *Boundary controllabilty for the nonlinear Korteweg-de Vries equation on any critical domain,* Annales Institut Henri Poincaré, Analyse non lineaire, 26 (2) (2009) pp 457–475.
4. Coron J.M., *Control and nonlinearity*, Mathematical Surveys and Monographs, 136, 2007, 427
5. Coron J.-M., Crépeau E., Exact boundary controllability of a nonlinear KdV equation with a critical lengths *J.Eur.Math.Soc.* vol 6, (2004), pp. 367–398.
6. Glass O., Guerrero S., *Some exact controllability results for the linear KdV equation and uniform controllability in the zero-dispersion limit*, Asymptotic Analysis 60 (2008) pp 61–100.
7. Ingham A. E., *Some trigonometrical inequality with applications to the theory of series*, Math.,Zeitschrift, 41,(1936), 367–379.
8. Jaffard S., Micu S., *Estimates of the constants in generalized Ingham's inequality and applications to the control of the wave equation*, Asymptotic Analysis Vol. 28, No.3–4, (2001), pp. 181–214
9. Lions J.L., *Contrôlabilité exacte, Perturbations et Stabilization de systèmes distribués.* Tome 1, Recherche en mathématiques appliquées [Research in Applied Mathematics], vol.8, (Masson, Paris), (1988).
10. Rosier L., *Exact boundary controllability for the Korteweg-de Vries equation on a bounded domain,* ESAIM Control Optim. Calc. Var.2 (1997), 33–55.
11. Rosier L., *Control of the surface of a fluid by a wavemaker,* ESAIM: Control, Optimisation and Calculus of Variations, vol 10, (2004), pp 346–380.

# Existence of solutions of a class of second order sweeping process in Banach spaces

**F. Aliouane and D. Azzam-Laouir**

**Abstract** In a previous work the authors proved in a separable Banach space under the assumption of the global upper semicontinuity of the perturbation, the existence of Lipschitz solutions for second order non convex sweeping processes in a separable reflexive uniformly smooth Banach space. In the present paper we prove the same results, where the perturbation is assumed to be separately measurable and separately upper semicontinuous.

## 1 Introduction

In [1], the authors proved the following theorem which is an extension of sweeping processes from Hilbert spaces to Banach spaces

**Theorem 1.** *Let* $I = [0,T] \quad (T > 0)$ *and* $E$ *be a separable reflexive uniformly smooth Banach space, which is I-smoothly weakly compact for an exponent* $p \in [2,\infty)$. *Let* $F : I \times E \times E \rightrightarrows E$ *be an upper semicontinuous set-valued mapping with nonempty closed convex values. We assume that there exists a constant* $m > 0$ *such that*

$$F(t,x,u) \subset m\overline{\mathbf{B}}_E, \ \forall (t,x,u) \in I \times E \times E. \tag{1}$$

*Let* $r > 0$ *and* $K : [0,T] \rightrightarrows E$ *be a set-valued mapping taking nonempty ball-compact and r-prox-regular values. We assume that* $K(.)$ *moves in a Lipschitz way, that is, there exists a constant* $k > 0$ *such that for all* $s,t \in I$,

$$\mathscr{H}(K(t),K(s)) \leq k|t-s|. \tag{2}$$

F. Aliouane (✉) • D. Azzam-Laouir
Jijel University, Jijel, Algeria
e-mail: faliouane@gmail.com; laouir.dalila@gmail.com



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_20

*Then for all* $x_0 \in E$ *and* $u_0 \in K(0)$*, the differential inclusion*

$$(\mathscr{P}_F)\begin{cases} u(0) = u_0; \\ x(t) = x_0 + \int_0^t u(s)ds, \ \forall t \in I; \\ u(t) \in K(t), \ \forall t \in I; \\ -\dot{u}(t) \in N_{K(t)}(u(t)) + F(t, x(t), u(t)), \ a.e. \ t \in I, \end{cases}$$

*has Lipschitz solutions* $u, x : I \to E$*. Moreover, we have for almost every* $t \in I$

$$\|\dot{u}(t)\| \leq 2m + k.$$

*In other words, the differential inclusion*

$$(\mathscr{P}_F)\begin{cases} -\ddot{x}(t) \in N_{K(t)}(\dot{x}(t)) + F(t, x(t), \dot{x}(t)), \ a.e. \ t \in I; \\ \dot{x}(t) \in K(t), \ \forall t \in I; \\ x(0) = x_0; \ \dot{x}(0) = u_0 \end{cases}$$

*has at least a Lipschitz solution* $x(.) \in \mathbf{C}^1_E(I)$*.*

In this paper, our main purpose is to obtain the existence of solutions of $(\mathscr{P}_F)$, in the case when the perturbation $F$ is assumed to be separately Lebesgue-measurable on $[0, T]$ and separately upper semicontinuous on $E \times E$. Before proving our main result in Theorem 2, we recall some needed concepts and definitions.

## 2 Notation and Preliminaries

Let $(E, \|.\|)$ be a separable Banach space, $E'$ its topological dual, and $\langle ., . \rangle$ their duality product. $\overline{\mathbf{B}}_E(0, r)$ is the closed ball of $E$ of center 0 and radius $r$, $\overline{\mathbf{B}}_E$ the closed unit ball and $\mathbf{S}_E$ is the unit sphere of $E$.

Let $\mathbf{C}_E([0, T])$ $(T > 0)$ be the Banach space of all continuous mappings $u : [0, T] \to E$, endowed with the sup-norm $\|.\|_{\mathbf{C}}$ and $\mathbf{C}^1_E([0, T])$ be the Banach space of all continuous mappings $u : [0, T] \to E$ with continuous derivative, equipped with the norm

$$\|u\|_{\mathbf{C}^1} = \max\{\max_{t\in[0,T]} \|u(t)\|, \max_{t\in[0,T]} \|\dot{u}(t)\|\}.$$

We denote by $\mathscr{L}([0, T])$ the $\sigma$-algebra of Lebesgue measurable subsets of $[0, T]$, $\lambda = dt$ is the Lebesgue measure on $[0, T]$, $(\mathbf{L}^1_E([0, T]), \|.\|_1)$ is the Banach space of Lebesgue-Bochner integrable $E$-valued mappings, and $(\mathbf{L}^\infty_E([0, T]), \|.\|_\infty)$ is the Banach space of essentially bounded $E$-valued mappings.

We said that a mapping $u : [0, T] \to E$ is absolutely continuous if there is a mapping $v \in \mathbf{L}^1_E([0, T])$ such that $u(t) = u(0) + \int_0^t v(s)ds$, for all $t \in [0, T]$, in this case $v = \dot{u}$ a.e.

For $A \subset E$, $co(A)$ denotes the convex hull of $A$ and $\overline{co}(A)$ its closed convex hull.

We denote by $\delta^*(x', A)$ the support function associated with $A$, i.e.,

$$\delta^*(x', A) = \sup_{y \in A} \langle x', y \rangle.$$

It is well known that the support function of an upper semicontinuous set-valued mapping is upper semicontinuous.

For closed subsets $A$ and $B$ of $E$, the Hausdorff distance between $A$ and $B$ is defined by

$$\mathscr{H}(A, B) = \sup(e(A, B), e(B, A))$$

where

$$e(A, B) = \sup_{a \in A} d(a, B)$$

stands for the excess of $A$ over $B$ and

$$d(a, B) = \inf_{x \in B} \|a - x\|.$$

We recall that for a closed convex subset $A$ of $E$, one has

$$d(x, A) = \sup_{x' \in \overline{\mathbf{B}}_{E'}} (\langle x', x \rangle - \delta^*(x', A)). \tag{3}$$

**Definition 1.** A subset $A \subset E$ is said to be ball-compact if for all closed ball $\overline{\mathbf{B}} = \overline{\mathbf{B}}(x, R)$ of $E$, the set $\overline{\mathbf{B}} \cap A$ is compact. Obviously, a ball-compact subset $A$ is closed.

**Definition 2.** Let $A$ be a closed subset of $E$. Then the set-valued projection operator $P_A$ is defined by

$$\forall x \in E,\ P_A(x) = \{y \in E,\ \|x - y\| = d(x, A)\}.$$

**Definition 3.** Let $A$ be a closed subset of $E$ and $x \in A$, we denote by $N_A(x)$ the proximal normal cone of $A$ at $x$, defined by

$$N_A(x) = \{v \in E,\ \exists s > 0,\ x \in P_A(x + sv)\}.$$

We now come to the main notion of prox-regularity. It was initially introduced by H. Federer [14] in spaces of finite dimension under the name of positively reached sets. Then, it was extended in Hilbert spaces by A. Canino in [7] and A. S. Shapiro in [16]. After, this notion was studied by F. H. Clarke, R. J. Stern, and P. R. Wolenski in [11] (see also [12]) and by R. A. Poliquin, R. T. Rockafellar and L. Thibault in [15]. Few years later, F. Bernard, L. Thibault, and N. Zlateva have defined this notion in Banach spaces (see [2] and [3, 4]).

**Definition 4.** Let $A$ be a closed subset of $E$ and $r > 0$. The set $A$ is said to be r- prox-regular if for all $x \in A$ and $v \in N_A(x)\backslash\{0\}$

$$\mathbf{B}(x + r\frac{v}{\|v\|}, r) \cap A = \emptyset.$$

Now we recall some useful definitions due to the geometric theory of Banach spaces (we refer the reader to [13] for these concepts and more details).

**Definition 5.** The vectorial normed space $(E, \|.\|)$ is said to be uniformly smooth if his norm is uniformly Fréchet differentiable away of 0, it means that for any two unit vectors $x_0, h \in E$, the limit

$$\lim_{t\to 0} \frac{\|x_0 + th\| - \|x_0\|}{t}$$

exists uniformly with respect to $h, x_0 \in \mathbf{S}_E$.

As we know that the norm could be non-differentiable at the origin 0, we study the function $x \longmapsto \|x\|^p$ for an exponent $p > 1$.

**Proposition 1.** *Let E be a uniformly smooth Banach space and* $p \in (1, \infty)$ *be an exponent. The function* $x \longmapsto \|x\|^p$ *is* $\mathbf{C}^1$ *over the whole space E.*

**Definition 6.** For $E$ a uniformly smooth Banach space and $p \in (1, \infty)$, we denote

$$J_p(x) := \frac{1}{p}(\nabla\|.\|^p)(x) \in E'.$$

**Definition 7.** Let $I$ be an interval of $\mathbb{R}$. A separable reflexive uniformly smooth Banach space $E$ is said to be "I-smoothly weakly compact" for an exponent $p \in (1, \infty)$ if for all bounded sequence $(x_n)_n$ of $\mathbf{L}^\infty_E(I)$, we can extract a subsequence $(y_n)_n$ weakly converging to a point $y \in \mathbf{L}^\infty_E(I)$ such that for all $z \in \mathbf{L}^\infty_E(I)$ and $\phi \in \mathbf{L}^1_{\mathbb{R}}(I)$,

$$\lim_{n\to\infty} \int_I \langle J_p(z(t) + y_n(t)) - J_p(y_n(t)), y_n(t)\rangle \phi(t)dt$$
$$= \int_I \langle J_p(z(t) + y(t)) - J_p(y(t)), y(t)\rangle \phi(t)dt. \quad (4)$$

The following proposition describes a useful property of weak continuity of the projection operator. For the proof, we refer the reader to [5].

**Proposition 2.** *Let* $(E, \|.\|)$ *be a separable reflexive and uniformly smooth Banach space. Let* $C_n, C : I \rightrightarrows E$ *be set-valued mappings taking nonempty closed values and satisfying*

$$\sup_{t\in I} \mathscr{H}(C_n(t), C(t)) \longrightarrow_{n\to\infty} 0.$$

*We assume that for an exponent $p \in [2, \infty)$ and a bounded sequence $(v_n)_n$ of $\mathbf{L}^\infty_E(I)$, we can extract a subsequence $(v_{k(n)})_n$ weakly converging to a point $v \in \mathbf{L}^\infty_E(I)$ such that for all $z \in \mathbf{L}^\infty_E(I)$ and $\phi \in \mathbf{L}^1_{\mathbb{R}}(I)$,*

$$\limsup_{n\to\infty} \int_I \langle J_p(z(t) + v_{k(n)}(t)) - J_p(v_{k(n)}(t)), v_{k(n)}(t)\rangle \phi(t)dt$$
$$= \int_I \langle J_p(z(t) + v(t)) - J_p(v(t)), v(t)\rangle \phi(t)dt. \tag{5}$$

*Then the projection $P_{C(.)}$ is weakly continuous in $\mathbf{L}^\infty_E(I)$ (relatively to the directions given by the sequence $(v_n)_n$) in the following sense: for all $r > 0$ and for any bounded sequence $(u_n)_n$ of $\mathbf{L}^\infty_E(I)$ satisfying*

$$\begin{cases} u_n \longrightarrow u \text{ in } \mathbf{L}^\infty_E(I); \\ u_n(t) \in P_{C_n(t)}(u_n(t) + rv_n(t))), \text{ a.e. } t \in I \end{cases}$$

*one has for almost every $t \in I$,*

$$u(t) \in P_{C(t)}(u(t) + rv(t)).$$

## 3 Main Results

Now, we are able to prove our main theorem.

**Theorem 2.** *The conclusion of Theorem 1 holds true if we replace the assumption of global upper semicontinuity of $F : I \times E \times E \rightrightarrows E$ by the following hypotheses*

$$\forall (x,u) \in E \times E,\ t \longmapsto F(t,x,u) \text{ is measurable}; \tag{6}$$

$$\forall t \in I,\ (x,u) \longmapsto F(t,x,u) \text{ is upper semicontinuous}. \tag{7}$$

*Proof.* By the Scorza-Dragoni's Theorem (e.g., [8, 9]), there is a multifunction $F_0 : I \times E \times E \rightrightarrows E$, which is measurable and has the following properties.

(**1**) There is a set $N \subset I$, independent of $(t,x,u)$ such that $\lambda(N) = 0$ and $F_0(t,x,u) \subset F(t,x,u)$, for all $t \in I \backslash N$ and for all $(x,u) \in E \times E$;
(**2**) if $u,\ x,\ z : I \to E$ are measurable mappings with $z(t) \in F(t,x(t),u(t))$ a.e., then $z(t) \in F_0(t,x(t),u(t))$ a.e.;
(**3**) for every $\epsilon > 0$, there is a compact subset $J_\epsilon \subset I$ such that $\lambda(I \backslash J_\epsilon) < \epsilon$, the restriction of $F_0$ on $J_\epsilon \times E \times E$ is upper semicontinuous and $\emptyset \neq F_0(t,x,u) \subset F(t,x,u)$; for all $(t,x,u) \in J_\epsilon \times E \times E$.

By the property (3), there exists a sequence of compact sets $J_n \subset I$ with
$\lambda(I \backslash J_n) = \epsilon_n \longrightarrow 0$ when $n \to \infty$ such that the restriction of $F_0$ to $J_n \times E \times E$ is upper semicontinuous and has nonempty values. We may also assume that $(J_n)$ is increasing. By Dugundji's Theorem (e.g., [6]), there is an upper semicontinuous extension $\tilde{F}_n$ of $F_0 \backslash_{J_n \times E \times E}$ to $I \times E \times E$, and

$$\tilde{F}_n(t, x, u) \subset m\overline{\mathbf{B}}, \ \forall (t, x, u) \in I \times E \times E. \tag{8}$$

So $\tilde{F}_n$ satisfies the hypotheses of Theorem 1. Thus, for every $x_0 \in E$ and $u_0 \in K(0)$, there are Lipschitz solutions $u_n, \ x_n : I \to E$ for the differential inclusion

$$\begin{cases} u_n(0) = u_0; \\ x_n(t) = x_0 + \int_0^t u_n(s)ds, \ \forall t \in I; \\ u_n(t) \in K(t), \ \forall t \in I; \\ -\dot{u}_n(t) \in N_{K(t)}(u_n(t)) + \tilde{F}_n(t, x_n(t), u_n(t)), \ a.e. \ t \in I. \end{cases}$$

Moreover, we have for almost every $t \in I$

$$\|\dot{u}_n(t)\| \leq 2m + k. \tag{9}$$

Consequently, for each $n \in \mathbb{N}$, there is a measurable mapping $z_n(.)$ such that

$$z_n(t) \in \tilde{F}_n(t, x_n(t), u_n(t)), \ \forall t \in I, \tag{10}$$

and

$$-\dot{u}_n(t) \in N_{K(t)}(u_n(t)) + z_n(t), \ a.e. \ t \in I. \tag{11}$$

In other words, the differential inclusion

$$\begin{cases} -\ddot{x}_n(t) \in N_{K(t)}(\dot{x}_n(t)) + \tilde{F}_n(t, x_n(t), \dot{x}_n(t)), \ a.e. \ t \in I; \\ \dot{x}_n(t) \in K(t), \ \forall t \in I; \\ x_n(0) = x_0; \ \dot{x}_n(0) = u_0 \end{cases}$$

has at least a Lipschitz solution $x_n(.) \in \mathbf{C}^1_E(I)$.

From the relation (9), we see that $(\dot{u}_n(.))$ is uniformly bounded by $(2m + k)$. So $(u_n(.))$ is a bounded sequence of $\mathbf{C}_E(I)$ since for every $t \in I$

$$\|u_n(t)\| \leq \|u_0\| + \int_0^t \|\dot{u}_n(s)\| ds \leq \|u_0\| + T(2m + k) := M. \tag{12}$$

Now, we will show that $(u_n(.))$ is relatively compact. Obviously, $(u_n(.))$ is equicontinuous. Let us prove that for every fixed $t$, the sequence $(u_n(t))$ is relatively compact. We have, for all $t \in I$,

$$u_n(t) \in K(t) \cap \overline{\mathbf{B}}(0, M) := \Delta(t). \tag{13}$$

Remark that the set $\Delta(t)$ is compact since $K(t)$ is ball-compact. Consequently $(u_n(t))$ is relatively compact. By Ascoli-Arzelà's Theorem, the sequence $(u_n(.))$ is relatively compact in $\mathbf{C}_E(I)$, by extracting a subsequence still denoted $(u_n(.))$ we may suppose the uniform convergence of $(u_n(.))$ to some mapping $u(.) \in \mathbf{C}_E(I)$. Obviously $u(0) = u_0,\ u(.)$ is a Lipschitz mapping and for all $t \in I$

$$u(t) \in K(t) \tag{14}$$

since $K(t)$ is closed.

Now, we will prove the convergence of $(x_n(.))$ in $\mathbf{C}_E(I)$. For all $t, s \in I$

$$\begin{aligned}\|x_n(t) - x_n(s)\| &\leq \|x_0 + \int_0^t u_n(\tau)d\tau - x_0 - \int_0^s u_n(\tau)d\tau\| \\ &\leq \int_s^t \|u_n(\tau)\|d\tau \leq M|t-s|.\end{aligned}$$

That is, $(x_n(.))$ is equicontinuous. Furthermore for all $t \in I$

$$\|x_n(t)\| = \|x_0 + \int_0^t u_n(s)ds\| \leq \|x_0\| + \int_0^t \|u_n(s)\|ds \leq \|x_0\| + MT.$$

On the other hand, as

$$x_n(t) = x_0 + \int_0^t u_n(s)ds,$$

by the relation (13), we get

$$x_n(t) \in x_0 + \int_0^t \overline{co}(\Delta(s))ds := \tilde{\Delta}(t),$$

which is a compact set since for all $t \in I,\ \overline{co}(\Delta(t))$ is a convex compact set (see [10] for more details). Therefore, $(x_n(.))$ is relatively compact. By, the Ascoli-Arzela's Theorem we conclude that $(x_n(.))$ has a subsequence (still denoted $(x_n(.))$) converging uniformly on $I$ to some mapping $x(.) \in \mathbf{C}_E(I)$. Obviously $x(0) = x_0$ and $x(.)$ is a Lipschitz mapping with ratio $M$. Observe that for all $t \in I$,

$$x(t) = \lim_{n\to\infty} x_n(t) = x_0 + \int_0^t \lim_{n\to\infty} u_n(s)ds = x_0 + \int_0^t u(s)ds \tag{15}$$

using Lebesgue's theorem since $(u_n(.))$ is equibounded (relation (12)), hence, $\dot{x}(.) = u(.)$ a.e. We see by the relation (9), that $(\dot{u}_n(.))_n$ is bounded in $\mathbf{L}^\infty_E(I)$, up to a subsequence, we may suppose that $(\dot{u}_n(.))_n$ weakly* converges in $\mathbf{L}^\infty_E(I)$ to some mapping $w(.)$ and that $w(.) = \dot{u}(.)$. Indeed, for all $y \in \mathbf{L}^1_{E'}(I)$,

$$\lim_{n\to\infty}\langle \dot{u}_n(.), y(.)\rangle = \langle w(.), y(.)\rangle,$$

i.e.,

$$\lim_{n\to\infty}\int_0^t \langle \dot{u}_n(s), y(s)\rangle ds = \int_0^t \langle w(s), y(s)\rangle ds,$$

in particular for $y(.) = \mathbf{1}_{[0,t]}(.)e_j$, with $t \in I$, $\mathbf{1}_{[0,t]}$ the characteristic function of the interval $[0, t]$, and $(e_j)$ a sequence of the space $E'$ which separates the points of $E$ (such a sequence exists since $E$ is separable), then we obtain

$$\langle \lim_{n\to\infty}\int_0^t \dot{u}_n(s)ds, e_j\rangle = \langle \int_0^t w(s)ds, e_j\rangle, \ \forall j,$$

which ensures

$$\lim_{n\to\infty}\int_0^t \dot{u}_n(s)ds = \int_0^t w(s)ds.$$

As $(u_n(.))$ is a sequence of absolutely continuous mappings, we have the following equality

$$\lim_{n\to\infty}(u_n(t) - u_n(0)) = \lim_{n\to\infty}\int_0^t \dot{u}_n(s)ds = \int_0^t w(s)ds,$$

then

$$u(t) = u(0) + \int_0^t w(s)ds,$$

so $u(.)$ is absolutely continuous, and hence $w(.) = \dot{u}(.)$.

By the relation (8) and (10), we deduce that $(z_n(.))$ is a bounded sequence in $\mathbf{L}^\infty_E(I)$, then we can extract a subsequence still denoted $(z_n(.))$ converging $\sigma(\mathbf{L}^\infty_E, \mathbf{L}^1_{E'})$ to $z(.)$ in $\mathbf{L}^\infty_E(I)$. We will show that for all $t \in I$, $z(t) \in F(t, x(t), u(t))$ a.e.

As $x_n$, $u_n$, and $z_n$ are three measurable mappings and satisfy the relation (10), then by the property (2), we get

$$z_n(t) \in F_0(t, x_n(t), u_n(t)), \text{ a.e.},$$

that is, for all $n \in \mathbb{N}$, there is a Lebesgue null set $N_n \subset J_n$ such that

$$z_n(t) \in F_0(t, x_n(t), u_n(t)), \ \forall t \in J_n\backslash N_n. \tag{16}$$

Let $N_0 := (I\backslash \cup_n J_n) \cup (\cup_n N_n)$ which is Lebesgue-negligible. Indeed;

$$\lambda(N_0) = \lambda((I\backslash \cup_n J_n) \cup (\cup_n N_n))$$

$$\leq \lambda(I\backslash \cup_n J_n) + \lambda(\cup_n N_n))$$
$$\leq \lambda(\cap_n (I\backslash J_n)) + \sum_n \lambda(N_n),$$

We have that the set $J_0$ has finite measure and the sequence $(I\backslash J_n)$ is decreasing because $(J_n)$ is increasing, thus

$$\lambda(\cap_n(I\backslash J_n)) = \lim_{n\to\infty} \lambda(I\backslash J_n) = \lim_{n\to\infty} \epsilon_n = 0,$$

and therefore

$$\lambda(N_0) \leq \lim_{n\to\infty} \lambda(I\backslash J_n) + \sum_n \lambda(N_n) = 0.$$

For all $t \in I\backslash N_0$, there is an integer $n_0 = n_0(t) \in N$ such that for all $n \geq n_0$, $t \in J_n\backslash N_n$, so by the relation (16), we obtain

$$z_n(t) \in F_0(t, x_n(t), u_n(t)),\ \forall n \geq n_0.$$

On the other hand, since $F_0$ is upper semicontinuous on $J_n \times E \times E$ and
$x_n(t) \to x(t),\ u_n(t) \to u(t)$ when $n \to \infty$, it follows that for all $x' \in E'$,

$$\limsup_{n\to\infty} \delta^*(x', F_0(t, x_n(t), u_n(t))) \leq \delta^*(x', F_0(t, x(t), u(t))).$$

For $t \notin N_0$ and $n \geq n_0$, we have

$$\langle x', z_n(t)\rangle \leq \delta^*(x', F_0(t, x_n(t), u_n(t))),$$

thus

$$\limsup_{n\to\infty} \langle x', z_n(t)\rangle \leq \limsup_{n\to\infty} \delta^*(x', F_0(t, x_n(t), u_n(t)))$$
$$\leq \delta^*(x', F_0(t, x(t), u(t))),$$

by Fatou's Lemma, we deduce that for every measurable set $A \subset I$ and every $x' \in E'$,

$$\int_A \langle x', z(t)\rangle dt = \lim_{n\to\infty} \int_A \langle x', z_n(t)\rangle dt$$
$$= \limsup_{n\to\infty} \int_A \langle x', z_n(t)\rangle dt$$
$$\leq \int_A \limsup_{n\to\infty} \langle x', z_n(t)\rangle dt$$

$$\leq \int_A \delta^*(x', F_0(t, x(t), u(t)))dt.$$

So,

$$\langle x', z(t)\rangle \leq \delta^*(x', F_0(t, x(t), u(t))) \text{ a.e.},$$

then

$$\sup_{x'\in E'} (\langle x', z(t)\rangle - \delta^*(x', F_0(t, x(t), u(t)))) = 0,$$

since $F_0$ has closed convex values, by the relation (3), we get
$d(z(t), F_0(t, x(t), u(t))) = 0$, this known to imply that, $z(t) \in F_0(t, x(t), u(t))$ a.e..
We have shown that there exists a negligible set $N_0 \subset I$ such that

$$z(t) \in F_0(t, x(t), u(t)),\ \ \forall t \in I\backslash N_0.$$

By property (1),

$$z(t) \in F(t, x(t), u(t)),\ \ \forall t \in I\backslash N_0,$$

this shows that

$$z(t) \in F(t, x(t), u(t)),\ \ \text{a.e. } t \in I. \tag{17}$$

Now, we have by (11)

$$-\dot{u}_n(t) \in N_{K(t)}(u_n(t)) + z_n(t),\ \ a.e.\ t \in I.$$

By the definition of the proximal normal cone, we deduce that there exists $\alpha > 0$ such that

$$u_n(t) \in P_{K(t)}(u_n(t) - \alpha(\dot{u}_n(t) + z_n(t)));\ \ a.e.\ t \in I. \tag{18}$$

Set $\Delta_n(t) = \dot{u}_n(t) + z_n(t)$. By the arguments given above we know that $(\Delta_n(.))_n$ weakly*-converges in $\mathbf{L}^{\infty}_E(I)$ to $\dot{u}(.) + z(.) := \Delta(.)$. Supplying the property
"I-smoothly weakly compact" supposed on the space $E$ to the sequence $(\alpha\Delta_n(.))_n$ we get that for all $y \in \mathbf{L}^{\infty}_E(I)$ and all $\phi \in \mathbf{L}^1_{\mathbb{R}}(I)$,

$$\lim_{n\to\infty} \int_I \langle J_p(y(t) - \alpha\Delta_n(t)) - J_p(-\alpha\Delta_n(t)), \Delta_n(t)\rangle\phi(t)dt$$
$$= \int_I \langle J_p(y(t) - \alpha\Delta(t)) - J_p(-\alpha\Delta(t)), \Delta(t)\rangle\phi(t)dt.$$

On the other side, we have that the sequence $(u_n(.))$ strongly converges in $\mathbf{L}_E^\infty(I)$ to $u(.)$ (because of its uniform convergence to $u(.)$ in $\mathbf{C}_E(I)$).

Then, by the relation (18) and the Proposition 2, we have for almost every $t \in I$

$$u(t) \in P_{K(t)}(u(t) - \alpha\Delta(t)),$$

that is, $-\Delta(t) \in N_{K(t)}(u(t))$, or equivalently

$$-\dot{u}(t) - z(t) \in N_{K(t)}(u(t)), \ a.e. \ t \in I,$$

and then by (17) we get

$$-\dot{u}(t) \in N_{K(t)}(u(t)) + F(t, x(t), u(t)), \ a.e. \ t \in I.$$

Finally, by the relation (14) and (15) we conclude that our problem $(\mathscr{P}_F)$ has at least a Lipschitz solution $x \in \mathbf{C}_E^1(I)$. Furthermore,

$$\|\ddot{x}(t)\| \leq 2m + k, \ a.e. \ t \in I.$$

This finishes the proof. □

## References


1. F. Aliouane and D. Azzam. Laouir, A second order differential inclusion with proximal normal cone in Banach spaces. Topol Methods in Nonlinear Analysis 44 (2014), 143–160.
2. F. Bernard and L. Thibault, Prox-regularity of functions and sets in Banach spaces. Set-valued Anal. 12 (2004), 25–47.
3. F. Bernard, L. Thibault and N. Zlateva, Characterizations of Prox-Regular sets in uniformaly convex Banach spaces. J. Convex Anal. 13 (2006), 525–560.
4. F. Bernard, L. Thibault and N. Zlateva, Prox-regular sets and epigraphs in uniformly convex Banach spaces: various regularities and other properties. Trans. Amer. Math. Soc. Volume 363, Number 4, (2010) 2211–2247.
5. F. Bernicot, J. Venel, Existence of sweeping process in Banach spaces under directional prox-regularity. J. Convex Anal. 17 (2010), 451–484.
6. H. Benabdellah and A. Faik, Perturbations convexes et nonconvexes des equations d'évolution, Portugal. Math, 53(2) (1996), 187–208.
7. A. Canino, On p-convex sets and geodesics. J. Diff. Euations 75 (1988), 118–157.
8. C. Castaing and M.D.P. Monteiro-Marques, Evolution problems associated with nonconvex closed moving sets with bounded variation. Portugal. Math. 53 (1996), 73–87.
9. C. Castaing and M. D. P. Monteiro Marques, A multivalued version of Scorza- Dragonis theorem with an application to normal integrands, Bull. Pol. Acad. Sci. Mathematics, 42 (1994) 133–140.
10. C. Castaing and M. Valadier, Convex analysis and measurable multifunctions. LNM 580, Springer Verlag, Berlin (1977).
11. F. H. Clarke, R. J. Stern and P. R. Wolenski, Proximal smoothness and the lower-$C^2$ property. J. Convex Anal. 2 (1995), 117–144.

12. F. H. Clarke, Y. S. Ledyaev, R. j. Stern and P. R. Wolenski, Nonsmooth Analysis and control theory. Springer-Verlag, (1998).
13. J. Diestel, Geometry of Banach spaces: selected topics, Springer-Verlag, New-York, (1975).
14. H. Federer, Curvature measures. Trans. Amer. Math. Soc. 93 (1959); 418–491.
15. R. A. Poliquin, R. T. Rockafellar and L. Thibault, Local differentiability of distance functions. Trans. Amer. Math. Soc. 352 (2000), 5231–5249.
16. A. S. Shapiro, Existence and differentiability of metric projections in Hilbert spaces. SIAM J. Optim. 4 (1994) 130–141.

# Concave and convex nonlinearities in nonstandard eigenvalue problems

**Nawel Benouhiba and Amina Bounouala**

**Abstract** This work deals with eigenvalues of the $p(x)$-Laplacian with a concave-convex nonlinearity in a bounded domain, subject to Dirichlet boundary conditions.

## 1 Introduction

In this paper we discuss the eigenvalue problem

$$\begin{cases} -\triangle_{p(x)}u + h(x)|u|^{s_1(x)-2}u = \lambda g(x)|u|^{q(x)-2}u + k(x)|u|^{s_2(x)-2}u & \text{in } \Omega, \\ u(x) = 0 \quad \text{on } \partial\Omega \end{cases} \tag{1}$$

where $\Omega$ is a bounded domain with smooth boundary $\partial\Omega$ in $\mathbb{R}^N, N \geq 2$. $g, h, k : \Omega \to \mathbb{R}^+$ are measurable functions, $p, q, s_1, s_2 : \Omega \to (1, +\infty[$ are variable exponents, and $\lambda$ is a real parameter. The operator $\triangle_{p(x)}u = div\left(|\nabla u|^{p(x)-2}\nabla u\right)$ is the $p(x)$-Laplacian which is the natural generalization of the $p$-Laplacian when $p$ is constant.

The study of the $p(x)$-Laplacian equations is being of an increasing interest in the recent years due to their applications in elasticity theory and electrorheological fluids. Further information, modeling and applications of the $p(x)$-Laplacian can be found in [4, 9] and [10].

Besides being of a nonstandard growth type, Problem 1 lies in the category of convex-concave problems. In the constant case, i.e. $p(x) = p > 1$, the authors in [6] establish the existence of infinitely many solutions that have negative energy. A similar problem is considered also in [8] when $p$ is not constant. More precisely, the considered problem was

N. Benouhiba (✉) • A. Bounouala
Laboratory of Applied Mathematics, Badji Mokhtar-Annaba University,
P.O. Box 12 El Hadjar 23000 Annaba, Algeria
e-mail: n_benouhiba@yahoo.fr; m2ma.bounouala@gmail.com



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings
in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_21

$$\begin{cases} -\triangle_{p(x)} u = A|u|^{a-2}u + B|u|^{b-2}u, & \text{in } \Omega \\ u = 0, & \text{on } \partial\Omega \end{cases} \tag{2}$$

where $a, b, A$, and $B$ are constants. Under the assumptions $1 < a < p^- < p^+ < b < \min(N, \frac{Np^-}{N-p^-})$ and $A, B > 0$, the author shows that there exists $\lambda > 0$ such that for any $A, B \in (0, \lambda)$ Problem 2 has at least two distinct nontrivial weak solutions.

In [7], the authors investigate the following problem

$$\begin{cases} -\triangle_{p(x)} u = \lambda(a(x)|u|^{q(x)-2}u + b(x)|u|^{h(x)-2}u), & \text{in } \Omega \\ u = 0, & \text{on } \partial\Omega. \end{cases} \tag{3}$$

Assuming that $1 < q^- < q^+ < p^- \leq p^+ < h^- < h^+ < \frac{Np^-}{N-p^-}$ and $p^+ < N$ they establish the existence of a nontrivial weak solution with positive energy for any $\lambda$ in some neighborhood of zero.

In this paper, we consider 1 as an eigenvalue problem. Using adequate variational methods we set the existence of $\overline{\lambda}$ such that for any $\lambda \in (-\infty, \overline{\lambda})$ there exists a weak solution of the considered problem with no positive energy.

We study Problem 1 under the assumptions

$(A_1)$ $1 < s_1, s_2, q, p$ are in $C(\overline{\Omega})$, $s_1(x), s_2(x), q(x) < p^*(x)$ in $\overline{\Omega}$ and

$$s_2^- < s_2^+ < q^- < q^+ < s_1^- < s_1^+ < p^- < p^+ < N.$$

$(A_2)$ The functions $g, h$, and $k$ are such that

- $(A_{2,g})$ $0 \leq g \in L^{r(x)}(\Omega)$ where $1 < r(x) \in C(\overline{\Omega})$ is such that $\frac{p^*(x)}{p^*(x)-q(x)} < r(x)$ in $\Omega$,
- $(A_{2,h})$ $0 \leq h \in L^{m(x)}(\Omega)$ where $1 < m(x) \in C(\overline{\Omega})$ is such that $\frac{p^*(x)}{p^*(x)-s_1(x)} < m(x)$ in $\Omega$ and
- $(A_{2,k})$ $0 \leq k \in L^{l(x)}(\Omega)$ where $1 < l(x) \in C(\overline{\Omega})$ is such that $\frac{p^*(x)}{p^*(x)-s_2(x)} < l(x)$ in $\Omega$.

**Organization of the paper.**

The rest of the paper is organized as follows. In section 2 we state some elementary properties of variable Lebesgue and Sobolev spaces. In Section 3 we prove the existence of a global minimum with no positive energy.

## 2 Notations and auxiliary results

Let $\Omega$ be an open bounded subset of $\mathbb{R}^N$. Write

$$L_+^\infty(\overline{\Omega}) = \{h | h \in L^\infty(\overline{\Omega}), \operatorname{ess\,inf}_{x\in\Omega} h(x) \geq 1\},$$

$$h^- = \text{ess}\inf_{x\in\Omega} h(x), \quad h^+ = \text{ess}\sup_{x\in\Omega} h(x).$$

For $p \in L^\infty_+(\overline{\Omega})$ the variable exponent Lebesgue space $L^{p(x)}(\Omega)$ is defined by

$$L^{p(x)}(\Omega) = \left\{ u | u \text{ is a measurable real-valued function}, \int_\Omega |u(x)|^{p(x)} dx < \infty \right\}$$

with the norm

$$|u|_{p(x)} = \inf \left\{ \mu > 0 : \int_\Omega \left| \frac{u(x)}{\mu} \right|^{p(x)} dx \le 1 \right\}.$$

Define the variable exponent Sobolev space $W^{1,p(x)}(\Omega)$

$$W^{1,p(x)}(\Omega) = \left\{ u \in L^{p(x)}(\Omega) : |\nabla u| \in L^{p(x)}(\Omega) \right\}$$

equipped with the norm

$$\|u\|_{1,p(x)} = |u|_{p(x)} + |\nabla u|_{p(x)}.$$

We also define the space $W_0^{1,p(x)}(\Omega)$ the closure of $C_0^\infty(\Omega)$. Assuming $p^- > 1$, the spaces $L^{p(x)}(\Omega)$, $W^{1,p(x)}(\Omega)$, and $W_0^{1,p(x)}(\Omega)$ are separable and reflexive Banach spaces [5] and we have the following properties.

**Proposition 1 ([5]).**

*(i)* *The conjugate space of $L^{p(x)}(\Omega)$ is $L^{p'(x)}(\Omega)$ where $p'(x) = \frac{p(x)}{p(x)-1}$.*

*(ii)* *If $p_1(x) \le p_2(x)$ for all $x \in \overline{\Omega}$, then $L^{p_2(x)}(\Omega) \hookrightarrow L^{p_1(x)}(\Omega)$ and the embedding is continuous if $|\Omega| < +\infty$.*

*(iii)* *For any $f \in L^{p(x)}(\Omega)$ and $g \in L^{q(x)}(\Omega)$ such that $\frac{1}{p(x)} + \frac{1}{q(x)} = 1$, we have the Hölder inequality*

$$\left| \int_\Omega fg dx \right| \le (\frac{1}{p^-} + \frac{1}{q^-}) |f|_{p(x)} |g|_{q(x)} \le 2 |f|_{p(x)} |g|_{q(x)}. \tag{4}$$

*(iv)* *For any $f \in L^{p(x)}(\Omega)$, $g \in L^{q(x)}(\Omega)$ and $k \in L^{r(x)}(\Omega)$ such that $\frac{1}{p(x)} + \frac{1}{q(x)} + \frac{1}{r(x)} = 1$, we have*

$$\left| \int_\Omega fgk dx \right| \le \left( \frac{1}{p^-} + \frac{1}{q^-} + \frac{1}{r^-} \right) |f|_{p(x)} |g|_{q(x)} |k|_{r(x)} \le 3 |f|_{p(x)} |g|_{q(x)} |k|_{r(x)}. \tag{5}$$

**Proposition 2 ([3, 5]).** *Define* $\varrho_{p(x)}(f) = \int_\Omega |f(x)|^{p(x)}dx$. *We have*

*(i) for any* $f \in L^{p(x)}(\Omega)$ *we have*

$$\begin{aligned} |f|_{p(x)}^{p^-} \le \varrho_{p(x)}(f) \le |f|_{p(x)}^{p^+} \quad \textit{if} \quad |f|_{p(x)} > 1, \\ |f|_{p(x)}^{p^+} \le \varrho_{p(x)}(f) \le |f|_{p(x)}^{p^-} \quad \textit{if} \quad |f|_{p(x)} \le 1. \end{aligned} \tag{6}$$

*(ii) For* $f_n, f \in L^{p(x)}(\Omega)$ *we have*

$$f_n \to f \;\; \textit{in} \;\; L^{p(x)}(\Omega) \textit{ if and only if } \varrho_{p(x)}(f_n - f) \to 0. \tag{7}$$

**Proposition 3 ([2]).** *Let p and q be measurable functions and* $1 \le p(x)q(x) < \infty$ *for a. e. in* $\Omega$. *Let* $f \in L^{q(x)}(\Omega)$, *then*

$$\begin{aligned} |f|_{p(x)q(x)}^{p^-} \le \left| |f|^{p(x)} \right|_{q(x)} \le |f|_{p(x)q(x)}^{p^+} \quad \textit{if} \quad |f|_{p(x)q(x)} > 1, \\ |f|_{p(x)q(x)}^{p^+} \le \left| |f|^{p(x)} \right|_{q(x)} \le |f|_{p(x)q(x)}^{p^-} \quad \textit{if} \quad |f|_{p(x)q(x)} \le 1. \end{aligned} \tag{8}$$

*In particular, if* $p(x) = p$ *is constant, then*

$$|\,|\,u\,|^p\,|_{q(x)} = |u|_{q(x)}^p.$$

Let us define the critical Sobolev exponent of $p$

$$p^*(x) = \begin{cases} \frac{Np(x)}{N-p(x)} & \text{if } p(x) < N, \\ +\infty & \text{if } p(x) \ge N. \end{cases} \tag{9}$$

**Proposition 4 ([3]).** *Let* $\Omega$ *be a bounded domain in* $\mathbb{R}^N$ *with Lipschitz boundary and* $p, q \in C(\overline{\Omega})$ *such that* $p(x) < N$ *and* $1 < q(x) < p^*(x)$ $\forall x \in \Omega$. *Then there is a compact and continuous embedding* $W_0^{1,p(x)}(\Omega) \to L^{q(x)}(\Omega)$.

On the Sobolev space $W_0^{1,p(x)}(\Omega)$ we can consider the equivalent norm

$$\|u\| = |\nabla u|_{p(x)}.$$

## 3 Global Minimum

We begin this section by giving the definition

**Definition 1.** We say that $\lambda$ is an eigenvalue of Problem1 if there exists $u \in W_0^{1,p(x)}(\Omega) \setminus \{0\}$ such that

$$\int_{\Omega} |\nabla u|^{p(x)-2} \nabla u \nabla v dx + \int_{\Omega} h(x) |u|^{s_1(x)-2} uv dx$$
$$- \lambda \int_{\Omega} g(x) |u|^{q(x)-2} uv dx - \int_{\Omega} k(x) |u|^{s_2(x)-2} uv dx = 0 \qquad (10)$$

for all $v \in W_0^{1,p(x)}(\Omega)$.

Let us recall that $u$ satisfying 10 is a critical point of the functional defined in $W_0^{1,p(x)}(\Omega$ by

$$I(u) = \int_{\Omega} \frac{1}{p(x)} |\nabla u|^{p(x)} dx +$$
$$+ \int_{\Omega} \frac{h(x)}{s_1(x)} |u|^{s_1(x)} dx - \lambda \int_{\Omega} \frac{g(x)}{q(x)} |u|^{q(x)} dx - \int_{\Omega} \frac{k(x)}{s_2(x)} |u|^{s_2(x)} dx. \qquad (11)$$

Standard arguments show that $I \in C^1(W_0^{1,p(x)}(\Omega), \mathbb{R})$ and that

$$\langle I'(u), v \rangle = \int_{\Omega} |\nabla u|^{p(x)-2} \nabla u . \nabla v dx +$$
$$+ \int_{\Omega} h(x) |u|^{s_1(x)-2} uv dx - \lambda \int_{\Omega} g(x) |u|^{q(x)-2} uv dx -$$
$$- \int_{\Omega} k(x) |u|^{s_2(x)-2} uv dx$$

for all $u, v \in W_0^{1,p(x)}(\Omega)$.

In order to prove that $I$ attains its minimum in $W_0^{1,p(x)}(\Omega)$ let us begin with the following Lemmas.

**Lemma 1.** *There exists $\overline{\lambda} > 0$ such that for every $\lambda < \overline{\lambda}$ there is a real $M_\lambda$ such that*

$$I(u) \geq M_\lambda$$

*for all $u \in W_0^{1,p(x)}(\Omega)$.*

*Proof.* We begin to prove the following inequality

$$at^{\alpha} - bt^{\beta} \leq a \left(\frac{a}{b}\right)^{\alpha/(\beta-\alpha)}, \quad \forall t \geq 0 \qquad (12)$$

for any $a, b > 0$ and $0 < \alpha < \beta$.

For $t \geq \left(\frac{a}{b}\right)^{\frac{1}{\beta-\alpha}}$ we have

$$t^{\alpha}(a - bt^{\beta-\alpha}) \leq 0 < a\left(\frac{a}{b}\right)^{\alpha/(\beta-\alpha)}.$$

Now if $0 < t < \left(\frac{a}{b}\right)^{\frac{1}{\beta-\alpha}}$ we get

$$at^{\alpha} - bt^{\beta} < at^{\alpha} < a\left(\frac{a}{b}\right)^{\alpha/(\beta-\alpha)}$$

since $t \mapsto t^{\alpha}$ is an increasing function.

We set

$$\lambda^* = \inf_{u \in W_0^{1,p(x)}(\Omega)\setminus\{0\}} \frac{\int_{\Omega} \frac{1}{p(x)} |\nabla u|^{p(x)} dx}{\int_{\Omega} \frac{g(x)}{q(x)} |u|^{q(x)} dx}. \tag{13}$$

It is known (see [1]) that $\lambda^*$ is a positive number. Then, we have for all $u \in W_0^{1,p(x)}(\Omega)$

$$\lambda^* \leq \frac{\int\limits_{\Omega} \frac{1}{p(x)} |\nabla u|^{p(x)} dx}{\int\limits_{\Omega} \frac{g(x)}{q(x)} |u|^{q(x)} dx}$$

so

$$\int\limits_{\Omega} g(x) |u|^{q(x)} dx \leq \frac{q^-}{\lambda^* p^-} \int\limits_{\Omega} |\nabla u|^{p(x)} dx.$$

For $\lambda \geq 0$ we have

$$I(u) \geq \frac{1}{p^+} \int\limits_{\Omega} |\nabla u|^{p(x)} dx - \frac{\lambda}{q^+} \int\limits_{\Omega} g(x) |u|^{q(x)} dx - \frac{1}{s_2^+} \int\limits_{\Omega} k(x) |u|^{s_2(x)} dx$$

and then

$$I(u) \geq \left(\frac{1}{p^+} - \frac{\lambda q^-}{\lambda^* p^- q^+}\right) \int\limits_{\Omega} |\nabla u|^{p(x)} dx - \frac{1}{s_2^+} \int\limits_{\Omega} k(x) |u|^{s_2(x)} dx.$$

Applying the Hölder inequality and Proposition 2 it yields

$$\int\limits_{\Omega} k(x) |u|^{s_2(x)} dx \leq 2 |k|_{l(x)} |u|_{s_2(x) l'(x)}^{s_2^i}, i = + or - .$$

We have by $(A_{2,k})$ that $1 < s_2(x)l'(x) < p^*(x)$. So the embedding $W_0^{1,p(x)}(\Omega) \hookrightarrow L^{s_2(x)l'(x)}(\Omega)$ is continuous. So there exists $c_1 > 0$ such that

$$\int_\Omega k(x)\,|u|^{s_2(x)}\,dx \le c_1\,|k|_{l(x)}\,\|u\|^{s_2^i}, i = +or-. \tag{14}$$

This together with 6 gives

$$I(u) \ge -\left[\frac{c_1}{s_2^+}\,|k|_{l(x)}\,\|u\|^{s_2^i} - \left(\frac{1}{p^+} - \frac{\lambda q^-}{\lambda^* p^- q^+}\right)\|u\|^{p^j}\right], i,j = +,-.$$

If we suppose that $\lambda < \frac{\lambda^* p^- q^+}{p^+ q^-}$ we get by relation 12

$$I(u) \ge -\left(\frac{c_1\,|k|_{l(x)}}{s_2^+}\right)\left[\frac{c_1\,|k|_{l(x)}}{s_2^+\left(\frac{1}{p^+} - \frac{\lambda q^-}{\lambda^* p^- q^+}\right)}\right]^{s_2^j/p^i - s_2^j}$$

This means that for $0 \le \lambda < \frac{\lambda^* p^- q^+}{p^+ q^-}$ there exists $M_\lambda = M(\lambda, \lambda^*, p, q, k, s_2) < 0$ such that

$$I(u) \ge M_\lambda$$

for any $u \in W_0^{1,p(x)}(\Omega)$.

Now if $\lambda < 0$ it yields by relations 6 and 14

$$I(u) \ge -\left(\frac{c_1}{s_2^+}|k|_{l(x)}\|u\|^{s^i} - \frac{1}{p^+}\|u\|^{p^j}\right), i,j = +,-.$$

Using again relation 12 we have

$$I(u) \ge -\frac{c_1}{s_2^+}|k|_{l(x)}\left(\frac{c_1|k|_{l(x)}}{s_2^+ p^+}\right)^{\frac{s_2^i}{p^j - s_2^i}}, i,j = +,-.$$

This means that $I$ is also bounded from below in the case where $\lambda < 0$. Hence the lemma is proved.

**Lemma 2.** *The functional $I$ is coercive and weakly lower semicontinuous on $W_0^{1,p(x)}(\Omega)$.*

*Proof.* For any $u \in W_0^{1,p(x)}(\Omega)$ we have by the Hölder inequality and Proposition 2

$$\int_{\Omega} g(x)\,|u|^{q(x)}\,dx \leq 2\,|g|_{r(x)}\,|u|^{q^i}_{q(x)r'(x)}\,, i = +, -. \quad (15)$$

By assumption $(A_{2,g})$ we get $1 < q(x)r'(x) < p^*(x)$ so there exists a constant $c_2 > 0$ from the continuous embedding of $W_0^{1,p(x)}(\Omega)$ into $L^{q(x)r'(x)}(\Omega)$ such that

$$\int_{\Omega} g(x)\,|u|^{q(x)}\,dx \leq c_2\,|g|_{r(x)}\,\|u\|^{q^i}, i = +, -. \quad (16)$$

If $\lambda > 0$ and $\|u\| > 1$, it follows by relations 6, 14, and 16 that

$$\begin{aligned} I(u) &\geq \frac{1}{p^+}\,\|u\|^{p^+} - \frac{\lambda c_2}{q^+}\,|g(x)|_{r(x)}\,\|u\|^{q^i} - \frac{c_1}{s_2^+}\,|k(x)|_{l(x)}\,\|u\|^{s_2^j} \\ &\geq \frac{1}{p^+}\,\|u\|^{q^i}\left(\|u\|^{p^+ - q^i} - c_3 - c_4\,\|u\|^{s_2^j - q^i}\right). \end{aligned}$$

Since we have by assumptions $s_2^- < s_2^+ < q^- < q^+ < p^- < p^+$ then by the last inequality $I(u) \to +\infty$ when $\|u\| \to +\infty$.

Same argument shows that $I$ is also coercive on $W_0^{1,p(x)}(\Omega)$ in the case where $\lambda < 0$.

For the second part of the lemma we put

$$J_1(u) = \int_{\Omega} \frac{1}{p(x)}|\nabla u|^{p(x)}dx + \int_{\Omega} \frac{h(x)}{s_1(x)}|u|^{s_1(x)}dx \quad \text{and}$$

$$J_2(u) = I(u) - J_1(u).$$

Since $J_1$ is a continuous convex functional on $W_0^{1,p(x)}(\Omega)$ (see [8]) it follows that it is weakly lower semicontinuous.

Let now $(u_n) \subset W_0^{1,p(x)}(\Omega)$ such that $u_n \rightharpoonup u$ for some $u$ in $W_0^{1,p(x)}(\Omega)$. So $u_n$ is bounded in $W_0^{1,p(x)}(\Omega)$. By assumption $(A_{2,g})$ and Proposition 4 the embedding of $W_0^{1,p(x)}(\Omega)$ in $L^{q(x)r'(x)}(\Omega)$ is compact. Then, we get a subsequence still denoted by $u_n$ that converges to $u$ in $L^{q(x)r'(x)}(\Omega)$. This together with 15 yields that $u \mapsto \int_{\Omega} \frac{g(x)}{q(x)}|u|^{q(x)}dx$ is a weakly strongly continuous functional.

Using similar arguments we get the same result for the functional $u \mapsto \int_{\Omega} \frac{k(x)}{s_2(x)}|u|^{s_2(x)}dx$. It follows that $J_2$ is weakly lower semicontinuous and the proof of the lemma is complete.

We state now the main result of this section.

**Theorem 1.** *For any* $\lambda < \frac{\lambda^* p^- q^+}{p^+ q^-}$ *where* $\lambda^*$ *is given by 13 there exists* $u_\lambda \in W_0^{1,p(x)}(\Omega)$ *and* $u_\lambda \neq 0$ *solution of Problem 1.*

*Proof.* By lemmas 6 and 7 we deduce the existence of a global minimum $u_\lambda \in W_0^{1,p(x)}(\Omega)$ of $I$ for any $\lambda < \frac{\lambda^* p^- q^+}{p^+ q^-}$.

Let now $\phi \in C_0^\infty(\Omega)$ be fixed and $0 < t < 1$. We have

$$I(t\phi) \leq t^{p^-} \int_\Omega \frac{1}{p(x)} |\nabla\phi|^{p(x)} dx - \lambda t^{q^i} \int_\Omega \frac{g(x)}{q(x)} |\phi|^{q(x)} dx$$
$$+ t^{s_1^-} \int_\Omega \frac{h(x)}{s_1(x)} |\phi|^{s_1(x)} dx - t^{s_2^+} \int_\Omega \frac{k(x)}{s_2(x)} |\phi|^{s_2(x)} dx$$

and then

$$I(t\phi) \leq t^{s_2^+} f(t)$$

where $f(t) = t^{p^- - s_2^+} \int_\Omega \frac{1}{p(x)} |\nabla\phi|^{p(x)} dx - \lambda t^{q^i - s_2^+} \int_\Omega \frac{g(x)}{q(x)} |\phi|^{q(x)} dx + t^{s_1^- - s_2^+} \int_\Omega \frac{h(x)}{s_1(x)} |\phi|^{s_1(x)} dx - \int_\Omega \frac{k(x)}{s_2(x)} |\phi|^{s_2(x)} dx$.

It is clear that $f$ is a continuous function and that $f(0) < 0$, then there exists $0 < t_0 < 1$ such that $f(t_0) < 0$. It follows that

$$I(u_0) < 0$$

where $u_0 = t_0\phi$, then $I(u_\lambda) < 0$ and this means that $u_\lambda$ is not trivial. The proof of the Theorem is complete.

## References


1. Benouhiba N.: On the eigenvalues of weighted p(x)-laplacian on $\mathbb{R}^N$. Nonlinear Analysis **74**, 235–243 (2011)
2. Edmunds D. E., Rákosník J.: Sobolev embedding with variable exponent. Studia Mathematica **143**, 267–293 (2000)
3. Fan X. L., Zhao D.: On the spaces $L^{p(x)}(\Omega)$ and $W^{1,p(x)}(\Omega)$. J. Math. Anal. Appl. **263**, 424–446 (2001)
4. Halsey T. C.: Electrorheological Fluids. Sciences **258**, 761–766 (1992)
5. Kovácik O., Rákosník J.: On spaces $L^{p(x)}(\Omega)$ and $W^{1,p(x)}(\Omega)$. Czechoslovak Math. J. **41**, 592–618 (1991)
6. Liu S., Li S. : An elliptic equation with concave and convex nonlinearities. Nonlinear Analysis **53**, 723–731 (2003)
7. Mashiyev R. A., Cekic B., Buhrii O. : Existence of solutions for p(x)-laplacian equations. E. J. of Qualitative Theory of Differential Equations **65**, 1–13 (2010)
8. M. Mihăilescu, Existence and multiplicity of solutions for an elliptic equation with p(x)-growth conditions. Glasgow Math. J. **48**, 411–418 (2006)
9. Rajagopal K. R., Růžička M.: Mathematical Modeling of Electrorheological Fluids. Continuum Mech. Thermdyn. **13**, 59–78 (2001)
10. Růžička M.: Electrorheological Fluids Modeling and Mathematical Theory. Springer-Verlag, Berlin (2002)

# On the time asymptotic behavior of a transport operator with bounce-back boundary condition

**Salma Charfi, Asrar Elleuch, and Aref Jeribi**

**Abstract** This paper deals with the spectral properties of multidimensional transport equations with bounce-back boundary conditions arising in $L_p$-spaces ($1 \leq p < \infty$). These properties are closely related to the large dependent solutions of transport equations. An adequate assumption allows us to investigate the uniform stability of solutions for the Cauchy problem without restriction on the initial data.

**Keywords** Cauchy problem • $C_0$-semigroup • transport operator • asymptotic behavior

## 1 Introduction

The main aim is to investigate the time asymptotic behavior of the solution of the following initial-boundary-value problem in $L_p$-spaces ($1 \leq p < \infty$).

$$\textbf{(I)} \qquad \begin{cases} \bullet \ \dfrac{\partial \psi}{\partial t}(x,v,t) = -v.\nabla_x \psi(x,v,t) - \Sigma(v)\psi(x,v,t) + \displaystyle\int_V \kappa(x,v,v')\,\psi(x,v',t)\,dv' \\ \qquad\qquad\qquad = T_H\psi(x,v,t) + K\psi(x,v,t), \quad (x,v) \in D \times V, \ t > 0 \\ \bullet \ \psi(x,v,0) = \psi_0(x,v). \end{cases}$$

Here $D \subset \mathbb{R}^N$ is a convex-bounded domain, $V$ be a symmetric bounded subset of $\mathbb{R}^N$, and $H$ denotes the boundary operator relating the outgoing $\psi^+$ and the incoming fluxes $\psi^-$. The collision frequency $\Sigma(.)$ is a non-negative function. The scattering kernel $\kappa(.,.,.)$ is non-negative and defines the linear operator $K$ called the collision operator, which is assumed to be bounded on $L_p\big(D \times V, dx \otimes dv\big)$

S. Charfi (✉) • A. Elleuch (✉) • A. Jeribi
Faculté des Sciences de Sfax, Département de Mathématiques, Université de Sfax,
Route de Soukra km 3.5, BP 1171, 3000, Sfax, Tunisie
e-mail: Aref.Jeribi@fss.rnu.tn; asrar_elleuch@yahoo.fr



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings
in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_22

($1 \leq p < \infty$). The operator $T_H$ appearing in (I) is called the streaming operator, whereas $T_H + K$ denotes the transport operator. The bounce-back boundary conditions are modeled by:

$$\psi^- = H(\psi^+),$$

where

$$H\psi^+(x, v) = \gamma\ \psi^+(x, -v),\ \text{for any } (x, v) \in \Gamma_-.$$

Here, $\gamma$ is a real constant belonging to $(0, 1)$ and $\Gamma_\pm$ represents the incoming and outgoing parts of the boundary of the phase space (see Section 2 for more details).

Then, Eq. (I) can be written formally, as the first order Cauchy problem

$$\begin{cases} \dfrac{\partial \psi}{\partial t} = A_H \psi := T_H \psi + K\psi \\ \\ \psi(0) = \psi_0, \end{cases} \tag{1}$$

where $\psi_0 \in L_p\big(D \times V, dx \otimes dv\big)$.

It is well known that, if $\|H\| < 1$, $T_H$ generates a $C_0$-semigroup of contractions $(U(t))_{t\geq 0}$ in $L_p\big(D \times V, dx \otimes dv\big)$ (see [7]). Since $A_H$ is a bounded perturbation of $T_H$, then, by the classical perturbation theory [6, Theorem 2.1, p. 495], it generates a $C_0$-semigroup $(V(t))_{t\geq 0}$.

The time asymptotic behavior of $V(t)$ was studied for the first time, in a general setting, by I. Vidav [12]. His approach relies on the spectra of perturbed semigroups and consists in expressing the solution $\psi(t)$ as an inverse Laplace transform of the resolvent of $A_H$. This technique was systematized in an abstract setting by M. Mokhtar-Kharroubi [9] and it was based on the following conditions:

$$(\mathscr{A}_0) \begin{cases} \bullet \ \text{There exists an integer } m \text{ such that } [(\lambda - T_H)^{-1}K]^m \text{ is compact for } Re\lambda > \eta, \\ \\ \bullet \ \text{There exists an integer } m \text{ such that} \\ \\ \qquad \lim\limits_{|Im\lambda| \to +\infty} \|[(\lambda - T_H)^{-1}K]^m\| = 0 \ \text{ uniformly on } \{\lambda : Re\lambda \geq w, w > \eta\}, \end{cases}$$

where $\eta$ is the type of $(U(t))_{t\geq 0}$.

The spectral analysis of the streaming operator subjected to bounce-back boundary conditions is studied in [8], where the authors showed that

$$\lim_{|Im\lambda| \to +\infty} \|K(\lambda - T_H)^{-1}K\| = 0 \ \text{ uniformly on } \{\lambda : Re\lambda \geq w, w > \eta\},$$

where $\eta$ is the type of $(U(t))_{t\geq 0}$.

Combining this result with the compactness of $[(\lambda - T_H)^{-1}K]^4$ (see Theorem 1 and 2), we remark that condition $(\mathscr{A}_0)$ is fulfilled then, $\sigma(A_H) \cap \{\lambda \in \mathbb{C}$ such that $Re\lambda > \eta\}$ consists at most of discrete eigenvalues with finite algebraic multiplicities $\{\lambda_1, \lambda_2, \ldots, \lambda_n, \lambda_{n+1}, \ldots\}$ which can be ordered in such a way that the real part decreases [6], i.e., $Re\lambda_1 > Re\lambda_2 > \ldots > Re\lambda_{n+1} > \ldots > \eta$ and $\{\lambda \in \mathbb{C}$ such that $Re\lambda > \eta\} \setminus \{\lambda_n : n = 1, \ldots,\} \subset \rho(A_H)$, where $\rho(A_H)$ is the resolvent set of $A_H$. Moreover, for any initial data $\psi_0 \in \mathscr{D}(A_H^2)$, the solution of the Cauchy problem (1) fulfills

$$\left\| \psi(t) - \sum_{i=1}^{n} e^{\lambda_i t} e^{D_i t} P_i \psi_0 \right\| = o(e^{\beta^* t}) \text{ where } \beta_1 < \beta^* < \beta_2 \tag{2}$$

$\beta_1 = \sup\{Re\lambda$ such that $\lambda \in \sigma(A_H), Re\lambda < w\}$, and $\beta_2 = \min\{Re\lambda_i, 1 \leq i \leq n\}$, $P_i$ and $D_i$ denote, respectively, the spectral projection and the nilpotent operator associated with $\lambda_i, i = 1, 2, \ldots, n$.

Hence, the estimate of $\|K(\lambda - T_H)^{-1}K\|$, for large $|Im\lambda|$, given in [8] leads to a characterization of the time asymptotic behavior of the solution of the transport equation with bounce-back boundary conditions only for $\psi_0 \in \mathscr{D}(A_H^2)$.

Our interest in this paper is to ameliorate the description of the time asymptotic behavior for large times of the solution without restriction to initial data $\psi_0 \in \mathscr{D}(A_H^2)$.

To this purpose, we are based on the spectral analysis given in [1] where the estimation (2) has been ameliorated and the time asymptotic behavior of solution of the abstract Cauchy problem (1) is given by an estimation similar to (2) when the stern condition $\psi_0 \in \mathscr{D}(A_H^2)$ is eliminated. This analysis is applied by S. Charfi et al. in [2, 3] for the study of solution of transport operators with diffuse reflection and Maxwell boundary conditions on $L_1$-spaces.

In our paper, referring to [1, 4], we have to show that for all $r \in [0, 1)$

$$\lim_{|Im\lambda| \to +\infty} |Im\lambda|^r \, \|K(\lambda - T_H)^{-1}K\| = 0 \text{ uniformly on } R_w,$$

where $R_w := \{\lambda \in \mathbb{C}$ such that $Re\lambda \geq -\lambda^* + w\}$.

This interesting result enables us to investigate the time asymptotic behavior of the solution of multidimensional transport equation and we prove that, for any initial data $\psi_0 \in \mathscr{D}(A_H)$:

($i$) For each $\varepsilon > 0$, there exists $M > 0$ such that

$$\left\| V(t) - \sum_{i=1}^{n} e^{\lambda_i t} e^{D_i t} P_i \right\|_{X_p} \leq M \, e^{(Re\lambda_{n+1} + \varepsilon)t}, \quad \forall\, t > 0 \text{ and } p \in [1, 2].$$

$(ii)$ Further, if $K$ is positive, then for each $p > 1$ there exists $M' > 0$ such that

$$\left\| V(t)(I - \tilde{P}) \right\|_{X_p} \leq M' \, e^{(s(A_H)-\varepsilon)t}, \quad \forall \, t > 0$$

and for every $\varepsilon \in (0, 2r(1 - p^{-1}))$ (resp. $\varepsilon \in (0, 2rp^{-1})$) if $p \leq 2$ (resp. $p > 2$), where $s(A_H)$ is the spectral bound of $A_H$ which is defined by

$$s(A_H) := \sup \left\{ Re\lambda : \lambda \in \sigma(A_H) \right\},$$

$\tilde{P}$ denotes the projection operator corresponding to

$$\left\{ \lambda \in \sigma(A_H) : Re\lambda = s(A_H) \right\},$$

and $r$ be the real defined by

$$r := s(A_H) - \sup \left\{ Re\lambda : \lambda \in \sigma(A_H), \ \lambda \neq s(A_H) \right\}.$$

This paper is organized as follows: in the next section, we fix notation and derive fundamental preliminary. In Section 3, we give crucial lemmas for future use. In Section 4, we prove under some conditions that, for all $r \in [0, 1)$

$$\lim_{|Im\lambda| \to +\infty} |Im\lambda|^r \, \|K(\lambda - T_H)^{-1} K\| = 0 \text{ uniformly on } R_w,$$

where $R_w := \{\lambda \in \mathbb{C} \text{ such that } Re\lambda \geq -\lambda^* + w\}$. This estimation will be useful in Section 5, in which we investigate the asymptotic spectrum of $A_H$ and we describe the time asymptotic behavior of the solution of our problem without restriction on the initial data.

## 2 Preliminary and Compactness Results

In this section, we gather the different notions and notation facts connected to our problem. Let

$$X_p := L_p\big(D \times V, dx \otimes dv\big) \qquad (1 \leq p < \infty),$$

where $D$ be a smooth-bounded open subset of $\mathbb{R}^N$ whilst $V \subset \mathbb{R}^N$ is a symmetric bounded subset. We define the partial Sobolev spaces as

$$W_p := \left\{ \psi \in X_p; \ v.\nabla_x \psi \in X_p \right\}.$$

We denote by $\Gamma_-$ (respectively $\Gamma_+$) the incoming (resp. outgoing) part of the boundary of the phase space $D \times V$ defined as

$$\Gamma_\pm := \Big\{(x,v) \in \partial D \times V;\ \pm v.n(x) \geq 0\Big\},$$

where $n(x)$ stands for the outward normal unit at $x \in \partial D$.

Suitable $L_p$-spaces for the traces on $\Gamma_\pm$ are defined as

$$L_p^\pm := L_p\big(\Gamma_\pm;\ |v.n(x)|\, d\gamma(x) \times dv\big),$$

where $d\gamma(.)$ being the Lebesgue measure on $\partial D$. For any $\psi \in W_p$, one can define the traces $\psi^\pm := \psi_{/\Gamma_\pm}$ on $\Gamma_\pm$; however, these traces do not belong to $L_p^\pm$ but to a certain weighted space. For this reason, one defines

$$\widetilde{W_p} := \Big\{\psi \in W_p;\ \psi_{/\Gamma_\pm} \in L_p^\pm\Big\}.$$

Moreover, these boundary spaces endowed with norm

$$\|\psi^-\|_{L_p^-} := \Big(\int_{\Gamma_-} |\psi(x,v)|^p |v.n(x)|\, d\gamma(x)\, dv\Big)^{\frac{1}{p}}$$

and

$$\|\psi^+\|_{L_p^+} := \Big(\int_{\Gamma_+} |\psi(x,v)|^p |v.n(x)|\, d\gamma(x)\, dv\Big)^{\frac{1}{p}}.$$

**Definition 1.** For any $(x,v) \in \overline{D} \times V$, define

$$\begin{aligned} t_\pm(x,v) &:= \sup\Big\{t > 0;\ x \pm sv \in D, \forall\, 0 < s < t\Big\}, \\ &:= \inf\Big\{s > 0;\ x \pm sv \notin D\Big\}. \end{aligned}$$

Furthermore, set

$$\tau(x,v) := t_-(x,v) + t_+(x,v) \quad \text{for any } (x,v) \in \overline{D} \times V.$$

◇

We define the streaming operator $T_H$ with bounce-back boundary condition:

$$\begin{cases} T_H : \mathscr{D}(T_H) \subseteq X_p \longrightarrow X_p \\ \qquad\qquad\qquad \psi \longrightarrow T_H\psi(x,v) = -v.\nabla_x\psi(x,v) - \Sigma(v)\psi(x,v) \\ \mathscr{D}(T_H) = \big\{\psi \in \widetilde{W_p} \text{ such that } \psi^- = H(\psi^+)\big\}, \end{cases}$$

where $H$ is the boundary operator defined by

$$\begin{cases} H : L_p^+ \longrightarrow L_p^- \\ \qquad \psi^+ \longrightarrow H\psi^+;\ H\psi^+(x,v) := \gamma\ \psi^+(x,-v) \text{ for any } (x,v) \in \Gamma_-, \end{cases}$$

where $0 < \gamma < 1$.

We will assume that the collision frequency is non-negative and bounded. Set

$$\lambda^* := \inf_{v \in V} \Sigma(v).$$

Now, let us investigate the resolvent of $T_H$. Let us define the following operators for $Re\lambda + \lambda^* > 0$ :

$$\begin{cases} M_\lambda : L_p^- \longrightarrow L_p^+ \\ \qquad u \longrightarrow M_\lambda u;\ M_\lambda u(x,v) = u(x - \tau(x,v)v, v)\ \exp\Big\{ - \tau(x,v)\big(\lambda + \Sigma(v)\big)\Big\}, \end{cases}$$

$$\begin{cases} B_\lambda : L_p^- \longrightarrow X_p \\ \qquad u \longrightarrow B_\lambda u;\ B_\lambda u(x,v) = u(x - t_-(x,v)v, v)\ \exp\Big\{ - t_-(x,v)\big(\lambda + \Sigma(v)\big)\Big\}, \end{cases}$$

$$\begin{cases} G_\lambda : X_p \longrightarrow L_p^+ \\ \qquad \varphi \longrightarrow G_\lambda \varphi;\ G_\lambda \varphi(x,v) = \displaystyle\int_0^{\tau(x,v)} \varphi(x - sv, v)\ \exp\Big\{ - \big(\lambda + \Sigma(v)\big)s\Big\}\, ds, \end{cases}$$

$$\begin{cases} C_\lambda : X_p \longrightarrow X_p \\ \qquad \varphi \longrightarrow C_\lambda \varphi;\ C_\lambda \varphi(x,v) = \displaystyle\int_0^{t_-(x,v)} \varphi(x - sv, v)\ \exp\Big\{ - \big(\lambda + \Sigma(v)\big)s\Big\}\, ds. \end{cases}$$

These operators are bounded in their respective spaces. In fact, for $Re\lambda > -\lambda^*$, the norms of the operators $M_\lambda$, $B_\lambda$, $G_\lambda$, and $C_\lambda$ are bounded, respectively, by 1, $[p(Re\lambda + \lambda^*)]^{-\frac{1}{p}}$, $[q(Re\lambda + \lambda^*)]^{-\frac{1}{q}}$ and $(Re\lambda + \lambda^*)^{-1}$ where $q^{-1} + p^{-1} = 1$.

Now, we have $\|H\| < 1$, then for any $\lambda$ satisfying $Re\lambda > -\lambda^*$, we have $\|M_\lambda H\| < 1$ hence $(I - M_\lambda H)^{-1}$ exists and the resolvent of the operator $T_H$ is given by

$$(\lambda - T_H)^{-1} = B_\lambda H (I - M_\lambda H)^{-1} G_\lambda + C_\lambda. \tag{3}$$

Next, the transport operator $A_H$ can be written as follows:

$$A_H := T_H + K,$$

where $K$ is a bounded linear operator on $X_p$ defined by

$$\begin{cases} K : X_p \longrightarrow X_p \\ \qquad \psi \longrightarrow \int_V \kappa(x, v, v') \; \psi(x, v') \, dv', \end{cases}$$

where the scattering Kernel $\kappa : \; D \times V \times V \to \mathbb{R}$ is assumed to be measurable.

We observe that the operator $K$ acts only on the velocity variable $v'$, so $x$ may be viewed as a parameter in $D$. Consequently, we can consider $K$ as a function

$$\begin{cases} K(.) : D \longrightarrow \mathscr{L}\big(L_p(V, dv)\big) \\ \qquad\quad x \longrightarrow K(x), \end{cases}$$

we consider the following assumption:

$$(\mathscr{A}) \begin{cases} \bullet \text{ The function } K(.) \text{ is strongly measurable,} \\ \bullet \text{ There exists a compact subset } C \subseteq \mathscr{L}\big(L_p(V, dv)\big) \text{ such that } K(x) \in C \text{ a.e on } D, \\ \bullet \; K(x) \in \mathscr{K}(L_p(V, dv)) \text{ a.e,} \end{cases}$$

where $\mathscr{K}(L_p(V, dv))$ is the subspace of compact operators.

Let us denote by $Z := \mathscr{L}\big(L_p(V, dv)\big)$. Using the second point of the above assumption, we obtain

$$K(.) \in L^\infty(D, Z).$$

If $\psi \in X_p$, then it is easy to see that $(K\psi)(x, v) = (K(x)\psi)(v)$; and so

$$\int_V |(K\psi)(x, v)|^p \, dv \le \|K(.)\|^p_{\mathscr{L}^\infty(D,Z)} \int_V |\psi(x, v)|^p \, dv.$$

Therefore,

$$\int_D \int_V |(K\psi)(x, v)|^p \, dv \, dx \le \|K(.)\|^p_{\mathscr{L}^\infty(D,Z)} \int_D \int_V |\psi(x, v)|^p \, dv \, dx.$$

Consequently,

$$\|K\|_{\mathscr{L}(X_p)} \le \|K(.)\|_{\mathscr{L}^\infty(D,Z)}.$$

In this paper we will use the concept of regular collision operators introduced by M. Mokhtar-Kharroubi.

**Definition 2.** A collision operator $K$ is regular if it satisfies the assumption $(\mathscr{A})$. $\diamond$

**Definition 3 ([8, Definition 4.1]).** If the collision operator $K$ is regular, then it can be approximated in the operator norm by operators of the form:

$$\varphi \in X_p \longrightarrow \sum_{i \in I} \alpha_i(x)\, \beta_i(v) \int_V \theta_i(w)\, \varphi(x,w)\, dw \in X_p,$$

where $I$ is finite, $\alpha_i \in L^\infty(D)$, $\beta_i \in L_p(V, dv)$ and $\theta_i \in L_q(V, dv)$; $p = q(p-1)$. $\diamond$

*Remark 2.1.* We can assume in the above definition that $\beta_i$ and $\theta_i$ are measurable simple functions with compact supports in $V$. $\diamond$

The compactness results are established in [7] and given by the following theorems:

**Theorem 2.1.** *Let $1 < p < \infty$. If $K$ is a regular operator, then for any complex number $\lambda$ satisfying $Re\lambda > -\lambda^*$, the operators $K(\lambda - T_H)^{-1}$ and $(\lambda - T_H)^{-1}K$ are compact on $X_p$.* $\diamond$

**Theorem 2.2.** *Let $K$ be a regular operator on $X_1$. If $H$ is weakly compact operator, then for $Re\lambda > -\lambda^*$ we have $K(\lambda - T_H)^{-1}K$ is weakly compact on $X_1$.* $\diamond$

In the next, we will prove the following result.

**Proposition 1.** *Let the boundary operator $H$ be non-negative, then $T_H$ generates a strongly continuous semigroup positive $(U(t))_{t \geq 0}$, satisfying*

$$\|U(t)\| \leq e^{-\lambda^* t}. \qquad \diamond$$

*Proof.* According to [7, Lemma 2.2] we have for any $\lambda$ satisfying $Re\lambda > -\lambda^*$,

$$\|(\lambda - T_H)^{-1}\| \leq \frac{1}{Re\lambda + \lambda^*}.$$

Combining this result together with [11, Corollary 3.8, p. 12], we may immediately deduce the result.

In all the sequel, we shall assume that $\Sigma(.)$ is an even function of the velocity, i.e. for any $v \in V$, $\Sigma(-v) = \Sigma(v)$.

## 3 Auxiliary Lemmas

The goal of this section is to establish some lemmas which we will use in the next section.

Let $w > 0$ and set

$$R_w := \{\lambda \in \mathbb{C} \text{ such that } Re\lambda \geq -\lambda^* + w\}.$$

Note that if $\lambda \in R_w$, then

$$Re\lambda + \lambda^* - \frac{w}{2} \geq \frac{w}{2} > 0$$

and for all $v \in V$

$$\Sigma(v) - \lambda^* + \frac{w}{2} \geq \frac{w}{2} > 0.$$

We consider

$$\begin{cases} \tilde{\varphi}_x : [-\frac{d}{a}, \frac{d}{a}] \longrightarrow \mathbb{C} \\ \qquad t \longrightarrow \tilde{\varphi}_x(t) = \exp\Big\{\big(\lambda^* - \Sigma(-\frac{x}{t}) - \frac{w}{2}\big)t\Big\} \exp\Big\{ - \big(\lambda + \Sigma(-\frac{x}{t})\big) \\ \qquad\qquad \times\big(2n\tau(x+z, -\frac{x}{t}) + 2t_-(x+z, -\frac{x}{t})\big)\Big\}, \end{cases}$$

where $x \in \overline{D}$ and $\tilde{\varphi}_x \in L_1([-\frac{d}{a}, \frac{d}{a}])$.

Indeed,

$$\begin{aligned} \int_{-\frac{d}{a}}^{\frac{d}{a}} |\tilde{\varphi}_x(t)| dt \leq \int_{-\frac{d}{a}}^{\frac{d}{a}} &\exp\Big\{\big(\lambda^* - \Sigma(-\frac{x}{t}) - \frac{w}{2}\big)t\Big\} \\ &\exp\Big\{ - \big(Re\lambda + \lambda^*\big)\big(2n\tau(x+z, -\frac{x}{t}) + 2t_-(x+z, -\frac{x}{t})\big)\Big\}\, dt \\ \int_{-\frac{d}{a}}^{\frac{d}{a}} |\tilde{\varphi}_x(t)| dt \leq \int_{-\frac{d}{a}}^{\frac{d}{a}} &\exp\Big\{\big(\lambda^* - \Sigma(-\frac{x}{t}) - \frac{w}{2}\big)t\Big\}\, dt \\ \leq \Big[&\frac{1}{\lambda^* - \sup_v \Sigma(v) - \frac{w}{2}}\Big(1 - \exp\Big\{ - \big(\lambda^* - \sup_v \Sigma(v) - \frac{w}{2}\big)\frac{d}{a}\Big\}\Big) \\ &-\frac{2}{w}\Big(\exp\Big\{ - \frac{wd}{2a}\Big\} - 1\Big)\Big]. \end{aligned}$$

Let $x \in \overline{D}$ and denote by $\Big(\tilde{\rho}_{x,p}(.)\Big)_{p\in\mathbb{N}}$ a sequence of continuous functions with compact support which converges to $\tilde{\varphi}_x(.)$ in $L_1([-\frac{d}{a}, \frac{d}{a}])$.

Set for $v \in V$

$$h(v) := \theta_1(v)\beta_1(-v).$$

Clearly, $h(.)$ is a simple measurable function with compact support.

Let $I := \Big( - t_-(x+z, -\frac{x}{t}), t_+(x+z, -\frac{x}{t})\Big)$ and we introduce

$$G_{p,\lambda}(x) = \sup_{z \in D - x} \Big| \int_{\mathbb{R}} h(-\frac{x}{t}) \exp\Big\{ - \Big(\lambda + \lambda^* - \frac{w}{2}\Big)t\Big\}\, \tilde{\rho}_{x,p}(t)\, \chi_I(t)\, \frac{dt}{|t^N|} \Big|$$

and

$$G_\lambda(x) = \sup_{z \in D-x} \Big| \int_{\mathbb{R}} h(-\frac{x}{t}) \, \exp\Big\{ - \Big(\lambda + \Sigma(-\frac{x}{t})\Big)\Big(t + 2n\tau(x+z, -\frac{x}{t}) \\ + 2t_-(x+z, -\frac{x}{t})\Big)\Big\} \, \chi_I(t) \, \frac{dt}{|t^N|}\Big|,$$

where $\chi_I(.)$ denotes the characteristic function of $I$.

**Lemma 1.** *Let $p \in \mathbb{N}$ and $\lambda \in R_w$, then:*

*(i)* $G_\lambda \in L_1(D)$.
*(ii)* $G_{p,\lambda} \in L_1(D)$.
*(iii)* *The sequence $(G_{p,\lambda})_{p \in \mathbb{N}}$ converges in $L_1(D)$ uniformly on $R_w$ to the function $G_\lambda$.*

◇

*Proof.* $(i)$ We have

$$\Big| \int_{\mathbb{R}} h(-\frac{x}{t}) \, \exp\Big\{ - \Big(\lambda + \Sigma(-\frac{x}{t})\Big)\Big(t + 2n\tau(x+z, -\frac{x}{t}) \\ + 2t_-(x+z, -\frac{x}{t})\Big)\Big\} \, \chi_I(t) \, \frac{dt}{|t^N|}\Big| \\ \leq \int_{\mathbb{R}} \Big|h(-\frac{x}{t})\Big| \, \exp\Big\{ - \Big(Re\lambda + \Sigma(-\frac{x}{t})\Big)t\Big\} \\ \exp\Big\{ - \Big(Re\lambda + \lambda^*\Big)\Big(2n\tau(x+z, -\frac{x}{t}) + 2t_-(x+z, -\frac{x}{t})\Big)\Big\} \, \frac{dt}{|t^N|}.$$

There is no loss of generality assuming that there exist two constants $a, b > 0$ such that

$$\text{Supp(h)} \subset \{v \in V : a \leq |v| \leq b\}.$$

In this case, in the above integral, one can see that $t \in \mathbb{R}$ is such that

$$a \leq \Big|\frac{x}{t}\Big| \leq b$$

which implies that $|t| \leq \frac{|x|}{a}$. This means that the above integral over $\mathbb{R}$ can be reduced actually to an integral over $[-\frac{d}{a}, \frac{d}{a}]$, where $d$ is the diameter of $D$.

$$G_\lambda(x) \leq \int_{-\frac{d}{a}}^{\frac{d}{a}} \Big|h(-\frac{x}{t})\Big| \, \exp\Big\{ - \Big(Re\lambda + \Sigma(-\frac{x}{t})\Big)t\Big\} \, \frac{dt}{|t^N|}.$$

Therefore,

$$\int_D G_\lambda(x)\,dx \le \int_D \int_{-\frac{d}{a}}^{0} \Big|h(-\frac{x}{t})\Big| \,\exp\Big\{-\Big(Re\lambda + \sup_v \Sigma(v)\Big)t\Big\}\,\frac{dt}{|t^N|}\,dx$$

$$+ \int_D \int_0^{\frac{d}{a}} \Big|h(-\frac{x}{t})\Big| \,\exp\Big\{-\Big(Re\lambda + \lambda^*\Big)t\Big\}\,\frac{dt}{|t^N|}\,dx.$$

The change of variable $v = \frac{x}{t}$ gives

$$\int_D G_\lambda(x)\,dx \le \int_V \int_{-\frac{d}{a}}^{0} |h(-v)| \,\exp\Big\{-\Big(Re\lambda + \sup_v \Sigma(v)\Big)t\Big\}\,dt\,dv$$

$$+ \int_V \int_0^{\frac{d}{a}} |h(-v)| \,\exp\Big\{-\Big(Re\lambda + \lambda^*\Big)t\Big\}\,dt\,dv.$$

$$\int_D G_\lambda(x)\,dx \le \int_V |h(-v)|\,dv\,\Big[\frac{-1}{Re\lambda + \sup_v \Sigma(v)}\Big(1 - \exp\Big\{\big(Re\lambda + \sup_v \Sigma(v)\big)\frac{d}{a}\Big\}\Big)\Big]$$

$$+ \int_V |\theta_1(-v)|\,|\beta_1(v)|\,dv\,\Big[\frac{1}{Re\lambda + \lambda^*}\Big(\exp\Big\{-(Re\lambda + \lambda^*)\frac{d}{a}\Big\} - 1\Big)\Big].$$

Using Hölder's inequality, we get

$$\int_D G_\lambda(x)\,dx \le \|\theta_1\|_{L_q(V)}\,\|\beta_1\|_{L_p(V)}\Big[\frac{-1}{Re\lambda + \sup_v \Sigma(v)}\Big(1 - \exp\Big\{\big(Re\lambda + \sup_v \Sigma(v)\big)\frac{d}{a}\Big\}\Big)$$

$$+ \frac{1}{Re\lambda + \lambda^*}\Big(\exp\Big\{-(Re\lambda + \lambda^*)\frac{d}{a}\Big\} - 1\Big)\Big] < \infty.$$

(*ii*) Since we have

$$\Big|\int_{\mathbb{R}} h(-\frac{x}{t})\,\exp\Big\{-\Big(\lambda + \lambda^* - \frac{w}{2}\Big)t\Big\}\,\tilde{\rho}_{x,p}(t)\,\chi_I(t)\,\frac{dt}{|t^N|}\Big|$$

$$\le \sup|\tilde{\rho}(.)| \int_{-\frac{d}{a}}^{\frac{d}{a}} \Big|h(-\frac{x}{t})\Big|\,\exp\Big\{-\Big(Re\lambda + \lambda^* - \frac{w}{2}\Big)t\Big\}\,\frac{dt}{|t^N|},$$

we get after making the change of variable $v = \frac{x}{t}$

$$\int_D G_{p,\lambda}(x)\,dx \le \sup|\tilde{\rho}(.)| \int_V \int_{-\frac{d}{a}}^{\frac{d}{a}} |h(-v)|\,\exp\Big\{-\Big(Re\lambda + \lambda^* - \frac{w}{2}\Big)t\Big\}\,dt\,dv.$$

Now using Hölder's inequality, we obtain

$$\int_D G_{p,\lambda}(x)\,dx \le \sup|\tilde{\rho}(.)|\ \ \|\theta_1\|_{L_q(V)}\ \|\beta_1\|_{L_p(V)}\Big[\frac{1}{(Re\lambda+\lambda^*-\frac{w}{2})}$$

$$\Big(\ \exp\Big\{\big(Re\lambda+\lambda^*-\tfrac{w}{2}\big)\tfrac{d}{a}\Big\}-\exp\Big\{-\big(Re\lambda+\lambda^*-\tfrac{w}{2}\big)\tfrac{d}{a}\Big\}\Big)\Big].$$

(*iii*) For all $p\in\mathbb{N}$ and $\lambda\in R_w$, we have

$$\begin{aligned}
&|G_{p,\lambda}(x)-G_\lambda(x)|\\
&\quad=\Big|\sup_{z\in D-x}\Big|\int_{\mathbb{R}}h(-\frac{x}{t})\ \exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)t\Big\}\ \tilde{\rho}_{x,p}(t)\ \chi_I(t)\ \frac{dt}{|t^N|}\Big|\\
&\qquad-\sup_{z\in D-x}\Big|\int_{\mathbb{R}}h(-\frac{x}{t})\ \exp\Big\{-\Big(\lambda+\Sigma(-\frac{x}{t})\Big)\\
&\qquad\Big(t+2n\tau(x+z,-\frac{x}{t})+2t_-(x+z,-\frac{x}{t})\Big)\Big\}\ \chi_I(t)\ \frac{dt}{|t^N|}\Big|\ \Big|\\
&\quad\le\sup_{z\in D-x}\Big|\int_{\mathbb{R}}\ h(-\frac{x}{t})\ \exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)t\Big\}\ \tilde{\rho}_{x,p}(t)\ \chi_I(t)\ \frac{dt}{|t^N|}\\
&\qquad-\int_{\mathbb{R}}h(-\frac{x}{t})\ \exp\Big\{-\Big(\lambda+\Sigma(-\frac{x}{t})\Big)\\
&\qquad\Big(t+2n\tau(x+z,-\frac{x}{t})+2t_-(x+z,-\frac{x}{t})\Big)\Big\}\ \chi_I(t)\ \frac{dt}{|t^N|}\Big|\\
&\quad\le\sup_{z\in D-x}\Big|\int_{\mathbb{R}}h(-\frac{x}{t})\ \exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)t\Big\}\ \chi_I(t)\ \Big[\tilde{\rho}_{x,p}(t)-\tilde{\varphi}_x(t)\Big]\ \frac{dt}{|t^N|}\Big|.
\end{aligned}$$

This implies

$$\begin{aligned}
&|G_{p,\lambda}(x)-G_\lambda(x)|\\
&\quad\le\exp\Big\{\Big(Re\lambda+\lambda^*-\frac{w}{2}\Big)\frac{d}{a}\Big\}\sup_{z\in D-x}\Big(\int_{-\frac{d}{a}}^{\frac{d}{a}}\Big|h(-\frac{x}{t})\Big|\ |\tilde{\rho}_{x,p}(t)-\tilde{\varphi}_x(t)|\frac{dt}{|t^N|}\Big).
\end{aligned}$$

Therefore

$$\|G_{p,\lambda}-G_\lambda\|_{L_1(D)}\le\exp\Big\{\Big(Re\lambda+\lambda^*-\frac{w}{2}\Big)\frac{d}{a}\Big\}\int_D\int_{-\frac{d}{a}}^{\frac{d}{a}}\Big|h(-\frac{x}{t})\Big|$$

$$\sup_{z\in D-x}|\tilde{\rho}_{x,p}(t)-\tilde{\varphi}_x(t)|\frac{dt}{|t^N|}dx$$

$$\leq \exp\left\{\left(Re\lambda + \lambda^* - \frac{w}{2}\right)\frac{d}{a}\right\} \int_D \int_{-\frac{d}{a}}^{\frac{d}{a}} \left|h(-\frac{x}{t})\right|$$

$$\sup_{x\in\overline{D}} \sup_{z\in D-x} |\tilde{\rho}_{x,p}(t) - \tilde{\varphi}_x(t)| \frac{dt}{|t^N|}\, dx.$$

The change of variable $v = \frac{x}{t}$ gives

$$\|G_{p,\lambda} - G_\lambda\|_{L_1(D)} \leq \exp\left\{\left(Re\lambda + \lambda^* - \frac{w}{2}\right)\frac{d}{a}\right\} \int_V \int_{-\frac{d}{a}}^{\frac{d}{a}} \left|h(-v)\right| |\tilde{\rho}_{x_1,p}(t) - \tilde{\varphi}_{x_1}(t)| dt\, dv.$$

Applying Hölder's inequality, we get

$$\|G_{p,\lambda} - G_\lambda\|_{L_1(D)} \leq \exp\left\{\left(Re\lambda + \lambda^* - \frac{w}{2}\right)\frac{d}{a}\right\} \|\theta_1\|_{L_q(V)}\ \|\beta_1\|_{L_p(V)}\|\tilde{\rho}_{x_1,p} - \tilde{\varphi}_{x_1}\|_{L_1([-\frac{d}{a},\frac{d}{a}])}.$$

Consequently,

$$\lim_{p\to+\infty} \|G_{p,\lambda} - G_\lambda\|_{L_1(D)} = 0 \text{ uniformly on } R_w.$$

**Lemma 2.** *Let $r \in [0, 1)$ and $p \in \mathbb{N}$, then*

$$\lim_{|Im\lambda|\to+\infty} |Im\lambda|^r \int_D G_{p,\lambda}(x)\, dx = 0 \textit{ uniformly on } R_w.$$

◇

*Proof.* Observe that

$$G_{p,\lambda}(x) \leq G^-_{p,\lambda}(x) + G^+_{p,\lambda}(x),$$

where

$$G^-_{p,\lambda}(x) = \sup_{z\in D-x} \left| \int_{-\infty}^{0} h(-\frac{x}{t})\ \exp\left\{-\left(\lambda + \lambda^* - \frac{w}{2}\right)t\right\} \tilde{\rho}_{x,p}(t)\ \chi_{I^-}(t)\ \frac{dt}{|t^N|}\right|$$

and

$$G^+_{p,\lambda}(x) = \sup_{z\in D-x} \left| \int_{0}^{+\infty} h(-\frac{x}{t})\ \exp\left\{-\left(\lambda + \lambda^* - \frac{w}{2}\right)t\right\} \tilde{\rho}_{x,p}(t)\ \chi_{I^+}(t)\ \frac{dt}{|t^N|}\right|,$$

with $I^- = I\cap]-\infty, 0]$ and $I^+ = I \cap [0, +\infty[$.

Consider now the sequence of operators $(G^+_{p,\lambda,\varepsilon_n})_{n\in\mathbb{N}}$ and $(G^-_{p,\lambda,\varepsilon_n})_{n\in\mathbb{N}}$ where

$$G^-_{p,\lambda,\varepsilon_n}(x) = \sup_{z\in D-x} \Big| \int_{-\infty}^{\varepsilon_n} h(-\frac{x}{t}) \ \exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)t\Big\} \ \tilde{\rho}_{x,p}(t) \ \chi_{I^-_{\varepsilon_n}}(t) \ \frac{dt}{|t^N|}\Big|,$$

$$G^+_{p,\lambda,\varepsilon_n}(x) = \sup_{z\in D-x} \Big| \int_{\varepsilon_n}^{+\infty} h(-\frac{x}{t}) \ \exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)t\Big\} \ \tilde{\rho}_{x,p}(t) \ \chi_{I^+_{\varepsilon_n}}(t) \ \frac{dt}{|t^N|}\Big|.$$

Clearly the sequence $(G^+_{p,\lambda,\varepsilon_n})_{n\in\mathbb{N}}$ (resp. $(G^-_{p,\lambda,\varepsilon_n})_{n\in\mathbb{N}}$) converges to $G^+_{p,\lambda}$ (resp. $G^-_{p,\lambda}$) uniformly on $R_w$ when $\varepsilon_n$ goes to zero. So, it suffices to show that, for all $\varepsilon > 0$

$$\lim_{|Im\lambda|\to+\infty} |Im\lambda|^r \int_D G^+_{p,\lambda,\varepsilon}(x)\,dx = 0 \text{ and } \lim_{|Im\lambda|\to+\infty} |Im\lambda|^r \int_D G^-_{p,\lambda,\varepsilon}(x)\,dx = 0.$$

Set

$$\begin{cases} L_{x,p} : [\varepsilon,+\infty[\longrightarrow \mathbb{C} \\ \qquad\qquad t \longrightarrow h(-\frac{x}{t}) \ \tilde{\rho}_{x,p}(t) \ \chi_{I^+_{\varepsilon}}(t) \ \frac{1}{|t^N|}. \end{cases}$$

$L_{x,p}$ is a simple function. Let $(t_i)_{1\le i\le m}$ denote a subdivision of its support satisfying $L_{x,p}(t) = L_{x,p}(t_i)$, for all $t\in[t_i,t_{i+1}[$, with $i\in\{1,\dots,m-1\}$. Hence

$$\begin{aligned} &\int_{\varepsilon}^{+\infty} \exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)t\Big\} \ L_{x,p}(t)\,dt \\ &= \sum_{i=1}^{m-1} L_{x,p}(t_i) \int_{t_i}^{t_{i+1}} \exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)t\Big\}\,dt, \\ &= \frac{1}{(\lambda+\lambda^*-\frac{w}{2})} \sum_{i=1}^{m-1} L_{x,p}(t_i)\Big(\exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)t_i\Big\} \\ &\quad - \exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)t_{i+1}\Big\}\Big). \end{aligned}$$

Since we have

$$\begin{cases} \Big|\exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)t_i\Big\} - \exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)t_{i+1}\Big\}\Big| \le 2 \\ \dfrac{1}{|\lambda+\lambda^*-\frac{w}{2}|} \le \dfrac{1}{|Im\lambda|} \\ |L_{x,p}(t_i)| \le \sup|h(.)| \ \dfrac{\sup|\tilde{\rho}(.)|}{\varepsilon^N}, \end{cases}$$

then

$$\Big|\int_{\varepsilon}^{+\infty} \exp\Big\{-\Big(\lambda+\lambda^{*}-\frac{w}{2}\Big)t\Big\}\, L_{x,p}(t)\, dt\Big| \leq \frac{2(m-1)}{|Im\lambda|\ \varepsilon^{N}} \sup|h(.)|\ \sup|\tilde{\rho}(.)|.$$

Therefore

$$\int_D G^{+}_{p,\lambda,\varepsilon}(x)\, dx \leq \frac{2(m-1)}{|Im\lambda|\ \varepsilon^{N}} \sup|h(.)|\ \sup|\tilde{\rho}(.)| \int_D dx.$$

So,

$$\lim_{|Im\lambda|\to+\infty} |Im\lambda|^{r} \int_D G^{+}_{p,\lambda,\varepsilon}(x)\, dx = 0 \text{ uniformly on } R_w.$$

As the same way, we prove that

$$\lim_{|Im\lambda|\to+\infty} |Im\lambda|^{r} \int_D G^{-}_{p,\lambda,\varepsilon}(x)\, dx = 0 \text{ uniformly on } R_w.$$

Let $\varepsilon > 0$, for $x \in \overline{D}$, we consider

$$\begin{cases} \phi_x : [\varepsilon, +\infty[\longrightarrow \mathbb{R}_+ \\ \qquad s \longrightarrow \frac{1}{s^N} \exp\Big\{-\Big(\Sigma(\frac{x}{s}) + \frac{w}{2} - \lambda^{*}\Big)s\Big\}. \end{cases}$$

$0 \leq \phi_x(.) \in L_1([\varepsilon, +\infty[)$. We denote by $\big(\tilde{\phi}_{x,n}(.)\big)_{n\in\mathbb{N}}$, an increasing sequence of non-negative step functions with compact support which converge to $\phi_x(.)$.

Let $A'_2(\lambda)$ be the operator defined by

$$\begin{cases} A'_2(\lambda) : L_p(D) \longrightarrow L_p(D) \\ \qquad \varphi \longrightarrow (A'_2(\lambda)\varphi) \end{cases}$$

$$(A'_2(\lambda)\varphi)(x) = \int_D \int_{\varepsilon}^{t_-(x,\frac{x-y}{s})} h_1\Big(\frac{x-y}{s}\Big)\, \varphi(y)\, \exp\Big\{-\Big(\lambda + \Sigma\Big(\frac{x-y}{s}\Big)\Big)s\Big\} \frac{ds}{s^N}\, dy.$$

We introduce the sequence $\big(A'_{2,n}(\lambda)\big)_{n\in\mathbb{N}}$ of operators defined, for all $n \in \mathbb{N}$, by

$$\begin{cases} A'_{2,n}(\lambda) : L_p(D) \longrightarrow L_p(D) \\ \qquad \varphi \longrightarrow (A'_{2,n}(\lambda)\varphi) \end{cases}$$

$$(A'_{2,n}(\lambda)\varphi)(x) = \int_D \int_{\varepsilon}^{t_-(x,\frac{x-y}{s})} h_1\Big(\frac{x-y}{s}\Big)\, \varphi(y)$$
$$\exp\Big\{-\Big(\lambda - \frac{\omega}{2} + \lambda^{*}\Big)s\Big\}\tilde{\phi}_{x-y,n}(s) ds\, dy.$$

**Lemma 3.** *The sequence of operators* $\left(A'_{2,n}(\lambda)\right)_{n\in\mathbb{N}}$ *converges uniformly on* $R_w$ *to* $A'_2(\lambda)$ *in* $\mathscr{L}(L_p(D))$. ◇

*Proof.* For all $n \in \mathbb{N}$ and $\lambda \in R_w$, we have

$$\left|(A'_{2,n}(\lambda)\varphi)x - (A'_2(\lambda)\varphi)x\right|^p = \Big|\int_D \int_\varepsilon^{t_-(x,\frac{x-y}{s})} h_1\Big(\frac{x-y}{s}\Big) \exp\Big\{-\Big(\lambda - \frac{\omega}{2} + \lambda^*\Big)s\Big\} \Big[\tilde{\phi}_{x-y,n}(s) - \phi_{x-y}(s)\Big] \varphi(y)\, ds\, dy\Big|^p.$$

This implies

$$\left|(A'_{2,n}(\lambda)\varphi)x - (A'_2(\lambda)\varphi)x\right|^p \le \Big(\int_D \int_\varepsilon^{+\infty} \Big|h_1\Big(\frac{x-y}{s}\Big)\Big| \left|\tilde{\phi}_{x-y,n}(s) - \phi_{x-y}(s)\right| |\varphi(y)|\, dsdy\Big)^p.$$

Then

$$\int_D \left|(A'_{2,n}(\lambda)\varphi)x - (A'_2(\lambda)\varphi)x\right|^p dx \le \int_D \Big[\int_D \int_\varepsilon^{+\infty} \Big|h_1\Big(\frac{x-y}{s}\Big)\Big| \left|\tilde{\phi}_{x-y,n}(s) - \phi_{x-y}(s)\right| \times |\varphi(y)|\, ds\, dy\Big]^p dx.$$

The change of variable $z = x - y \in D$ gives

$$\|A'_{2,n}(\lambda)\varphi - A'_2(\lambda)\varphi\|^p \le \int_D \Big[\int_D \int_\varepsilon^{+\infty} \Big|h_1\Big(\frac{z}{s}\Big)\Big| \left|\tilde{\phi}_{z,n}(s) - \phi_z(s)\right| |\varphi(y)|\, ds\, dy\Big]^p dz.$$

Therefore

$$\|A'_{2,n}(\lambda)\varphi - A'_2(\lambda)\varphi\|^p \le \int_D \Big(\int_D (|\varphi(y)|^p + 1)dy\Big)^p \sup |h_1(.)|^p \Big(\int_\varepsilon^{+\infty} \sup_{z\in D} \left|\tilde{\phi}_{z,n}(s) - \phi_z(s)\right| ds\Big)^p dz.$$

Thus

$$\|A'_{2,n}(\lambda)\varphi - A'_2(\lambda)\varphi\| \le \Big(\int_D dz\Big)^{\frac{1}{p}} \Big(\|\varphi\|^p + \int_D dy\Big) \sup |h_1(.)| \left\|\tilde{\phi}_{z_1,n} - \phi_{z_1}\right\|_{L_1([\varepsilon,+\infty[)}.$$

Consequently

$$\|A'_{2,n}(\lambda) - A'_2(\lambda)\| \le \Big(\int_D dz\Big)^{\frac{1}{p}} \Big(1 + \int_D dy\Big) \sup |h_1(.)| \left\|\tilde{\phi}_{z_1,n} - \phi_{z_1}\right\|_{L_1([\varepsilon,+\infty[)}.$$

## 4 Estimation of the Resolvent

The objective of this section is to establish Theorem 4.1 which is required in the proof of our main result.

**Theorem 4.1.** *Let $p \in [1, +\infty[$, assume that the collision operator $K$ is regular. Then for all $r \in [0, 1)$, we have*

$$\lim_{|Im\lambda| \to +\infty} |Im\lambda|^r \, \|K(\lambda - T_H)^{-1}K\| = 0 \text{ uniformly on } R_w. \qquad \diamond$$

The proof of Theorem 4.1 is very technical, so it will be decomposed into several lemmas. For any $\lambda$ belonging to the half-plane $\left\{\lambda \in \mathbb{C} \text{ such that } Re\lambda > -\lambda^*\right\}$, we have

$$K(\lambda - T_H)^{-1}K = KB_\lambda H(I - M_\lambda H)^{-1}G_\lambda K + KC_\lambda K.$$

*Remark 4.1.* According to Lemma 3 and Remark 2.1, it suffices to establish the result for a one rank collision operator $K$ with kernel in the form

$$\kappa(x, v, v') = \alpha_1(x)\, \beta_1(v)\, \theta_1(v')$$

where $\alpha_1(.) \in L^\infty(D)$, while $\beta_1(.)$ and $\theta_1(.)$ are measurable simple functions with compact supports in $V$. $\diamond$

**Lemma 4.** *According to the hypotheses of Theorem 4.1, we have*

$$\lim_{|Im\lambda| \to +\infty} |Im\lambda|^r \, \|KB_\lambda H(I - M_\lambda H)^{-1}G_\lambda K\| = 0 \text{ uniformly on } R_w.$$

$\diamond$

*Proof.* It is shown in [8] that

$$B_\lambda H(I - M_\lambda H)^{-1}G_\lambda = \sum_{n \geq 0} \mathscr{J}_n(\lambda),$$

where $\left[\mathscr{J}_{2n+1}(\lambda)\varphi\right](x, v) = \int_0^{+\infty} \exp\{-\lambda t\} \, \left[U_{2n+1}(t)\varphi\right](x, v)\, dt$

and

$$\left[\mathscr{J}_{2n}(\lambda)\varphi\right](x, v) = \int_0^{+\infty} \exp\{-\lambda t\} \, \left[U_{2n+2}(t)\varphi\right](x, v)\, dt \;\; \forall\, n \geq 0.$$

Where for any fixed $t \geq 0$, $[U_0(t)\varphi](x, v) = \varphi(x - tv, v)\, \exp\left\{-\Sigma(v)t\right\}\, \chi_{\{t < t_-(x,v)\}}$ $\forall\, (x, v) \in D \times V$. While for any $n \geq 0$

$$[U_{2n+2}(t)\varphi](x,v) = \gamma^{2n+2}\ \exp\Big\{-\Sigma(v)t\Big\}\ \chi_{I_{2n+1}(x,v)}(t)$$
$$\varphi(x-tv+(2n+2)\tau(x,v)v,v)$$

and

$$[U_{2n+1}(t)\varphi](x,v) = \gamma^{2n+1}\ \exp\Big\{-\Sigma(v)t\Big\}\ \chi_{I_{2n}(x,v)}(t)$$
$$\varphi(x+tv-2t_{-}(x,v)v-2n\tau(x,v)v,-v),$$

with

$$I_k(x,v) = \Big[k\tau(x,v)+t_{-}(x,v);\ \ (k+1)\tau(x,v)+t_{-}(x,v)\Big] \text{ for any } k\in\mathbb{N}.$$

So it suffices to establish the result for $\mathscr{J}_{2n+1}(\lambda)$ and $\mathscr{J}_{2n}(\lambda)$ for all $n \in \mathbb{N}$.

Let $\varphi \in X_p$,

$$\begin{aligned}
&\Big[K\mathscr{J}_{2n+1}(\lambda)K\varphi\Big](x,v)\\
&\quad = K\mathscr{J}_{2n+1}(\lambda)\Big(\alpha_1(x)\beta_1(v)\int_V \theta_1(w)\varphi(x,w)\,dw\Big)\\
&\quad = K\Big(\int_0^{+\infty}\gamma^{2n+1}\exp\{-\lambda t\}\ \exp\{-\Sigma(v)t\}\ \chi_{I_{2n}(x,v)}(t)\ \alpha_1(x+tv-2t_{-}(x,v)v\\
&\qquad -2n\tau(x,v)v)\beta_1(-v)\int_V \theta_1(w)\varphi(x+tv-2t_{-}(x,v)v-2n\tau(x,v)v,w)dw\,dt\Big)\\
&\quad = \alpha_1(x)\ \beta_1(v)\int_V \theta_1(w')\int_0^{+\infty}\gamma^{2n+1}\exp\{-\lambda t\}\exp\{-\Sigma(w')t\}\times\alpha_1(x+tw'\\
&\qquad -2t_{-}(x,w')w'-2n\tau(x,w')w')\ \chi_{I_{2n}(x,w')}(t)\ \beta_1(-w')\\
&\int_V \theta_1(w)\varphi(x+tw'-2t_{-}(x,w')w'-2n\tau(x,w')w',w)\ dw\ dt\ dw'.
\end{aligned}$$

As a result, we can see that

$$\Big[K\mathscr{J}_{2n+1}(\lambda)K\varphi\Big](x,v) = \Big[A_3A_2(\lambda)A_1\varphi\Big](x,v),$$

where

$$\begin{cases} A_1 : X_p \longrightarrow L_p(D)\\ \qquad \varphi \longrightarrow (A_1\varphi);\ (A_1\varphi)(x) = \alpha_1(x)\displaystyle\int_V \varphi(x,w)\theta_1(w)\,dw\ \forall\ x\in D,\end{cases}$$

$$\begin{cases} A_3 : L_p(D) \longrightarrow X_p\\ \qquad \psi \longrightarrow (A_3\psi);\ (A_3\psi)(x,v)=\gamma^{2n+1}\ \alpha_1(x)\ \beta_1(v)\ \psi(x)\ \ \forall\ (x,v)\in D\times V\end{cases}$$

and

$$\begin{cases} A_2(\lambda) : L_p(D) \longrightarrow L_p(D) \\ \qquad\qquad\quad \varphi \longrightarrow (A_2(\lambda)\varphi) \end{cases}$$

such that for all $x \in D$ we have

$$(A_2(\lambda)\varphi)(x) = \int_V \beta_1(-w')\, \theta_1(w') \int_0^{+\infty} \exp\Big\{ - (\lambda + \Sigma(w'))t\Big\}$$
$$\varphi(x + tw' - 2t_-(x, w')w' - 2n\tau(x, w')w')\ \chi_{I_{2n}(x,w')}(t)\ dt\ dw'.$$

$A_1$ and $A_3$ are bounded. Indeed

$$\int_D |(A_1\varphi)(x)|^p\ dx \le \|\alpha_1\|^p_{L^\infty(D)} \int_D \Big(\int_V |\varphi(x, w)|\ |\theta_1(w)|\ dw\Big)^p dx$$
$$\le \|\alpha_1\|^p_{L^\infty(D)}\ \ \|\theta_1\|^p_{L_q(V)}\ \ \|\varphi\|^p_{X_p}.$$

Consequently,

$$\|A_1\| \le \|\alpha_1\|_{L^\infty(D)}\ \ \|\theta_1\|_{L_q(V)}.$$

Furthermore,

$$\|A_3\| < \|\alpha_1\|_{L^\infty(D)}\ \ \|\beta_1\|_{L_p(V)}.$$

Since $A_1$ and $A_3$ are independent of $\lambda$, it suffices to show that

$$\lim_{|Im\lambda| \to +\infty} |Im\lambda|^r\ \|A_2(\lambda)\| = 0 \text{ uniformly on } R_w. \tag{4}$$

To do so, let $\varphi \in L_p(D)$,

$$(A_2(\lambda)\varphi)(x) = \int_V \beta_1(-w')\, \theta_1(w') \int_{2n\tau(x,w')+t_-(x,w')}^{(2n+1)\tau(x,w')+t_-(x,w')} \exp\Big\{ - (\lambda + \Sigma(w'))t\Big\}$$
$$\times \varphi(x + tw' - 2t_-(x, w')w' - 2n\tau(x, w')w')\ \ dt\ dw'.$$

The change of variable $s = t - 2n\tau(x, w') - t_-(x, w')$ gives

$$(A_2(\lambda)\varphi)(x)$$
$$= \int_V \beta_1(-w')\, \theta_1(w') \int_0^{\tau(x,w')} \exp\Big\{ - (\lambda + \Sigma(w'))(s + 2n\tau(x, w') + t_-(x, w'))\Big\}$$
$$\varphi(x + sw' - t_-(x, w')w')\ \ ds\ dw'.$$

Set $\mu(w') = \lambda + \Sigma(w')$ and make the change of variable $t = s - t_-(x, w')$, then we get

$$(A_2(\lambda)\varphi)(x) = \int_V h(w') \exp\Big\{ - \mu(w')\Big(2n\tau(x, w') + 2t_-(x, w')\Big)\Big\} \times \int_{-t_-(x,w')}^{t_+(x,w')} \exp\Big\{ - \mu(w')t\Big\} \varphi(x + tw') \; dt \, dw'.$$

Since for all $(x, w') \in D \times V$ we have

$$t \in (-t_-(x, w'), t_+(x, w')) \iff y = x + tw' \in D,$$

so the change of variable $y = x + tw'$ gives

$$\begin{aligned}(A_2(\lambda)\varphi)(x) &= \int_D \int_{-t_-(x, \frac{y-x}{t})}^{t_+(x, \frac{y-x}{t})} h\Big(\frac{y-x}{t}\Big) \exp\Big\{ - \mu\Big(\frac{y-x}{t}\Big)\Big(t + 2n\tau\Big(x, \frac{y-x}{t}\Big) \\ &\quad + 2t_-(x, \frac{y-x}{t})\Big)\Big\} \varphi(y) \; \frac{dt}{|t^N|} \, dy \\ &= \int_D \kappa(\lambda, x, y) \, \varphi(y) \, dy\end{aligned}$$

where

$$\begin{aligned}\kappa(\lambda, x, y) &= \int_{\mathbb{R}} h\Big(\frac{y-x}{t}\Big) \exp\Big\{ - \mu\Big(\frac{y-x}{t}\Big)\Big(t + 2n\tau\Big(x, \frac{y-x}{t}\Big) \\ &\quad + 2t_-(x, \frac{y-x}{t})\Big)\Big\} \times \; \chi_{\left(-t_-(x, \frac{y-x}{t}),\; t_+(x, \frac{y-x}{t})\right)}(t) \; \frac{dt}{|t^N|}.\end{aligned}$$

Notice that the very rough estimate

$$\|A_2(\lambda)\| \leq \; \Big(\int_D \Big(\int_D |\kappa(\lambda, x, y)|^q \, dy\Big)^{\frac{p}{q}} \, dx\Big)^{\frac{1}{p}}$$

apparently does not lead to (4). We have to estimate the norm of $A_2(\lambda)$ more carefully.

The difficulty in estimating $\|A_2(\lambda)\|$ is that $A_2(\lambda)$ is not a convolution operator. To overcome this difficulty, we set

$$\begin{aligned}N_\lambda(x, z) &= \int_{\mathbb{R}} h\Big( - \frac{x}{t}\Big) \exp\Big\{ - \mu\Big( - \frac{x}{t}\Big)\Big(t + 2n\tau\Big(x + z, -\frac{x}{t}\Big) \\ &\quad + 2t_-(x + z, -\frac{x}{t})\Big)\Big\} \times \; \chi_{\left(-t_-(x+z, -\frac{x}{t}),\; t_+(x+z, -\frac{x}{t})\right)}(t) \; \frac{dt}{|t^N|},\end{aligned}$$

where $(x,z) \in D \times D$ with $x+z \in D$. This approach is inspired from [8]. Therefore

$$G_\lambda(x) = \sup_{z \in D-x} |N_\lambda(x,z)|.$$

Observe that $N_\lambda(x-y,y) = \kappa(\lambda,x,y)$ and denote by $\overline{G}_\lambda$ (resp. $\overline{\varphi}$) the trivial extension to $\mathbb{R}^N$, then we have

$$\begin{aligned}
|(A_2(\lambda)\varphi)(x)| &\leq \int_D |N_\lambda(x-y,y)|\, |\varphi(y)|\, dy \\
&\leq \int_D G_\lambda(x-y)\, |\varphi(y)|\, dy \\
&\leq (\overline{G}_\lambda * |\overline{\varphi}|)(x).
\end{aligned}$$

This yields

$$\begin{aligned}
\int_{\mathbb{R}^N} |\overline{A_2(\lambda)\varphi}(x)|^p\, dx &\leq \int_{\mathbb{R}^N} |(\overline{G}_\lambda * |\overline{\varphi}|)(x)|^p\, dx \\
\|\overline{A_2(\lambda)\varphi}\|_{L_p(\mathbb{R}^N)} &\leq \|\overline{G}_\lambda * |\overline{\varphi}|\|_{L_p(\mathbb{R}^N)} \\
&\leq \|\overline{G}_\lambda\|_{L_1(\mathbb{R}^N)}\, \|\overline{\varphi}\|_{L_p(\mathbb{R}^N)}.
\end{aligned}$$

Consequently

$$\begin{aligned}
\|A_2(\lambda)\varphi\|_{L_p(D)} &\leq \|G_\lambda\|_{L_1(D)}\, \|\varphi\|_{L_p(D)}, \\
\|A_2(\lambda)\| &\leq \|G_\lambda\|_{L_1(D)}.
\end{aligned}$$

Using Lemma 2, we obtain

$$\lim_{|Im\lambda| \to +\infty} |Im\lambda|^r \|G_\lambda\|_{L_1(D)} = 0 \text{ uniformly on } R_w.$$

So

$$\lim_{|Im\lambda| \to +\infty} |Im\lambda|^r\, \|K \mathscr{J}_{2n+1}(\lambda) K\| = 0 \text{ uniformly on } R_w.$$

A similar reasoning allows us to prove that

$$\lim_{|Im\lambda|\to+\infty} |Im\lambda|^r \; \|K\mathscr{J}_{2n}(\lambda)K\| = 0 \text{ uniformly on } R_w.$$

This ends the proof of Lemma 4.

**Lemma 5.** *With same hypotheses as Theorem 4.1, we have*

$$\lim_{|Im\lambda|\to+\infty} |Im\lambda|^r \; \|KC_\lambda K\| = 0 \textit{ uniformly on } R_w.$$

◇

*Proof.* Consider the sequence of operators $(C_{\lambda,\varepsilon_n})_{n\in\mathbb{N}}$, where

$$(C_{\lambda,\varepsilon_n}\varphi)(x,v) = \int_{\varepsilon_n}^{t_-(x,v)} \varphi(x-sv,v) \; \exp\Big\{-\big(\lambda+\Sigma(v)\big)s\Big\}\, ds.$$

This sequence converges to $C_\lambda$ uniformly on $R_w$ when $\varepsilon_n$ goes to zero. Hence, it suffices to prove that, for $\varepsilon > 0$

$$\lim_{|Im\lambda|\to+\infty} |Im\lambda|^r \; \|KC_{\lambda,\varepsilon}K\| = 0 \text{ uniformly on } R_w.$$

Let $\varphi \in X_p$, an immediate calculation shows that

$$\begin{aligned}\Big[KC_{\lambda,\varepsilon}K\varphi\Big](x,v) &= \alpha_1(x)\beta_1(v)\int_V \theta_1(w')\int_\varepsilon^{t_-(x,w')} \alpha_1(x-sw')\beta_1(w') \\ &\qquad \int_V \theta_1(w)\varphi(x-sw',w)\times\exp\Big\{-(\lambda+\Sigma(w'))s\Big\}\, dw\, ds\, dw' \\ &= \Big[A_3'A_2'(\lambda)A_1'\varphi\Big](x,v),\end{aligned}$$

where

$$\begin{cases} A_1' : X_p \longrightarrow L_p(D) \\ \qquad \varphi \longrightarrow (A_1'\varphi);\; (A_1'\varphi)(x) = \alpha_1(x)\displaystyle\int_V \varphi(x,w)\theta_1(w)\, dw \;\forall\; x\in D, \end{cases}$$

$$\begin{cases} A_3' : L_p(D) \longrightarrow X_p \\ \qquad \varphi \longrightarrow (A_3'\varphi);\; (A_3'\varphi)(x,v) = \alpha_1(x)\;\beta_1(v)\;\varphi(x) \;\;\forall\; (x,v)\in D\times V \end{cases}$$

and

$$\begin{cases} A_2'(\lambda) : L_p(D) \longrightarrow L_p(D) \\ \qquad\qquad \varphi \longrightarrow (A_2'(\lambda)\varphi) \end{cases}$$

such that $(A_2'(\lambda)\varphi)(x) = \int_V \theta_1(w')\ \beta_1(w') \int_\varepsilon^{t_-(x,w')} \varphi(x - sw')\ \exp\Big\{ - (\lambda + \Sigma(w'))s\Big\}\ ds\, dw'$.

$A_1'$ and $A_3'$ are bounded and independent of $\lambda$, so it suffices to show that

$$\lim_{|Im\lambda| \to +\infty} |Im\lambda|^r\ \|A_2'(\lambda)\| = 0 \text{ uniformly on } R_w.$$

Let $\varphi \in L_p(D)$,

$$(A_2'(\lambda)\varphi)(x) = \int_V \int_\varepsilon^{t_-(x,w')} h_1(w')\ \varphi(x - sw')\ \exp\Big\{ - (\lambda + \Sigma(w'))s\Big\}\ ds\, dw',$$

where $h_1(w') = \theta_1(w')\ \beta_1(w')$.

We have

$$s \in [\varepsilon, t_-(x, w')[ \Longleftrightarrow y = x - sw' \in D.$$

The change of variable $y = x - sw' \in D$ gives

$$(A_2'(\lambda)\varphi)(x) = \int_D \int_\varepsilon^{t_-(x, \frac{x-y}{s})} h_1\Big(\frac{x-y}{s}\Big)\ \varphi(y)\ \exp\Big\{ - \Big(\lambda + \Sigma\Big(\frac{x-y}{s}\Big)\Big)s\Big\} \frac{ds}{|s^N|}\ dy.$$

According to Lemma 3, it remains to show that, for all $n \in \mathbb{N}$,

$$\lim_{|Im\lambda| \to +\infty} |Im\lambda|^r\ \|A_{2,n}'(\lambda)\| = 0 \text{ uniformly on } R_w,$$

where

$$(A_{2,n}'(\lambda)\varphi)(x) = \int_D \int_\varepsilon^{t_-(x, \frac{x-y}{s})} h_1\Big(\frac{x-y}{s}\Big)\ \varphi(y)$$
$$\exp\Big\{ - \Big(\lambda - \frac{\omega}{2} + \lambda^*\Big)s\Big\}\tilde{\phi}_{x-y,n}(s) ds\, dy.$$

Set

$$\begin{cases} f_{x,y} : [\varepsilon, +\infty[ \longrightarrow \mathbb{R} \\ \qquad\qquad s \longrightarrow h_1\Big(\frac{x-y}{s}\Big)\ \tilde{\phi}_{x-y,n}(s)\ \chi_{[\varepsilon, t_-(x, \frac{x-y}{s})[}(s) \end{cases}$$

is a simple function. Let $(s_i)_{1\le i\le m}$ denote a subdivision of its support satisfying $f_{x,y}(s)=f_{x,y}(s_i)$, for all $s\in[s_i,s_{i+1}[$ with $i\in\{1,\dots,m-1\}$. So,

$$\begin{aligned}&\int_{\varepsilon}^{+\infty} f_{x,y}(s)\ \exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)s\Big\}\,ds\\ &\quad=\sum_{i=1}^{m-1} f_{x,y}(s_i)\int_{s_i}^{s_{i+1}}\exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)s\Big\}ds\\ &\quad=\frac{1}{\big(\lambda+\lambda^*-\frac{w}{2}\big)}\sum_{i=1}^{m-1} f_{x,y}(s_i)\Big(\exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)s_i\Big\}\\ &\qquad-\exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)s_{i+1}\Big\}\Big).\end{aligned}$$

Since

$$\begin{cases}\Big|\exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)s_i\Big\}-\exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)s_{i+1}\Big\}\Big|\le 2\\[2mm] \dfrac{1}{\Big|\lambda+\lambda^*-\frac{w}{2}\Big|}\le\dfrac{1}{|Im\lambda|}\\[2mm] |f_{x,y}(s_i)|\le\sup|h_1(.)|\ \sup|\tilde{\phi}(.)|,\end{cases}$$

then

$$\Big|\int_{\varepsilon}^{+\infty} f_{x,y}(s)\ \exp\Big\{-\Big(\lambda+\lambda^*-\frac{w}{2}\Big)s\Big\}\ ds\Big|\le\frac{2(m-1)}{|Im\lambda|}\sup|\tilde{\phi}(.)|\sup|h_1(.)|.$$

This yields

$$\Big|(A'_{2,n}(\lambda)\varphi)(x)\Big|^p\le\Big(\frac{2(m-1)}{|Im\lambda|}\sup|h_1(.)|\sup|\tilde{\phi}(.)|\Big)^p\Big(\int_D|\varphi(y)|dy\Big)^p.$$

Therefore

$$\begin{aligned}&|Im\lambda|^{pr}\int_D\Big|(A'_{2,n}(\lambda)\varphi)(x)\Big|^p\,dx\\ &\quad\le\Big(\frac{2(m-1)}{|Im\lambda|^{1-r}}\sup|h_1(.)|\ \sup|\tilde{\phi}(.)|\Big)^p\Big(\int_D(|\varphi(y)|^p+1)dy\Big)^p\int_D dx.\end{aligned}$$

Consequently,

$$|Im\lambda|^r\ \|A'_{2,n}(\lambda)\|\le\frac{2(m-1)}{|Im\lambda|^{1-r}}\ \sup|h_1(.)|\ \sup|\tilde{\phi}(.)|\Big(1+\int_D dy\Big)\Big(\int_D dx\Big)^{\frac{1}{p}}.$$

## 5 Asymptotic Behavior of the solution

In this section we will discuss the part of the spectrum of the transport operator $A_H$ in the half-plane $\left\{\lambda \in \mathbb{C} : Re\lambda > -\lambda^*\right\}$, such that

$$P(A_H) = \sigma(A_H) \cap \left\{\lambda \in \mathbb{C} :\ Re\lambda > -\lambda^*\right\}.$$

**Lemma 6.** *Let $p \in [1, +\infty[$ and assume that the collision operator $K$ is regular on $X_p$. If $H$ is weakly compact operator, then*

*(i) $P(A_H)$ consists of, at most, isolated eigenvalues with finite algebraic multiplicity.*
*(ii) If $w > 0$, then the set $\sigma(A_H) \cap R_w$ is finite.*
*(iii) If $w > 0$, then $\|(\lambda - A_H)^{-1}\|$ is uniformly bounded on $R_w$ for large $|Im\lambda|$.* ◇

*Proof.* Let $\lambda$ such that $Re\lambda > -\lambda^*$. Since $K$ is regular, then the use of Theorems 2.1 and 2.2 implies the compactness of $\left[(\lambda - T_H)^{-1}K\right]^4$ on $X_p$ for $1 \leq p < \infty$. Furthermore, applying Theorem 4.1 for $r = 0$, we obtain

$$\lim_{|Im\lambda| \to +\infty} \left\| \left[(\lambda - T_H)^{-1}K\right]^4 \right\| = 0 \text{ uniformly on } R_w.$$

We have $(i)$, $(ii)$, and $(iii)$ follow immediately from Lemma 1.1 in [9].

In all the sequel, we denote by $T_{H,p}$ and $A_{H,p}$, the streaming operator and the transport operator on the space $X_p$, respectively. Then, we designate by $\sigma_p(.)$ the point spectrum and $r_\sigma(.)$ the spectral radius.

**Lemma 7.** *Let $\lambda \in \mathbb{C}$ such that $Re\lambda > -\lambda^*$ and $p \in [1, +\infty[$. Assume that the collision operator $K$ is regular on $X_p$. If $H$ is weakly compact operator, then*

*(i) $\lambda \in \sigma_p(A_{H,p})$ if and only if $1 \in \sigma_p((\lambda - T_{H,p})^{-1}K)$ and the corresponding eigensubspaces coincide.*
*(ii) $\lambda \in \rho(A_{H,p})$ if and only if $1 \in \rho((\lambda - T_{H,p})^{-1}K)$.* ◇

The above lemma can be checked as the same way as [1, Lemma 7.2]. Now, we state the following result:

**Lemma 8.** *Let $E := \left\{\lambda \in \mathbb{C} : Re\lambda > -\lambda^*\right\}$ and $p \in [1, +\infty[$. Then, we have*

$$r_\sigma\left[\left((\lambda - T_{H,p})^{-1}K\right)^4\right] < 1 \textit{ for some } \lambda \in E.$$

◇

*Proof.* Let $\lambda \in E$, so we can write

$$\left\| \left[ (\lambda - T_{H,p})^{-1} K \right]^4 \right\| \leq \left( \frac{\|K\|}{Re\lambda + \lambda^*} \right)^4.$$

Then, there exists $\alpha > 0$ such that for every $\lambda$ satisfying $Re\lambda > \alpha$ we have:

$$r_\sigma \left[ \left( (\lambda - T_{H,p})^{-1} K \right)^4 \right] < 1.$$

In the remainder of this section, we denote by $P_{\lambda,p}$ the spectral projection associated with $\lambda$ which is an eigenvalue of $A_{H,p}$ with finite algebraic multiplicity. Let $\delta > 0$ be such that $\{\mu \in \mathbb{C} : 0 < |\lambda - \mu| \leq \delta\} \cap P(A_{H,p}) = \emptyset$, so

$$P_{\lambda,p} = \frac{1}{2i\pi} \int_{|\mu-\lambda|=\delta} (\mu - A_{H,p})^{-1} \, d\mu.$$

**Lemma 9.** *The spectral projection $P_{\lambda,p}$ satisfying for every $k \geq 1$,*

$$N[(\lambda - A_{H,p})^k] \subseteq R(P_{\lambda,p}),$$

*where $N[(\lambda - A_{H,p})^k]$ and $R(P_{\lambda,p})$ designate, respectively, the null space of $(\lambda - A_{H,p})^k$ and the range of $P_{\lambda,p}$.* ◇

*Proof.* We argue by recurrence, we have:
for $k = 1$, $N[(\lambda - A_{H,p})] \subseteq R(P_{\lambda,p})$.
We assume that the inclusion is true until the order $k - 1$ and we prove that it remains true to the order $k$. Let $f \in N[(\lambda - A_{H,p})^k] \backslash N[(\lambda - A_{H,p})]$, where $N[(\lambda - A_{H,p})]$ is a subspace of finite dimension on $X_p$. Then, there is a closed subspace $H$ satisfying $N(\lambda - A_{H,p}) \oplus H = N[(\lambda - A_{H,p})^k]$. Set, $f = f_1 + f_H$. Since $(\lambda - A_{H,p})f \in N[(\lambda - A_{H,p})^{k-1}]$ then $P_{\lambda,p}[(\lambda - A_{H,p})f_H] = (\lambda - A_{H,p})f_H$. Therefore, $P_{\lambda,p}f_H - f_H \in R(P_{\lambda,p})$. Consequently, $f_H \in R(P_{\lambda,p})$.

Now, we are ready to prove the following result.

**Theorem 5.1.** *Let $p \in [1, +\infty[$ and assume that the collision operator $K$is regular on $X_p$. If $H$ is weakly compact, then:*

*(i)* *$P(A_{H,p})$ is independent of $p$.*
*(ii)* *For each $\lambda \in P(A_{H,p})$ and $k \in \mathbb{N}^*$, $N[(\lambda - A_{H,p})^k]$ is independent of $p$.* ◇

*Proof.* Recall that $E := \left\{ \lambda \in \mathbb{C} : Re\lambda > -\lambda^* \right\}$. Let $\lambda \in E$ and $K$ is regular operator. Using Lemma 8, we obtain that $I - [(\lambda - T_{H,p})^{-1}K]^4$ is invertible for some $\lambda \in E$. We know from Theorems 2.1 and 2.2 that the operator $[(\lambda - T_{H,p})^{-1}K]^4$ is compact on $X_p$ for $p \in [1, +\infty[$. Since $E$ is connected open subset of $\mathbb{C}$, applying the Gohberg-Shmul'yan theorem [5, Theorem 11.4, p. 258] we infer that, for $p \geq 1$, $\left[ I - \left( (\lambda - T_{H,p})^{-1} K \right)^4 \right]$ is invertible on $E$ except a set $S_p$ of isolated points which are poles of finite orders. Since

$$\left[I-(\lambda-T_{H,p})^{-1}K\right]^{-1}$$
$$=\left[I+(\lambda-T_{H,p})^{-1}K\right]\left[I+((\lambda-T_{H,p})^{-1}K)^2\right]\left[I-((\lambda-T_{H,p})^{-1}K)^4\right]^{-1},$$

the function $\left[I-(\lambda-T_{H,p})^{-1}K\right]^{-1}$ has a similar behavior as $\left[I-((\lambda-T_{H,p})^{-1}K)^4\right]^{-1}$ in the half-plane $Re\lambda > -\lambda^*$. Using Lemma 7, we obtain that $P(A_{H,p}) = S_p$ for every $p \geq 1$. In [1], it is shown that $S_p$ is independent of $p$, hence the proof of this assertion is completed.

$(ii)$ Since

$$(\lambda-A_{H,p})^{-1} = [I-(\lambda-T_{H,p})^{-1}K]^{-1}\,(\lambda-T_{H,p})^{-1},$$

then $P_{\lambda,1}\diagup X_p = P_{\lambda,p}$. Furthermore, $\overline{P_{\lambda,p}(C_0^\infty(D\times V))} = R(P_{\lambda,p})$ this implies that $R(P_{\lambda,p}) = P_{\lambda,p}(C_0^\infty(D\times V))\ \forall\ p \geq 1$. Thus, the use of $R(P_{\lambda,1}) = R(P_{\lambda,p})$ and Lemma 9 leads to $N[(\lambda-A_{H,1})^k] \subseteq X_p$. Consequently, $N[(\lambda-A_{H,p})^k] = N[(\lambda-A_{H,1})^k] \subseteq X_p$.

Since $T_H$ is an infinitesimal generator of a $C_0$-semigroup $(U(t))_{t\geq 0}$ acting on $X_p$, $p \in [1,+\infty[$ and $K$ is a bounded linear operator, then by the classical perturbation theory [6, Theorem 2.1, p. 495], the operator $A_H = T_H + K$ generates also a $C_0$-semigroup $(V(t))_{t\geq 0}$ on $X_p$ given by the Dyson-Phillips expansion

$$V(t) = \sum_{j=0}^{n-1} U_j(t) + R_n(t),$$

where $U_0(t) = U(t), U_j(t) = \int_0^t U(s)KU_{j-1}(t-s)\,ds$, $j = 1,2,\ldots,$ and the $n^{th}$ order remainder term $R_n(t)$ can be expressed by

$$R_n(t) = \int_{s_1+\ldots s_n\leq t, s_i\geq 0} U(s_1)K\ldots U(s_n)KV(t-s_1-\ldots-s_n)\,ds_1\ldots ds_n.$$

We suppose that $K$ is regular on $X_p$ and $H$ is a weakly compact operator. Then, from Lemma 6 $(i)$, the asymptotic spectrum $P(A_H)$ consists of, at most, isolated eigenvalues with finite algebraic multiplicity $\{\lambda_1,\lambda_2,\ldots,\lambda_n,\lambda_{n+1},\ldots\}$ which can be ordered in such way that the real part decreases [6], i.e., $Re\lambda_1 > Re\lambda_2 > \ldots > Re\lambda_n > Re\lambda_{n+1} > \ldots > -\lambda^*$, and $\{\lambda \in \mathbb{C}$ such that $Re\lambda > -\lambda^*\} \setminus \{\lambda_n : n = 1,2,3,\ldots\} \subset \rho(A_H)$.

Now, using the spectral decomposition theorem corresponding to the set $\{\lambda_1,\lambda_2,\ldots,\lambda_n\}$ and $\sigma(A_H) \setminus \{\lambda_1,\lambda_2,\ldots,\lambda_n\}$ which was established in [10], we can write

$$V(t) = \tilde{V}(t) + \sum_{i=1}^{n} e^{\lambda_i t} e^{D_i t} P_i,$$

where $P_i$ and $D_i$ denote, respectively, the spectral projection and the nilpotent operator associated with $\lambda_i$, $i = 1, 2, \ldots, n$. $P = \sum_{i=1}^{n} P_i$ is the spectral projection of the compact set $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ which commutes with $V(t)$ and $\tilde{V}(t) := V(t)(I-P)$ is a $C_0$-semigroup on the Banach space $(I-P)X_p$ with generator $\tilde{A}_H = A_H(I-P)$.

In the same way as in [1], we prove the following lemma.

**Lemma 10.** *Let $K$ be a regular collision operator. If $H$ is weakly compact, then for any $\varepsilon > 0$ we have:*

*(i) For all $q \in \{0, \ldots, 3\}$, $\left\| \left[(\lambda - T_H)^{-1}K\right]^q \right\|$ is uniformly bounded on $\left\{\lambda \in \mathbb{C} : Re\lambda \geq Re\lambda_{n+1} + \varepsilon\right\}$.*

*(ii) $\|(\lambda - A_H)^{-1}(I-P)\|$ is uniformly bounded on $\left\{\lambda \in \mathbb{C} : Re\lambda \geq Re\lambda_{n+1} + \varepsilon\right\}$.* ◇

Now, we will prove the main results of this paper.

**Theorem 5.2.** *Let $p = 1$ and assume that the hypotheses of Lemma 10 are satisfied, then for each $\varepsilon > 0$, there exists $M > 0$ such that*

$$\left\| V(t) - \sum_{i=1}^{n} e^{\lambda_i t} e^{D_i t} P_i \right\| \leq M\, e^{(Re\lambda_{n+1} + \varepsilon)t}, \quad \forall\, t > 0. \qquad \diamond$$

*Proof.* From Proposition 1, the type of $U(t)$ is less than or equal to $-\lambda^*$. Let us first recall that by Theorems 2.1 and 2.2 we obtain $\left[(\lambda - T_H)^{-1}K\right]^4$ is compact on $X_p$ for $1 \leq p < \infty$. Furthermore, inspired on [1, Lemma 4.3], it is easy to check that $\{\lambda \in \mathbb{C}$ such that $Re\lambda \geq \|K\| - \lambda^* + 1\} \subset \rho(A_H)$ and

$$\|(\lambda - A_H)^{-1}\| \leq 1.$$

Now, according to Theorem 4.1, we have

$$\begin{cases} (i) \|(\lambda - A_H)^{-1}\| \text{ is uniformly bounded on } \left\{\lambda \in \mathbb{C} : Re\lambda \geq \|K\| - \lambda^* + 1\right\}, \\ (ii) \text{ a real } r_0 \text{ and for } w > -\lambda^*, \text{ there exists } C(w) \text{ such that} \\ |Im\lambda|^{r_0} \left\| \left[(\lambda - T_H)^{-1}K\right]^4 \right\| \text{ is uniformly bounded on } \left\{\lambda \in \mathbb{C}; Re\lambda \geq w,\ |Im\lambda| \geq C(w)\right\} \end{cases}$$

Hence, the result follows from Lemma 10 and [1, Theorem 1.1].

Let $r$ be the real defined by

$$r := s(A_H) - \sup\Big\{Re\lambda : \lambda \in \sigma(A_H),\ \lambda \neq s(A_H)\Big\}.$$

By virtue of Lemma 6, we can see that $r > 0$. Let $\tilde{P}$ denote the projection operator corresponding to $\Big\{\lambda \in \sigma(A_H) : Re\lambda = s(A_H)\Big\}$ which is compact. Therefore, the spectral decomposition theorem [10, Theorem 3.3, p. 70] can be applied.

We close this paper with the following theorem which can be checked in a similar way as [1, Theorem 1.2].

**Theorem 5.3.** *Let $K$ be a regular collision operator on $X_p$, $1 \leq p < \infty$. If $H$ is weakly compact operator, then*

*(i) For each $\varepsilon > 0$, there exists $M > 0$ such that*

$$\left\| V(t) - \sum_{i=1}^{n} e^{\lambda_i t} e^{D_i t} P_i \right\|_{X_p} \leq M\, e^{(Re\lambda_{n+1}+\varepsilon)t}, \quad \forall\, t > 0 \text{ and } p \in [1,2].$$

*(ii) Further, if $K$ is positive, then for each $p > 1$ there exists $M' > 0$ such that*

$$\left\| V(t)(I - \tilde{P}) \right\|_{X_p} \leq M'\, e^{(s(A_H)-\varepsilon)t}, \quad \forall\, t > 0$$

*and for every $\varepsilon \in (0, 2r(1-p^{-1}))$ (resp. $\varepsilon \in (0, 2rp^{-1})$) if $p \leq 2$ (resp. $p > 2$).*

◇

## References


1. B. Abdelmoumen, A. Jeribi and M. Mnif, *Time asymptotic description of the Solution to an abstract Cauchy problem and application to transport equation*, Math.Z., 268, 837–869, (2011).
2. S. Charfi, *On the time Asymptotic behavior of a transport operator with diffuse reflection boundary condition*, Transport Theory and Statistical Physics, 41, 7, 529–551, (2012).
3. S. Charfi, A. Jeribi, N. Moalla, *Time asymptotic behavior of the solution of an abstract Cauchy problem given by a one-velocity transport operator with Maxwell boundary condition*, in Collectanea Mathematica, 64, 97–109, (2013).
4. A. Jeribi, Spectral Theory and Applications of Linear Operators and Block Operator Matrices, Springer-Verlag, New-York, (2015)
5. H. G. Kapper, C. G. Lekkerkerker and J. Hejtmanex, *Spectral Methods in Transport Theory*, Birkhäuser Verlag, (1982).
6. T. Kato, *Perturbation theory for linear operators*, Berlin Heidelberg New York, Springer, (1966).
7. K. Latrach, *Compactness results for transport equations and applications*, Math. Models Methods Appl. Sci., 11, 1181–1202, (2001).

8. K. Latrach and B. Lods, *Spectral analysis of transport equations with bounce-back boundary conditions*, Math. Models Methods Appl. Sci., 11, 1325–1344, (2009).
9. M. Mokhtar-Kharroubi, *Time asymptotic behaviour and compactness in neutron transport theory,* Europ. J. Mech. B Fluid., 11, 39–68, (1992).
10. R. Nagel (ed.), *One-parameter semigroups of positive operators,* Lect. Notes Math., Springer Verlag, 1184, (1986).
11. A. Pazy, *Semigroups of linear operators and applications to differential equations,* Springre Verlag, (1983).
12. I. Vidav, *Spectra of perturbed semigroups with applications to transport theory,* J. Math. Anal. and Appl. 30, 264–279, (1970).

# Construction of MATLAB adaptative step ODE solvers

**E. Alberdi Celaya and J.J. Anza Aguirrezabala**

**Abstract** MATLAB software package offers a set of open source adaptative step functions for solving Ordinary Differential Equations (ODEs) which are easy to use by non-experts. Two of these functions are the well-known *ode45* and *ode15s*. The *ode45* is adequate when solving nonstiff problems, while the *ode15s* is recommended for stiff problems. Due to the wide utilization of MATLAB in science and engineering and taking into account that some issues of interest are found, we have studied the numerical methods in which these two ode solvers are based, describing the error estimation and the step size control implemented in the codes. First and second order linear ODEs are solved as two extreme examples to characterize the functions response and finally, some conclusions related to the stability of these solvers and the inefficiency of *ode15s* in stiff problems with pure imaginary eigenvalues are presented.

**Keywords** BDF • NDF • DOPRI(5 • 4) • Runge Kutta methods • stiff ODEs • stability • ode45 • ode15s

**2000 AMS subject classifications:** 65L05, 65L06, 65L20

E.A. Celaya (✉)
Department of Applied Mathematics, EUIT de Minas y Obras Públicas, Universidad del País Vasco UPV/EHU, Vizcaya, Spain
e-mail: elisabete.alberdi@ehu.es

J.J.A. Aguirrezabala
Department of Applied Mathematics, ETS de Ingeniería de Bilbao, Universidad del País Vasco UPV/EHU, Vizcaya, Spain
e-mail: juanjose.anza@ehu.es



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_23

## 1 Introduction

A set of codes to solve initial value problems (IVPs) [1, 12] is implemented in MATLAB:

$$y'(t) = f\left(t, y(t)\right), \quad y\left(t_0\right) = y_0 \tag{1}$$

Some of these ode solvers are recommended for nonstiff problems and others for stiff problems. The *ode45* based on the Dormand Prince (5,4) pair [3] is one of the codes recommended for nonstiff problems, and the *ode15s* for stiff problems, which uses the Backward Differentiation Formulae (BDF) [4] and the Numerical Differentiation Formulae (NDF) [14].

The term stiffness has been defined in different ways [2, 5, 7, 8, 10, 15], sometimes leading to confusion. The definition used in this work is given in [6, 9], which says that the greater the ratio of the eigenvalues of the Jacobian matrix, the more stiff becomes the system of ODEs.

We have considered the first order ODE given by:

$$y' = \lambda \cdot y, \;\; y(0) = 1, \;\; \text{where} \;\; \lambda < 0 \tag{2}$$

And the second order ODE:

$$y'' = \lambda^2 \cdot y, \;\; (y(0), y'(0)) = (1, 0), \;\; \text{where} \;\; \lambda^2 < 0 \tag{3}$$

which has been reduced to a system of two first order ODEs:

$$\begin{pmatrix} y' \\ z' \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \lambda^2 & 0 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix}, \quad (y(0), z(0)) = (1, 0) \tag{4}$$

being $z = y'$.

The eigenvalues of (2) and (4) are, respectively, $\lambda$ and $\pm |\lambda|\, i$, which represent two extreme cases: the problem with real eigenvalues and the one with pure imaginary eigenvalues. Problems (2) and (4) have been solved in the time interval $T = [0, 10]$ using the ode solvers *ode45* and *ode15s*, for different values of $\lambda$ and $\lambda^2$, with default values for relative and absolute tolerances. The number of steps given by each ode solver is listed in Tables 1 and 2, respectively.

For the smallest eigenvalues, the *ode45* takes less steps than the *ode15s* when solving problem (2); while the *ode15s* results more efficient (less steps) with the largest ones, as expected for a specialize function for stiff systems. Nevertheless, this does not happen in problem (4), where the *ode45* is more efficient than the *ode15s*, even for the large eigenvalues. In order to understand the low efficiency of the *ode15s* when solving second order stiff systems associated with vibratory or wave-type problems, this work studies the algorithms that support the functions *ode45* and *ode15s*: the methods on which they are based, the local error estimation,

**Table 1** Number of steps given by the *ode45* and the *ode15s* when solving problem (2).

| $\lambda$ | *ode45* | *ode15s* NDF | *ode15s* BDF |
|---|---|---|---|
| $-1$ | 13 | 42 | 42 |
| $-2$ | 19 | 55 | 55 |
| $-4$ | 24 | 62 | 64 |
| $-8$ | 36 | 68 | 70 |
| $-20$ | 73 | 75 | 74 |
| $-100$ | 314 | 80 | 78 |

**Table 2** Number of steps given by the *ode45* and the *ode15s* when solving problem (4).

| $\pm\lvert\lambda\rvert i$ | *ode45* | *ode15s* NDF | *ode15s* BDF |
|---|---|---|---|
| $\pm 1i$ | 18 | 46 | 51 |
| $\pm 2i$ | 31 | 81 | 82 |
| $\pm 4i$ | 57 | 151 | 146 |
| $\pm 8i$ | 109 | 291 | 281 |
| $\pm 20i$ | 262 | 711 | 689 |
| $\pm 100i$ | 1281 | 3515 | 3406 |

the implementation of the adaptative step size and the stability regions of each method as the eigenvalues multiplied by the step sizes may lie inside these regions in order to obtain an accurate solution. When a stiff ODE is solved using a numerical method with small stability region, the step sizes will be smaller in order to retain stability requirements [10].

Characteristics of the *ode45* and the *ode15s* have been studied in Sections 2 and 3, respectively. In Section 4, both codes have been used to solve the ODEs (2) and (3) and among other features, the step sizes, the error estimations, and the positions of the eigenvalues multiplied by the step sizes with respect to the stability regions have been analysed. Conclusions derived from this study are presented in Section 5.

## 2 The ode solver *ode45*

The ode solver *ode45* is based on an embedded Runge-Kutta method, the Dormad-Prince pair known as DOPRI(5,4) [3]. The Butcher table of an embedded Runge-Kutta method is given by Table 3.

The general s-stage Runge-Kutta method is given by:

$$y_{n+1} = y_n + h \sum_{i=1}^{s} b_i k_i \tag{5}$$

where the coefficients $k_i$ are given by:

$$k_i = f(t_n + c_i h, y_n + h \sum_{j=1}^{s} a_{ij} k_j), \quad i = 1, 2, \ldots, s \tag{6}$$

**Table 3** Butcher table of an embedded Runge-Kutta method.

| | |
|---|---|
| **c** | $A$ |
| | $\mathbf{b}^T$ |
| | $\hat{\mathbf{b}}^T$ |
| | $\mathbf{E}^T$ |

When the vector $\mathbf{b}^T$ of DOPRI(5,4) is used to calculate (5), the method is 4-order accurate, while the method obtained using $\hat{\mathbf{b}}^T$ is 5-order accurate. The values obtained with the 4th and 5th order methods will be denoted by $y_{n+1}$ and $\hat{y}_{n+1}$, respectively. It is the 5-order formula the one used by the *ode45* to advance:

$$y_{n+1} = y_n + h \sum_{i=1}^{s} b_i k_i, \quad \hat{y}_{n+1} = y_n + h \sum_{i=1}^{s} \hat{b}_i k_i \tag{7}$$

## 2.1 Error estimation in the ode45

The *ode45* uses the difference between the values obtained with the 5-order and the 4-order formulae as the local error estimation. This value results the 4th order method local truncation error [13] :

$$\begin{aligned} est &= \hat{y}_{n+1} - y_{n+1} \\ &= L_{n,4} + O\left(h^{5+1}\right) \end{aligned} \tag{8}$$

being $L_{n,4}$ the local truncation error of the 4th-order formula. By substituting the expressions of (7) in (8), the local error estimation results:

$$est = \hat{y}_{n+1} - y_{n+1} = h \sum_{i=1}^{s} \left(\hat{b}_i - b_i\right) k_i \tag{9}$$

## 2.2 Step size control in the ode45

Suppose that after having given a step of size $h$, we want to calculate the next step size. This step size could be for a new step or for the repetition of the previous step because the requirement of the tolerance has not been verified. In both cases, the next step size is calculated by multiplying the previous step size by a constant $\sigma$:

$$\sigma = \left(\frac{Rtol}{\|est\|}\right)^{1/5} \tag{10}$$

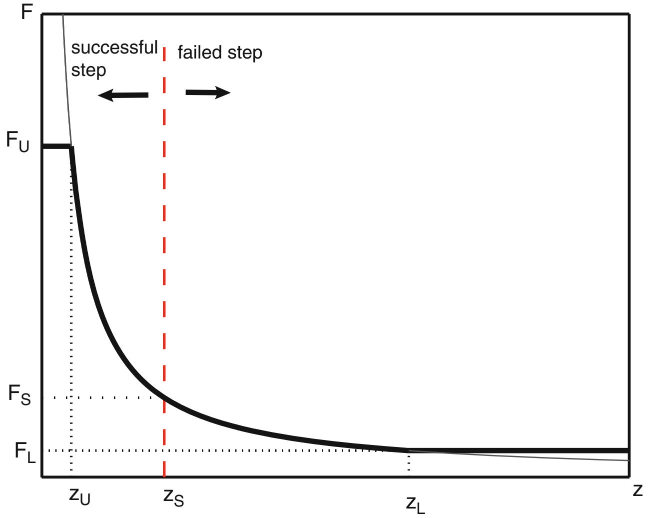


**Fig. 1** The factor $F$ as function of the variable $z$.

being *Rtol* the specified tolerance. A safety factor of $F_S = 0.8$ is used to guarantee that the next step will be under tolerance [13]. So the new step is calculated as follows:

$$h_{new} = F \cdot h \tag{11}$$

where $F = F_S \cdot \sigma = 0.8 \cdot \sigma$. Introducing a new variable $z$ defined by:

$$z = \frac{1}{0.8 \cdot \sigma} = 1.25 \left( \frac{\|est\|}{Rtol} \right)^{1/5} \tag{12}$$

the factor $F$ takes the form $F = 1/z$ which represents the hyperbola of Figure 1. In this figure two regions can be distinguished, separated by $\|est\| = Rtol$, where $z_S = 1.25$ is verified and which marks the limit between successful and failed steps. The factor $F$ takes the value of the safety factor, $F_S = 0.8$, when $z_S = 1.25$. In each step, the *ode45* computes the values of $\|est\|$ and $z$, and the step is considered successful if $z \leq z_S = 1.25$ is verified. Otherwise, the step is considered failed. When the given step has been successful, that is to say $z \in [0, 1.25]$, or equivalently, $F = \frac{1}{z} \in [0.8, \infty)$, an upper threshold $F_U$ is introduced to avoid excessively long steps, see Figure 1. The *ode45* uses the value $F_U = 5$. This makes $F \in [0.8, 5]$ and the next step size, $h_{new}$, verifies:

$$0.8 \cdot h \leq h_{new} \leq 5 \cdot h \tag{13}$$

Hence, the next step size is defined as:

$$h_{new} = \begin{cases} F_U \cdot h, & z \le z_U \\ F \cdot h, & z_U < z \le z_S \end{cases} \tag{14}$$

being $z_U = \frac{1}{F_U} = 0.2$, $z_S = \frac{1}{F_S} = 1.25$. When the given step results failed, that is to say $z > 1.25$, or equivalently, $F < 0.8$, the step is repeated. For the first trial a lower threshold $F_L$ is set to avoid excessively short steps, which in the *ode45* takes the value $F_L = 0.1$. This makes $F \in [0.1, 0.8]$ and the repetition of the step verifies:

$$0.1 \cdot h \le h_{new} \le 0.8 \cdot h \tag{15}$$

And the expression used in the first trial after a failed step is:

$$h_{new} = \begin{cases} F \cdot h, & z_S < z \le z_L \\ F_L \cdot h, & z > z_L \end{cases} \tag{16}$$

where $z_L = \frac{1}{F_L} = 10$, $z_S = \frac{1}{F_S} = 1.25$. The expressions of (16) are replaced by $h/2$ in second or posterior trials after an unsuccessful step.

## 3 The ode solver ode15s

The *ode15s* is based on the BDF methods [4], and it is possible to use the BDFs of orders $1 - 5$. The final *s* of the *ode15s* indicates that the algorithm is usually used to solve stiff differential equations [13]. By default *ode15s* uses NDF methods [14] which based on BDF methods, anticipate a backward difference of order $(k + 1)$ when working in order $k$. The code always starts solving in order $k = 1$, and the maximum order to be reached can be given to the code as data. In Figure 2 the stability regions of DOPRI (5,4), BDFs and NDFs can be seen.

### *3.1 Error estimation in the ode15s*

The *ode15s* uses the local truncation error as the error estimation:

$$est \approx LTE = Ch^{k+1}y^{k+1}(t_n) + O\left(h^{k+2}\right) \tag{17}$$

being $C$ the error constant of the method. Backward differences are used to calculate an approximation of $y^{k+1}(t_n)$. An approximation of $y(t)$ is obtained by using the

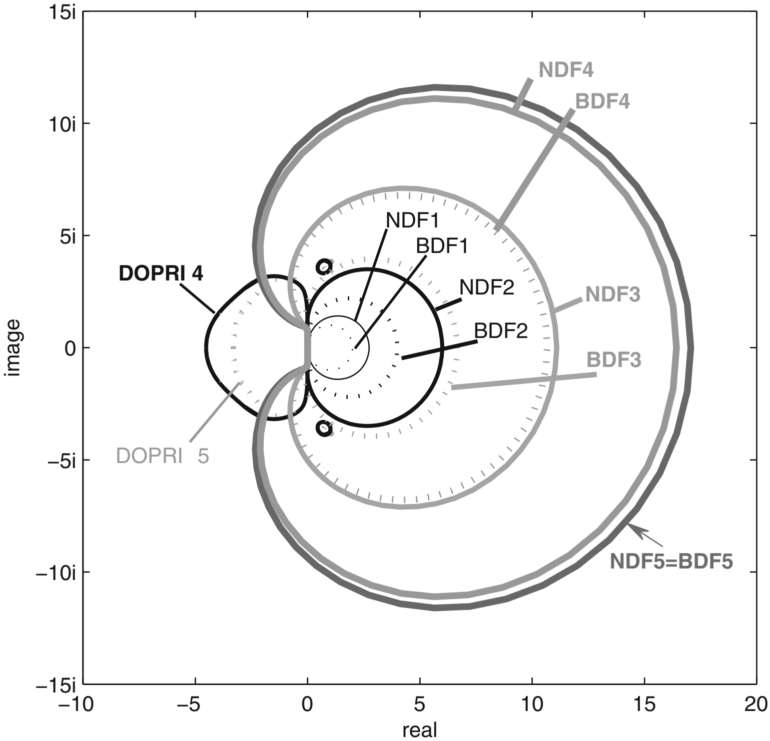


**Fig. 2** Stability regions of DOPRI(5,4) (interior to curves), BDFs and NDFs (exterior to curves).

backward interpolating polynomial of Newton that passes from the $(k+2)$ points $\{(t_{n+i}, y_{n+i})\}$ for $i = -1, 0, 1, 2, \ldots, k$:

$$y(t) \approx Q(t) = y_{n+k} + \sum_{j=1}^{k+1} \nabla^j y_{n+k} \frac{1}{j!h^j} \prod_{m=0}^{j-1} (t - t_{n+k-m}) \tag{18}$$

And the $(k+1)$th derivative of expression (18) is calculated:

$$Q^{(k+1)}(t) = \nabla^{k+1} y_{n+k} \frac{1}{(k+1)!h^{k+1}} (k+1)! = \nabla^{k+1} y_{n+k} \frac{1}{h^{k+1}} \tag{19}$$

Obtaining:

$$y^{(k+1)}(t) \approx Q^{(k+1)}(t) = \nabla^{k+1} y_{n+k} \frac{1}{h^{k+1}} \tag{20}$$

Substituting the approximation (20) in (17), the error estimation of the *ode15s* in terms of backward differences is obtained:

$$LTE \approx C \cdot \nabla^{k+1} y_{n+k} = est \tag{21}$$

### 3.2 Step size control in the ode15s

The *ode15s* is not allowed to change either the order or the step size until a minimum of $(k+2)$ consecutive steps are given with the same order and step size. If one of the steps results failed, the order of the method or the step size is reduced. When the compulsory $(k+2)$ successful steps are given, it is possible to change the order and the step size. In this case, the step sizes which correspond to orders $(k-1)$ (for $k > 1$), $k$ and $(k+1)$ (whenever the maximum order defined has not been reached) are calculated. If we are solving in order $k$, the *ode15s* calculates the next step $h_k$ which corresponds to order $k$ by multiplying the actual step size by a constant $\sigma$:

$$\sigma = \left( \frac{Rtol}{\|est\|} \right)^{\frac{1}{k+1}} \tag{22}$$

The *ode15s* uses a safety factor $F_S = \frac{5}{6}$ and the new step size is:

$$h_k = F \cdot h \tag{23}$$

where $F = F_S \cdot \sigma = \frac{5}{6} \cdot \sigma$. Again, a new variable $z$ is defined as follows:

$$z = 1.2 \left( \frac{\|est\|}{Rtol} \right)^{1/(k+1)} \tag{24}$$

resulting the factor $F = 1/z$. When the given step is successful, that is to say $z \in [0, 1.2]$, or equivalently $F \in \left[\frac{5}{6}, \infty\right]$, an upper threshold $F_U = 10$ is set and the step size of the $k$-order method is defined as:

$$h_k = \begin{cases} F_U \cdot h, & z \leq z_U \\ F \cdot h, & z_S < z \leq z_U \end{cases} \tag{25}$$

where $F_U = 10$, $z_U = \frac{1}{F_U} = 0.1$ and $z_S = 1.2$. In a similar way, the step sizes $h_{k-1}$ and $h_{k+1}$ which correspond to orders $(k-1)$ and $(k+1)$ are calculated, being the safety factors $\frac{10}{13}$ and $\frac{10}{14}$, respectively. The error estimations of the methods of order $(k-1)$ and $(k+1)$ are calculated, $\|est_{k-1}\|$ and $\|est_{k+1}\|$. And the actual step size is multiplied by the factor $F_{k-1} = \frac{1}{z_{k-1}}$ or $F_{k+1} = \frac{1}{z_{k+1}}$, respectively, where:

$$z_{k-1} = 1.3 \cdot \left( \frac{\|est_{k-1}\|}{Rtol} \right)^{1/k}, \quad z_{k+1} = 1.4 \cdot \left( \frac{\|est_{k+1}\|}{Rtol} \right)^{1/(k+2)} \tag{26}$$

Depending on the values of $z_{k-1}$ and $z_{k+1}$, the step sizes $h_{k-1}$ and $h_{k+1}$ with upper threshold $F_{k-1,U} = F_{k+1,U} = 10$ are defined:

$$h_{k-1} = \begin{cases} F_{k-1,U} \cdot h, & z_{k-1} \leq z_{k-1,U} \\ F_{k-1} \cdot h, & z_{k-1,S} < z_{k-1} \leq z_{k-1,U} \end{cases} \tag{27}$$

$$h_{k+1} = \begin{cases} F_{k+1,U} \cdot h, & z_{k+1} \leq z_{k+1,U} \\ F_{k+1} \cdot h, & z_{k+1,S} < z_{k+1} \leq z_{k+1,U} \end{cases} \tag{28}$$

where $z_{k-1,U} = 0.1$ and $z_{k-1,S} = 1.3$ and $z_{k+1,U} = 0.1$ and $z_{k+1,S} = 1.4$. Once the step sizes that correspond to orders $k$, $(k-1)$ and $(k+1)$ are available, the process that the *ode15s* follows to set the next step size is:

1. It compares the step size of order $(k-1)$ and the one of order $k$. If $h_{k-1} > h_k$ is verified, the value $h_{k-1}$ is saved in $h_{new}$ and the order that corresponds to $h_{k-1}$ is considered: $k_{new} = (k-1)$. If $h_{k-1} > h_k$ is not verified, $h_{new} = h_k$ and $k_{new} = k$ are considered.
2. Next, the step size that corresponds to $(k+1)$ is compared with $h_{new}$. If $h_{k+1} > h_{new}$, the value $h_{k+1}$ is stored in $h_{new}$ and the order $k_{new}$ increases one unit.
3. Finally, the value $h_{new}$ is compared with the step size of the last step $h$. If $h_{new} > h$ is verified, the next step size will be $h_{new}$ and it will be given with order $k_{new}$. If not, the order and the step size of the last step are maintained: order $k$ and step size $h$.

When the given step results failed, the step is repeated. The step size which corresponds to order $k$ is calculated using a lower threshold $F_L = 0.1$ as follows:

$$h_k = \begin{cases} F \cdot h, & z_S < z \leq z_L \\ F_L \cdot h, & z > z_L \end{cases} \tag{29}$$

where $z_L = \frac{1}{F_L} = 10$ and $z_S = 1.2$ has been defined before. In the case in which $k > 1$, the step size for order $(k-1)$ is also calculated as:

$$h_{k-1} = \begin{cases} F_{k-1} \cdot h, & z_{k-1,S} < z_{k-1} \leq z_{k-1,L} \\ F_{k-1,L} \cdot h, & z_{k-1} > z_{k-1,L} \end{cases} \tag{30}$$

where $F_{k-1,L} = 0.1$, $z_{k-1,L} = \frac{1}{F_{k-1,L}} = 10$ and $z_{k-1,S} = 1.3$ has been defined before. If $h_{k-1} > h_k$, the next step will be given in order $(k-1)$ and the step size $h_{new}$ will be the minimum value of the present step $h$ and $h_{k-1}$. In second or posterior trials after an unsuccessful step, the new step size is calculated by dividing by 2 the actual step.

## 4 Some numerical results

In this section we will analyse the performance of the ode solvers *ode45* and *ode15s* in the ODEs (2) and (4).

### 4.1 First order ODEs

Consider the first order initial value ODE given by (2). The solution of (2) in the time interval $T = [0, 10]$ has been found using *ode45* and *ode15s*. The problem has been solved for different values of $\lambda$ and using the default values defined in the codes ($Rtol = 10^{-3}$ and being $10^{-6}$ all the components of the vector **Atol**). The number of steps given by each ode solver to solve (2) is listed in Table 1, resulting the *ode45* more efficient with the smallest values of $\lambda$ and the *ode15s* with the largest. In particular, we have studied the value $\lambda = -100$ where the *ode15s* works better. The step size and the error estimations of both codes have been calculated. For the *ode15s* the order in which each step has been given has been represented too and for the *ode45* positions of the values $\hat{h}$ in the complex plane have been drawn. These positions have not been drawn for *ode15s*, as $\mathbb{R}^-$ belongs to the stability regions of BDFs and NDFs. The graphics of these section correspond to use NDFs in the *ode15s*. For $\lambda = -100$ and *ode45*, some values $\hat{h}$ are outside the 5th order stability region, Figure 3. A step size $h$ for which $\hat{h}$ lies outside the 5th order stability region is generally followed by a smaller step size that reduces the error estimation and which makes $\hat{h}$ be inside the stability region or closer to its exterior limit. The steps given by the *ode45* are very regular although slightly oscillatory, see Figure 4, and the smaller the error the greater the next step size, see Figure 5. In the *ode15s*, the error estimation decreases a lot as long as we advance in the time interval and this makes the steps longer, Figures 4 and 5. The greatest steps of the *ode15s* are given when the function solution has not significant variations (values that are near to zero).

### 4.2 Second order ODEs

Consider the second order initial value ODE given by (3), which is reduced to a system of two first order ODEs (4). The eigenvalues of the problem (4) are $\pm\,|\lambda|\,i$. Problem (4) has been solved in the time interval $T = [0, 10]$, using the ode solvers *ode45* and *ode15s*, and using the default values defined in both codes. Again, we have analysed the case $\lambda^2 = -100^2$. In the case of the *ode15s*, the problem has been solved using BDFs. For $\lambda^2 = -100^2$, the *ode15s* gives 3406 steps and the *ode45* 1281 steps. The result obtained by the *ode15s* is worse than the one obtained

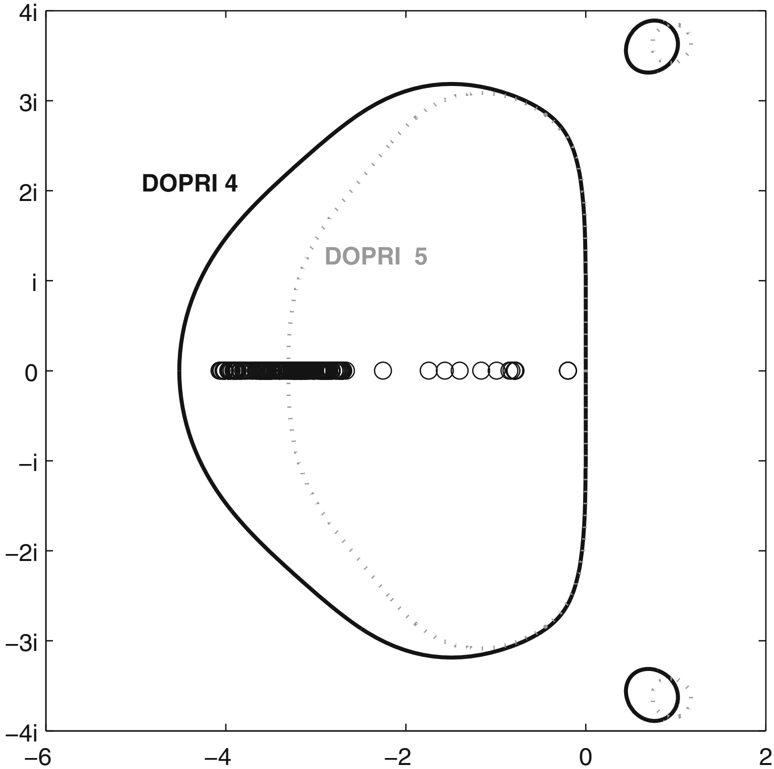


**Fig. 3** Positions of $\hat{h}$ in problem (2) in the stab. region DOPRI(5,4), being $\lambda = -100$.

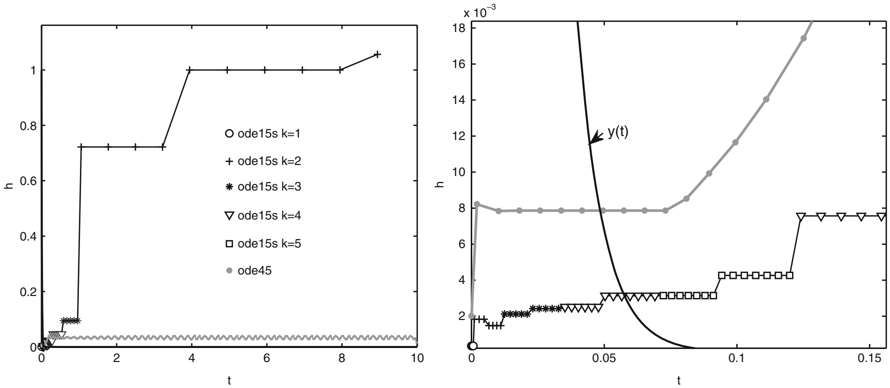


**Fig. 4** Step sizes in problem (2), being $\lambda = -100$ (right detail).

by the *ode45*, Figure 6. The main reason is that for the *ode45*, the values $\hat{h}$ are all inside the 5th order DOPRI(5,4) stability region, Figure 7. But this does not happen for the *ode15s*. In the first image of Figure 8 all the positions of the values $\hat{h}$ in

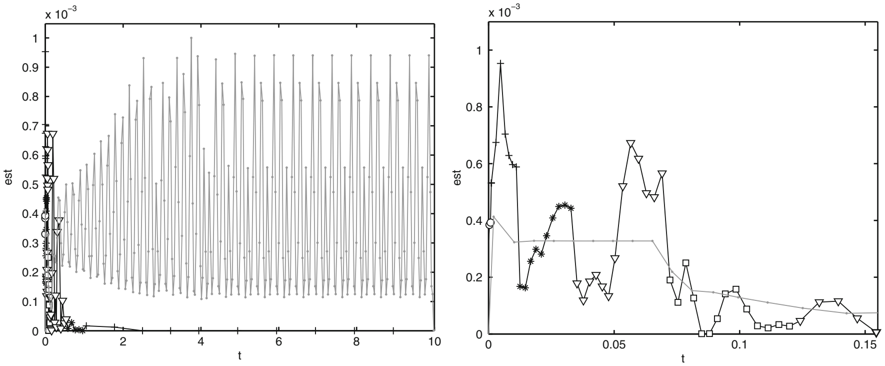


**Fig. 5** Error estimations in problem (2), being $\lambda = -100$ (right detail).

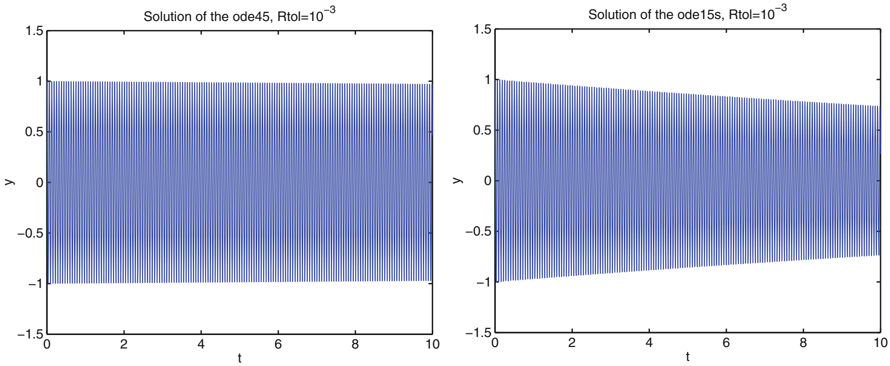


**Fig. 6** Solutions of the problem (4), being $\lambda^2 = -100^2$.

the BDF stability regions can be seen and the other figures show the detail. With the exception of the values corresponding to orders $k = 1, 2$ and 5, all the rest are outside the stability region.

Figure 9 and Figure 10 show the step sizes and the error estimations of the two solvers when $Rtol = 10^{-3}$. It can be observed that the error estimations of the *ode15s* are bigger than in the case in which the first order ODE (2) was solved.

## 5 Conclusions

From this analysis we conclude that:

- The ode solvers *ode45* and *ode15s* do not do additional calculations to control the eigenvalues of the jacobian matrix [11]. This means that the values $\hat{h} = h\lambda$ may lie outside the stability regions of the method, although the control of the local error and the subsequent adaptative step size will avoid an unstable solution.

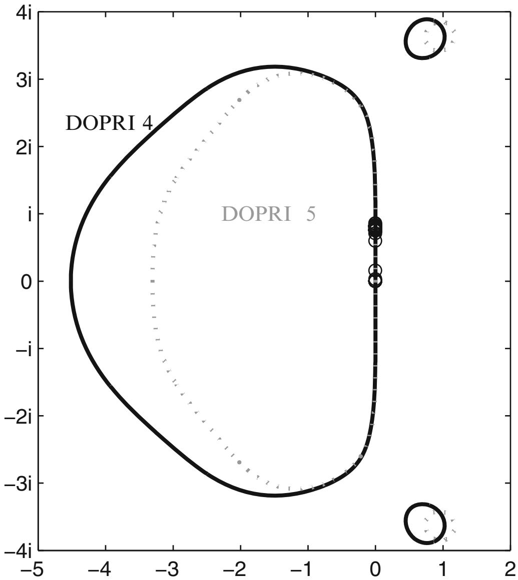


**Fig. 7** Positions of $\hat{h}$ in problem (4) in the stab. region DOPRI(5,4), being $\lambda^2 = -100^2$.

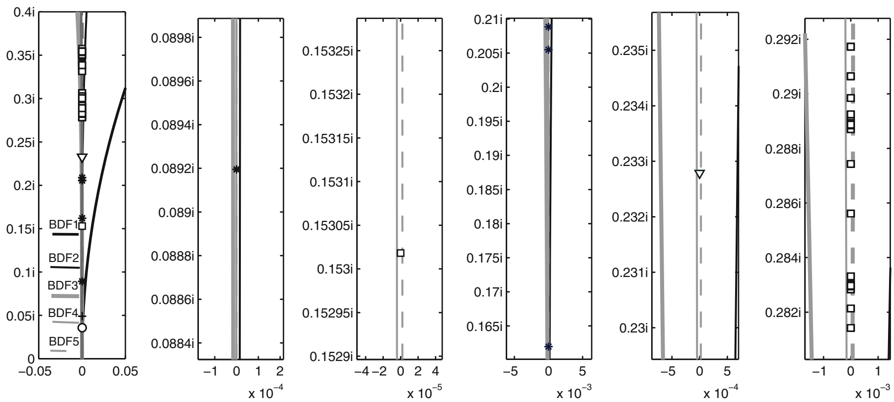


**Fig. 8** Positions of $\hat{h}$ in problem (4) in the stab. regions BDF, being $\lambda^2 = -100^2$.

- The local error estimation of the *ode15s* results very small when the solution decreases exponentially. This is the reason why the *ode15s* results very efficient when solving first order ODEs, mainly when $|\lambda|$ increases. Nevertheless, when solving second order ODEs with pure imaginary eigenvalues, the strong variations of the oscillatory solution produce a greater error estimation and the advantage that the *ode15s* has in first order ODEs disappears.
- The construction of each ode solver (their local error estimations, the formulae for the new step sizes, and so on) is what makes each of the codes more efficient when solving some type of problems.

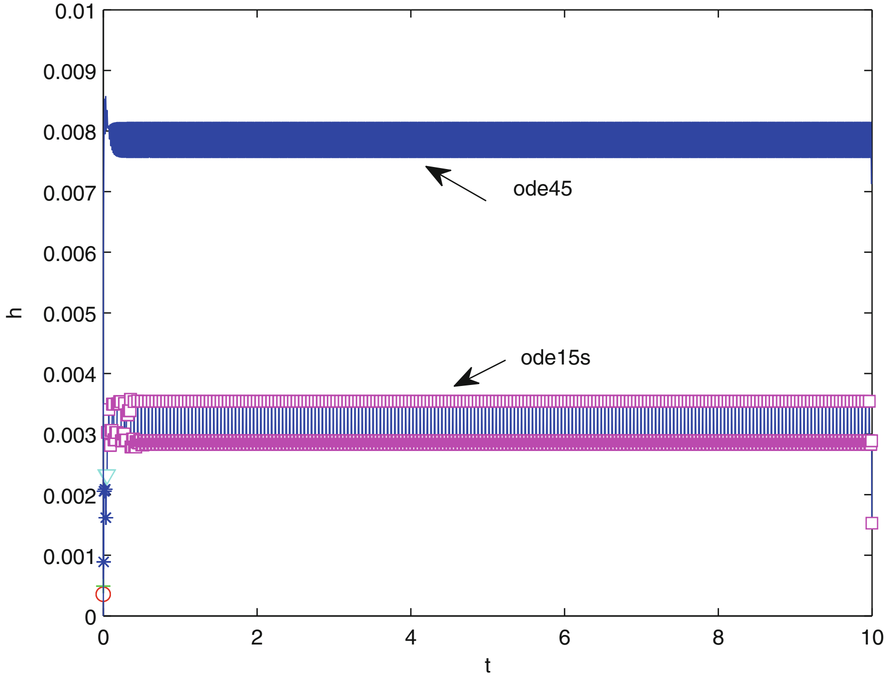


**Fig. 9** Step sizes in problem (4), being $\lambda^2 = -100^2$.

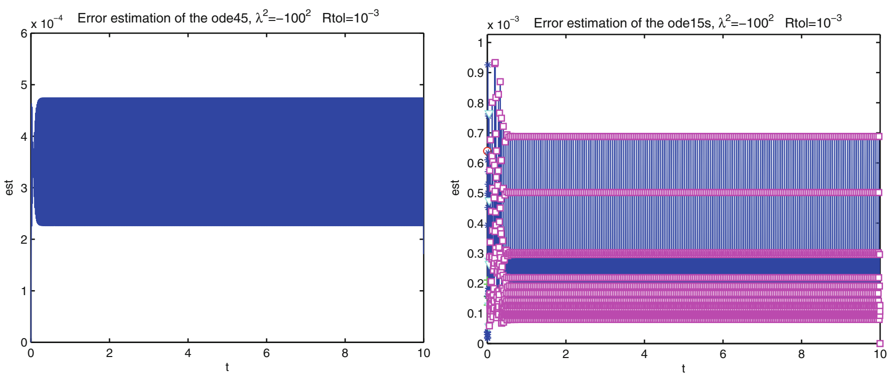


**Fig. 10** Error estimations in problem (4), being $\lambda^2 = -100^2$.

**Acknowledgements** The first author was partially funded by the Basque Government Consolidated Research Group Grant IT649-13 on "Mathematical Modeling, Simulation, and Industrial Applications (M2SI)".

## References


1. The Math Works, Inc.: http://www.mathworks.com
2. Aiken, R. C.: Stiff Computation. Oxford University Press, New York (1985).
3. Dormand, J. R., Prince, P. J.: A family of embedded Runge-Kutta formulae. J. Comput. Appl. Math. **6**, 19–26 (1980).
4. Gear, C. W.: Numerical initial value problems in Ordinary Differential Equations. Prentice Hall, New Jersey (1971).
5. Hairer, E., Wanner, G.: Solving ordinary differential equations, II, Stiff and Differential-Algebraic Problems. Springer, Berlin (1991).
6. Heath, M. T.: Scientific Computing. An introductory survey. Mc Graw Hill, New York (1997).
7. Hoffman, J. D.: Numerical methods for engineers and scientists. Marcel Dekker Inc., New York (2001).
8. Kostelich, E. J., Armbruster, D.: Introductory differential equations: From linearity to chaos. Addison Wesley, Massachusetts (1996).
9. Lambert, J. D.: Computational Methods in Ordinary Differential Equations. John Wiley, Chichester (1973).
10. Lambert, J. D.: Numerical methods for ordinary differential systems: the initial value problem. John Wiley & Sons, New York (1991).
11. Shampine, L. F.: Error estimation and control for ODEs. J. Sci. Comput. **15**, 3–16 (2005).
12. Shampine, L. F., Corless, R. M.: Initial value problems for ODEs in problem solving environments. J. Comput. Appl. Math **125**, 31–40 (2000).
13. Shampine, L. F., Gladwell, I., Thompson, S.: Solving ODEs with Matlab. Cambridge University Press, NewYork (2003).
14. Shampine, L. F., Reichelt, M. W.: The MATLAB ODE suite. SIAM J. Sci. Comput. **18**, 1–22 (1997).
15. Spijker, M. N.: Stiffness in numerical initial-value problems. J. Comput. Appl. Math **72**, 393–406 (1996).

# An overview on bounded elements in some partial algebraic structures

**Giorgia Bellomonte**

**Abstract** The notion of bounded element is fundamental in the framework of the spectral theory. Before implanting a spectral theory in some algebraic or topological structure it is needed to establish which are its bounded elements. In this paper, we want to give an overview on bounded elements of some particular algebraic and topological structures, summarizing our most recent results on this matter.

## 1 Introduction

Though the notion of bounded element has been considered, in different forms, within the theory regarding the structure of (topological) *-algebras, it is not so for the algebraic structures that do not possess a multiplication or possess just a partial one. Indeed, for (topological) *-algebras, the notion of bounded element is strictly linked to the operation of multiplication. In 1965, Allan wanted to construct a spectral theory for locally convex algebras. He judged natural to mimic the spectral theory of a closed operator on a Banach space: it is well known that if $A$ is a closed operator on a Banach space $\mathscr{B}$, then its spectrum is the set of the complex numbers $\lambda$ such that the operator $A - \lambda I$ has no bounded inverse. It became fundamental for him, therefore, to fix the concept of bounded element for a locally convex algebra. He defined (see [1, Def. 2.1]) *bounded* those and only those elements $a$ of the locally convex algebra $\mathfrak{A}[\tau]$ for which there exists a complex number $\lambda \neq 0$ such that the set $\{(\lambda x)^n;\ n = 1, 2, \dots\}$ is a bounded subset of $\mathfrak{A}[\tau]$. This definition does not apply to the algebraic structures we will examine in this overview: in general, neither a partial *-algebra nor a C*-inductive locally convex space possesses an everywhere defined multiplication, hence powers of a given element need not be defined.

Another notion of bounded element of a *-algebra is due to Vidav [11, Definition in Section 2] and involves a convex pointed cone $P$ of positive elements of the algebra (which are all and only those elements that can be written as the finite sums

G. Bellomonte (✉)
Dipartimento di Matematica e Informatica, Università degli Studi di Palermo, Via Archirafi 34, I-90123 Palermo, Italy
e-mail: giorgia.bellomonte@unipa.it



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_24

of elements of the form $a^*a$, with $a \in \mathfrak{A}$): an element $a \in \mathfrak{A}$ is bounded if there exists a positive number $\xi$ such that $e \in \mathfrak{A}$ the identity of $\mathfrak{A}$.

In order to extend the notion of bounded element to our case we have at disposal more than one possibility: we can define bounded elements taking into account both the topological structure and the algebraic structure of the set where we pick elements.

The paper is organized as follows. In Section 2 we will summarize the definitions and results we gave in [2] about bounded elements in a *-semisimple topological partial *-algebra: in that paper we considered the elements that are bounded with respect to a sufficient family $\mathfrak{M}$ of invariant positive sesquilinear (ips) forms (see also [4]) and elements that are bounded with respect to some positive cone, hence defined in purely algebraic terms. The outcome is that, under appropriate conditions, order bounded elements reduce to $\mathfrak{M}$-bounded ones. In Section 3, in the setting of C*-inductive locally convex spaces, we consider both bounded elements defined starting from the C*-inductive structure and those we have defined by means of an order cone and finally prove the equivalence of the two different notions we have given in [5].

## 2 Bounded elements in *-semisimple partial *-algebras

This section summarizes the results showed in [2] by J-P. Antoine, C. Trapani, and the author. We refer to that paper for the proofs and further readings. Before going forth, let us recall, for convenience of the reader, the main definitions we need.

A *partial *-algebra* $\mathfrak{A}$ is a complex vector space with conjugate linear involution $^*$ and a distributive partial multiplication $\cdot$, defined on a subset $\Gamma \subset \mathfrak{A}\times\mathfrak{A}$, satisfying the property that $(x, y) \in \Gamma$ if, and only if, $(y^*, x^*) \in \Gamma$ and $(x \cdot y)^* = y^* \cdot x^*$. From now on we will write simply $xy$ instead of $x \cdot y$ whenever $(x, y) \in \Gamma$. For every $y \in \mathfrak{A}$, the set of left (resp. right) multipliers of $y$ is denoted by $L(y)$ (resp. $R(y)$), i.e., $L(y) = \{x \in \mathfrak{A} : (x, y) \in \Gamma\}$ (resp. $R(y) = \{x \in \mathfrak{A} : (y, x) \in \Gamma\}$). We denote by $L\mathfrak{A}$ (resp. $R\mathfrak{A}$) the space of universal left (resp. right) multipliers of $\mathfrak{A}$ (for more details, we refer to [3]).

In general, a partial *-algebra is not associative, but in several situations a weaker form of associativity holds. More precisely, we say that $\mathfrak{A}$ is *semi-associative* if $y \in R(x)$ implies $yz \in R(x)$, for every $z \in R\mathfrak{A}$, and

$$(xy)z = x(yz).$$

The partial *-algebra $\mathfrak{A}$ has a unit if there exists an element $e \in \mathfrak{A}$ such that $e = e^*$, $e \in R\mathfrak{A} \cap L\mathfrak{A}$ and $xe = ex = x$, for every $x \in \mathfrak{A}$.

Let $\mathfrak{A}$ be a partial *-algebra. We assume that $\mathfrak{A}$ is a locally convex Hausdorff vector space under the topology $\tau$ defined by a (directed) set $\{p_\alpha\}_{\alpha\in\mathscr{I}}$ of seminorms. Assume that (cl) for every $x \in \mathfrak{A}$, the linear map $\mathsf{L}_x : R(x) \mapsto \mathfrak{A}$ with $\mathsf{L}_x(y) = xy$, $y \in R(x)$, is closed with respect to $\tau$, in the sense that, if $\{y_\alpha\} \subset R(x)$ is a net such that $y_\alpha \to y$ and $xy_\alpha \to z \in \mathfrak{A}$, then $y \in R(x)$ and $z = xy$.

in this case, $\mathfrak{A}$ is said to be a *topological partial *-algebra*. If the involution $x \mapsto x^*$ is continuous, we say that $\mathfrak{A}$ is a **-topological* partial *-algebra.

Starting from the family of seminorms $\{p_\alpha\}_{\alpha \in \mathscr{I}}$, we can define a second topology $\tau^*$ on $\mathfrak{A}$ by introducing the set of seminorms $\{p_\alpha^*(x)\}_{\alpha \in \mathscr{I}}$, where

$$p_\alpha^*(x) = \max\{p_\alpha(x), p_\alpha(x^*)\}, \quad x \in \mathfrak{A}.$$

The involution $x \mapsto x^*$ is automatically $\tau^*$-continuous. By (cl) it follows that, for every $x \in \mathfrak{A}$, both maps $\mathsf{L}_x$, $\mathsf{R} = (\mathsf{L}_{x^*})^*$ are $\tau^*$-closed. Hence, $\mathfrak{A}[\tau^*]$ is a *-topological partial *-algebra.

Let $\mathscr{H}$ be a complex Hilbert space and $\mathscr{D}$ a dense subspace of $\mathscr{H}$. We denote by $\mathscr{L}^\dagger(\mathscr{D}, \mathscr{H})$ the set of all (closable) linear operators $X$ such that $D(X) = \mathscr{D}$, $D(X^*) \supseteq \mathscr{D}$. The set $\mathscr{L}^\dagger(\mathscr{D}, \mathscr{H})$ is a partial *-algebra with respect to the following operations: the usual sum $X_1 + X_2$, the scalar multiplication $\lambda X$, the involution $X \mapsto X^\dagger := X^* \upharpoonright \mathscr{D}$ and the *(weak)* partial multiplication $X_1 \,\Box\, X_2 := {X_1}^{\dagger *} X_2$, defined whenever $X_2$ is a weak right multiplier of $X_1$ (we shall write $X_2 \in R^{\mathrm{w}}(X_1)$ or $X_1 \in L^{\mathrm{w}}(X_2)$), that is, whenever $X_2\mathscr{D} \subset D({X_1}^{\dagger *})$ and $X_1^*\mathscr{D} \subset D(X_2^*)$.

It is easy to check that $X_1 \in L^{\mathrm{w}}(X_2)$ if and only if there exists $Z \in \mathscr{L}^\dagger(\mathscr{D}, \mathscr{H})$ such that

$$\langle X_2\xi | X_1^\dagger \eta\rangle = \langle Z\xi | \eta\rangle, \quad \forall \xi, \eta \in \mathscr{D}. \tag{1}$$

In this case $Z = X_1 \,\Box\, X_2$. $\mathscr{L}^\dagger(\mathscr{D}, \mathscr{H})$ is neither associative nor semi-associative. If $I$ denotes the identity operator of $\mathscr{H}$, $I_{\mathscr{D}} := I \upharpoonright \mathscr{D}$ is the unit of the partial *-algebra $\mathscr{L}^\dagger(\mathscr{D}, \mathscr{H})$. We will indicate by $\mathsf{t}_s$ the *strong topology* on $\mathscr{L}^\dagger(\mathscr{D}, \mathscr{H})$, defined by the seminorms

$$p_\xi(X) = \|X\xi\|, \quad X \in \mathscr{L}^\dagger(\mathscr{D}, \mathscr{H}),\ \xi \in \mathscr{D}.$$

Let $\mathfrak{A}[\tau]$ be a topological partial *-algebra with locally convex topology $\tau$. Then a subspace $\mathfrak{B}$ of $R\mathfrak{A}$ is called a *multiplication core* [2, Definition 2.3] if

($\mathsf{d}_1$) $e \in \mathfrak{B}$ if $\mathfrak{A}$ has a unit $e$;
($\mathsf{d}_2$) $\mathfrak{B} \cdot \mathfrak{B} \subseteq \mathfrak{B}$;
($\mathsf{d}_3$) $\mathfrak{B}$ is $\tau^*$-dense in $\mathfrak{A}$;
($\mathsf{d}_4$) for every $b \in \mathfrak{B}$, the map $x \mapsto xb$ , $x \in \mathfrak{A}$, is $\tau$-continuous;
($\mathsf{d}_5$) one has $b^*(xc) = (b^*x)c$, $\forall\, x \in \mathfrak{A}, b, c \in \mathfrak{B}$.

$\mathfrak{A}[\tau]$ is called $\mathfrak{A}_0$-*regular* if it possesses a multiplication core $\mathfrak{A}_0$ which is a *-algebra and, for every $b \in \mathfrak{A}_0$, the map $x \mapsto bx$ , $x \in \mathfrak{A}$, is $\tau$-continuous [4, Def. 4.1].

A **-representation* of a partial *-algebra $\mathfrak{A}$ in the Hilbert space $\mathscr{H}$ is a linear map $\pi : \mathfrak{A} \to \mathscr{L}^\dagger(\mathscr{D}(\pi), \mathscr{H})$ such that: (i) $\pi(x^*) = \pi(x)^\dagger$, for every $x \in \mathfrak{A}$;

(ii) $x \in L(y)$ in $\mathfrak{A}$ implies $\pi(x) \in L^{\mathrm{w}}(\pi(y))$ and $\pi(x) \,\Box\, \pi(y) = \pi(xy)$. The subspace $\mathscr{D}(\pi)$ is called the *domain* of the *-representation $\pi$. The *-representation $\pi$ is said to be *bounded* if $\overline{\pi(x)} \in \mathscr{B}(\mathscr{H})$ for every $x \in \mathfrak{A}$. We will denote by $\mathrm{Rep}_c(\mathfrak{A})$ the set of all $(\tau, \mathsf{t}_s)$-continuous *-representations of $\mathfrak{A}$. Let $\varphi$ be a positive sesquilinear form on $D(\varphi) \times D(\varphi)$, where $D(\varphi)$ is a subspace of $\mathfrak{A}$. Then we have

$$\varphi(x, y) = \overline{\varphi(y, x)}, \quad \forall\, x, y \in D(\varphi), \tag{2}$$

$$|\varphi(x, y)|^2 \leqslant \varphi(x, x)\varphi(y, y), \quad \forall\, x, y \in D(\varphi). \tag{3}$$

We put

$$N_\varphi = \{x \in D(\varphi) : \varphi(x, x) = 0\}.$$

By (3), we have

$$N_\varphi = \{x \in D(\varphi) : \varphi(x, y) = 0, \quad \forall\, y \in D(\varphi)\},$$

and so $N_\varphi$ is a subspace of $D(\varphi)$ and the quotient space $D(\varphi)/N_\varphi := \{\lambda_\varphi(x) \equiv x + N_\varphi; x \in D(\varphi)\}$ is a pre-Hilbert space with respect to the inner product

$$\langle \lambda_\varphi(x) | \lambda_\varphi(y) \rangle = \varphi(x, y), \quad x, y \in D(\varphi).$$

We denote by $\mathscr{H}_\varphi$ the Hilbert space obtained by completion of $D(\varphi)/N_\varphi$.

Our overview on bounded elements starts focusing on the so-called *-semisimple topological partial *-algebras. A topological partial *algebras. $\mathfrak{A}[\tau]$ is called **-semisimple* [2, Definition 3.5] if, for every $x \in \mathfrak{A} \setminus \{0\}$ there exists $\pi \in \mathrm{Rep}_c(\mathfrak{A})$ such that $\pi(x) \neq 0$ or, equivalently, if the **-radical* of $\mathfrak{A}$

$$\mathscr{R}^*(\mathfrak{A}) := \{x \in \mathfrak{A} : \pi(x) = 0, \text{ for all } \pi \in \mathrm{Rep}_c(\mathfrak{A})\}$$

is equal to $\{0\}$.

A positive sesquilinear form $\varphi$ on $\mathfrak{A} \times \mathfrak{A}$ is said to be *invariant*, and called an *ips-form*, if there exists a subspace $B(\varphi)$ of $\mathfrak{A}$ (called a *core* for $\varphi$) with the properties

(ips$_1$) $B(\varphi) \subset R\mathfrak{A}$;
(ips$_2$) $\lambda_\varphi(B(\varphi))$ is dense in $\mathscr{H}_\varphi$;
(ips$_3$) $\varphi(xa, b) = \varphi(a, x^*b)$, $\forall\, x \in \mathfrak{A}, \forall\, a, b \in B(\varphi)$;
(ips$_4$) $\varphi(x^*a, yb) = \varphi(a, (xy)b)$, $\forall\, x \in L(y), \forall\, a, b \in B(\varphi)$.

We will denote by $\mathscr{P}_{\mathfrak{B}}(\mathfrak{A})$ the set of *all* $\tau$-continuous ips-forms with core $\mathfrak{B}$.

A family $\mathfrak{M}$ of continuous ips-forms on $\mathfrak{A} \times \mathfrak{A}$ is *sufficient* if $x \in \mathfrak{A}$ and $\varphi(x, x) = 0$, for every $\varphi \in \mathscr{M}$ imply $x = 0$.

**Proposition 1.** *Let $\mathfrak{A}[\tau]$ be a topological partial *-algebra with unit e. Let $\mathfrak{B}$ be a multiplication core. For an element $x \in \mathfrak{A}$ the following statements are equivalent.*

*(i)* $x \in \mathscr{R}^*(\mathfrak{A})$.
*(ii)* $\varphi(x,x) = 0$, *for every* $\varphi \in \mathscr{P}_{\mathfrak{B}}(\mathfrak{A})$.

*Remark 1.* By Proposition 1, $\mathfrak{A}[\tau]$ is *-semisimple if, and only if, for some multiplication core $\mathfrak{B}$, the family $\mathscr{P}_{\mathfrak{B}}(\mathfrak{A})$ is *sufficient*.

If the family $\mathfrak{M}$ is sufficient, any larger family $\mathfrak{M}' \supset \mathfrak{M}$ is sufficient too. In this case, the maximal sufficient family (having $\mathfrak{B}$ as core) is obviously $\mathscr{P}_{\mathfrak{B}}(\mathfrak{A})$. Hence if a sufficient family $\mathfrak{M} \subseteq \mathscr{P}_{\mathfrak{B}}(\mathfrak{A})$ exists, then $\mathfrak{A}[\tau]$ is *-semisimple.

We say that the *weak* multiplication $x \Box y$ is well defined (with respect to $\mathfrak{M}$) if there exists $z \in \mathfrak{A}$ such that:

$$\varphi(ya, x^*b) = \varphi(za, b), \ \forall\, a, b \in \mathfrak{B}, \forall\, \varphi \in \mathfrak{M}.$$

In this case, we put $x \Box y := z$ and the sufficiency of $\mathfrak{M}$ guarantees that $z$ is unique. The weak multiplication $\Box$ clearly depends on $\mathfrak{M}$: the larger is $\mathfrak{M}$, the stronger is the weak multiplication, in the sense that if $\mathfrak{M} \subseteq \mathfrak{M}' \subseteq \mathscr{P}_{\mathfrak{B}}(\mathfrak{A})$ and $x \Box y$ exists w.r. to $\mathfrak{M}'$, then $x \Box y$ exists with respect to $\mathfrak{M}$ too.

Since it may be difficult to identify in practice such a sufficient family of continuous ips-forms that guarantees the *-semisimplicity of $\mathfrak{A}[\tau]$, we examine in what sense ips-forms may be replaced by a special class of continuous linear functionals, called *representable*.

**Definition 1.** Let $\omega$ be a linear functional on $\mathfrak{A}$ and $\mathfrak{B}$ a subspace of $R\mathfrak{A}$. We say that $\omega$ is *representable* (with respect to $\mathfrak{B}$) if the following requirements are satisfied:

($\mathsf{r}_1$) $\omega(a^*a) \geq 0$ for all $a \in \mathfrak{B}$ ($\mathfrak{B}$-positiveness);
($\mathsf{r}_2$) $\omega(b^*(x^*a)) = \overline{\omega(a^*(xb))}, \ \forall\ a, b \in \mathfrak{B}, x \in \mathfrak{A}$;
($\mathsf{r}_3$) $\forall x \in \mathfrak{A}$ there exists $\gamma_x > 0$ such that $|\omega(x^*a)| \leq \gamma_x\, \omega(a^*a)^{1/2}$, for all $a \in \mathfrak{B}$.

We will denote by $\mathscr{R}_c(\mathfrak{A}, \mathfrak{B})$ the set of $\tau$-continuous linear functionals that are representable (with respect to $\mathfrak{B}$).

In this case, one can prove that there exists a triple $(\pi_\omega^{\mathfrak{B}}, \lambda_\omega^{\mathfrak{B}}, \mathscr{H}_\omega^{\mathfrak{B}})$ such that

(a) $\pi_\omega^{\mathfrak{B}}$ is a *-representation of $\mathfrak{A}$ in $\mathscr{H}_\omega^{\mathfrak{B}}$;
(b) $\lambda_\omega^{\mathfrak{B}}$ is a linear map of $\mathfrak{A}$ into $\mathscr{H}_\omega^{\mathfrak{B}}$ with $\lambda_\omega^{\mathfrak{B}}(\mathfrak{B}) = \mathscr{D}(\pi_\omega^{\mathfrak{B}})$ and $\pi_\omega^{\mathfrak{B}}(x)\lambda_\omega^{\mathfrak{B}}(a) = \lambda_\omega^{\mathfrak{B}}(xa)$, for every $x \in \mathfrak{A}$, $a \in \mathfrak{B}$;
(c) $\omega(b^*(xa)) = \langle \pi_\omega^{\mathfrak{B}}(x)\lambda_\omega^{\mathfrak{B}}(a) | \lambda_\omega^{\mathfrak{B}}(b)\rangle$, for every $x \in \mathfrak{A}$, $a, b \in \mathfrak{B}$.

In particular, if $\mathfrak{A}$ has a unit $e$ and $e \in \mathfrak{B}$, we have:

($\mathrm{a}_1$) $\pi_\omega^{\mathfrak{B}}$ is a cyclic *-representation of $\mathfrak{A}$ with cyclic vector $\xi_\omega$;
($\mathrm{b}_1$) $\lambda_\omega^{\mathfrak{B}}$ is a linear map of $\mathfrak{A}$ into $\mathscr{H}_\omega^{\mathfrak{B}}$ with $\lambda_\omega^{\mathfrak{B}}(\mathfrak{B}) = \mathscr{D}(\pi_\omega^{\mathfrak{B}})$, $\xi_\omega = \lambda_\omega^{\mathfrak{B}}(e)$ and $\pi_\omega^{\mathfrak{B}}(x)\lambda_\omega^{\mathfrak{B}}(a) = \lambda_\omega^{\mathfrak{B}}(xa)$, for every $x \in \mathfrak{A}$, $a \in \mathfrak{B}$;
($\mathrm{c}_1$) $\omega(x) = \langle \pi_\omega^{\mathfrak{B}}(x)\xi_\omega | \xi_\omega\rangle$, for every $x \in \mathfrak{A}$.

For what we have already noted, it is interesting to identify a class of topological partial *-algebras for which representable linear functionals and ips-forms can be freely replaced by one another, since every representable linear functional comes (as for *-algebras with unit) from an ips-form. These partial *-algebras are called *fully representable*: a topological partial *-algebra $\mathfrak{A}[\tau]$, with multiplication core $\mathfrak{B}$ is fully representable if

(fr) $D(\overline{\varphi_\omega}) = \mathfrak{A}$, for every continuous linear functional $\omega$ on $\mathfrak{A}$ which is representable w.r. to the same core $\mathfrak{B}$.

The following definitions and results can be found in [2, Subsection 5.1].

**Definition 2.** Let $\mathfrak{A}$ be a topological partial *-algebra with multiplication core $\mathfrak{B}$ and a sufficient family $\mathfrak{M}$ of continuous ips-forms with core $\mathfrak{B}$. An element $x \in \mathfrak{A}$ is called $\mathfrak{M}$-*bounded* if there exists $\gamma_x > 0$ such that

$$|\varphi(xa, b)| \leq \gamma_x \, \varphi(a, a)^{1/2} \varphi(b, b)^{1/2}, \ \forall \, \varphi \in \mathfrak{M}, \ a, b \in \mathfrak{B} \, .$$

A useful characterization of $\mathfrak{M}$-bounded elements is given by the following proposition.

**Proposition 2.** *Let $\mathfrak{A}[\tau]$ be a topological partial *-algebra with multiplication core $\mathfrak{B}$. Then, an element $x \in \mathfrak{A}$ is $\mathfrak{M}$-bounded if, and only if, there exists $\gamma_x \in \mathbb{R}$ such that $\varphi(xa, xa) \leq \gamma_x^2 \, \varphi(a, a)$ for all $\varphi \in \mathfrak{M}$ and $a \in \mathfrak{B}$.*

If $x, y$ are $\mathfrak{M}$-bounded elements and their weak product $x \Box y$ exists, then $x \Box y$ is also $\mathfrak{M}$-bounded.

## *2.1 Order bounded elements*

Before giving the definition of order bounded element of a topological partial *-algebra $\mathfrak{A}[\tau]$ with unit and endowed multiplication core, we need to introduce an order structure in $\mathfrak{A}[\tau]$. We have done it by defining several order cones or wedges of $\mathfrak{A}[\tau]$.

### 2.1.1 Order structure of $\mathfrak{A}[\tau]$

Let $\mathfrak{A}[\tau]$ be a topological partial *-algebra with multiplication core $\mathfrak{B}$. If $\mathfrak{A}[\tau]$ is *-semisimple, there is a natural order on $\mathfrak{A}$ defined by the family $\mathscr{P}_{\mathfrak{B}}(\mathfrak{A})$ or by any sufficient subfamily $\mathfrak{M}$ of $\mathscr{P}_{\mathfrak{B}}(\mathfrak{A})$, and this order can be used to define a different notion of *boundedness* of an element $x \in \mathfrak{A}$ [8, 10, 11].

**Definition 3.** Let $\mathfrak{A}[\tau]$ be a topological partial *-algebra and $\mathfrak{B}$ a subspace of $R\mathfrak{A}$. A subset $\mathfrak{K}$ of $\mathfrak{A}_h := \{x \in \mathfrak{A} : \ x = x^*\}$ is called a $\mathfrak{B}$-*admissible wedge* if

(1) $e \in \mathfrak{K}$, if $\mathfrak{A}$ has a unit $e$;
(2) $x + y \in \mathfrak{K}, \ \forall \, x, y \in \mathfrak{K}$;
(3) $\lambda x \in \mathfrak{K}, \ \forall \, x \in \mathfrak{K}, \ \lambda \geq 0$;

(4) $(a^*x)a = a^*(xa) =: a^*xa \in \mathfrak{K}$, $\forall\, x \in \mathfrak{K}$, $a \in \mathfrak{B}$.

As usual, $\mathfrak{K}$ defines an order on the real vector space $\mathfrak{A}_h$ by $x \le y \Leftrightarrow y - x \in \mathfrak{K}$.

In the rest of this section, we will suppose that the partial *-algebras under consideration are *semi-associative*. Under this assumption, the first equality in (4) of Definition 3 is automatically satisfied.

Now, let us define a series of admissible cones with respect to some subspace of $R\mathfrak{A}$.

- Let $\mathfrak{A}$ be a topological partial *-algebra with multiplication core $\mathfrak{B}$. We put

$$\mathfrak{B}^{(2)} = \left\{ \sum_{k=1}^{n} x_k^* x_k,\ x_k \in \mathfrak{B},\ n \in \mathbb{N} \right\}.$$

  If $\mathfrak{B}$ is a *-algebra, this is nothing but the set (wedge) of positive elements of $\mathfrak{B}$. The $\mathfrak{B}$-*strongly positive* elements of $\mathfrak{A}$ are then defined as the elements of $\mathfrak{A}^+(\mathfrak{B}) := \overline{\mathfrak{B}^{(2)}}^{\tau}$. Since $\mathfrak{A}$ is semi-associative, the set $\mathfrak{A}^+(\mathfrak{B})$ of $\mathfrak{B}$-strongly positive elements is a $\mathfrak{B}$-admissible wedge.
- We also define

$$\mathfrak{A}^+_{\rm alg} = \left\{ \sum_{k=1}^{n} x_k^* x_k,\ x_k \in R\mathfrak{A},\ n \in \mathbb{N} \right\},$$

  the set (wedge) of positive elements of $\mathfrak{A}$ and we put $\mathfrak{A}^+_{\rm top} := \overline{\mathfrak{A}^+_{\rm alg}}^{\tau}$. The semi-associativity implies that $R\mathfrak{A} \cdot R\mathfrak{A} \subseteq R\mathfrak{A}$ and then $\mathfrak{A}^+_{\rm top}$ is $R\mathfrak{A}$-admissible.
- Let $\mathfrak{M} \subseteq \mathscr{P}_{\mathfrak{B}}(\mathfrak{A})$. An element $x \in \mathfrak{A}$ is called $\mathfrak{M}$-*positive* if

$$\varphi(xa, a) \ge 0, \quad \forall \varphi \in \mathfrak{M},\ a \in \mathfrak{B}.$$

  It can be proved that an $\mathfrak{M}$-positive element is automatically hermitian. We denote by $\mathfrak{A}^+_{\mathfrak{M}}$ the set of all $\mathfrak{M}$-positive elements. Clearly $\mathfrak{A}^+_{\mathfrak{M}}$ is a $\mathfrak{B}$-admissible wedge.

As can be easily checked, the following inclusions hold

$$\mathfrak{A}^+(\mathfrak{B}) \subseteq \mathfrak{A}^+_{\rm top} \subseteq \mathfrak{A}^+_{\mathfrak{M}}, \quad \forall\, \mathfrak{M} \subseteq \mathscr{P}_{\mathfrak{B}}(\mathfrak{A}). \tag{4}$$

Moreover, it can be proved that, if the family $\mathfrak{M}$ is sufficient, then $\mathfrak{A}^+_{\mathfrak{M}}$ is a cone, i.e., $\mathfrak{A}^+_{\mathfrak{M}} \cap (-\mathfrak{A}^+_{\mathfrak{M}}) = \{0\}$; this automatically implies that $\mathfrak{A}^+(\mathfrak{B})$ is a cone too.

Put $\mathfrak{A}^+_{\mathscr{P}} := \mathfrak{A}^+_{\mathscr{P}_{\mathfrak{B}}(\mathfrak{A})}$. It can be proved that, if $\mathfrak{A}$ is a fully-representable *-semisimple *-topological partial *-algebra with multiplication core $\mathfrak{B}$ and unit $e \in \mathfrak{B}$ and if $\mathfrak{A}[\tau]$ is a Fréchet space and the following property holds

(P) $y \in \mathfrak{A}$ and $\omega(a^*ya) \ge 0$, for every $\omega \in \mathscr{R}_c(\mathfrak{A}, \mathfrak{B})$ and $a \in \mathfrak{A}_0$, imply $y \in \mathfrak{A}^+(\mathfrak{B})$

then the chain of inclusions (4) collapses: $\mathfrak{A}^+(\mathfrak{B}) = \mathfrak{A}^+_{\mathscr{P}}$ (see [2, Propositions 5.13, 5.14 and Corollary 5.16]).

The following statement shows that $\mathscr{P}_{\mathfrak{B}}(\mathfrak{A})$-positivity is exactly what is needed if we want the order to be preserved under any continuous *-representation.

**Proposition 3.** *Let $\mathfrak{A}$ be a topological partial *-algebra with multiplication core $\mathfrak{B}$ and unit $e \in \mathfrak{B}$. Let $x \in \mathfrak{A}$. Then, the following are equivalent:*

1. $x \in \mathfrak{A}^+_{\mathscr{P}}$ ;
2. *the operator $\pi(x)$ is positive for every $(\tau, \mathsf{t}_s)$-continuous *-representation $\pi$ with $\pi(e) = I_{\mathscr{D}(\pi)}$.*

### 2.1.2 Order bounded elements

Let $\mathfrak{A}[\tau]$ be a topological partial *-algebra with multiplication core $\mathfrak{B}$ and unit $e \in \mathfrak{B}$. As we have seen in Section 2.1.1, $\mathfrak{A}[\tau]$ has several natural orders, all related to the topology $\tau$. Each of them can be used to define *bounded* elements.

Let $x \in \mathfrak{A}$; put $\Re(x) = \frac{1}{2}(x + x^*)$, $\Im(x) = \frac{1}{2i}(x - x^*)$. Then $\Re(x), \Im(x) \in \mathfrak{A}_h$ (the set of self-adjoint elements of $\mathfrak{A}$) and $x = \Re(x) + i\Im(x)$.

Let now $\mathfrak{K}$ be an arbitrary $\mathfrak{B}$-admissible cone.

**Definition 4.** An element $x \in \mathfrak{A}$ is called $\mathfrak{K}$-*bounded* if there exists $\gamma \geq 0$ such that

$$\pm\Re(x) \leq \gamma e; \qquad \pm\Im(x) \leq \gamma e.$$

We denote by $\mathfrak{A}_b(\mathfrak{K})$ the family of $\mathfrak{K}$-bounded elements.

The following statements are easily checked.

(1) $\alpha x + \beta y \in \mathfrak{A}_b(\mathfrak{K}), \quad \forall x, y \in \mathfrak{A}_b(\mathfrak{K}),\ \alpha, \beta \in \mathbb{C}.$
(2) $x \in \mathfrak{A}_b(\mathfrak{K}) \Leftrightarrow x^* \in \mathfrak{A}_b(\mathfrak{K}).$

For $x \in \mathfrak{A}_h$, put

$$\|x\|_b := \inf\{\gamma > 0 :\ -\gamma e \leq x \leq \gamma e\}.$$

$\|\cdot\|_b$ is a seminorm on the real vector space $(\mathfrak{A}_b(\mathfrak{K}))_h$.

Let $\mathfrak{A}[\tau]$ be a *-semisimple topological partial *-algebra with multiplication core $\mathfrak{B}$. We can then specify the wedge $\mathfrak{K}$ as one of those defined above. Take first $\mathfrak{K} = \mathfrak{A}^+_{\mathfrak{M}}$, where $\mathfrak{M} = \mathscr{P}_{\mathfrak{B}}(\mathfrak{A})$ is the sufficient family of all continuous ips-forms with core $\mathfrak{B}$. For simplicity, we write again $\mathscr{P} := \mathscr{P}_{\mathfrak{B}}(\mathfrak{A})$, hence $\mathfrak{A}^+_{\mathscr{P}} := \mathfrak{A}^+_{\mathscr{P}_{\mathfrak{B}}(\mathfrak{A})}$ and $\mathfrak{A}_b(\mathscr{P}) := \mathfrak{A}_b(\mathfrak{A}^+_{\mathscr{P}})$.

**Proposition 4.** *If $x \in \mathfrak{A}_b(\mathscr{P})$, then $\pi(x)$ is a bounded operator, for every $(\tau, \mathsf{t}_s)$-continuous *-representation of $\mathfrak{A}$. Moreover, if $x = x^*$, $\|\pi(x)\| \leq \|x\|_b$.*

Hence, as it is natural, the $\mathfrak{A}_b(\mathscr{P})$-bounded elements are those that are represented by a bounded operator in any $(\tau, \mathsf{t}_s)$-continuous *-representation of $\mathfrak{A}$.

The following theorem states the equivalence, under opportune hypothesis, of the notions of order bounded element and of element bounded with respect to a sufficient family of ips-forms.

**Theorem 1.** *Let $\mathfrak{A}[\tau]$ be a *-semisimple topological partial *-algebra with multiplication core $\mathfrak{B}$ and unit $e \in \mathfrak{B}$. For $x \in \mathfrak{A}$, the following statements are equivalent.*

*(i)* $x$ *is* $\mathscr{P}_{\mathfrak{B}}(\mathfrak{A})$*-bounded;*
*(ii)* $x \in \mathfrak{A}_b(\mathscr{P})$*;*
*(iii)* $\pi(x)$ *is bounded, for every* $\pi \in \mathrm{Rep}_c(\mathfrak{A})$*, and*

$$\sup\{\|\overline{\pi(x)}\|,\ \pi \in \mathrm{Rep}_c(\mathfrak{A})\} < \infty.$$

Another possible choice for the order cone is, for instance, $\mathfrak{A}^+(\mathfrak{B})$. It is clear that $\mathfrak{A}_b(\mathfrak{A}^+(\mathfrak{B})) \subseteq \mathfrak{A}_b(\mathscr{P})$; it can be proved also that the two wedges coincide if $\mathfrak{A}[\tau]$ is a Fréchet space which is also a fully representable, semi-associative *-topological partial *-algebra, with multiplication core $\mathfrak{B}$ and unit $e \in \mathfrak{B}$ and the property (P) (see Subsection 2.1.1) holds.

## 3 Bounded elements for a C*-inductive locally convex space

In this section we recap what S. Di Bella, C. Trapani and the author have done in [5], i.e. extending the notion of bounded element to the case of C*-inductive locally convex spaces; for this reason, we refer to that paper for the proofs of every result we report on.

Before going forth, we recall the notions of directed system of C*-algebras and of C*-inductive locally convex space we introduced in [7].

Let $\mathfrak{A}$ be a vector space over $\mathbb{C}$. Let $\mathbb{F}$ be a set of indices directed upward and consider, for every $\alpha \in \mathbb{F}$, a Banach space $\mathfrak{A}_\alpha \subset \mathfrak{A}$ such that:

(I.1) $\mathfrak{A}_\alpha \subseteq \mathfrak{A}_\beta$, if $\alpha \leq \beta$;
(I.2) $\mathfrak{A} = \bigcup_{\alpha\in\mathbb{F}} \mathfrak{A}_\alpha$;
(I.3) $\forall\alpha \in \mathbb{F}$, there exists a C*-algebra $\mathfrak{B}_\alpha$ (with unit $e_\alpha$ and norm $\|\cdot\|_\alpha$) and a norm-preserving isomorphism of vector spaces $\phi_\alpha : \mathfrak{B}_\alpha \to \mathfrak{A}_\alpha$;
(I.4) $x_\alpha \in \mathfrak{B}_\alpha^+ \Rightarrow x_\beta = (\phi_\beta^{-1}\phi_\alpha)(x_\alpha) \in \mathfrak{B}_\beta^+$, for every $\alpha, \beta \in \mathbb{F}$ with $\beta \geq \alpha$.

We put $j_{\beta\alpha} = \phi_\beta^{-1}\phi_\alpha$, if $\alpha, \beta \in \mathbb{F},\ \beta \geq \alpha$.

If $x \in \mathfrak{A}$, there exist $\alpha \in \mathbb{F}$ such that $x \in \mathfrak{A}_\alpha$ and (a unique) $x_\beta \in \mathfrak{B}_\beta$ such that $x = \phi_\beta(x_\beta)$, for all $\beta \geq \alpha$.

Then, we put

$$j_{\beta\alpha}(x_\alpha) := x_\beta \qquad \text{if } \alpha \leq \beta.$$

By (I.4), it follows easily that $j_{\beta\alpha}$ preserves the involution; i.e., $j_{\beta\alpha}(x_\alpha^*) = (j_{\beta\alpha}(x_\alpha))^*$.

The family $\{\mathfrak{B}_\alpha, j_{\beta\alpha}, \beta \geq \alpha\}$ is a *directed system of C*-algebras*, in the sense that:

(J.1) for every $\alpha, \beta \in \mathbb{F}$, with $\beta \geq \alpha$, $j_{\beta\alpha} : \mathfrak{B}_\alpha \to \mathfrak{B}_\beta$ is a linear and injective map; $j_{\alpha\alpha}$ is the identity of $\mathfrak{B}_\alpha$,
(J.2) for every $\alpha, \beta \in \mathbb{F}$, with $\alpha \leq \beta$, $\phi_\alpha = \phi_\beta j_{\beta\alpha}$.
(J.3) $j_{\gamma\beta} j_{\beta\alpha} = j_{\gamma\alpha}$, $\alpha \leq \beta \leq \gamma$.

We assume that, in addition, the $j_{\beta\alpha}$'s are Schwarz maps (see, e.g., [9]); i.e.,

(sch) $j_{\beta\alpha}(x_\alpha)^* j_{\beta\alpha}(x_\alpha) \leq j_{\beta\alpha}(x_\alpha^* x_\alpha), \quad \forall x_\alpha \in \mathfrak{B}_\alpha,\ \alpha \leq \beta.$

For every $\alpha, \beta \in \mathbb{F}$, with $\alpha \leq \beta$, $j_{\beta\alpha}$ is continuous [9] and, moreover,

$$\|j_{\beta\alpha}(x_\alpha)\|_\beta \leq \|x_\alpha\|_\alpha, \quad \forall x_\alpha \in \mathfrak{B}_\alpha.$$

An involution in $\mathfrak{A}$ is defined as follows. Let $x \in \mathfrak{A}$. Then $x \in \mathfrak{A}_\alpha$, for some $\alpha \in \mathbb{F}$, i.e., $x = \phi_\alpha(x_\alpha)$, for a unique $x_\alpha \in \mathfrak{B}_\alpha$. Put $x^* := \phi_\alpha(x_\alpha^*)$. Then if $\beta \geq \alpha$, we have

$$\phi_\beta^{-1}(x^*) = \phi_\beta^{-1}(\phi_\alpha(x_\alpha^*)) = j_{\beta\alpha}(x_\alpha^*) = (j_{\beta\alpha}(x_\alpha))^* = x_\beta^*.$$

It is easily seen that the map $x \mapsto x^*$ is an involution in $\mathfrak{A}$. Moreover, by the definition itself, it follows that every map $\phi_\alpha$ *preserves the involution*; i.e., $\phi_\alpha(x_\alpha^*) = (\phi_\alpha(x_\alpha))^*$, for all $x_\alpha \in \mathfrak{B}_\alpha$, $\alpha \in \mathbb{F}$.

**Definition 5.** A locally convex vector space $\mathfrak{A}$, with involution $^*$, is called a *C*-inductive locally convex space* if

(i) there exists a family $\{\{\mathfrak{B}_\alpha, \phi_\alpha\}, \alpha \in \mathbb{F}\}$, where $\mathbb{F}$ is a direct set and, for every $\alpha \in \mathbb{F}$, $\mathfrak{B}_\alpha$ is a C*-algebra and $\phi_\alpha$ is a linear injective map of $\mathfrak{B}_\alpha$ into $\mathfrak{A}$, satisfying the above conditions (I.1)–(I.4) and (sch), with $\mathfrak{A}_\alpha = \phi_\alpha(\mathfrak{B}_\alpha)$, $\alpha \in \mathbb{F}$;
(ii) $\mathfrak{A}$ is endowed with the locally convex inductive topology $\tau_{\text{ind}}$ generated by the family $\{\{\mathfrak{B}_\alpha, \phi_\alpha\}, \alpha \in \mathbb{F}\}$.

The family $\{\{\mathfrak{B}_\alpha, \phi_\alpha\}, \alpha \in \mathbb{F}\}$ is called *the defining system of* $\mathfrak{A}$. We notice that the involution is automatically continuous in $\mathfrak{A}[\tau_{\text{ind}}]$.

A C*-inductive locally convex space has a natural positive cone.

An element $x \in \mathfrak{A}$ is called *positive* if there exists $\gamma \in \mathbb{F}$ such that $\phi_\alpha^{-1}(x) \in \mathfrak{B}_\alpha^+$, $\forall \alpha \geq \gamma$.

We denote by $\mathfrak{A}^+$ the set of all positive elements of $\mathfrak{A}$.

Then,

(i) Every positive element $x \in \mathfrak{A}$ is hermitian; i.e., $x \in \mathfrak{A}_h := \{y \in \mathfrak{A} : y^* = y\}$.
(ii) $\mathfrak{A}^+$ is a nonempty convex pointed cone; i.e., $\mathfrak{A}^+ \cap (-\mathfrak{A}^+) = \{0\}$.
(iii) If $\alpha \in \mathbb{F}$ and $x_\alpha \in \mathfrak{B}_\alpha^+$, $\phi_\alpha(x_\alpha)$ is positive.

Moreover, every hermitian element $x = x^*$ is the difference of two positive elements, i.e. there exist $x^+, x^- \in \mathfrak{A}^+$ such that $x = x^+ - x^-$.

Now, let $\mathfrak{A}$ be a C*-inductive locally convex space with defining family of C*-algebras $\{\mathfrak{B}_\alpha; \alpha \in \mathbb{F}\}$ ($\mathbb{F}$ is an index set directed upward). There are also in this case several possibilities: the first one consists in taking elements that have *representatives* in every C*-algebra $\mathfrak{B}_\alpha$ of the family whose norms are uniformly bounded; the second one consists in taking into account the order structure of $\mathfrak{A}$, in the same spirit of the quoted papers of Vidav and Schmüdgen.

## *3.1 Bounded elements and the C*-inductive structure of* $\mathfrak{A}$

In this section we will report definitions and results that can be found in [5], regarding bounded elements defined through the C*-inductive structure of the space.

**Definition 6.** Let $\mathfrak{A}$ be a C*-inductive locally convex space. An element $x \in \mathfrak{A}$ is called *bounded* if $x \in \mathfrak{A}_\alpha$, for every $\alpha \in \mathbb{F}$ and $\sup_{\alpha\in\mathbb{F}} \|x_\alpha\|_\alpha < \infty$.

The set of bounded elements of $\mathfrak{A}$ is denoted by $\mathfrak{A}_b$.

It is easy to see that the set $\mathfrak{A}_b$ is a Banach space under the norm $\|x\|_b = \sup_{\alpha\in\mathbb{F}} \|x_\alpha\|_\alpha$.

In what follows we will consider *-representations of a C*-inductive locally convex space. We recall the basic definitions.

Let $\mathbb{F}$ be a set directed upward by $\leq$. A family $\{\mathscr{H}_\alpha, U_{\beta\alpha}, \alpha, \beta \in \mathbb{F}, \beta \geq \alpha\}$, where each $\mathscr{H}_\alpha$ is a Hilbert space (with inner product $\langle\cdot|\cdot\rangle_\alpha$ and norm $\|\cdot\|_\alpha$) and, for every $\alpha, \beta \in \mathbb{F}$, with $\beta \geq \alpha$, $U_{\beta\alpha}$ is a linear map from $\mathscr{H}_\alpha$ into $\mathscr{H}_\beta$, is called a *directed contractive system of Hilbert spaces* if the following conditions are satisfied

(i) $U_{\beta\alpha}$ is injective;
(ii) $\|U_{\beta\alpha}\xi_\alpha\|_\beta \leq \|\xi_\alpha\|_\alpha, \quad \forall \xi_\alpha \in \mathscr{H}_\alpha$;
(iii) $U_{\alpha\alpha} = I_\alpha$, the identity of $\mathscr{H}_\alpha$;
(iv) $U_{\gamma\alpha} = U_{\gamma\beta}U_{\beta\alpha}, \alpha \leq \beta \leq \gamma$.

A directed contractive system of Hilbert spaces defines a conjugate dual pair $(\mathscr{D}^\times, \mathscr{D})$ which is called the *joint topological limit* [6] of the directed contractive system $\{\mathscr{H}_\alpha, U_{\beta\alpha}, \alpha, \beta \in \mathbb{F}, \beta \geq \alpha\}$ of Hilbert spaces.

**Definition 7.** Let $\mathfrak{A}$ be the C*-inductive locally convex space defined by the system $\{\{\mathfrak{B}_\alpha, \Phi_\alpha\}, \alpha \in \mathbb{F}\}$ as in Definition 5. For each $\alpha \in \mathbb{F}$, let $\pi_\alpha$ be a *-representation of $\mathfrak{B}_\alpha$ in Hilbert space $\mathscr{H}_\alpha$. The collection $\pi := \{\pi_\alpha\}$ is said to be a *-representation of $\mathfrak{A}$ if

(i) for every $\alpha, \beta \in \mathbb{F}$ there exists a linear map $U_{\beta\alpha} : \mathscr{H}_\alpha \to \mathscr{H}_\beta$ such that the family $\{\mathscr{H}_\alpha, U_{\beta\alpha}, \alpha, \beta \in \mathbb{F}, \beta \geq \alpha\}$ is a directed contractive system of Hilbert spaces;

(ii) the following equality holds

$$\pi_\beta(j_{\beta\alpha}(x_\alpha)) = U_{\beta\alpha}\pi_\alpha(x_\alpha)U^*_{\beta\alpha}, \quad \forall x_\alpha \in \mathfrak{B}_\alpha,\ \beta \geq \alpha. \tag{5}$$

In this case we write $\pi(x) = \varinjlim \pi_\alpha(x_\alpha)$ for every $x = (x_\alpha) \in \mathfrak{A}$ or, for short, $\pi = \varinjlim \pi_\alpha$.

The *-representation $\pi$ is said to be *faithful* if $x \in \mathfrak{A}^+$ and $\pi(x) = 0$ imply $x = 0$ (of course, $\pi(x) = 0$ means that there exists $\gamma \in \mathbb{F}$ such that $\pi_\alpha(x_\alpha) = 0$, for $\alpha \geq \gamma$ ).

*Remark 2.* With this definition (which is formally different from that given in [7] but fully equivalent), $\pi(x)$, $x \in \mathfrak{A}$, is not an operator but rather a collection of operators. However, as it was shown in [7], $\pi(x)$ can be regarded as an operator acting on the joint topological limit $(\mathscr{D}^\times, \mathscr{D})$ of $\{\mathscr{H}_\alpha, U_{\beta\alpha}, \alpha, \beta \in \mathbb{F}, \beta \geq \alpha\}$ (see [6]). The corresponding space of operators was denoted by $\mathsf{L}_\mathsf{B}(\mathscr{D}, \mathscr{D}^\times)$; it behaves in the very same way as the space $\mathfrak{L}_\mathsf{B}(\mathscr{D}, \mathscr{D}^\times)$ studied in [5, Section 3] and reduces to it when the family of Hilbert spaces is exactly $\{\mathscr{H}_A; A \in \mathscr{L}^\dagger(\mathscr{D})\}$. The main difference consists in the fact that the $\mathscr{H}_\alpha$'s need not be all subspaces of a certain Hilbert space $\mathscr{H}$.

Let $\pi = \varinjlim \pi_\alpha$ be a faithful representation. Then, for every $\alpha \in \mathbb{F}$, $\pi_\alpha$ is a faithful *-representation of $\mathfrak{B}_\alpha$.

As shown in [7, Proposition 3.16], if a C*-inductive locally convex space $\mathfrak{A}$ fulfills the following conditions

($\mathsf{r}_1$) if $x_\alpha \in \mathfrak{B}_\alpha$ and $j_{\beta\alpha}(x_\alpha) \geq 0$, $\beta \geq \alpha$, then $x_\alpha \geq 0$;
($\mathsf{r}_2$) $e_\beta \in j_{\beta\alpha}(\mathfrak{B}_\alpha)$, $\quad \forall \alpha, \beta \in \mathbb{F}, \beta \geq \alpha$;
($\mathsf{r}_3$) every positive linear functional $\omega = \varinjlim \omega_\alpha$ on $\mathfrak{A}$ satisfies the following property

- if $\alpha \in \mathbb{F}$ and $\omega_\beta(j_{\beta\alpha}(x^*_\alpha)j_{\beta\alpha}(x_\alpha)) = 0$, for some $\beta \geq \alpha$ and $x_\alpha \in \mathfrak{B}_\alpha$, then $\omega_\alpha(x^*_\alpha x_\alpha) = 0$;

then, $\mathfrak{A}$ admits a faithful representation. These conditions, in fact, guarantee that $\mathfrak{A}$ possesses sufficiently many positive linear functionals, in the sense that for every $x \in \mathfrak{A}^+$, $x \neq 0$, there exists a positive linear functional $\omega$ such that $\omega(x) > 0$ [7, Theorem 3.14].

The following theorem provides a relation between the bounded elements of $\mathfrak{A}$ and its *bounded* representations.

**Theorem 2.** *Let $\mathfrak{A}$ be a C*-inductive locally convex space and $x = (x_\alpha) \in \mathfrak{A}$.*

*(i) If $x \in \mathfrak{A}_b$, then, for every representation $\pi = \varinjlim \pi_\alpha$ of $\mathfrak{A}$, one has*

$$\sup_{\alpha \in \mathbb{F}} \|\pi_\alpha(x_\alpha)\|_{\alpha\alpha} < \infty,$$

*where $\|\cdot\|_{\alpha\alpha}$ denote the norm of $\mathfrak{B}(\mathscr{H}_\alpha)$.*

*(ii) Conversely, if* $\mathfrak{A}$ *admits a faithful *-representation* $\pi^f = \varinjlim \pi_\alpha^f$ *and*

$$\sup_{\alpha \in \mathbb{F}} \|\pi_\alpha^f(x_\alpha)\|_{\alpha\alpha} < \infty,$$

*then* $x \in \mathfrak{A}_b$.

### 3.2 Bounded elements and the order structure of $\mathfrak{A}$

Here we collect a series of definitions and results given in [5] about bounded elements of a C*-inductive locally convex space defined by an order cone. As before, we refer to that paper for the proofs.

The reader will immediately realize that the following definitions are very similar to those given in Subsection 2.1.2, however we search here a characterization of bounded elements that originates from the bounded elements of the C*-algebras that give raise to the C*-inductive locally convex space.

Let $\mathfrak{A}$ be a C*-inductive locally convex space. If $x \in \mathfrak{A}$, we put, as before,

$$\Re(x) = \frac{x + x^*}{2} \text{ and } \Im(x) = \frac{x - x^*}{2i}.$$

Both $\Re(x)$ and $\Im(x)$ are symmetric elements of $\mathfrak{A}$.

Assume that $\mathfrak{A}$ has an element $u = u^*$ such that $\|u_\alpha\|_\alpha \leq 1$, for every $\alpha \in \mathbb{F}$, and there exists $\gamma \in \mathbb{F}$ such that $u_\beta = j_{\beta\gamma}(e_\gamma)$ $\forall \beta \geq \gamma$ ($e_\gamma$ is the unit of $\mathfrak{B}_\gamma$). For shortness we call the element $u$ a *pre-unit* of $\mathfrak{A}$. It is not difficult to prove that the pre-unit $u \in \mathfrak{A}$, if any, is unique.

**Definition 8.** Let $\mathfrak{A}$ be a C*-inductive locally convex space with pre-unit $u$. We say that $x \in \mathfrak{A}$ is *order bounded (with respect to u)* if there exists $\lambda > 0$ such that

$$-\lambda u \leq \Re(x) \leq \lambda u \qquad -\lambda u \leq \Im(x) \leq \lambda u.$$

The following theorem shows that the notions of bounded element and of order bounded element we gave within the present section are equivalent.

**Theorem 3.** *Let* $\mathfrak{A}$ *be a C*-inductive locally convex space satisfying condition* ($r_1$)*. Assume that* $\mathfrak{A}$ *has a pre-unit u. Then,* $x \in \mathfrak{A}_b$ *if, and only if, x has a representative for every* $\alpha \in \mathbb{F}$ *(i.e., for every* $\alpha \in \mathbb{F}$*, there exists* $x_\alpha \in \mathfrak{B}_\alpha$ *such that* $x = \phi_\alpha(x_\alpha)$*) and x is order bounded with respect u.*

Now, recalling that the set $\mathfrak{A}_b$ is a Banach space under the norm $\|x\|_b = \sup_{\alpha\in\mathbb{F}} \|x_\alpha\|_\alpha$, we can draw a consequence of Theorem 3.

**Proposition 5.** *Let* $x = x^* \in \mathfrak{A}_b$ *and put*

$$p(x) = \inf\{\lambda > 0; -\lambda u \le x \le \lambda u\}.$$

*Then,* $p(x) = \|x\|_b$.

## References


1. Allan, G.R.: A spectral theory for locally convex algebras. Proc. London. Math. Soc. **15**, 399–421 (1965).
2. Antoine, J-P., Bellomonte, G., Trapani, C.: Fully representable and *-semisimple topological partial *-algebras, Studia Math. **208**, 167–194 (2012).
3. Antoine, J-P., Inoue A., Trapani, C. Partial *-algebras and their operator realizations, Kluwer, Dordrecht, (2002).
4. Antoine, J-P., Trapani, C., Tschinke, F.:Bounded elements in certain topological partial *-algebras, Studia Math. **203**, 223–251, (2011).
5. Bellomonte, G., Di Bella, S. Trapani, C.: Bounded elements of C*-inductive locally convex spaces. Annali di Matematica Pura e Applicata, DOI: 10.1007/s10231-014-0466-7.
6. Bellomonte, G., Trapani, C.: Rigged Hilbert spaces and contractive families of Hilbert spaces, Monatshefte f. Math.**164** (3), 271–285 (2011).
7. Bellomonte, G., Trapani, C.: Quasi *-algebras and generalized inductive limits of C*-algebras. Studia Math. **202**, 165–190 (2011).
8. Fragoulopoulou, M., Trapani, C., Triolo, S., Locally convex quasi *-algebras with sufficiently many *-representations, J. Math. Anal. Appl. **388**, 1180–1193 (2012).
9. Palmer, T.W.: Banach Algebras and the General Theory of *-Algebras, Volume 2, Cambridge Univ. Press, Cambridge (2001).
10. Schmüdgen, K.: A strict Positivstellensatz for the Weyl algebra, Math. Ann. **331**, 779–794 (2005).
11. Vidav, I.: On some *regular rings, Acad. Serbe Sci. Publ. Inst. Math. **13**, 73–80 (1959).

# Some spectral properties for operators acting on Rigged Hilbert spaces

**Salvatore Di Bella**

**Abstract** Operators on Rigged Hilbert spaces have been considered from the 80s of the 20th century on as good ones for describing several physical models whose observable set didn't turn out to be a $C^*$-algebra.

A notion of resolvent set for an operator acting in a rigged Hilbert space $\mathscr{D} \subset \mathscr{H} \subset \mathscr{D}^\times$ is proposed. This set depends on a family of intermediate locally convex spaces living between $\mathscr{D}$ and $\mathscr{D}^\times$, called interspaces. Some properties of the resolvent set and of the corresponding multivalued resolvent function are derived and some examples are discussed.

## 1 Introduction

Spaces of linear maps acting on a *rigged Hilbert space* (RHS, for short)

$$\mathscr{D} \subset \mathscr{H} \subset \mathscr{D}^\times$$

have often been considered in the literature both from a pure mathematical point of view and for their applications to quantum theory [2, 4, 5, 11–13]. One of the motivations about the introduction of partial $^*$-algebras (that came up during the 80s) is within their possible applications in several physical models. In 1964 R. Haag and D. Kastler proposed the *algebraic approach* to quantum systems with infinite degrees of freedom. In this it was supposed that the set of the observable of the system were a $C^*$-algebra and that all the operations concerning the topology could be done without any problem. Nevertheless, this didn't take physicists long to show the existence of quantum models in which this assumption was no longer true; then a new approach was needed. A first step was to introduce *topological partial* $^*$*-algebras* (G. Lassner 1980), the use of which showed a good approach to describe some *spin-systems*. Since then, the development of topological partial $^*$-algebras went straight both in mathematical and physical applications [3, 4, 6, 7, 14].

In this paper we consider the case of $\mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$.

S. Di Bella (✉)
Dipartimento di Matematica e Informatica, Università di Palermo, I-90123 Palermo, Italy
e-mail: svt.dibella@gmail.com



A. Jeribi et al. (eds.), *Applied Mathematics in Tunisia*, Springer Proceedings in Mathematics & Statistics 131, DOI 10.1007/978-3-319-18041-0_25

Let $\mathscr{D}$ be a dense linear subspace of Hilbert space $\mathscr{H}$ and $t$ a locally convex topology on $\mathscr{D}$, finer than the topology induced by the Hilbert norm. Then the space $\mathscr{D}^\times$ of all continuous conjugate linear functionals on $\mathscr{D}[t]$, i.e., the conjugate dual of $\mathscr{D}[t]$, is a linear vector space and *contains* $\mathscr{H}$, in the sense that $\mathscr{H}$ can be identified with a subspace of $\mathscr{D}^\times$. These identifications imply that the sesquilinear form $B(\cdot,\cdot)$ that puts $\mathscr{D}$ and $\mathscr{D}^\times$ in duality is an extension of the inner product of $\mathscr{D}$; i.e., $B(\xi,\eta) = \langle \xi | \eta \rangle$, for every $\xi, \eta \in \mathscr{D}$ (to simplify notations we adopt the symbol $\langle \cdot | \cdot \rangle$ for both of them). The space $\mathscr{D}^\times$ will always be considered as endowed with the *strong dual topology* $t^\times = \beta(\mathscr{D}^\times, \mathscr{D})$. The Hilbert space $\mathscr{H}$ is dense in $\mathscr{D}^\times[t^\times]$.

We get in this way a *Gelfand triplet* or *rigged Hilbert space* (RHS)

$$\mathscr{D}[t] \hookrightarrow \mathscr{H} \hookrightarrow \mathscr{D}^\times[t^\times], \tag{1}$$

where $\hookrightarrow$ denotes a continuous embedding with dense range. Clearly this includes the well-known triplets of distribution spaces

$$\mathscr{D}(\Omega) \hookrightarrow L^2(\Omega, d^n x) \hookrightarrow \mathscr{D}^\times(\Omega)$$

where $\Omega$ is an open of $\mathbb{R}^n$, or

$$\mathscr{S}(\mathbb{R}^n) \hookrightarrow L^2(\mathbb{R}^n) \hookrightarrow \mathscr{S}^\times(\mathbb{R}^n).$$

Let $\mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ denote the vector space of all continuous linear maps from $\mathscr{D}[t]$ into $\mathscr{D}^\times[t^\times]$. In $\mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ an involution $X \mapsto X^\dagger$ can be introduced by the equality

$$\langle X\xi | \eta \rangle = \overline{\langle X^\dagger \eta | \xi \rangle}, \quad \forall \xi, \eta \in \mathscr{D}.$$

Hence $\mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ is a *-invariant vector space. As it is shown in [3, 6, 7, 10], $\mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ can be made into a partial *-algebra by selecting an appropriate family of intermediate spaces between $\mathscr{D}$ and $\mathscr{D}^\times$.

## 2 Operators in RHS

Let $\mathscr{D}[t] \subset \mathscr{H} \subset \mathscr{D}^\times[t^\times]$ be a rigged Hilbert space and $\mathscr{E}[t_\mathscr{E}]$ a locally convex space such that

$$\mathscr{D}[t] \hookrightarrow \mathscr{E}[t_\mathscr{E}] \hookrightarrow \mathscr{D}^\times[t^\times]. \tag{2}$$

Let $\mathscr{E}^\times$ be the conjugate dual of $\mathscr{E}[t_\mathscr{E}]$ endowed with its own strong dual topology $t_{\mathrm{e}}^\times$. Then by duality, $\mathscr{E}^\times$ is continuously embedded in $\mathscr{D}^\times$ and the embedding has dense range. Also $\mathscr{D}$ is continuously embedded in $\mathscr{E}$, but in this case the image of $\mathscr{D}$

is not necessarily dense in $\mathscr{E}^\times$ [4, Example 10.2.21], unless $\mathscr{E}$ is endowed with the Mackey topology $\tau(\mathscr{E}, \mathscr{E}^\times) =: \tau_{\mathscr{E}}$, [8]; in which case we say that $\mathscr{E}$ is an *interspace*. If $\mathscr{E}, \mathscr{F}$ are interspaces and $\mathscr{E} \subset \mathscr{F}$, then $\tau_{\mathscr{F}}$ is coarser than $\tau_{\mathscr{E}}$.

Let $\mathscr{E}, \mathscr{F}$ be interspaces. Let us define

$$\mathscr{C}(\mathscr{E}, \mathscr{F}) := \{X \in \mathfrak{L}(\mathscr{D}, \mathscr{D}^\times) : \exists Y \in \mathfrak{L}(\mathscr{E}, \mathscr{F}),\ Y\xi = X\xi, \forall \xi \in \mathscr{D}\},$$

where $\mathfrak{L}(\mathscr{E}, \mathscr{F})$ denotes the vector space of all continuous linear maps from $\mathscr{E}[\tau_{\mathscr{E}}]$ into $\mathscr{F}[\tau_{\mathscr{F}}]$. It is clear that $X \in \mathscr{C}(\mathscr{E}, \mathscr{F})$ if and only if it has a continuous extension $X_{\mathscr{E}} : \mathscr{E}[\tau_{\mathscr{E}}] \to \mathscr{F}[\tau_{\mathscr{F}}]$. In particular, if $X \in \mathscr{C}(\mathscr{E}, \mathscr{F})$, then $X \in \mathscr{C}(\mathscr{E}, \mathscr{D}^\times)$. The continuous extension of $X$ from $\mathscr{E}$ into $\mathscr{D}^\times$ clearly coincides with $X_{\mathscr{E}}$. Obviously, if $X, Y \in \mathscr{C}(\mathscr{E}, \mathscr{D}^\times)$, then $(X + Y)_{\mathscr{E}} = X_{\mathscr{E}} + Y_{\mathscr{E}}$.

Let now $X, Y \in \mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ and assume there exists an interspace $\mathscr{E}$ such that $Y \in \mathfrak{L}(\mathscr{D}, \mathscr{E})$ and $X \in \mathscr{C}(\mathscr{E}, \mathscr{D}^\times)$; it would then be natural to define

$$X \cdot Y\xi = X_{\mathscr{E}}(Y\xi), \quad \xi \in \mathscr{D}. \tag{3}$$

However, this product is not well defined, because it may depend on the choice of the interspace $\mathscr{E}$. There are in fact examples, due to Kürsten [9, 10], showing that this situation may really occur.

**Definition 2.1.** A family $\mathfrak{F}$ of interspaces in the rigged Hilbert space $(\mathscr{D}[t], \mathscr{H}, \mathscr{D}^\times[t^\times])$ is called a *multiplication framework* if

1. $\mathscr{D} \in \mathfrak{F}$;
2. $\forall\, \mathscr{E} \in \mathfrak{F}$, its conjugate dual $\mathscr{E}^\times$ also belongs to $\mathfrak{F}$;
3. $\forall\, \mathscr{E}, \mathscr{F} \in \mathfrak{F}$, $\mathscr{E} \cap \mathscr{F} \in \mathfrak{F}$.

**Definition 2.2.** Let $\mathfrak{F}$ be a multiplication framework in the rigged Hilbert space $(\mathscr{D}[t], \mathscr{H}, \mathscr{D}^\times[t^\times])$. The product $X \cdot Y$ of two elements of $\mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ is defined, with respect to $\mathfrak{F}$, if there exist three interspaces $\mathscr{E}, \mathscr{F}, \mathscr{G} \in \mathfrak{F}$ such that $X \in \mathscr{C}(\mathscr{F}, \mathscr{G})$ and $Y \in \mathscr{C}(\mathscr{E}, \mathscr{F})$. In this case, the multiplication $X \cdot Y$ is defined by

$$X \cdot Y = (X_{\mathscr{F}} Y_{\mathscr{E}}) \upharpoonright \mathscr{D}$$

or, equivalently, by

$$X \cdot Y\xi = X_{\mathscr{F}} Y\xi, \quad \xi \in \mathscr{D}.$$

Actually, the product so defined does not depend on the particular choice of the interspaces $\mathscr{E}, \mathscr{F}, \mathscr{G} \in \mathfrak{F}$ but it may depend on the choice of $\mathfrak{F}$.

As shown in [4, Theorem 10.2.30], we have

**Theorem 2.3.** *Let $\mathfrak{F}$ be a multiplication framework in the rigged Hilbert space $(\mathscr{D}[t], \mathscr{H}, \mathscr{D}^\times[t^\times])$. Then $\mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$, with the multiplication defined above, is a (non-associative) partial *-algebra.*

## 3 Resolvent and spectrum

In this section we introduce and discuss a notion of spectrum for operators in $\mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$. Even though it would be natural to define the inverse of an injective and surjective $X \in \mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ as the operator $X^{-1} : \mathscr{D}^\times \to \mathscr{D}$ such that $XX^{-1} = I_{\mathscr{D}^\times}$ and $X^{-1}X = I_{\mathscr{D}}$, this approach, as shown in the next proposition, turns out to be too restrictive.

**Proposition 3.1.** *Let $\mathscr{D}[t] \hookrightarrow \mathscr{H} \hookrightarrow \mathscr{D}^\times[t^\times]$ be a rigged Hilbert space and $X \in \mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ a linear bijection. Then there exists a triplet of Hilbert spaces $\mathscr{H}_X \hookrightarrow \mathscr{H} \hookrightarrow \mathscr{H}_X^\times$ such that $\mathscr{D} \subseteq \mathscr{H}_X$ and $\mathscr{D}^\times \subseteq \mathscr{H}_X^\times$.*

Since the existence of global inverses of operators of $\mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ is a so strong condition, one may try to exploit the intermediate structure between $\mathscr{D}$ and $\mathscr{D}^\times$ for a more appropriate definition of the inversion procedure. The fact that once fixed a multiplication framework $\mathfrak{F}$, $\mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ becomes a partial *-algebra [Theorem 2.3] suggests an *algebraic* definition: $Y \in \mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ is the inverse of $X \in \mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ if

$$X \cdot Y \text{ and } Y \cdot X \text{ are well defined and } X \cdot Y\xi = Y \cdot X\xi = \xi, \forall \xi \in \mathscr{D}. \tag{4}$$

This equality, however, does not define $Y$ uniquely, because of possible lack of associativity.

Actually, as we are going to see, this lack of uniqueness will play a fundamental rule on our definition of spectrum, [1].

Let $X \in \mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ and $\mathfrak{F}_0$ a family of interspaces. Assume that there exist $\mathscr{E}, \mathscr{F} \in \mathfrak{F}_0$ such that $X \in \mathscr{C}(\mathscr{E}, \mathscr{F})$. If the extension $X_{\mathscr{E}}$ is bijective from $\mathscr{E}$ into $\mathscr{F}$, then $X_{\mathscr{E}}^{-1}$ exists. If $X_{\mathscr{E}}^{-1}$ is continuous from $\mathscr{F}$ onto $\mathscr{E}$, then its restriction to $\mathscr{D}$ is automatically continuous from $\mathscr{D}[t]$ into $\mathscr{D}^\times[t^\times]$; i.e. $X_{\mathscr{E}}^{-1} \upharpoonright_{\mathscr{D}} \in \mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ and, moreover, $X_{\mathscr{E}}^{-1} \upharpoonright_{\mathscr{D}} \in \mathscr{C}(\mathscr{F}, \mathscr{E})$. If this is the case, and if $\mathfrak{F}_0$ is a multiplication framework, then (4) holds. So that $X_{\mathscr{E}}^{-1} \upharpoonright_{\mathscr{D}}$ is the algebraic inverse of $X$. The converse may fail to be true. For this reason there is no need, in what follows, to consider $\mathfrak{F}_0$ as a multiplication framework.

**Definition 3.2.** Let $X \in \mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ and $\lambda \in \mathbb{C}$. We say that $\lambda$ is a *generalized eigenvalue* of $X$ if there exists an interspace $\mathscr{E}$ such that $X$ has a continuous extension $X_{\mathscr{E}}$ from $\mathscr{E}[t_{\mathscr{E}}]$ into $\mathscr{D}^\times[t^\times]$ and $X_{\mathscr{E}} - \lambda I_{\mathscr{E}}$ is not injective. Any nonzero vector $\xi \in N(X_{\mathscr{E}} - \lambda I_{\mathscr{E}}) \subset \mathscr{E}$ is called a *generalized eigenvector*. If $\mathscr{E} = \mathscr{D}$, we say that $\lambda$ is an eigenvalue of $X$ and elements of $N(X_{\mathscr{D}} - \lambda I_{\mathscr{D}})$ are called eigenvectors.

**Definition 3.3.** Let $X \in \mathfrak{L}(\mathscr{D}, \mathscr{D}^\times)$ and $\mathfrak{F}_0$ be a family of interspaces. The $\mathfrak{F}_0$-*resolvent set* of $X$, $\varrho^{\mathfrak{F}_0}(X)$, consists of the set of complex numbers $\lambda$ satisfying the following conditions: there exist $\mathscr{E}, \mathscr{F} \in \mathfrak{F}_0$, with $\mathscr{E} \subseteq \mathscr{F}$, such that

1. $X \in \mathscr{C}(\mathscr{E}, \mathscr{F})$ and $(X_{\mathscr{E}} - \lambda I_{\mathscr{E}})\mathscr{E} = \mathscr{F}$;
2. $(X_{\mathscr{E}} - \lambda I_{\mathscr{E}})^{-1}$ exists and it is continuous from $\mathscr{F}[\tau_{\mathscr{F}}]$ onto $\mathscr{E}[\tau_{\mathscr{E}}]$.

For consistency of notations we put $\varrho_{\mathscr{E},\mathscr{F}}(X) = \emptyset$ if $X \notin \mathscr{C}(\mathscr{E},\mathscr{F})$. With this convention, one has

$$\varrho^{\mathfrak{F}_0}(X) = \bigcup_{\mathscr{E},\mathscr{F}\in\mathfrak{F}_0} \varrho_{\mathscr{E},\mathscr{F}}(X). \tag{5}$$

The set $\sigma^{\mathfrak{F}_0}(X) := \mathbb{C} \setminus \varrho^{\mathfrak{F}_0}(X)$ will be called the $\mathfrak{F}_0$-*spectrum* of $X$.

*Remark 3.4.* The assumption of continuity in condition (1) can be omitted if we suppose that $\mathscr{E},\mathscr{F}$ are Banach spaces; in this case, in fact, the inverse mapping theorem guarantees the continuity of $(X_{\mathscr{E}} - \lambda I_{\mathscr{E}})^{-1}$.

In particular, as it is shown in [1], if we suppose that $\mathfrak{F}_0$ is a family of interspaces whose elements are Hilbert spaces (in this case we will prefer the notation $\mathscr{B}(\mathscr{E},\mathscr{F})$ to $\mathfrak{L}(\mathscr{E},\mathscr{F})$ ), we get the following properties:

**Theorem 3.5.** *Let $\mathscr{E},\mathscr{F} \in \mathfrak{F}_0$ and $A \in \mathscr{B}(\mathscr{E},\mathscr{F})$. Then*

1. *the set $G(\mathscr{B}(\mathscr{E},\mathscr{F}))$ of all invertible elements of $\mathscr{B}(\mathscr{E},\mathscr{F})$ is open;*
2. *the map $A \in G(\mathscr{B}(\mathscr{E},\mathscr{F})) \to A^{-1} \in \mathscr{B}(\mathscr{F},\mathscr{E})$ is continuous;*
3. *$\varrho_{\mathscr{E},\mathscr{F}}(A)$ is open;*
4. *the function $\lambda \in \varrho_{\mathscr{E},\mathscr{F}}(A) \to (A - \lambda I_{\mathscr{E}})^{-1} \in \mathscr{B}(\mathscr{F},\mathscr{E})$ is analytic on every connected component of $\varrho_{\mathscr{E},\mathscr{F}}(A)$.*

**Proposition 3.6.** *Let $X \in \mathfrak{L}(\mathscr{D},\mathscr{D}^\times)$, $\mathscr{E} \subset \mathscr{F}$ and $\lambda_0 \in \varrho_{\mathscr{E},\mathscr{F}}(X)$. Then there exists $\delta > 0$ such that, for every $\lambda \in \mathbb{C}$ with $|\lambda - \lambda_0| < \delta$, $\lambda \in \varrho_{\mathscr{E},\mathscr{F}}(X)$ and*

$$R_\lambda^{\mathscr{E},\mathscr{F}}(X) = \sum_{n=0}^{+\infty} (\lambda - \lambda_0)^n R_{\lambda_0}^{\mathscr{E},\mathscr{F}}(X)^{(n+1)},$$

*where the series converges in the operator norm of $\mathscr{B}(\mathscr{F},\mathscr{E})$.*

By the definition itself of resolvent for an operator $X \in \mathfrak{L}(\mathscr{D},\mathscr{D}^\times)$, it may happen that a complex number belongs to more than one local resolvent, i.e., $\lambda_0 \in \varrho_{\mathscr{E},\mathscr{F}}(X_{\mathscr{E}}) \cap \varrho_{\mathscr{E}',\mathscr{F}'}(X_{\mathscr{E}'})$ for some $\mathscr{E},\mathscr{E}',\mathscr{F},\mathscr{F}' \in \mathfrak{F}_0$. In this case we denote $(\mathbf{X} - \lambda_0\mathbf{I})^{-1}$ the collection of all resolvent operators corresponding to $\lambda_0$ and by $\lambda \in \varrho^{\mathfrak{F}_0}(X) \to (\mathbf{X} - \mathbf{–I})^{-1}$ the corresponding multivalued resolvent function, where its restriction to any $\mathscr{E} \in \mathfrak{F}_0$ can be seen as a single valued branch.

**Definition 3.7.** Let $\mathscr{E},\mathscr{E}',\mathscr{F},\mathscr{F}' \in \mathfrak{F}_0$ and $B \in \mathscr{B}(\mathscr{E},\mathscr{F})$, $C \in \mathscr{B}(\mathscr{E}',\mathscr{F}')$. We say that $B$ and $C$ are equivalent, and write $B \equiv C$, if $B \upharpoonright_{\mathscr{D}} = C \upharpoonright_{\mathscr{D}}$.

Some favorable situations are given by the two following propositions:

**Proposition 3.8.** *Let $\lambda_0 \in \varrho_{\mathscr{E},\mathscr{F}}(X_{\mathscr{E}}) \cap \varrho_{\mathscr{E}',\mathscr{F}'}(X_{\mathscr{E}'})$ for some $\mathscr{E},\mathscr{E}',\mathscr{F},\mathscr{F}' \in \mathfrak{F}_0$. Let us assume that $\mathscr{E} \subset \mathscr{E}'$. The corresponding resolvent functions are equivalent on some open neighborhood of $\lambda_0$ and they are direct analytic continuations of each other.*

**Proposition 3.9.** *Let $\lambda_0 \in \varrho_{\mathscr{E},\mathscr{F}}(X_{\mathscr{E}}) \cap \varrho_{\mathscr{E}',\mathscr{F}'}(X_{\mathscr{E}'})$ for some $\mathscr{E}, \mathscr{E}', \mathscr{F}, \mathscr{F}' \in \mathfrak{F}_0$. If there exist $\mathscr{G}$ and $\mathscr{G}'$ in $\mathfrak{F}_0$ such that $\mathscr{G} \subseteq \mathscr{E} \cap \mathscr{E}'$ and $\lambda_0 \in \varrho_{\mathscr{G},\mathscr{G}'}(X)$, then the functions $\lambda \to (X_{\mathscr{E}} - \lambda I_{\mathscr{E}})^{-1}$ and $\lambda \to (X_{\mathscr{E}'} - \lambda I_{\mathscr{E}'})^{-1}$ are analytic continuations of each other in some open connected set containing $\lambda_0$.*

## 3.1 Examples

Let $A$ be a self-adjoint operator in Hilbert space $\mathscr{H}$. The space $\mathscr{D} = \mathscr{D}^\infty(A)$, endowed with its natural topology $t_A$, defined by the seminorms $p_n(\xi) = \|A^n\xi\|$, $n \in \mathbb{N}$, generates in canonical way an RHS, with $\mathscr{D}$ a Fréchet space. For every $n \in \mathbb{N}$ we denote by $\mathscr{H}_n$ the Hilbert space obtained by endowing $D(A^n)$ with its graph norm $\|\cdot\|_n := \|(I + A^{2n})^{1/2} \cdot \|$ and by $\mathscr{H}_{-n}$ the space obtained by completing $\mathscr{H}$ with respect to the norm $\|\cdot\|_{-n} := \|(I + A^{2n})^{-1/2} \cdot \|$. Put $\mathscr{H}_0 := \mathscr{H}$. Then, the family of spaces $\{\mathscr{H}_n;\, n \in \mathbb{Z}\}$ is totally ordered; namely,

$$\cdots \mathscr{H}_{n+1} \subset \mathscr{H}_n \subset \cdots \subset \mathscr{H} = \mathscr{H}_0 \subset \mathscr{H}_{-n} \subset \mathscr{H}_{-n-1} \cdots$$

Let us put $S = A \upharpoonright_{\mathscr{D}}$ and take $\mathfrak{F}_0 = \{\mathscr{H}_n; n \in \mathbb{Z}\}$. The operator $A$ (or its extension by duality denoted by the same symbol) maps $\mathscr{H}_n$ in $\mathscr{H}_{n-1}$, $\forall n \in \mathbb{Z}$ continuously; hence, $S \in C(\mathscr{H}_n, \mathscr{H}_{n-1})$, for every $n \in \mathbb{Z}$. Let us denote by $\varrho_{\mathscr{H}}(A)$ the usual resolvent of $A$. For shortness, we will put $\varrho_{n,m}(S) := \varrho_{\mathscr{H}_n,\mathscr{H}_m}(S)$.

**Proposition 3.10.** *Let $A$ be a self-adjoint operator, $\mathscr{D}$ and $\mathfrak{F}_0$ as above. Then $\varrho^{\mathfrak{F}_0}(S) = \varrho_{\mathscr{H}}(A)$.*

*Example 1.* Let $S$ be a closed symmetric operator with equal and finite defect indices. Again we put

$$\mathscr{D}^\infty(S) = \bigcap_{n \geq 0} \mathrm{d}(S^n)$$

and, also in this case, $\mathscr{D}^\infty(S)$ is dense in $\mathscr{H}$ [13, Prop. 1.6.1]. If $S'$ is a self-adjoint extension of $S$, we clearly have

$$\mathscr{D}(S^n) \subset \mathscr{D}(S'^n), \quad \forall n \geq 1$$

and then

$$\mathscr{D}^\infty(S) \subset \mathscr{D}^\infty(S').$$

Let us assume that $S$ has a family $\{S_\alpha\}_{\alpha \in I}$ of self-adjoint extensions. We put $\mathscr{H}_{\alpha,n} = \mathscr{D}(S_\alpha^n)$ endowed with the graph norm as before and consider

$$\mathfrak{F}_0 = \{\mathscr{H}_{\alpha,n}; \alpha \in I, n \in \mathbb{N}\}$$

Then $S \in \mathscr{C}(\mathscr{H}_{\alpha,n}, \mathscr{H}_{\beta,m})$ if and only if $\alpha = \beta$ and $m \leq n-1$. By the previous result, it follows that

$$\varrho_{\mathscr{H}_{\alpha,n},\mathscr{H}_{\alpha,n-1}}(S) = \varrho_{\mathscr{H}}(S_\alpha).$$

Hence $\varrho^{\mathfrak{F}_0}(S) = \cup_{\alpha \in I} \; \varrho_{\mathscr{H}}(S_\alpha)$.

*Example 2.* Let us consider the operator $H_0 = -\frac{d^2}{dx^2}$. A case of interest arises if we impose a boundary condition by taking, for instance, $\mathscr{D} := \mathscr{S}_y := \{f \in \mathscr{S}; f(y) = 0\}$, $y \in \mathbb{R}$, with the topology induced by $\mathscr{S}$. This domain is used when *perturbing* the free Hamiltonian with a $\delta$-interaction centered at $y$ [2]. The domain of the closure $H_1$ of $H_0$ is $W^{2,2}_y(\mathbb{R}) = \{f \in W^{2,2}(\mathbb{R}); f(y) = 0\}$. As shown in [2, Th. 3.1.1], the operator $H_1$ is no longer self-adjoint; it has, in fact, defect indices (1,1) and, for each $\alpha \in \mathbb{R}$, it possesses a self-adjoint extension $H_\alpha$. The domain of $H_\alpha$ is

$$D(H_\alpha) = \left\{g \in W^{1,2}(\mathbb{R}) \cap W^{2,2}(\mathbb{R} \setminus \{y\}) : \; g'(y^+) - g'(y^-) = \alpha g(y)\right\}.$$

As for the spectrum, we have

$$\sigma_{\mathscr{H}}(H_\alpha) = \begin{cases} \mathbb{R}^+ \cup \{0\} & \text{if } \alpha \geq 0 \\ \mathbb{R}^+ \cup \{-\frac{\alpha^2}{4}, 0\} & \text{if } \alpha < 0, \end{cases}$$

since for $\alpha < 0$, $-\frac{\alpha^2}{4}$ is an eigenvalue of $H_\alpha$.

Then, proceeding as in Example 1, we get that

$$\varrho^{\mathfrak{F}_0}(H) = \bigcup_{\alpha \in \mathbb{R}} \varrho_{\mathscr{H}}(H_\alpha) = \mathbb{C} \setminus \{\mathbb{R}^+ \cup \{0\}\}$$

where $\mathfrak{F}_0 = \{\mathscr{H}_{\alpha,n}; \alpha \in I, n \in \mathbb{N}\}$ (with $\mathscr{H}_{\alpha,n} = \mathscr{D}(H^n_\alpha)$ endowed with the graph norm, as before). Hence, also in this case, we get

$$\sigma^{\mathfrak{F}_0}(H) = \mathbb{R}^+ \cup \{0\}.$$

*Example 3.* As it is well known, the Hermite functions defined by $\phi_0(x) = \pi^{-1/4} e^{-x^2/2}$ and

$$\phi_n(x) = (2^n n!)^{-1/2} (-1)^n \pi^{-1/4} e^{x^2/2} \left(\frac{d}{dx}\right)^n e^{-x^2}$$

constitute an orthonormal basis of $L^2(\mathbb{R})$. If $f \in \mathscr{S}$, then $f$ has the expansion

$$f = \sum_{n=0}^{\infty} c_n \phi_n, \text{ with } \sup_n |c_n| n^m < \infty, \; \forall m \in \mathbb{N} \tag{6}$$

and the series converges in the topology of $\mathcal{S}$. The space of sequences $\{c_n\}$ satisfying, for a given $m \in \mathbb{N}$,

$$\sup_n |c_n| n^m < \infty,$$

will be denoted by $\mathsf{s}_m$. We will indicate with $\mathsf{s}$ the so-called space of *rapidly decreasing sequences*; i.e., $\mathsf{s} = \bigcap_{m\in\mathbb{N}} \mathsf{s}_m$.

An element $F \in \mathcal{S}^\times$ can be represented as

$$F = \sum_{n=0}^{\infty} b_n \phi_n, \text{ with } |b_n| \leq M(1+n)^s, \quad \text{for some } M > 0,\ s \in \mathbb{N}, \tag{7}$$

the series being weakly convergent.

Let now $\{a_n\}$ be a sequence of complex numbers such that

$$\forall \{c_n\} \in \mathsf{s},\ \exists\, m \in \mathbb{N} \text{ such that } \sup_n \frac{|a_n||c_n|}{(1+n)^m} < \infty. \tag{8}$$

Then

$$f = \sum_{n=0}^{\infty} c_n \phi_n \mapsto Af := \sum_{n=0}^{\infty} a_n c_n \phi_n$$

defines a linear map from $\mathcal{S}$ into $\mathcal{S}^\times$. Since $\mathcal{S}$ is a reflexive Fréchet space, it is sufficient to check that $A$ is continuous from $\mathcal{S}[\sigma(\mathcal{S}, \mathcal{S}^\times)]$ into $\mathcal{S}^\times[\sigma(\mathcal{S}^\times, \mathcal{S})]$. Continuity follows immediately from the fact that the map

$$A^\dagger : f = \sum_{n=0}^{\infty} d_n \phi_n \mapsto A^\dagger f := \sum_{n=0}^{\infty} \overline{a_n} d_n \phi_n, \quad \{d_n\} \in \mathsf{s},$$

is the adjoint of $A$. Hence $A \in \mathfrak{L}(\mathcal{S}, \mathcal{S}^\times)$. A natural choice of $\mathfrak{F}_0$ consists in taking the spaces $\mathcal{S}_m$ whose elements are all $F \in S^\times$ for which the expansion (7) has coefficients in $\mathsf{s}_m$ and their dual $\mathcal{S}_m^\times$. It is easy to check that every $a_n$ is an eigenvalue of $A$. Thus, if $\lambda \notin \overline{\{a_n,\, n \in \mathbb{N}\}}$, the sequence $\left\{\frac{1}{a_n - \lambda}\right\}$ is bounded. For these values of $\lambda$, the operator $(A - \lambda I)^{-1}$ maps $\mathcal{S}_m$ into $\mathcal{S}_m$ (and $\mathcal{S}_m^\times$ into $\mathcal{S}_m^\times$) continuously. Hence $\sigma^{\mathfrak{F}_0}(A) = \overline{\{a_n,\, n \in \mathbb{N}\}}$.

**Acknowledgements** The author wishes to thank both the organizers of ICAAM 2013 and the PhD School of the Department of Mathematics and Computer Science of Palermo for having given him the opportunity of participating and discussing the results of this paper.

## References

1. Bellomonte,G., Di Bella,S., Trapani,C. : Operators in rigged Hilbert spaces: some spectral properties, J. Math. Anal. Appl., **411**, 931–946 (2014).
2. Albeverio,S., Gesztesy,F., Høegh-Krohn,R., Holden,H.: Solvable Models in Quantum Mechanics. Springer Verlag, New York, (1988).
3. Antoine,J-P., Bellomonte,G., Trapani,C.: Fully representable and *-semisimple topological partial *-algebras, Studia Mathematica, **208** 167–194 (2012)
4. Antoine,J-P., Inoue, A., Trapani,C.: Partial *-algebras and their operator realizations. Kluwer, Dordrecht, (2002)
5. Antoine,J-P., Trapani,C.: Partial Inner Product Spaces — Theory and Applications, Springer Lecture Notes in Mathematics, vol. 1986, Berlin, Heidelberg, (2009)
6. Bellomonte,G., Trapani,C.: Rigged Hilbert spaces and contractive families of Hilbert spaces, Monatshefte f. Math., **164**, 271–285 (2011)
7. Bellomonte,G.,Trapani, C.: Quasi *-algebras and generalized inductive limits of C*-algebras Studia Mathematica, **202**,165–190 (2011)
8. Köthe,G.: Topological Vector Spaces II, Springer, Heidelberg (1979)
9. Kürsten,K-D.: The completion of the maximal $O_p$*-algebra on a Fréchet domain, Publ. Res. Inst. Math. Sci., Kyoto Univ. **22**, 151–175 (1986)
10. Kürsten,K-D.,Läuter, M.: An extreme example concerning factorization products on the Schwartz space $\mathfrak{S}(\mathbb{R}^n)$, Note Mat. **25**,31–38 (2005/06)
11. Lassner,G.: Topological algebras and their applications in Quantum Statistics, Wiss. Z. KMU-Leipzig, Math.-Naturwiss. R. **30**,572–595 (1981)
12. Lassner,G.: Algebras of unbounded operators and quantum dynamics, **A 124**, 471–480 (1984)
13. Schmüdgen,K.: Unbounded operator algebras and representation theory, Birkhäuser Verlag, Basel (1990)
14. Trapani,C., Tschinke,F.: Partial multiplication of operators in rigged Hilbert spaces, Integral Equations Operator Theory **51**, 583–600 (2005)