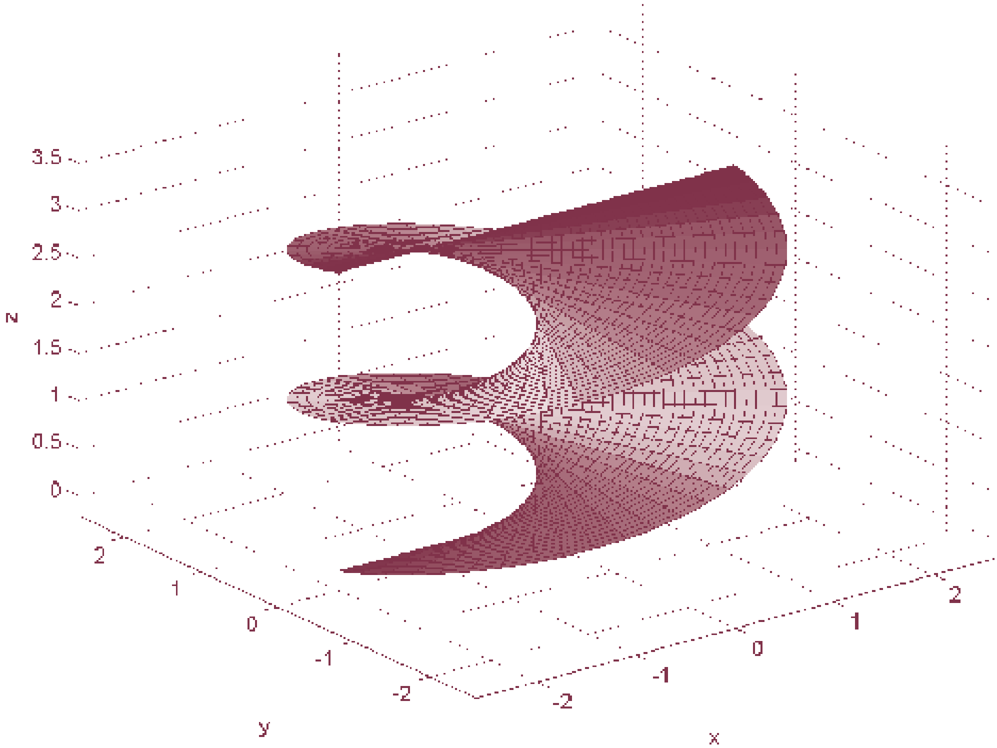
STUDIA UNIVERSITATIS
BABEŞ-BOLYAI



MATHEMATICA

1/2021

STUDIA
UNIVERSITATIS BABEŞ-BOLYAI
MATHEMATICA

1/2021

YEAR	**(LXVI) 2021**
MONTH	**MARCH**
ISSUE	**1**

S T U D I A
UNIVERSITATIS BABEŞ-BOLYAI

MATHEMATICA
1

Redacţia: 400084 Cluj-Napoca, str. M. Kogălniceanu nr. 1
Telefon: 0264 405300

CONTENTS

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 3–4

Gheorghe Moroşanu – On the occasion of his 70th birthday

Adrian Petruşel

Gheorghe Moroşanu was born on April 30, 1950, in Darabani, Botoşani County, Romania. After a 12-year period of education, from primary to high school (1957-1969), in 1969, Gheorghe Moroşanu started studying Mathematics at "Alexandru Ioan Cuza" University in Iaşi, Romania. In 1981, under the joint supervision of Adolf Haimovici and Viorel Barbu, he obtained his Ph.D. in Mathematics with a dissertation entitled *Qualitative Problems for Nonlinear Differential Equations of Accretive Type in Banach Spaces*.

Regarding his teaching or research positions, we note that Gheorghe Moroşanu was, between 1991 and 2004, Full Professor of Mathematics at "Alexandru Ioan Cuza" University in Iaşi, after previously holding positions (since 1975) at the same university. Between 2001 and 2002, Professor Moroşanu held a research position at the University of Stuttgart, Germany. In 2002, Professor Moroşanu joined the Department of Mathematics and its Applications, Central European University, Budapest, Hungary. He served this institution from 2002 to 2020, acting as Head of Department between 2004 and 2012. Since 2015, Gheorghe Moroşanu has been Invited Professor of the Babeş-Bolyai University in Cluj-Napoca. He was also a visiting professor at the University of Jyväskylä, Finland (1989) and at Ohio University, Athens, Ohio, U.S.A. (1998 and 2000).

Professor Moroşanu was a prolific Ph.D. supervisor; the following individuals (in alphabetical order) have completed their Ph.D. theses under his supervision or co-supervision: Muhammad Ahsan (2013), Panait Anghel (1999), Narcisa Apreutesei (1999), Luminiţa Barbu (1998), Oganeditse A. Boikanyo (2011), Christian Coclici (1998), Nicuşor Costea (2015), Paul Georgescu (2008), Tihomir Gyulov (2010), Alexandru Kristály (2010), Gabriela Lorelai Liţcanu (2001), Rodica Luca (1995), Mihai Mihăilescu (2010), Viorica Venera Motreanu (2003), and Andras Sereny (2008).

Concerning his rich teaching and research activities, the main fields of interest of Professor Gheorghe Moroşanu were: Ordinary and Partial Differential Equations, Difference Equations, Calculus of Variations, Evolution Equations in Banach Spaces, Fluid Mechanics, Singular Perturbation Theory, and various topics in Applied Mathematics. Gheorghe Moroşanu is the author or co-author of 16 books (monographs and textbooks) as well as of more than 150 research papers in top-ranked journals or proceedings. Gheorghe Moroşanu was director or (main) investigator for several research grants in Romania and abroad.

His research achievements have had a great impact on the mathematical community; his publications have collected until now over 1500 citations, while the current Google Scholar H-index of Gheorghe Moroşanu is 19. In 1983, he was awarded the Gheorghe Lazar Prize of the Romanian Academy in recognition of his outstanding contributions to the theory of hyperbolic partial differential equations. Along the same lines, Professor Gheorghe Moroşanu is Doctor Honoris Causa of Ovidius University in Constanţa and of Craiova University as well as Professor Honoris Causa of Babeş-Bolyai University, Cluj-Napoca, Romania. In 2020, he became a corresponding member of the Academy of Romanian Scientists, Romania. During all these years, Professor Moroşanu visited many institutions for research purposes, such as: International Centre for Theoretical Physics Trieste, University of Jyväskylä, University of Stuttgart, Ohio University, Athens, Ohio, Technical University München, University of Rousse, Babeş-Bolyai University Cluj-Napoca, University of Iowa, Ovidius University in Constanţa, Craiova University, Simion Stoilow Institute of Mathematics of the Romanian Academy, and many others. He cooperated with many researchers on various topics within his areas of interest or neighboring areas and disciplines, including biology, chemistry, economics, engineering, mechanics and physics.

The authors included in this issue are happy and honored to dedicate their papers to Professor Gheorghe Moroşanu, for his long and outstanding career and for the remarkable achievements in the field of mathematics.

The editors of the journal would like to thank all authors who contributed to this special issue and the reviewers who kindly accepted the invitation to provide their expertise and gave constructive comments.

Adrian Petruşel
Babeş-Bolyai University, Faculty of Mathematics and Computer Science
Kogălniceanu Str. no. 1, 400084 Cluj-Napoca, Romania
e-mail: petrusel@math.ubbcluj.ro

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 5–15
DOI: 10.24193/subbmath.2021.1.01

Applications of implicit parametrizations

Dan Tiba

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. We review several applications of the implicit parametrization theorem in optimization. In nonlinear programming, we discuss both new forms, with less multipliers, of the known optimality conditions, and new algorithms of global type. For optimal control problems, we analyze the case of mixed equality constraints and indicate an algorithm, while in shape optimization problems the emphasis is on the new penalization approach.

Mathematics Subject Classification (2010): 34A34, 49K21, 49Q10, 26B10.

Keywords: Implicit functions and implicit parametrizations, optimization, optimal control, shape optimization.

1. Introduction

In several papers [17, 10, 19], a constructive extension was proposed for the classical implicit functions theorem, involving implicit parametrizations in finite dimensional spaces. While it is intuitive that implicit parametrizations offer, in general, a more advantageous representation of implicitly defined manifolds, the representation is even global in some important cases. For instance, in dimension two, in the general setting of the Poincaré-Bendixson theorem [4, 14], the implicit parametrization that we construct, is always global [18]. In dimension three, we quote the example of the torus, from [10]. In fact, this is an important question for the applications: when have the implicit parametrizations a global character, also in dimension three?

This new representation of manifolds (of arbitrary dimension and codimension) was intended for applications in geometric optimization problems and we quote [18, 7, 21, 22] for recent results in this respect. It turns out that it is also useful in mathematical programming and in optimal control as shown in [23, 20, 25].

In this paper, we briefly review such results and their possible extensions. The Section 2 is devoted to the implicit parametrization question. In Section 3, applications in nonlinear programming and optimal control are briefly discussed. The last section includes some relevant properties obtained in shape optimization and their possible generalizations.

2. Implicit parametrizations

We consider a system of $l < d$ equalities defined in some bounded open set $\Omega \subset R^d$, with l, d natural numbers:

$$F_1(\overline{x}) = F_2(\overline{x}) = \ldots = F_l(\overline{x}) = 0,\ \overline{x} \in \Omega. \tag{2.1}$$

Above we assume that $F_1, F_2, \ldots, F_l \in C^1(\overline{\Omega})$ and there is $\overline{x}_0 \in \Omega$ such that (2.1) is satisfied and

$$\frac{D(F_1, F_2, \ldots, F_l)}{D(x_1, x_2, \ldots, x_l)} \neq 0 \text{ in } \overline{x}_0 = (x_1^0, x_2^0, \ldots, x_d^0). \tag{2.2}$$

Notice that (2.2) remains valid on a neighbourhood V of $\overline{x}_0$, $V \subset \Omega$. In V, we define the linear algebraic system

$$v(\overline{x}) \cdot \nabla F_j(\overline{x}) = 0,\ j = \overline{1, l}, \tag{2.3}$$

where the unknown vector $v(\overline{x}) \in R^d$.

It is known that (2.1), under condition (2.2), defines a $d-l$ dimensional manifold contained in Ω and $\nabla F_1(\overline{x}), \nabla F_2(\overline{x}), \ldots, \nabla F_l(\overline{x})$ are a basis in the normal space at $\overline{x}$, to this manifold.

Therefore, any solution to (2.3) is a vector in the tangent space to this manifold and we fix $v_1, v_2, \ldots, v_{d-l}$ as continuous (in V) independent solutions of (2.3) that is a basis in the tangent space to the manifold. The choice of $v_1, v_2, \ldots, v_{d-l}$ is not unique [19].

We associate to them a system of nonlinear partial derivatives of order one:

$$\frac{\partial y_1(t_1)}{\partial t_1} = v_1(y_1(t_1)),\ t_1 \in I_1 \subset R,\ y_1(0) = \overline{x}_0; \tag{2.4}$$

$$\frac{\partial y_2(t_1, t_2)}{\partial t_2} = v(y_2(t_1, t_2)),\ t_2 \in I_2(t_1) \subset R,\ y_2(t_1, 0) = y_1(t_1); \tag{2.5}$$

$$\ldots$$

$$\ldots$$

$$\begin{aligned} \frac{\partial y_{d-l}(t_1, t_2, \ldots, t_{d-l})}{\partial t_{d-l}} &= v_{d-l}(y_{d-l}(t_1, t_2, \ldots, t_{d-l})), \\ t_{d-l} &\in I_{d-l}(t_1, t_2, \ldots, t_{d-l}), \\ y_{d-l}(t_1, t_2, \ldots, t_{d-l-1}, 0) &= y_{d-l-1}(t_1, t_2, \ldots, t_{d-l-1}). \end{aligned} \tag{2.6}$$

The system (2.4) - (2.6) has an iterated character.
Each equation has one supplementary independent variable and the initial condition corresponding to it is given by the solution of the previous equation (by $\overline{x}_0$ in the first one). Moreover, each equation includes just one derivative, therefore (2.4) - (2.6) is in fact a system of $d-l$ ordinary differential subsystems, each of dimension d. By $I_1, I_2(t_1), \ldots, I_{d-l}(t_1, t_2, \ldots, t_{d-l})$ we denote the corresponding existence intervals,

around the origin. The existence is ensured by the Peano theorem, due to the continuity of $v_1, v_2, \ldots, v_{d-l}$. The independent variables not involved in derivation, play the role of parameters and they enter just via the initial condition. The numerical solution via Matlab is standard and easy.

Furthermore, each of the system (2.4), (2.5), ..., (2.6) solves an inverse problem: given $F_1, F_2, \ldots, F_l \in C^1(\overline{\Omega})$ with conditions (2.1), (2.2), the mentioned systems are constructed in such a way that $F_1, F_2, \ldots, F_l$ are prime integrals for any of them (see [19]).

Theorem 2.1. *For every $k = \overline{1, l}$, $j = \overline{1, d-l}$, we have*

$$F_k(y_j(t_1, t_2, \ldots, t_j)) = 0, \tag{2.7}$$

for any $(t_1, t_2, \ldots, t_j) \in I_1 \times I_2(t_1) \times \ldots \times I_j(t_1, t_2, \ldots, t_{j-1})$.

Due to the conservation property in (2.7) and to the examples in [17], [10], we call such systems to be of Hamiltonian type. They have unexpected properties.

Theorem 2.2. *Under condition (2.2), each system has the uniqueness property in V, the intervals $I_2(t_1), \ldots, I_{d-l}(t_1, t_2, \ldots, t_{d-l-1})$ may be chosen independently of the parameters and the unique solutions of (2.4), (2.5), ..., (2.6) are of class C^1 in each of their arguments and*

$$\frac{\partial y_{d-l}}{\partial t_k}(t_1, t_2, \ldots, t_{d-l}) = v_k(y_{d-l}(t_1, \ldots, t_{d-l})), k = \overline{1, d-l}. \tag{2.8}$$

Relation (2.8) is immediately extended to $y_1, \ldots, y_{d-l-1}$ due to the initial conditions in (2.4), (2.5), ..., (2.6).

The most important property obtained via (2.4), (2.5), ..., (2.6) is the following.

Theorem 2.3. *Under the above assumptions, the mapping*

$$y_{d-l} : I_1 \times I_2 \times \ldots \times I_{d-l} \to R^d$$

is regular and one-to-one on its image.

That is, y_{d-l} gives a parametrization of the manifold (2.1) around $\overline{x}$. In [19], it is also shown that the classical implicit functions theorem may be obtained as well as a special case of the above constructive approach. However, as we have already argued, parametrizations offer a more complete description of the manifold.

We also recall that the classical hypothesis (2.2) may be dropped and a generalized solution of the system (2.1) may be introduced and studied according to [17], [19].

In dimension two, the iterated system (2.4), (2.5), ..., (2.6) becomes the simplest Hamiltonian system associated to some $g \in C^1(\overline{\Omega})$, a new notation of F_1, such that $g(\overline{x}_0) = 0$, $\nabla g(\overline{x}_0) \neq 0$, $\overline{x}_0 \in \Omega \subset R^2$, which correspond to the conditions (2.1), (2.2):

$$\begin{aligned} x_1'(t) &= -\frac{\partial g}{\partial x_2}(x_1(t), x_2(t_1)),\ t \in I, \\ x_2'(t) &= \frac{\partial g}{\partial x_1}(x_1(t), x_2(t_1)),\ t \in I, \end{aligned} \tag{2.9}$$

$$(x_1(0), x_2(0)) = \overline{x}_0. \tag{2.10}$$

Obviously, the Theorem 2.1, Theorem 2.2, Theorem 2.3 remain valid for the system (2.9), (2.10), including relations (2.7), (2.8). We introduce now the hypothesis

$$|\nabla g(x_1, x_2)| > 0 \text{ on } G = \{(x_1, x_2) \in \Omega;\ g(x_1, x_2) = 0\}, \tag{2.11}$$

which is a reformulation of the hypothesis in the Poincaré - Bendixson theorem [4], [14], for (2.9), (2.10).

For convenience, we also assume that

$$g(x_1, x_2) > 0 \text{ on } \partial\Omega. \tag{2.12}$$

Theorem 2.4. *Under conditions* (2.11), (2.12), *G is a finite union of disjoint closed curves, without self intersections and not intersecting $\partial\Omega$, parametrized by the solution of* (2.9), (2.10), *when some initial condition $\overline{x}_0$ is chosen on each of its components.*

This result was proved in [18] and gives the global existence and the periodicity of the solution for the Hamiltonian system (2.9), (2.10). It has an important role in the analysis of shape optimization problems in dimension two, which is a case of interest [18], [7].

Remark 2.5. The question of the extension of Thm. 2.4 to dimension three or higher, is open, [10]. This is mainly due to the fact that the Poincare-Bendixson theorem is valid just in dimension two. The extension (of interest in the setting of shape optimization problems) refers to the iterated Hamiltonian systems (2.4), (2.5), ..., (2.6) and consists of finding reasonable sufficient conditions ensuring that the obtained manifold is closed and the representation via (2.4), (2.5), ..., (2.6) is global.

3. Optimization and optimal control

We discuss here the general constrained nonlinear programming problem in R^d:

$$\mathrm{Min}\{h(x_1, \dots, x_d)\} \tag{3.1}$$

subject to (2.1) and to inequality constraints

$$G_j(x_1, \dots, x_d) \le 0,\ j = \overline{1,m}, \tag{3.2}$$

where $h, F_i, i = \overline{1,l}, G_j, j = \overline{1,m}$ are in $C^1(R^d)$ and the classical Mangasarian - Fromowitz assumption, see [1], is valid.

That is (2.2) is assumed and there is $e \in R^d$ such that

$$\nabla F_i(\overline{x}_0)e = 0,\ i = \overline{1,l}, \nabla G_j(\overline{x}_0)e < 0, j \in I(\overline{x}_0). \tag{3.3}$$

Here, $I(\overline{x}_0)$ is the set of indices $j = \overline{1,m}$ of the active inequality constraints at $\overline{x}_0$. By using Thm. 2.3 (the special case of implicit functions, when the parametrization y_{d-l} has the last $d-l$ components given by the coordinates in R^d, see [19]) we obtain the reduced optimization problem that involves just inequality constraints:

$$\mathrm{Min}\{h(y^1_{d-l}, y^2_{d-l}, \dots, y^l_{d-l}, t_1 + x^0_{l+1}, \dots, t_{d-l} + x^0_d)\} \tag{3.4}$$

subject to

$$G_j(y^1_{d-l}, y^2_{d-l}, \dots, y^l_{d-l}, t_1 + x^0_{l+1}, \dots, t_{d-l} + x^0_d) \le 0, j = \overline{1,m}, \tag{3.5}$$

where $(t_1, \ldots, t_{d-l})$ is in a neighbourhood of the origin in R^{d-l}.

It turns out (see [23]) that the reduced problem (3.4), (3.5) also satisfies the Mangasarian - Fromowitz condition (3.3) (adapted to this setting) in the origin of R^{d-l}. Using derivation formulas as in (2.8), we get

Theorem 3.1. *Let $\overline{x}_0$ be a local solution of the problem* (3.1), (2.1), (3.2). *Then, there are $\beta_j \geq 0$, $j = \overline{1, m}$ such that*

$$0 = \nabla h(\overline{x}_0) \cdot v_s(\overline{x}_0) + \sum_{j=1}^{m} \beta_j \nabla G_j(\overline{x}_0) \cdot v_s(\overline{x}_0), s = \overline{1, d-l}, \tag{3.6}$$

$$0 = \beta_j G_j(\overline{x}_0), j = \overline{1, m}. \tag{3.7}$$

This is a simplified version of the KKT optimality conditions since the multipliers associated to (2.1) are eliminated. If we consider just the optimization problem with equality constraints (3.1), (2.1), then (finally) we obtain the optimality conditions in Fermat form by taking $\beta_j = 0, j = \overline{1, m}$, in Thm. 3.1.

Notice that the first term in (3.6) is the tangential derivative of the cost in $\overline{x}_0$ (the components of $\nabla h(\overline{x}_0)$ from the tangent plane to the manifold of equality constraints). This allows to formulate an algorithm of gradient type with projection for the problem (3.1), (2.1). The novelty here is that the projection can be effectively computed as y_{d-l} in each step of this algorithm.

Details in [23], and we underline that the question of the computation of the projection is this main drawback for this type of numerical methods, Ciarlet [2]. Moreover, in the general case of the problem (3.1), (2.1), (3.2) we notice that the inequality constraints that are not active at $\overline{x}_0$ define a neighbourhood of $\overline{x}_0$. One can reformulate equivalently the problem on this neighbourhood and involving just the equality constraints (2.1) and the equality constraints obtained from the active inequality constraints. Therefore, the conditions (3.6), (3.7) reduce to (3.6) with $\beta_j = 1, j \in I(\overline{x}_0)$ and $\beta_j = 0$ otherwise and $\{v_s\}$ restricted to include just the basis of the tangent space to the manifold defined by all these equality constraints (assumed independent in $\overline{x}_0$).

Using the maximal (in space) description offered by Thm. 2.3, of the manifold defined by (2.1), we introduce a class of algorithms of "global type". Namely, they search for solution in a maximal admissible neighbourhood of $\overline{x}_0$ (which is just an admissible point here) and may find all the solutions from this admissible set.

The basic observation is that the solution of (2.4) , (2.5), . . . , (2.6) exists (roughly speaking), up to the "moment" of meeting a critical point. And the computation of y_{d-l} on this maximal existence interval, for a finer and finer discretization, provides a dense set of points in the manifold defined by (2.1). We denote it by A_n, in the n-th discretization step. The constraints (3.2) are to be just checked on this points in A_n. We may impose even supplementary abstract constraints in the problem (3.1), (2.1), (3.2), of the form:

$$\overline{x} \in D, \ D \subset R^d \text{ closed subset.} \tag{3.8}$$

We denote by C_n the discrete admissible set of points, in step n, obtained after checking the points in A_n for (3.2) and (3.8). The algorithm is as follows:

Algorithm 3.2

1) choose $n = 1$, the discretization step $\frac{1}{n}$ in (3.4) - (3.6) and the solution intervals $I_1^n, \dots, I_{d-l}^n$, the small parameter δ.

2) compute A_n and C_n.

3) find in C_n, the minimum of $h(\cdot)$, by direct computation, denoted by x_n.

4) test $|h(x_n) - h(x_{n-1})| < \delta$.

5) If YES, then STOP; if NO, then GO TO step 1).

In this setting, it is enough to assume h and $G_j, j = \overline{1,m}$ to be in $C(R^d)$, $F_i, i = \overline{1,l}$ satisfy (2.2) and D has nonvoid interior. By density, we get

Theorem 3.2. *The algorithm is convergent for $n \to \infty$.*

The admissible set for the problem (3.1), (2.1) may have several connected components, see [23].

Then, it is necessary to know an initial point $\overline{x}_0$ on each component, for the algorithm to work. If hypothesis (2.2) is not fulfilled, we suggest to work with generalized solutions of (2.4), (2.5), ..., (2.6). This subject is not yet investigated in the literature.

Finally, we mention the recent paper [16], that proposes an alternative approach in similar situation. We have reworked in [23] their main numerical example (in R^6) by employing implicit parametrizations and starting from the solutions they found. Our investigation allows a very consistent decrease of the optimal value for the performance index. The Algorithm 3.2 easily allows to extend the search region, simply by increasing the computations intervals for (2.4), (2.5), ..., (2.6). Some high dimensional applications are also reported in [9].

Similar ideas work for constrained optimal control problems.

Here, we briefly discuss the difficult case of mixed equality constraints, following [23]. The problem, of Mayer type, is the following:

$$\text{Min}\{l(x(0), x(1))\}, \tag{3.9}$$

$$x'(t) = f(t, x(t), u(t)),\ t \in [0,1], \tag{3.10}$$

$$h(x(t), u(t)) = 0,\ t \in [0.1], \tag{3.11}$$

and it is inspired from the recent works [3], [12], where the maximum principle is discussed.

Here $l(\cdot,\cdot), f(t,\cdot,\cdot), h(\cdot,\cdot)$ are defined in finite dimensional spaces $X \times X, X \times U$ of appropriate dimension and (3.10) is assumed to be uniquely and globally solvable, as it is standard in optimal control theory. Later, we shall add to it initial conditions. We require l continuous, f continuous and locally Lipschitzian in (x,u), h of class C^1, with locally Lipschitzian gradient for each component, such that there is the vector (x^0, u^0) in $X \times U$ satisfying

$$h(x^0, u^0) = 0,\ \nabla h(x^0, u^0) \text{ of maximal rank.} \tag{3.12}$$

The finite dimensional nonlinear algebraic system $h(x,u) = 0$, under condition (3.12), defines a manifold $M \subset X \times U$ that can be parametrized and discretized via the system (2.4), (2.5), ..., (2.6). Moreover, the relations (3.10), (3.11) can be interpreted as a DAE system and we differentiate (3.11) and replace it by:

$$\nabla_x h(x(t), u(t) f(t, x(t)), u(t)) + \nabla_u h(x(t), u(t)) u'(t) = 0. \tag{3.13}$$

The important remark is that the manifold M provides consistent initial conditions for (3.10), (3.13).

Theorem 3.3. *Any trajectory of* (3.10), (3.13) *starting from a point in M, remains in M.*

This gives a characterization of the admissible global trajectories for the constrained control problem (3.9) - (3.11). Consequently, it can be shown that the discretization of M provided by (2.4), (2.5), ..., (2.6), generates a dense family of admissible trajectories and an algorithm of global type, similar to Algorithm 3.2, can be formulated and its convergence remains valid [23].

Some academic examples can be found in [23] as well.

4. Shape optimization

Shape optimization problems have a similar structure with optimal control problems, for instance:

$$\operatorname*{Min}_{\Omega} \int_{\Omega} j(x, y(x)) dx, \tag{4.1}$$

$$-\Delta y = f \text{ in } \Omega, \tag{4.2}$$

$$y = 0 \text{ on } \partial\Omega. \tag{4.3}$$

Here $f \in L^2(D)$ and $j(\cdot,\cdot)$ is a Caratheodory mapping, the admissible domains satisfy $\Omega \subset D$, D a given domain bounded in R^d.

Other elliptic operators (or even evolution operators, [24]), other boundary conditions or cost functionals (defined on $\partial\Omega$, or on some given subset $E \subset \Omega$, for any Ω admissible, or depending as well on $\nabla y(x)$, etc.), more constraints (for instance, on the state y) may be considered in (4.1) - (4.3).

An important choice is the admissible family of geometries in R^d, denoted by $\mathcal{O}$. In one of the first approaches in shape optimization, due to Murat and Simon [6], the family $\mathcal{O}$ is given as the image of some fixed domain $B \subset R^d$ (for instance, a ball) via a family $\mathcal{F}$ of mapping $T : B \to R^d$, of class C^2, one-to-one on their image $T(B)$ and T^{-1} of class C^2 as well.

Then, (4.1) - (4.3) may be transported on B and the transformation $T \in \mathcal{F}$ will enter, together with its derivatives, in the coefficients of the transformed elliptic operator in B. The geometric optimization problem (4.1) - (4.3) is then equivalent with a control by the coefficients problem, if $\mathcal{O}$ is defined as above. The drawback of this approach is that all the admissible domains $\Omega \in \mathcal{O}$ have to be simply connected (when B is a ball), that is this family $\mathcal{O}$ is not general enough.

A similar discussion may be pursued in the case of the speed method of Zolesio, [26]. See [13], [15], [8] for information in this respect.

A more far reaching point of view is to assume that the admissible domains $\Omega \in \mathcal{O}$ are given via an implicit representation, using a family of functions $\Phi \in \mathcal{F}$:

$$\Omega = \Omega_\Phi = \{x \in R^d;\ \Phi(x) < 0,\ \Phi \in \mathcal{F}\}, \tag{4.4}$$

where $\mathcal{F}$ is now a subset in $C(\overline{D})$ with $D \subset R^d$ some given bounded domain. Obviously, relation (4.4) defines an open set and supplementary information should be given in order to select some connected component of interest, not necessarily simply connected. In this way, both topological and boundary variations may be considered in the problem (4.1) - (4.3).

This point of view was introduced by Osher and Sethian [11] in the setting of free boundary problems and the treatment is based on the Hamilton - Jacobi equation.

In shape optimization, implicit representation of domains were considered independently, already in [5]. Recently, it was shown that iterated Hamiltonian type systems, as discussed in §2, play a fundamental role in this setting [18]. Such ordinary differential systems are much easier to handle as Hamilton - Jacobi equations and Thm. 2.4 is the key result in dimension two, which is a case of interest in shape optimization.

A frequently met supplementary constraint on the admissible $\Omega \in \mathcal{O}$ is $E \subset \Omega$, where $E \subset D$ is another given open subset.

This geometric condition, under definition (4.4), is expressed as $\Phi(x) < 0$ in E, a very simple algebraic condition. Notice that it also selects the connected component of Ω_Φ, that is the domain of interest in the optimization problem. Similarly, one may ask that, for a given point $x_0 \in D$, we have $x_0 \in \partial\Omega_\Phi$ (or for some given submanifold $\Gamma \subset D$, we have $\Gamma \subset \partial\Omega_\Phi$). This is expressed algebraically as $\Phi(x_0) = 0$ (or $\Phi(x) = 0$ on Γ) and again selects in (4.4) the connected component of interest of Ω_Φ.

We underline that, if we assume just $\mathcal{F} \subset C(\overline{D})$, then $\partial\Omega_\Phi$ may have positive measure. Under condition (2.11), this cannot happen and the above examples are clearly defined, while the facility to translate geometric constraints in simple algebraic conditions is remarkable.

It turns out that the geometric optimization problem (4.1) - (4.3) is equivalent with a state constraint optimal control problem in D, for $\mathcal{O}$ given by (4.4).

Theorem 4.1. *Assume* (2.11), (2.12) *and let* Ω_Φ *be defined by* (4.4). *For any* $\Phi \in \mathcal{F}$, *there is* $u_\Phi \in L^2(D)$ *(not unique) such that the solution of*

$$-\Delta y = f + H(\Phi)u_\Phi \text{ in } D, \tag{4.5}$$

$$y = 0 \text{ on } \partial D, \tag{4.6}$$

coincides in Ω_Φ *with the solution of* (4.2), (4.3) *and satisfies the constraint*

$$\int_{\partial\Omega_\Phi} |y(\sigma)|^2 d\sigma = 0. \tag{4.7}$$

The cost (4.1) *corresponding to* Ω_Φ *is identical with the cost associated to* $[y_\phi, \Phi, u_\Phi]$ *given in* (4.5), (4.6).

Here, $H(\cdot) : R \to R$ is the Heaviside function.

The condition (4.7) may be expressed in the form (independently of the geometry):

$$\int_0^{T_\Phi} |y(x_1(t), x_2(t))|^2 \sqrt{x_1'(t)^2 + x_2'(t)^2} dt = 0, \tag{4.8}$$

where $(x_1(t), x_2(t))$ solves (2.9), (2.10) on the period $[0, T_\Phi]$ and $\overline{x}_0$ is some fixed given point on $\partial\Omega_\Phi$. Moreover, the cost functional (4.1) may be also rewritten in an "independent of the geometry" form:

$$\int_D (1 - H(\Phi)) j(x, y(x)) dx \tag{4.9}$$

and, in fact, $H(\Phi)$ is the characteristic function of $D \setminus \Omega_\Phi$, under hypotheses (2.11), (2.12).

Consequently, the optimal control problem (4.5), (4.6), (4.8), (4.9) (with controls $\Phi \in \mathcal{F}$, $u \in L^2(D)$) is independent of the geometry and is equivalent with the shape optimization problem (4.1) - (4.3), on $\mathcal{O}$ defined by $\mathcal{F}$ via (4.4).

A standard procedure in state constrained control problem is the penalization of the constraint in the cost ($\varepsilon > 0$):

$$\int_D (1 - H(\Phi)) j(x, y(x)) dx + \frac{1}{\varepsilon} \int_0^{T_\Phi} |y(x_1(t), x_2(t))|^2 \sqrt{x_1'(t)^2 + x_2'(t)^2} dt. \tag{4.10}$$

General approximation properties of the problem (4.5), (4.6), (4.10) with respect to the constrained problem (4.5), (4.6), (4.8), (4.9) or to the original shape optimization problem (4.1), (4.2), (4.3), are discussed in [18]. We indicate here just one property, when the cost integrand in (4.1) depends as well on ∇y, $j(x, y(x), \nabla y(x))$.

Theorem 4.2. *Assume that $j(\cdot, \cdot, \cdot)$ is Caratheodory on $D \times R \times R^2$ and satisfies the coercivity assumption*

$$j(x, y, v) \geq \alpha_1 |v|^2 + \beta_1 |y|^2 + \gamma, \ \alpha_1 > 0, \beta_1 > 0, \gamma \in R$$

and $j(x, y, \cdot)$ is convex. Then, if $[y_n^\varepsilon, \Phi_n^\varepsilon, u_n^\varepsilon]$ denote a minimizing sequence in the penalized problem (4.5), (4.6), (4.10) and y^*, Ω^* are cluster points of the sequence $[y_n^\varepsilon, \Omega_{\Phi_n^\varepsilon}]$ in the weak topology of $L^2(D)$, respectively in the Hausdorff - Pompeiu complementary topology, then $[y^*, \Omega^*]$ is an optimal pair for the problem (4.1) - (4.3).

The technique employed in [18] includes as well a modification of $\{y_n^\varepsilon\}$ outside $\Omega_{\Phi_n^\varepsilon}$. In the paper [7], a differentiable variant of this approach is studied. The implicit parametrization theorem gives a global representation of the boundary and allows to compute integrals as in (4.7), (4.8), (4.10). It also allows to discuss boundary observation problems [22].

Remark 4.3. One question of interest, in this context, is to obtain efficient gradient algorithms, in general shape optimization problems. Certain results of this type are reported in [7], for Dirichlet boundary conditions. Another question is related to the possibility to use just one control in the "extension" (4.5), (4.6) of the state system, while preserving all the other properties.

References

[1] Bonnans, J.F., Shapiro, A., *Perturbation Analysis of Optimization Problems*, Springer Verlag, New York, 2000.

[2] Ciarlet, P.G., *Introduction to Numerical Linear Algebra and Optimization*, Cambridge Univ. Press, Cambridge, 1989.

[3] Clarke, F., de Pinho, M.R., *Optimal control problems with mixed constraints*, SIAM J. Control Optim., **48**(2011), 4500-4524.

[4] Hirsch, M.W., Smale, S., Devaney, R.L., *Differential Equations, Dynamical Systems and Introduction to Chaos*, 3rd ed., Elsevier, Academic Press, New York, 2013.

[5] Mäkinen, R., Neittaanmäki, P., Tiba, D., *On a fixed domain approach for a shape optimization problem*, in: Computational and applied Mathematics II: Differential equations (W.F. Ames, P.J. van Houwen, eds.), North Holland, Amsterdam, 1992, 317-326.

[6] Murat, F., Simon, J., *Etude de problèmes d'optimal design*, in: Optimization techniques. Modeling and optimization in the service of man. Part 2 (J. Céa, ed.), LNCS41, Springer Verlag, Berlin, 1976, 54-62.

[7] Murea, C., Tiba, D., *Topological optimization via cost penalization*, TMNA, **54**(2019), no. 2B, 1023-1050.

[8] Neittaanmäki, P., Sprekels, J., Tiba, D., *Optimization of Elliptic Systems. Theory and Applications*, Springer Verlag, New York, 2006.

[9] Nicolai, M.R., *High dimensional applications of implicit parametrizations in nonlinear programming*, Ann. Acad. Rom. Sci. Ser. Math. Appl., **8**(2016), 44-55.

[10] Nicolai, M.R., Tiba, D., *Implicit functions and parametrization in dimension three: generalized solutions*, Discrete Contin. Dyn. Syst., **35**(2015), 2701-2710.

[11] Osher, S., Sethian, J.A., *Fronts propagating with curvature-dependent speed; algorithms based on Hamiltonian-Jacobi formulation*, Journal of Computational Physics, **79**(1988), no. 1, 12-49.

[12] de Pinho, M.R., *On necessary conditions for implicit control systems*, Pure Appl. Funct. Anal., **1**(2016), 185-196.

[13] Pironneau, O., *Optimal Shape Design for Elliptic Systems*, Springer Verlag, Berlin, 1984.

[14] Pontryagin, L.S., *Equation Differentielles Ordinaires*, MIR, Moscow, 1968.

[15] Sokolovski, J., Zolesio, J.-P., *Introduction to Shape Optimization. Shape Sensitivity Analysis*, Springer Verlag, Berlin, 1992.

[16] Stuber, M.D., Scott, J.K., Bartar, P.I., *Convex and concave relaxations of implicit functions*, Optimization Methods and Software, **30**(2015), 424-460.

[17] Tiba, D., *The implicit functions theorem and implicit parametrizations*, Ann. Acad. Rom. Sci. Ser. Math. Appl., **5**(2013), 193-208.

[18] Tiba, D., *A penalization approach in shape optimization*, Atti. Acad. Pelorit. Pericol. Cl. Sci. Fis. Mat. Nat., **16**(2018), no. 1, **A8**.

[19] Tiba, D., *Iterated Hamiltonian type systems and applications*, JDE, **264**(2018), 5465-5479.

[20] Tiba, D., *Some remarks on state constraints and mixed constraints*, Ann. Acad. Rom. Sci. Ser. Math. Appl., **10**(2018), 25-40.

[21] Tiba, D., *Neumann boundary conditions in shape optimization*, PAFA, **3**(2018), 241-253.

[22] Tiba, D., *Boundary observation in shape optimization*, in: New Trends in Differential Equations, Control Theory and Optimization (V. Barbu, C. Lefter, I. Vrabie, Eds.), World Scientific, Singapore, 2016, 301-315.

[23] Tiba, D., *Implicit parametrizations and applications in optimization and control*, MCRF, **10**(2020), 455-470.

[24] Tiba, D., Yamamoto, M., *A parabolic shape optimization problem*, Ann. Acad. Rom. Sci. Ser. Math. Appl., **12**(2020), 312-328.

[25] Tiba, D., Zălinescu, C., *On the necessity of some constraint qualification conditions in convex programming*, J. Convex Anal., **11**(2004), 95-110.

[26] Zolesio, J.-P., *Identification de Domaines par Deformation*, Thèse, Univ. of Nice, France, 1979.

Dan Tiba
"Simion Stoilow" Institute of Mathematics of the Romanian Academy
and
Academy of Romanian Scientists,
Bucureşti, Romania
e-mail: dan.tiba@imar.ro

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 17–27
DOI: 10.24193/subbmath.2021.1.02

On some evolution inclusions in non separable Banach spaces

Aurelian Cernea

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. We study a Cauchy problem of a class of nonconvex second-order integro-differential inclusions and a boundary value problem associated to a semi-linear evolution inclusion defined by nonlocal conditions in non-separable Banach spaces. The existence of mild solutions is established under Filippov type assumptions.

Mathematics Subject Classification (2010): 34A60.

Keywords: Lusin measurable multifunctions, selection, mild solution.

1. Introduction

In this note we study two classes of evolution differential inclusions. First we consider the problem

$$x''(t) \in A(t)x(t) + \int_0^t K(t,s)F(s,x(s))ds, \quad x(0) = x_0, x'(0) = y_0, \tag{1.1}$$

where $F : [0,T] \times X \to \mathcal{P}(X)$ is a set-valued map lipschitzian with respect to the second variable, X is a Banach space, $\{A(t)\}_{t\geq 0}$ is a family of linear closed operators from X into X that genearates an evolution system of operators $\{G(t,s)\}_{t,s\in[0,T]}$, $\Delta = \{(t,s) \in [0,T] \times [0,T]; t \geq s\}$, $K(.,.) : \Delta \to \mathbb{R}$ is continuous and $x_0, y_0 \in X$. The general framework of evolution operators $\{A(t)\}_{t\geq 0}$ that define problem (1.1) has been developed by Kozak ([19]) and improved by Henriquez ([17]).

Existence results and some qualitative properties of the mild solutions of problem (1.1) may be found in [14] in the case when X is a separable Banach space.

De Blasi and Pianigiani ([15]) obtained the existence of mild solutions for semi-linear differential inclusions on an arbitrary, not necessarily separable, Banach space X. Even if Filippov's ideas ([16]) are still present, the approach in [15] is fundamental different: it consists in the construction of the measurable selections of the multifunction. This construction does not use classical selection theorems such as Kuratowski and Ryll-Nardzewski's ([20]) or Bressan and Colombo's ([7]).

The aim of this note is to obtain an existence result for problem (1.1) similar to the one in [15]. We will prove the existence of solutions for problem (1.1) in an arbitrary space X under Filippov-type assumptions on F.

In several recent papers ([2, 3, 5, 12, 13, 17, 18]) existence results and qualitative properties of mild solutions have been obtained for the following problem

$$x''(t) \in A(t)x(t) + F(t, x(t)), \quad x(0) = x_0, x'(0) = y_0, \tag{1.2}$$

with $A(.)$ and $F(.,.)$ as above.

On one hand, the result in the present paper extends to the integro-differential framework (1.1) the result in [12] obtained for problem (1.2) and, on the other hand, this paper extends to second-order integro-differential inclusions a similar result in [10] obtained for a class of first-order integro-differential inclusions.

The second class of evolution inclusions that we are considering is

$$x' \in Ax + F(t, x) \quad a.e. \ ([0, T]), \tag{1.3}$$

$$x(0) + \sum_{i=1}^{m} a_i x(t_i) = x_0, \tag{1.4}$$

where X is a real separable Banach space, $a_i \in \mathbb{R}$, $a_i \neq 0$, $i = \overline{1, m}$, $x_0 \in X$, $0 < t_1 < t_2 < ... < t_m < T$, $F : [0, T] \times X \to \mathcal{P}(X)$ is a set-valued map and A is the infinitesimal generator of a linear semigroup $\{\mathcal{G}(t);\ t \geq 0\}$.

The nonlocal condition (1.4) was used by Byszewski ([8, 9]). If $a_i \neq 0$, $i = \overline{1, m}$ the results can be applied in kinematics to determine the evolution $t \to x(t)$ of the location of a physical object for which the positions $x(0), x(t_1), ..., x(t_m)$ are unknown but it is known the condition (1.4). Consequently, to describe some physical phenomena the nonlocal condition may be more useful than the standard initial condition $x(0) = x_0$. Obviously, when $a_i = 0$, $i = \overline{1, m}$, one has the classical initial condition.

Existence of mild solutions of problem (1.3)-(1.4) has been obtained in [4, 6] for convex as well as nonconvex set-valued maps. All these results are based on some suitable theorems of fixed point theory. In our recent paper [11] it is shown that Filippov's ideas ([1, 16]) can be suitably adapted in order to prove the existence of solutions to problem (1.3)-(1.4) provided the Banach space X is separable.

The result that we established in non separable Banach spaces for problem (1.3)-(1.4) may be interpreted as extension of the result in [15] from Cauchy problems to boundary value problems defined by nonlocal conditions and as an extension of the result in [11] to non separable Banach spaces.

The paper is organized as follows: in Section 2 we present the notations, definitions and preliminary results to be used in the sequel and in Section 3 we prove the main results.

2. Preliminaries

Consider X, an arbitrary real Banach space with norm $|.|$ and with the corresponding metric $d(.,.)$. Let $\mathcal{P}(X)$ be the space of all bounded nonempty subsets of X endowed with the Hausdorff pseudometric

$$\mathrm{d}_H(A,B) = \max\{\mathrm{d}^*(A,B), \mathrm{d}^*(B,A)\}, \quad \mathrm{d}^*(A,B) = \sup_{a\in A} \mathrm{d}(a,B),$$

where $\mathrm{d}(x,A) = \inf_{a\in A} |x-a|$, $A \subset X, x \in X$.

Let $\mathcal{L}$ be the σ-algebra of the (Lebesgue) measurable subsets of R and, for $A \in \mathcal{L}$, let $\mu(A)$ be the Lebesgue measure of A.

Let X be a Banach space and Y be a metric space. An open (resp., closed) ball in Y with center y and radius r is denoted by $B_Y(y,r)$ (resp., $\overline{B}_Y(y,r)$. In what follows, $B = B_X(0,1)$.

A multifunction $F : Y \to \mathcal{P}(X)$ with closed bounded nonempty values is said to be d_H-continuous at $y_0 \in Y$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for any $y \in B_Y(y_0, r)$ there is $\mathrm{d}_H(F(y), F(y_0)) \leq \varepsilon$. F is called d_H-continuous if it is so at each point $y_0 \in Y$.

Let $A \in \mathcal{L}$, with $\mu(A) < \infty$. A multifunction $F : Y \to \mathcal{P}(X)$ with closed bounded nonempty values is said to be *Lusin measurable* if for every $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset A$, with $\mu(A\backslash K_\varepsilon) < \varepsilon$ such that F restricted to K_ε is d_H-continuous.

It is clear that if $F, G : A \to \mathcal{P}(X)$ and $f : A \to X$ are Lusin measurable, then so are F restricted to B ($B \subset A$ measurable), $F+G$ and $t \to \mathrm{d}(f(t), F(t))$. Moreover, the uniform limit of a sequence of Lusin measurable multifunctions is Lusin measurable, too.

Let I stand for the interval $[0,T]$, $T > 0$, $C(I,X)$ is the Banach space of all continuous functions from I to X with the norm $||x||_C = \sup_{t\in I} |x(t)|$ and $L^1(I,X)$ is the Banach space of (Bochner) integrable functions $u(.) : I \to X$ endowed with the norm $||u||_1 = \int_0^T |u(t)|dt$. Denote by $B(X)$ the Banach space of bounded linear operators from X into X with the norm $||N|| = \sup\{|N(y)|;\ |y| = 1\}$.

In what follows $\{A(t)\}_{t\geq 0}$ is a family of linear closed operators from X into X that genearates an evolution system of operators $\{G(t,s)\}_{t,s\in I}$. By hypothesis the domain of $A(t)$, $D(A(t))$ is dense in X and is independent of t.

Definition 2.1. ([17, 19]) A family of bounded linear operators $G(t,s) : X \to X$, $(t,s) \in \Delta := \{(t,s) \in I \times I; s \leq t\}$ is called an evolution operator of the equation

$$x''(t) = A(t)x(t) \tag{2.1}$$

if

i) For any $x \in X$, the map $(t,s) \to G(t,s)x$ is continuously differentiable and

a) $G(t,t) = 0$, $t \in I$.

b) If $t \in I, x \in X$ then $\frac{\partial}{\partial t} G(t,s)x|_{t=s} = x$ and $\frac{\partial}{\partial s} G(t,s)x|_{t=s} = -x$.

ii) If $(t,s) \in \Delta$, then $\frac{\partial}{\partial s} G(t,s)x \in D(A(t))$, the map $(t,s) \to G(t,s)x$ is of class C^2 and

a) $\frac{\partial^2}{\partial t^2} G(t,s)x \equiv A(t)G(t,s)x$.

b) $\frac{\partial^2}{\partial s^2}G(t,s)x \equiv G(t,s)A(t)x$.

c) $\frac{\partial^2}{\partial s \partial t}G(t,s)x|_{t=s} = 0$.

iii) If $(t,s) \in \Delta$, then there exist $\frac{\partial^3}{\partial t^2 \partial s}G(t,s)x$, $\frac{\partial^3}{\partial s^2 \partial t}G(t,s)x$ and

a) $\frac{\partial^3}{\partial t^2 \partial s}G(t,s)x \equiv A(t)\frac{\partial}{\partial s}G(t,s)x$ and the map $(t,s) \to A(t)\frac{\partial}{\partial s}G(t,s)x$ is continuous.

b) $\frac{\partial^3}{\partial s^2 \partial t}G(t,s)x \equiv \frac{\partial}{\partial t}G(t,s)A(s)x$.

As an example for equation (2.1) one may consider the problem (e.g., [19])

$$\frac{\partial^2 z}{\partial t^2}(t,\tau) = \frac{\partial^2 z}{\partial \tau^2}(t,\tau) + a(t)\frac{\partial z}{\partial t}(t,\tau), \quad t \in [0,T], \tau \in [0,2\pi],$$

$$z(t,0) = z(t,\pi) = 0, \quad \frac{\partial z}{\partial \tau}(t,0) = \frac{\partial z}{\partial \tau}(t,2\pi),\ t \in [0,T],$$

where $a(.) : I \to \mathbb{R}$ is a continuous function. This problem is modeled in the space $X = L^2(\mathbb{R},\mathbb{C})$ of 2π-periodic 2-integrable functions from $\mathbb{R}$ to $\mathbb{C}$, $A_1 z = \frac{d^2 z(\tau)}{d\tau^2}$ with domain $H^2(\mathbb{R},\mathbb{C})$ the Sobolev space of 2π-periodic functions whose derivatives belong to $L^2(\mathbb{R},\mathbb{C})$. It is well known that$A_1$ is the infinitesimal generator of strongly continuous cosine functions $C(t)$ on X. Moreover, A_1 has discrete spectrum; namely the spectrum of A_1 consists of eigenvalues $-n^2$, $n \in \mathbb{Z}$ with associated eigenvectors

$$z_n(\tau) = \frac{1}{\sqrt{2\pi}}e^{in\tau},\ n \in \mathbb{N}.$$

The set $z_n,\ n \in \mathbb{N}$ is an orthonormal basis of X. In particular,

$$A_1 z = \sum_{n\in\mathbb{Z}} -n^2 < z, z_n > z_n,\ z \in D(A_1).$$

The cosine function is given by

$$C(t)z = \sum_{n\in\mathbf{Z}} \cos(nt) < z, z_n > z_n$$

with the associated sine function

$$S(t)z = t < z, z_0 > z_0 + \sum_{n\in\mathbf{Z}^*} \frac{\sin(nt)}{n} < z, z_n > z_n.$$

For $t \in I$ define the operator $A_2(t)z = a(t)\frac{dz(\tau)}{d\tau}$ with domain $D(A_2(t)) = H^1(\mathbb{R},\mathbb{C})$. Set $A(t) = A_1 + A_2(t)$. It has been proved in [19] that this family generates an evolution operator as in Definition 2.1.

Definition 2.2. A continuous mapping $x(.) \in C(I,X)$ is called a mild solution of problem (1.1) if there exists a (Bochner) integrable function $f(.) \in L^1(I,X)$ such that

$$f(t) \in F(t,x(t)) \quad a.e.\,(I), \tag{2.2}$$

$$x(t) = -\frac{\partial}{\partial s}G(t,0)x_0 + G(t,0)y_0 + \int_0^t G(t,s)\int_0^s K(s,\tau)f(\tau)d\tau,\ t \in I. \tag{2.3}$$

We shall call $(x(.), f(.))$ a *trajectory-selection pair* of (1.1) if $f(.)$ verifies (2.2) and $x(.)$ is defined by (2.3).

We note that condition (2.3) can be rewritten as

$$x(t) = -\frac{\partial}{\partial s}G(t,0)x_0 + G(t,0)y_0 + \int_0^t U(t,s)f(s)ds \quad \forall t \in I, \tag{2.4}$$

where $U(t,s) = \int_s^t G(t,\tau)K(\tau,s)d\tau$.

Hypothesis H1. i) There exists an evolution operator $\{G(t,s)\}_{t,s\in I}$ associated to the family $\{A(t)\}_{t\geq 0}$.

ii) There exist $M, M_0 \geq 0$ such that $|G(t,s)|_{B(X)} \leq M$, $|\frac{\partial}{\partial s}G(t,s)| \leq M_0$, for all $(t,s) \in \Delta$.

iii) $K(.,.) : \Delta \to \mathbb{R}$ is continuous.

Hypothesis H2. i) A is the infinitesimal generator of a strongly continuous and compact semigroup $\{\mathcal{G}(t);\ t \geq 0\}$ in X.

ii) There exists an operator $C : X \to X$ defined by

$$C = [I + \sum_{i=1}^{m} a_i \mathcal{G}(t_i)]^{-1}.$$

Let $m_0 \geq 0$ be such that $|\mathcal{G}(t)| \leq m_0\ \forall t \in I$.

According to [4] if we assume that $\sum_{i=1}^m |a_i| < \frac{1}{m_0}$ then there exists C as in Hypothesis H2 ii).

Definition 2.3. A continuous mapping $x(.) \in C(I,X)$ is called a mild solution of problem (1.3)-(1.4) if there exists a (Bochner) integrable function $f(.) \in L^1(I,X)$ such that

$$f(t) \in F(t, x(t)) \quad a.e.\, (I) \tag{2.5}$$

$$x(t) = \mathcal{G}(t)Cx_0 - \sum_{i=1}^{m} a_i \mathcal{G}(t)C \int_0^{t_i} \mathcal{G}(t_i - u)f(u)du + \int_0^t \mathcal{G}(t-u)f(u)du, t \in I. \tag{2.6}$$

Remark 2.4. If we denote

$$H(t,s) = \sum_{i=1}^{m} a_i \mathcal{G}(t)C\mathcal{G}(t_i - s)\chi_{[0,t_i]}(s) + \mathcal{G}(t-s)\chi_{[0,t]}(s),$$

where $\chi_S(\cdot)$ is the characteristic function of the set S, then the solution $x(\cdot)$ in Definition 2.3 may be written as

$$x(t) = \mathcal{G}(t)Cx_0 - \int_0^T H(t,s)f(s)ds. \tag{2.7}$$

Obviously,

$$|H(t,s)| \leq \sum_{i=1}^{m} |a_i| m_0^2 ||C|| + m_0 =: m \quad \forall\, t,s \in I.$$

In what follows X is a real Banach space and we assume the following hypotheses.

Hypothesis H3. i) *$F(.,.) : I \times X \to \mathcal{P}(X)$ has nonempty closed bounded values and for any $x \in X$ $F(.,x)$ is Lusin measurable on I.*

ii) *There exists $l(.) \in L^1(I,(0,\infty))$ such that, $\forall t \in I$*

$$\mathrm{d}_H(F(t,x_1),F(t,x_2)) \leq l(t)|x_1 - x_2|, \quad \forall\, x_1,x_2 \in X.$$

iii) *There exists $q(.) \in L^1(I,(0,\infty))$ such that $\forall t \in I$ we have*

$$F(t,0) \subset q(t)B.$$

Denote $L = \displaystyle\int_0^T l(s)ds$.

The technical results summarized in the following lemma are essential in the proof of our results. For the proof, we refer the reader to [15].

Lemma 2.5. *i) Let $F_i : I \to \mathcal{P}(X)$, i=1,2 be two Lusin measurable multifunctions and let $\varepsilon_i > 0$, i=1,2 be such that*

$$H_1(t) := (F_1(t) + \varepsilon_1 B) \cap (F_2(t) + \varepsilon_2 B) \neq \emptyset, \quad \forall t \in I.$$

Then the multifunction $H_1 : I \to \mathcal{P}(X)$ has a Lusin measurable selection $h : I \to X$.

ii) Assume that Hypothesis H3 is satisfied. Then for any continuous $x(.) : I \to X$, $u(.) : I \to X$ measurable and any $\varepsilon > 0$ one has

a) the multifunction $t \to F(t,x(t))$ is Lusin measurable on I.

b) the multifunction $H_2 : I \to \mathcal{P}(X)$ defined by

$$H_2(t) := (F(t,x(t)) + \varepsilon B) \cap B_X(u(t), d(u(t),F(t,x(t))) + \varepsilon)$$

has a Lusin measurable selection $g : I \to X$.

3. The results

Set $n(t) = \int_0^t l(u)\mathrm{d}u$, $t \in I$, denote $K_0 := \sup_{(t,s)\in\Delta} |K(t,s)|$ and note that

$$|U(t,s)| \leq MK_0(t-s) \leq MK_0T.$$

Theorem 3.1. *We assume that Hypotheses H1 and H3 are satisfied. Then, for every $x_0, y_0 \in X$, Cauchy problem (1.1) has a mild solution $x(.) \in C(I,X)$.*

Proof. Let us first note that if $z(.) : I \to X$ is continuous, then every Lusin measurable selection $u : I \to X$ of the multifunction $t \to F(t,z(t)) + B$ is Bochner integrable on I. More precisely, for any $t \in I$, there holds

$$\begin{aligned}|u(t)| &\leq d_H(F(t,z(t)) + B, 0) \leq d_H(F(t,z(t)),F(t,0)) + d_H(F(t,0),0) + 1\\ &\leq l(t)|z(t)| + q(t) + 1.\end{aligned}$$

Let $0 < \varepsilon < 1$, $\quad \varepsilon_n = \frac{\varepsilon}{2^{n+2}}$.

Consider $f_0(.) : I \to X$, an arbitrary Lusin measurable, Bochner integrable function, and define

$$x_0(t) = -\frac{\partial}{\partial s}G(t,0)x_0 + G(t,0)y_0 + \int_0^t U(t,s)f_0(s)ds, \quad t \in I.$$

Since $x_0(.)$ is continuous, by Lemma 2.5 ii) there exists a Lusin measurable function $f_1(.) : I \to X$ which, for $t \in I$, satisfies

$$f_1(t) \in (F(t, x_0(t)) + \varepsilon_1 B) \cap B(f_0(t), d(f_0(t), F(t, x_0(t))) + \varepsilon_1)$$

Obviously, $f_1(.)$ is Bochner integrable on I. Define $x_1(.) : I \to X$ by

$$x_1(t) = -\frac{\partial}{\partial s}G(t,0)x_0 + G(t,0)y_0 + \int_0^t U(t,s)f_1(s)ds, \quad t \in I.$$

By induction, we construct a sequence $x_n : I \to X$, $n \geq 2$ given by

$$x_n(t) = -\frac{\partial}{\partial s}G(t,0)x_0 + G(t,0)y_0 + \int_0^t U(t,s)f_n(s)ds, \quad t \in I, \tag{3.1}$$

where $f_n(.) : I \to X$ is a Lusin measurable function which, for $t \in I$, satisfies:

$$f_n(t) \in (F(t, x_{n-1}(t)) + \varepsilon_n B) \cap B(f_{n-1}(t), d(f_{n-1}(t), F(t, x_{n-1}(t))) + \varepsilon_n). \tag{3.2}$$

At the same time, as we saw at the beginning of the proof, $f_n(.)$ is also Bochner integrable.
From (3.2), for $n \geq 2$ and $t \in I$, we obtain

$$\begin{aligned}
|f_n(t) - f_{n-1}(t)| &\leq d(f_{n-1}(t), F(t, x_{n-1}(t))) + \varepsilon_n \\
&\leq d(f_{n-1}(t), F(t, x_{n-2}(t))) + d_H(F(t, x_{n-2}(t)), F(t, x_{n-1}(t))) + \varepsilon_n \\
&\leq \varepsilon_{n-1} + l(t)|x_{n-1}(t) - x_{n-2}(t)| + \varepsilon_n.
\end{aligned}$$

Since $\varepsilon_{n-1} + \varepsilon_n < \varepsilon_{n-2}$, for $n \geq 2$, we deduce that

$$|f_n(t) - f_{n-1}(t)| \leq \varepsilon_{n-2} + l(t)|x_{n-1}(t) - x_{n-2}(t)|. \tag{3.3}$$

Denote $p_0(t) := d(f_0(t), F(t, x_0(t))), t \in I$. We next prove by recurrence, that for $n \geq 2$ and $t \in I$

$$\begin{aligned}
|x_n(t) - x_{n-1}(t)| \leq \sum_{k=0}^{n-2} \int_0^t \varepsilon_{n-2-k} \frac{(MK_0T)^{k+1}(n(t) - n(u))^k}{k!} du \\
+ \varepsilon_0 \int_0^t \frac{(MK_0T)^n (n(t) - n(u))^{n-1}}{(n-1)!} du \\
+ \int_0^t \frac{(MK_0T)^n (n(t) - n(u))^{n-1}}{(n-1)!} p_0(u) du.
\end{aligned} \tag{3.4}$$

We start with $n = 2$. In view of (3.1), (3.2) and (3.3), for $t \in I$, there is

$$|x_2(t) - x_1(t)| \leq \int_0^t |U(t,s)| \cdot |f_2(s) - f_1(s)| ds$$

$$\leq \int_0^t MK_0T[\varepsilon_0 + l(s)|x_1(s) - x_0(s)|]ds$$

$$\leq \varepsilon_0 MK_0Tt + \int_0^t \left[MK_0Tl(s) \int_0^s |U(s,r)| \cdot |f_1(r) - f_0(r)| dr \right] ds$$

$$\leq \varepsilon_0 MK_0Tt + \int_0^t \left[(MK_0T)^2 l(s) \int_0^s (p_0(u) + \varepsilon_1) du \right] ds$$

$$\leq \varepsilon_0 MK_0Tt + \int_0^t \left[(MK_0T)^2(p_0(u)+\varepsilon_1)\int_u^t l(s)ds\right]du$$
$$= \varepsilon_0 MK_0Tt + \int_0^t (MK_0T)^2(n(t)-n(s))[p_0(s)+\varepsilon_0]ds,$$

i.e, (3.4) is verified for $n = 2$.
Using again (3.3) and (3.4), we conclude

$$|x_{n+1}(t)-x_n(t)| \leq \int_0^t |U(t,s)|.|f_{n+1}(s)-f_n(s)|ds$$
$$\leq \int_0^t MK_0T[\varepsilon_{n-1}+l(s)|x_n(s)-x_{n-1}(s)|]ds$$
$$\leq \varepsilon_{n-1}MK_0Tt + \int_0^t l(s)\left[\sum_{k=0}^{n-2}\int_0^s \varepsilon_{n-2-k}\frac{(MK_0T)^{k+2}(n(s)-n(u))^k}{k!}du\right.$$
$$\left. + \int_0^s \frac{(MK_0T)^{n+1}(n(s)-n(u))^{n-1}}{(n-1)!}(p_0(u)+\varepsilon_0)du\right]ds$$
$$= \varepsilon_{n-1}MK_0Tt + \sum_{k=0}^{n-2}\varepsilon_{n-2-k}\int_0^t\left[\int_0^s \frac{(MK_0T)^{k+2}(n(s)-n(u))^k}{k!}l(s)du\right]ds$$
$$+ \int_0^t l(s)\left(\int_0^s \frac{(MK_0T)^{n+1}(n(s)-n(u))^{n-1}}{(n-1)!}l(s)[p_0(u)+\varepsilon_0]du\right)ds$$
$$= \varepsilon_{n-1}MK_0Tt + \sum_{k=0}^{n-2}\varepsilon_{n-2-k}\int_0^t\left(\int_u^t \frac{(MK_0T)^{k+2}(n(s)-n(u))^k}{k!}l(s)ds\right)du$$
$$+ \int_0^t \left(\int_u^t \frac{(MK_0T)^{n+1}(n(s)-n(u))^{n-1}}{(n-1)!}l(s)ds\right)[p_0(u)+\varepsilon_0]du$$
$$= \varepsilon_{n-1}MK_0Tt + \sum_{k=0}^{n-2}\varepsilon_{n-2-k}\int_0^t \frac{(MK_0T)^{k+2}(n(s)-n(u))^{k+1}}{(k+1)!}du$$
$$+ \int_0^t \frac{(MK_0T)^{n+1}(n(s)-n(u))^n}{n!}[p_0(u)+\varepsilon_0]du$$
$$= \sum_{k=0}^{n-1}\varepsilon_{n-1-k}\cdot\int_0^t \frac{(MK_0T)^{k+1}(n(s)-n(u))^k}{k!}du$$
$$+ \int_0^t \frac{(MK_0T)^{n+1}(n(s)-n(u))^n}{n!}[p_0(u)+\varepsilon_0]du$$

and statement (3.8) it is true for $n+1$.

From (3.8) it follows that for $n \geq 2$ and $t \in I$

$$|x_n(t)-x_{n-1}(t)| \leq a_n, \tag{3.5}$$

where

$$a_n = \sum_{k=0}^{n-2}\varepsilon_{n-2-k}\frac{(MK_0T)^{k+1}n(T)^k}{k!} + \frac{(MK_0T)^n n(T)^{n-1}}{(n-1)!}\left[\int_0^1 p_0(u)du+\varepsilon_0\right],$$

Obviously, the series whose n-th term is a_n converges. So, from (3.5) we infer that $x_n(.)$ converges to a continuous function, $x(.) : I \to X$, uniformly on I.

On the other hand, in view of (3.3) there is

$$|f_n(t) - f_{n-1}(t)| \leq \varepsilon_{n-2} + l(t)a_{n-1}, \quad t \in I, n \geq 3$$

which implies that the sequence $f_n(.)$ converges to a Lusin measurable function $f(\cdot) : I \to X$.

Since $x_n(.)$ is bounded and

$$|f_n(t)| \leq l(t)|x_{n-1}(t)| + q(t) + 1,$$

we infer that $f(.)$ is also Bochner integrable.

Passing with $n \to \infty$ in (3.1) and using the Lebesgue dominated convergence theorem, we obtain

$$x(t) = -\frac{\partial}{\partial s}G(t,0)x_0 + G(t,0)y_0 + \int_0^t U(t,s)f(s)ds, \quad t \in I.$$

On the other hand, from (3.2) we get

$$f_n(t) \in F(t, x_n(t)) + \varepsilon_n B, \quad t \in I, n \geq 1$$

and letting $n \to \infty$ we obtain

$$f(t) \in F(t, x(t)), \quad t \in I,$$

which completes the proof. □

Theorem 3.2. *Assume that Hypotheses H2 and H3 are satisfied and $mL < 1$. Then, for every $x_0 \in X$ problem (1.3)-(1.4) has a solution $x(.) : I \to X$.*

Proof. The proof follows the same pattern as in the proof of Theorem 3.1. This time

$$x_n(t) = \mathcal{G}(t)Cx_0 - \int_0^T H(t,s)f_n(s)\mathrm{d}s, \quad \forall t \in I,$$

with $f_n(\cdot)$ as before and

$$|x_n(t) - x_{n-1}(t)| \leq \sum_{j=0}^{n-2} \varepsilon_{n-2-j}m^{j+1}L^jT + m^nL^{n-1}\int_0^T (p_0(s) + \varepsilon_0)ds$$

for $n \geq 2$ and $t \in I$. The estimate in (3.5) becames

$$|x_n(t) - x_{n-1}(t)| \leq a_n,$$

where

$$a_n = \sum_{j=0}^{n-2} \varepsilon_{n-2-j}m^{j+1}L^jT + m^nL^{n-1}\int_0^T (p_0(s) + \varepsilon_0)ds$$

Taking into account the fact that $mL < 1$, we deduce that the series whose n-th term is a_n is convergent. □

References

[1] Aubin, J.P., Frankowska, H., *Set-valued Analysis*, Birkhauser, 1990.

[2] Baliki, A., Benchohra, M., Graef, J.R., *Global existence and stability of second order functional evolution equations with infinite delay*, Electronic J. Qual. Theory Differ. Equations, **2016**(2016), no. 23, 1-10.

[3] Baliki, A., Benchohra, M., Nieto, J.J., *Qualitative analysis of second-order functional evolution equations*, Dynamic Syst. Appl., **24**(2015), 559-572.

[4] Benchohra, M., Gatsori, E.P., Ntouyas, S.K., *Controllability results for semilinear evolution inclusions with nonlocal conditions*, J. Optim. Theory Appl., **118**(2003), 493-513.

[5] Benchohra, M., Medjadj, I., *Global existence results for second order neutral functional differential equations with state-dependent delay*, Comment. Math. Univ. Carolin., **57**(2016), 169-183.

[6] Benchohra, M., Ntouyas, S.K., *Existence of mild solutions of semilinear evolution inclusions with nonlocal conditions*, Georgian Math. J., **7**(2000), 221-230.

[7] Bressan, A., Colombo, G., *Extensions and selections of maps with decomposable values*, Studia Math., **90**(1988), 69-86.

[8] Byszewski, L., *Application of properties of the right-hand sides of evolution equations to an investigation of nonlocal evolution problems*, Nonlin. Anal., **33**(1998), 413-426.

[9] Byszewski, L., *Existence and uniqueness of a classical solution to a functional-differential abstract nonlocal Cauchy problem*, J. Appl. Math. Stochastic Anal., **12**(1999), 91-97.

[10] Cernea, A., *On an evolution inclusion in non separable Banach spaces*, Opuscula Math., **29**(2009), 131-138.

[11] Cernea, A., *A note on a semilinear evolution inclusion with nonlocal conditions*, ROMAI J., **12**(2016), 33-38.

[12] Cernea, A., *A note on the solutions of a second-order evolution inclusion in non separable Banach spaces*, Comment. Math. Univ. Carolin., **58**(2017), 307-314.

[13] Cernea, A., *Some remarks on the solutions of a second-order evolution inclusion*, Dynamic Syst. Appl., **27**(2018), 319-330.

[14] Cernea, A., *On the mild solutions of a class of second-order integro-differential inclusions*, J. Nonlinear Variational Anal., **3**(2020), 247-256.

[15] De Blasi F.S., Pianigiani, G., *Evolution inclusions in non separable Banach spaces*, Comment. Math. Univ. Carolin., **40**(1999), 227-250.

[16] Filippov, A.F., *Classical solutions of differential equations with multivalued right-hand side*, SIAM J. Control Optim., **5**(1967), 609-621.

[17] Henriquez, H.R., *Existence of solutions of nonautonomous second order functional differential equations with infinite delay*, Nonlinear Anal., **74**(2011), 3333-3352.

[18] Henriquez, H.R., Poblete, V., Pozo, J.C., *Mild solutions of non-autonomous second order problems with nonlocal initial conditions*, J. Math. Anal. Appl., **412**(2014), 1064-1083.

[19] Kozak, M., *A fundamental solution of a second-order differential equation in a Banach space*, Univ. Iagel. Acta. Math., **32**(1995), 275-289.

[20] Kuratowski, K., Ryll-Nardzewski, C., *A general theorem on selectors*, Bull. Acad. Pol. Sci. Math. Astron. Phys., **13**(1965), 397-403.

Aurelian Cernea
University of Bucharest,
Faculty of Mathematics and Computer Sciences
14, Academiei Street,
010014 Bucharest,

Academy of Romanian Scientists,
Splaiul Independenţei 54,
050094 Bucharest, Romania
e-mail: acernea@fmi.unibuc.ro

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 29–46
DOI: 10.24193/subbmath.2021.1.03

Theoretical and numerical considerations on Bratu-type problems

Adrian Petruşel, Ioan A. Rus and Marcel Adrian Şerban

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. In this paper we present an heuristic introduction to Bratu problem and we give some variants of Bratu's theorem (G. Bratu, Sur les équations intégrales non linéaires, Bulletin Soc. Math. France, 42(1914), 113-142). Using the positivity of Green's function, the monotone iterations technique and the contraction principle, some generalizations of Bratu's result are also given. Numerical aspects are also considered.

Mathematics Subject Classification (2010): 34B18, 47H10, 65R20, 34B27, 45G10, 35K58.

Keywords: Bratu problem, Bratu theorem, Bratu-type equation, Cauchy problem, Nicoletti problem, boundary value problem, Green function, integral equations with exponential nonlinearity, numerical aspects of Bratu's problem.

1. Introduction

The classical Bratu problem is the following boundary value problem

$$\begin{cases} -y''(x) = \lambda e^{y(x)}, \ x \in [0,1] \\ y(0) = 0, \ y(1) = 0, \end{cases} \tag{1.1}$$

where $\lambda > 0$ is a parameter.

Bratu's problem has both theoretical and applicative relevance.

It was proved that Bratu's problem in one-dimensional planar coordinates has analytical solution in the following form:

$$y(x) = -2\log\left(\frac{\cosh[(x-\frac{1}{2})\frac{\theta}{2}]}{\cosh\frac{\theta}{4}}\right),$$

where θ is the solution of

$$\theta = \sqrt{2\lambda}\cosh\frac{\theta}{4}.$$

Notice that y has the maximum value (denoted by μ) at $x = \frac{1}{2}$ and there is an analytical expression between μ and λ discovered by Liouville in 1853, see [37]. Moreover, Bratu's problem has at most two solutions and the distribution of the solutions depends on a critical value of λ, denoted by λ_c. The critical value λ_c satisfies the equation

$$1 = \frac{1}{4}\sqrt{2\lambda}\sinh\frac{\theta}{4}$$

and it was approximated as $\lambda_c \approx 3.513830719$, see, e.g. [9]. More precisely, if $0 < \lambda < \lambda_c$ then (1.1) has two solutions, if $\lambda = \lambda_c$ there is one solution for (1.1), while for $\lambda > \lambda_c$ there is no solution for Bratu's problem.

Bratu's problem governs several important real life problems, such as the fuel ignition model in the thermal combustion theory, the model of thermal reaction process, the Chandrasekhar model related to the expansion of the universe and to some relativity theory models and it is connected to models from chemical reaction theory, radiative heat transfer theory and nanotechnologies (see [23], [19], [20], [1], [15], [22],...).

In the last two decades, many published papers have focused on solving (1.1) by analytical (e.g., Adomian decomposition method, homotopy analysis method, variational iteration methods, Laplace transform decomposition method or differential transformation method) and numerical (e.g., B-spline method, the finite difference method, weighted residual method, the shooting method, multigrid-based methods, the Sinc-Galerkin method, collocation methods based on B-spline basis functions, Bessel collocation method) methods, see [13], [14], [15], [51], [52],[7], [8], [9], [16], [17], [29], [32], [35],...

An extension of Bratu's problem is the following boundary value problem, so-called Liouville-Bratu-Gelfand problem (see [26], [21], [27], [28], [30], [35], [41]) :

$$\begin{cases} -\Delta u(x) = \lambda e^{u(x)}, \ x \in \Omega \\ u(x) = 0, \ x \in \partial\Omega \end{cases} \tag{1.2}$$

where $\lambda > 0$ is a parameter and $\Omega \subset \mathbb{R}^n$ is a bounded domain.

The aim of our paper is to we give some variants of a Bratu's theorem (G. Bratu, Sur les équations intégrales non linéaires, Bulletin de la Soc. Math. France, 42(1914), 113-142) using the positivity of Green's function, monotone iteration technique and the contraction principle. Some generalizations of Bratu's result are also given.

The structure of the paper is the following one:

1. Introduction
2. Preliminaries
3. Heuristic considerations on particular solutions of Bratu equation
4. Some variants of Bratu theorem
5. Bratu-type problems
6. Other generalizations
7. Numerical aspects of Bratu-type problem.

The Reference list will conclude the paper.

2. Preliminaries

2.1. Linear two point boundary value problem

Let $L_0 := -\dfrac{d^2}{dx^2} + p(x)\dfrac{d}{dx}$ and $L := L_0 + q(x)$, where $p, q \in C[a,b]$. We consider the following two-point boundary value problem

$$L(u) = f \tag{2.1}$$

$$l_1(y;a) := a_{10}y(a) - a_{11}y'(a) = r_1 \tag{2.2}$$

$$l_2(y;b) := a_{20}y(b) + a_{21}y'(b) = r_2 \tag{2.3}$$

where $f \in C[a,b]$, $a_{ij} \geq 0$, $i = 1,2$, $j = 0,1$, $a_{10} \cdot a_{20} > 0$ and $r_1, r_2 \in \mathbb{R}$.

It is well known that if $q(x) \geq 0$, for $x \in [a,b]$, then the Green function for this problem exists and is positive. However, the assumption $q(x) \geq 0$ is not a necessary condition for the positivity of the Green's function. Moreover, we have the following theorem of equivalent statements concerning the positivity of Green's function.

Theorem 2.1. (I.A. Rus [47]) *The following statements are equivalent:*

(i) There exists a function $v \in C^2(]a,b[) \cap C^1[a,b]$ such that: $v > 0$ on $[a,b]$, $(L_0 + q)(v) > 0$ on $[a,b]$, $l_1(y;a) > 0$ and $l_2(y;b) > 0$.

(ii) The following implication holds:

$$y \in C^2(]a,b[) \cap C^1[a,b],\ (L_0 + q_1)(y) = 0,\ l_1(y;a) = 0,\ l_2(y;b) = 0 \Rightarrow y = 0,$$

for each $q_1 \in C[a,b]$ with $q_1(x) \geq q$.

(iii) The following implication holds:

$$y \in C^2(]c,d[) \cap C^1[c,d],\ (L_0 + q)(y) = 0,\ l_1(y;c) = 0,\ l_2(y;d) = 0 \Rightarrow y = 0,$$

for each $[c,d] \subset [a,b]$.

(iv) There exists the Green function $G(x,s)$, corresponding to problem (2.1), (2.2), (2.3), *and $G(x,s) \geq 0$, $\forall\ x, s \in [a,b]$.*

(v) The first eigenvalue of the Sturm-Liouville problem

$$(L_0 + q)(y) = \lambda y,$$

$$l_1(y;a) = 0,\ l_2(y;b) = 0,$$

is positive.

By definition, we have the strong uniqueness property for the problem (2.1), (2.2), (2.3) if one (i.e., all) of the statements, in the above Theorem 2.1, is a theorem. In this case, we call the interval $[a,b]$, a strong uniqueness interval.

In many results on boundary value problems, the condition $q(x) \geq 0$ appears. The problem is in which of them we can put a strong uniqueness condition instead of $q(x) \geq 0$ condition ?

In deep connection with this problem is the following notion. Let us consider the second order linear differential equation

$$Ly := -y'' + py' + qy = 0, \text{ for } x \in [a,b], \text{ where } p, q \in C[a,b].$$

We suppose that $[a,b]$ is not a strong uniqueness interval with respect to (L, l_1, l_2), where $l_1(y)(a) = y(a)$ and $l_2(y)(b) = y(b)$.

By definition, an interval $[\alpha, \beta[\subset [a, b]$ is a maximum strong uniqueness interval in $[a, b]$ if $[\alpha, \beta]$ is not a uniqueness interval and each interval $[c, d] \subset [\alpha, \beta[$ is a strong uniqueness interval.
Let $h(p, q) := \min\{\beta - \alpha : [\alpha, \beta[$ is a maximum strong uniqueness interval in $[a, b]\}$. It is clear that $h(p, q) > 0$. An interesting problem is to give estimates for $h(p, q)$ in terms of p and q, see [41], [42], [4],[5], [47], [18], [46] (pp. 99-112).

Remark 2.2. For the Green function technique in nonlinear boundary value problems, see [5], [24], [44], [45], [48], [4], [11], [30], [36], [40], [42], [38],...

2.2. Saturated contraction principle

In our paper, we shall use the following variant of the contraction principle.

Theorem 2.3. [49] *Let (X, d) be a complete metric space and $f : X \to X$ be an l-contraction. Then we have:*

(i) There exists $x^* \in X$ such that

$$F_{f^n} = \{x^*\}, \ \forall \ n \in \mathbb{N}^*.$$

(ii) For all $x \in X$, $f^n(x) \to x^*$ as $n \to \infty$.

(iii) $d(x, x^*) \leq \psi(d(x, f(x)))$, $\forall \ x \in X$, where $\psi(t) = \dfrac{t}{1-l}$, $t \geq 0$.

(iv) If $\{y_n\}_{n\in\mathbb{N}}$ is a sequence in X such that

$$d(y_n, f(y_n)) \to 0 \ \text{as } n \to \infty,$$

then $y_n \to x^*$ as $n \to \infty$.

(v) If $\{y_n\}_{n\in\mathbb{N}}$ is a sequence in X such that

$$d(y_{n+1}, f(y_n)) \to 0 \ \text{as } n \to \infty,$$

then $y_n \to x^*$ as $n \to \infty$.

2.3. Fixed point of increasing operators

In this section, as a tool for the monotone iteration technique, two fixed point theorems for increasing operators on an ordered Banach space $(\mathbb{B}, +, \mathbb{R}, \leq)$ are presented.

Theorem 2.4. [2] *Let $(\mathbb{B}, +, \mathbb{R}, \leq)$ be an ordered Banach space and $X \subset \mathbb{B}$ be an order convex subset of $\mathbb{B}$. Let $f : X \to \mathbb{B}$ be an operator. We suppose that:*

(1) *f is increasing and continuous;*

(2) *f is relatively compact on every order interval in X;*

(3) *there exist $\overline{x}, \widehat{x} \in X$ with $\overline{x} < \widehat{x}$ such that $\overline{x} \leq f(\overline{x})$ and $f(\widehat{x}) \leq \widehat{x}$.*

Then:

(a) f has a minimum and a maximum fixed point in $[\overline{x}, \widehat{x}]$. Moreover, we have

$$x_{min} = \lim_{n\to\infty} f^n(\overline{x}) \ \text{and} \ x_{max} = \lim_{n\to\infty} f(\widehat{x}).$$

(b) If, additionally, $x_{min} \geq x_{max}$, then $f\big|_{[\overline{x}, \widehat{x}]}$ is a PO.

Theorem 2.5. *Let* $(\mathbb{B}, +, \mathbb{R}, \|\cdot\|, \leq)$ *be an ordered Banach space and*

$$P := \{x \in \mathbb{B} \mid x \geq 0\}.$$

Let $f, g : P \to P$ *be two operators. We suppose that:*

(i) f *and* g *are increasing and continuous;*

(ii) $\overline{f([0,x])}$ *and* $\overline{g([0,x])}$ *are compact subset for each* $x \in P$*;*

(iii) $f \leq g$*;*

(iv) $F_g = \{x^*\}$.

Then:

(1) *the interval* $[0, x^*]$ *is invariant for* f *and* g*;*

(2) $g : [0, x^*] \to [0, x^*]$ *is a Picard operator;*

(3) $\{f^n(0)\}_{n\in\mathbb{N}}$ *converges to the minimum fixed point of* f *in* $[0, x^*]$ *and* $\{f^n(x^*)\}$ *converges to the maximum fixed point of* f *in* $[0, x^*]$*;*

(4) *if* f *has a unique fixed point in* $[0, x^*]$*, then* $f : [0, x^*] \to [0, x^*]$ *is a Picard operator.*

Proof. (1) The fact that the interval $[0, x^*]$ is invariant with respect to g follows immediately by (i) and (iv). Let $x \in [0, x^*]$. Then $0 \leq x \leq x^*$. By (i) and (iii) we have

$$0 \leq f(0) \leq f(x) \leq f(x^*) \leq g(x^*) = x^*.$$

Thus, $[0, x^*]$ is invariant with respect to f.

(2) Take any $x \in [0, x^*]$. Then, by (i) and (iii), we have, for any $n \in \mathbb{N}^*$, that

$$0 \leq g^n(x) \leq x^*.$$

Consequently, the sequence $\{g^n(x)\}_{n\in\mathbb{N}}$ is contained in the compact set $\overline{g([0,x])}$, and thus, it has at least one limit point. By induction, it is easily seen that the sequence $\{g^n(0)\}_{n\in\mathbb{N}}$ is increasing. This implies that it has exactly one limit point and that the whole sequence converges to this point. Since g is continuous, $\{g^n(0)\}_{n\in\mathbb{N}}$ converges to x^*. Thus, for any $x \in [0, x^*]$, we have that

$$g^n(0) \leq g^n(x) \leq x^*, \text{ for any } n \in \mathbb{N}.$$

By passing to the limit we get the desired conclusion.

(3) The third conclusion follows by Theorem 2.4 (a).

(4) The last conclusion follows by Theorem 2.4 (b). □

Remark 2.6. For the fixed point theory in ordered sets and ordered Banach spaces see [2], [3], [33], [25], [39], [50],...

3. Heuristic considerations on particular solutions of Bratu equation

Let us consider Bratu's equation

$$-y'' = \lambda e^y, \ \lambda > 0. \tag{B_λ}$$

We start this section with some remarks on the solutions $y \in C^2(\mathbb{R})$ of this equation.

Remark 3.1. If y is a solution of (B_λ), then y is a strictly concave function. This implies that y' is a strictly decreasing function.

Remark 3.2. If y is a solution of (B_λ), then:

(1) $y(x+c)$, $x \in \mathbb{R}$ is a solution of (B_λ), $\forall\, c \in \mathbb{R}$;

(2) $y(-x+c)$, $x \in \mathbb{R}$ is a solution of (B_λ), $\forall\, c \in \mathbb{R}$;

(3) $c + y\left(e^{\frac{c}{2}} x\right)$, $x \in \mathbb{R}$ is a solution of (B_λ), $\forall\, c \in \mathbb{R}$.

Remark 3.3. Let y be a solution of (B_λ) such that there exists $x_0 \in \mathbb{R}$, with $y'(x_0) = 0$. Let $z(x) = y(2x_0 - x)$. We observe that: $z(x_0) = y(x_0)$ and $z'(x_0) = -y'(x_0) = 0$. By the uniqueness of the solution of Cauchy problem, we have that $y(x) = z(x)$, $\forall\, x \in \mathbb{R}$, i.e., $y(x) = y(2x_0 - x)$, $\forall\, x \in \mathbb{R}$. From this, it follows that $y(x_0 - x) = y(x + x_0)$, $\forall\, x \in \mathbb{R}$, i.e., the graphic of y is symmetric with respect to the line, $x = x_0$.

Now let us make the change of the function y, by $e^y = \dfrac{1}{u^2}$. Then for u we have the equation

$$2(u''u - u'^2) = \lambda.$$

If u is such that $u'' = u$ and $u^2 - u'^2 = 1$, then $y = \ln \dfrac{1}{u^2}$ is a solution of (B_2). Such a function u is, for example, $u(x) = \cosh(x)$. Therefore, the function $y(x) = -2\ln(\cosh x)$, $x \in \mathbb{R}$ is a solution of (B_2).

In order to find a solution for (B_λ), let us try with

$$y(x) = -2\ln(c_1 \cosh c_2 x),\ x \in \mathbb{R},\ c_1 \in \mathbb{R}_+^*,\ c_2 \in \mathbb{R}.$$

Such a function is a solution of (B_λ), if $c_1 c_2 = \sqrt{\dfrac{\lambda}{2}}$. By Remark 3.2(1), if $c_1 c_2 = \sqrt{\dfrac{\lambda}{2}}$, the function

$$y(x) = -2\ln[c_1 \cosh(c_2(x - x_0) + c_3)],\ x, x_0, c_3 \in \mathbb{R}, c_1, c_2 \in \mathbb{R}^*$$

is a solution of (B_λ).

Now we shall prove that, for each $x_0 \in \mathbb{R}$, this is the general solution of (B_λ). For to do this, let us consider the Cauchy problem

$$-y'' = \lambda e^y,\ y(x_0) = y_0,\ y'(x_0) = y_0',\ x_0, y_0 \in \mathbb{R}.$$

From the Cauchy problem, we have for c_1, c_2, c_3, the following system of equations

$$\begin{cases} c_1 c_2 = \sqrt{\dfrac{\lambda}{2}} \\ c_1 \cosh c_3 = e^{-\frac{y_0}{2}} \\ c_2 \tanh c_3 = -\dfrac{y_0'}{2} \end{cases}$$

Since this system has a unique solution, the conclusion is obvious. Moreover, we have the following result.

Theorem 3.4. *The Cauchy problem for (B_λ) has a unique saturated solution defined on $\mathbb{R}$.*

From this theorem it follows that:

If $y \in C^2[0,b]$ or $y \in C^2(]0,b[) \cap C[0,b]$ is a solution of Bratu problem

$$-y'' = \lambda e^y,\ \lambda > 0,\ y(0) = 0,\ y(b) = 0,\ b > 0 \tag{$B_{\lambda,b}$}$$

then there exists a unique solution $\widetilde{y} \in C^\infty(\mathbb{R})$ of (B_λ) such that $\widetilde{y}\big|_{[0,b]} = y$.

4. Some variants of Bratu theorem

We start by considering the following problems (for $\lambda > 0$):

Bratu problem:

$$-y'' = \lambda e^y,\ y(0) = 0,\ y(b) = 0,\ b > 0 \tag{$B_{\lambda,b}$}$$

Gelfand problem:

$$-y'' = \lambda e^y,\ y(-a) = y(a) = 0,\ a > 0 \tag{$G_{\lambda,a}$}$$

Cauchy problem:

$$-y'' = \lambda e^y,\ y(0) = 0,\ y'(0) = \mu > 0 \tag{$C_{\lambda,\mu}$}$$

Nicoletti problem:

$$-y'' = \lambda e^y,\ y(0) = 0,\ y(x_0) = a,\ x_0 > 0 \tag{N_{λ,x_0}}$$

For the problem $(B_{\lambda,b})$ the following result is well known.

Bratu Theorem. ([10], [12]) *For each $\lambda > 0$, there exists $b^*(\lambda) > 0$ such that:*
(1) for $0 < b < b^*(\lambda)$, the problem $(B_{\lambda,b})$ has two solutions;
(2) the problem $(B_\lambda, b^*(\lambda))$ has a unique solution;
(3) for $b > b^*(\lambda)$, the problem $(B_{\lambda,b})$ has no solution.
For each $b > 0$, there exists $\lambda^*(b)$ such that:
(1′) for $0 < \lambda < \lambda^*(b)$, the problem $(B_{\lambda,b})$ has two solutions;
(2′) the problem $(B_{\lambda^*(b),b})$ has a unique solution;
(3′) for $\lambda > \lambda^*(b)$, the problem $(B_{\lambda,b})$ has no solutions.

There exist some deep relations between the problems $(B_{\lambda,b})$, $(G_{\lambda,a})$, $(C_{\lambda,\mu})$ and (N_{λ,x_0}).

For example, from Bratu's Theorem we have:

Gelfand Theorem. ([23]) *For each $\lambda > 0$ there exists $a^*(\lambda)$ such that:*
(1) for $0 < a < a^*(\lambda)$, the problem $(G_{\lambda,a})$ has two solutions;
(2) the problem $(G_\lambda, a^*(\lambda))$ has a unique solution;
(3) for $a > a^*(\lambda)$, the problem $(G_{\lambda,a})$ has no solutions.

Proof. $y \in C^2(\mathbb{R})$ is a solution of $(B_{\lambda,b})$ if and only if $y\left(x + \dfrac{b}{2}\right)$ is a solution of $\left(G_{\lambda,\frac{b}{2}}\right)$. See Remark 3.2(1) and Theorem 3.4.

From our remarks in Section 3, we also have:

Theorem 4.1. (1) *If* $y \in C^2(\mathbb{R})$ *is a solution of* $(B_{\lambda,b})$, *then* y *is a solution of* $\left(N_{\lambda,\frac{b}{2}}\right)$.

(2) *If* $y \in C^2(\mathbb{R})$ *is a solution of* (N_{λ,x_0}), *then* y *is a solution of* $(B_{\lambda,2x_0})$.

(3) *If* y^* *is the unique solution of* $(B_{\lambda^*(b),b})$ *then there exists a unique* $\mu^* > 0$ *such that* y^* *is a solution of* $(C_{\lambda^*(b),\mu^*})$. *If* $0 < \lambda < \lambda^*(b)$, *then there exists* $\mu_1 < \mu^* < \mu_2$ *such that if* y_i *is the unique solution of* (C_{λ,μ_i}), *then the solution set of* $(B_{\lambda,b})$ *is* $\{y_1, y_2\}$. *Moreover,* $y_1 < y^* < y_2$.

In what follow, we shall study the problem $(B_{\lambda,b})$, where $0 < \lambda < \lambda^*(b)$. From Theorem 4.1 it is clear that the problem $(B_{\lambda,b})$ has a unique solution in the order interval $[0, y^*]$. On the other hand, the problem $(B_{\lambda,b})$ is equivalent to the fixed point equation

$$y(x) = \lambda \int_0^b G(x,s)e^{y(s)}ds, \; x \in [0,b]. \tag{4.1}$$

Let $P_\lambda : C([0,b],\mathbb{R}_+) \to C([0,b],\mathbb{R}_+)$, be defined by

$$P_\lambda(y)(x) := \lambda \int_0^b G(x,s)e^{y(s)}ds, \; x \in [0,b]. \tag{4.2}$$

Notice that the operator P_λ is completely continuous, increasing and $P_\lambda([0,y^*]) \subset P_\lambda([0,y^*])$, for $0 < \lambda < \lambda^*(b)$. By Theorem 2.4 we have the following result.

Theorem 4.2. *For* $0 < \lambda < \lambda^*(b)$, *the mapping* $P_\lambda : [0,y^*] \to [0,y^*]$ *defined by* (4.2) *is a Picard operator.*

Proof. By Theorem 4.1, we get that $F_{P_\lambda} = \{y_\lambda\}$. By Theorem 2.4 we obtain

$$P_\lambda^n(0) \to y_\lambda \text{ as } n \to \infty \text{ and } P_\lambda^n(y^*) \to y_\lambda \text{ as } n \to \infty.$$

Since P_λ is increasing, if $y \in [0,y^*]$, then $P^n(0) \le P^n(y) \le P^n(y^*)$. This implies that $P^n(y) \to y_\lambda$ as $n \to \infty$. □

On the other hand, since $\lambda > 0$ and $b > 0$, then

$$\|P_\lambda(y)\|_\infty \le \frac{\lambda b^2}{8} e^{\|y\|_\infty}, \quad \text{for all } y \in C([0,b],\mathbb{R}_+).$$

Let $M > 0$. If λ and b are such that $\dfrac{\lambda b^2}{8}e^M \le M$, then the order interval $[0,M] \subset C([0,b],\mathbb{R}_+)$ is invariant subset of P_λ. If, in addition, $\dfrac{\lambda b^2}{8}e^M < 1$, then $P_\lambda : [0,M] \to [0,M]$ is a contraction. Thus, in terms of the Saturated Contraction Principle (see Theorem 2.3 or Theorem 1.1 in [49]) we can obtain more information with respect to the solution of $(B_{\lambda,b})$ in $[0,M]$. We have the following result.

Theorem 4.3. *Let us consider the problem* $(B_{\lambda,b})$. *For* $0 < \lambda < \lambda^*(b)$ *and* $\lambda b^2 < \frac{8}{e}$, *take any* $M \in]0, \ln \frac{8}{\lambda b^2}[$. *Then, the following conclusions hold:*

(i) the problem $(B_{\lambda,b})$ *has a unique solution* y^* *in the order interval* $[0,M] \subset C([0,b],\mathbb{R}_+)$;

(ii) the sequence $(y_n)_{n\in\mathbb{N}}$ defined by

$$y_{n+1}(x) := \lambda \int_0^b G(x,s)e^{y_n(s)}ds,\ x \in [0,b],\ n \in \mathbb{N},$$

(where y_0 is arbitrary in the order interval $[0,M] \subset C([0,b],\mathbb{R}_+)$) converges to y^*;

(iii) for every y from the order interval $[0,M] \subset C([0,b],\mathbb{R}_+)$ we have

$$\|y - y^*\|_\infty \le \frac{1}{1-K}\|y - P_\lambda y\|_\infty,$$

where $P_\lambda(y)(x) := \lambda \int_0^b G(x,s)e^{y(s)}ds$ and $K := \dfrac{\lambda b^2}{8}e^M$;

(iv) if $(u_n)_{n\in\mathbb{N}}$ is a sequence in the order interval $[0,M] \subset C([0,b],\mathbb{R}_+)$ such that

$$\|u_n - P_\lambda u_n\|_\infty \to 0 \text{ as } n \to \infty,$$

then $u_n \to y^*$ as $n \to \infty$;

(iv) if $(u_n)_{n\in\mathbb{N}}$ is a sequence in the order interval $[0,M] \subset C([0,b],\mathbb{R}_+)$ such that

$$\|u_{n+1} - P_\lambda u_n\|_\infty \to 0 \text{ as } n \to \infty,$$

then $u_n \to y^*$ as $n \to \infty$.

Proof. Consider the fixed point equation equation (4.1) and the operator P_λ defined by (4.2). By the above assumptions, we have that

$$\frac{\lambda b^2}{8}e^M \le M \text{ and } \frac{\lambda b^2}{8}e^M < 1.$$

Thus, $P_\lambda : [0,M] \to [0,M]$ and it is a contraction. The rest of the conclusions follow from Theorem 1.1 in [49]. □

Remark 4.4. For a better understanding of this result it is useful to compare it with Theorem 1 in [29].

5. Bratu-type problems

From the above considerations (see Section 3) on Bratu's equation, we are motivated to adopt the following notions.

Let us consider the equation

$$-y'' = \lambda f(y) \qquad (E_{f,\lambda})$$

where $\lambda > 0$, $f \in C^2(\mathbb{R})$ and $f^{(k)}(t) > 0$, for all $t \in \mathbb{R}$ and $k \in \{0,1,2\}$. Since f is locally Lipschitz, each Cauchy problem associated to $(E_{f,\lambda})$ has a unique saturated solution $y \in C^2(]x_-,x_+[)$. We suppose that: $x_- = -\infty$ and $x_+ = +\infty$.

By definition, the equation $(E_{f,\lambda})$ is of Bratu-type if the above conditions are satisfied. In this case, we denote it by $(BT(f,\lambda))$.

As in Section 3, we have:

Remark 5.1. If y is a solution of $(BT(f,\lambda))$ then:

(1) y is strictly concave function;

(2) the function $x \mapsto y(x+c)$, $x \in \mathbb{R}$, is a solution of $(BT(f,\lambda))$, for all $c \in \mathbb{R}$;

(3) the function $x \mapsto y(-x+c)$, $x \in \mathbb{R}$, is a solution of $(BT(f,\lambda))$, for all $c \in \mathbb{R}$;

(4) if $y'(x_0) = 0$, then $y(x_0 - x) = y(x_0 + x)$, $\forall\ x \in \mathbb{R}$;

(5) if $y(0) = 0$, $y'(x_0) = 0$, $0 < x_0$, then $y(2x_0) = 0$;

(6) if $y(a) = 0$, $y(b) = 0$, $a < b$, then $y'(a) > 0$, $y'(b) < 0$ and $y(x) > 0$, $\forall\ x \in]a,b[$.

By definition, we call the problem

$$\begin{cases} -y'' = \lambda f(y) \\ y(0) = 0,\ y(b) = 0,\ 0 < b, \end{cases}$$

the Bratu-type problem. We denote it by $(BT(f,\lambda,b))$.

For the Bratu-type problem, we have the following result.

Theorem 5.2. *For each $\lambda > 0$, there exists $b^*(\lambda) > 0$ such that:*

(1) for $0 < b < b^*(\lambda)$, the problem $(BT(f,\lambda,b))$ has two solutions;

(2) the problem $(BT(f,\lambda,b^*(\lambda)))$ has a unique solution;

(3) for $b > b^*(\lambda)$, the problem $(BT(f,\lambda,b))$ has no solution.

For each $b > 0$, there exists $\lambda^*(b)$ such that:

(1′) for $0 < \lambda < \lambda^*(b)$, the problem $(BT(f,\lambda,b))$ has two solutions;

(2′) the problem $(BT(f,\lambda^*(b),b))$ has a unique solution;

(3′) for $\lambda > \lambda^*(b)$, the problem $(BT(f,\lambda,b))$ has no solutions.

Proof. Let $g(t) = \displaystyle\int_0^t f(s)ds + 1$. In terms of g, the Bratu-type problem takes the following form

$$-y'' = \lambda g'(y),\ y(0) = 0,\ y(b) = 0.$$

If y is a solution of this problem, then $y'(0) = \mu > 0$, and from

$$-2y'y'' = 2\lambda y' g'(y),$$

we have that

$$-y'^2(x) + \mu^2 = 2\lambda g(y(x)) - 2\lambda,\ \forall\ x \in [0,b].$$

From now on, we follow Bratu's proof of his theorem. □

6. Other generalizations

In this section, we shall consider the following boundary value problem with increasing nonlinearity (see $ 2.2), denoted by (BVP):

$$\begin{cases} L(y) := -y'' + p(x)y' + q(x)y = f(x,y) \\ l_1(y)(a) = 0,\ l_2(y)(b) = 0 \end{cases}$$

where $p, q \in C[a,b]$, $f \in C([a,b] \times \mathbb{R}_+)$, $f(x,t) > 0$, for all $x \in [a,b]$, $t \in \mathbb{R}_+$, and the interval $[a,b]$ is a strong uniqueness interval with respect to (L, l_1, l_2).

Let us denote

$$S_+(BVP) := \{y \in C^2([a,b], \mathbb{R}_+) \mid y \text{ is a solution of } (BVP)\}.$$

The problem (BVP) is equivalent to the fixed point equation (in $C([a,b], \mathbb{R}_+)$,

$$y(x) = \int_a^b G(x,s) f(s, y(s)) ds, \; x \in [a,b], \tag{IE}$$

where $G(x,s)$ is the Green function corresponding to (L, l_1, l_2).

Since $[a,b]$ is a strong uniqueness interval, hence $G(x,s) \geq 0$, $\forall\, x, s \in [a,b]$.

We consider the operator $P : C([a,b], \mathbb{R}_+) \to C([a,b], \mathbb{R}_+)$ defined by

$$P(y)(x) := \text{second part of } (IE).$$

It is clear that $S_+(BVP) = F_P$.

For the problem (BVP), we have the following result:

Theorem 6.1. *In addition, we suppose that:*

(1) $f(x, \cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ *is strictly increasing,* $\forall\, x \in [a,b]$*;*

(2) *the* (BVP) *has a positive strict supersolution denoted by* $\widehat{y}$*.*

In these conditions we have that:

(i) the ordered set $(S_+(BVP), \leq)$ *has a minimum element* y_{min} *(see* [39]*);*

(ii) $y_{min} = \lim\limits_{n \to \infty} P^n(0)$*, in* $(C([a,b], \mathbb{R}_+, \|\cdot\|_\infty))$*;*

(iii) if $\lim P^n(\widehat{y}) = y_{min}$*, then* $P\big|_{[0,\widehat{y}]}$ *is PO.*

Proof. First, we remark that the operator P is completely continuous and strictly increasing. Moreover, 0 is a strict lower fixed point of P. Now the proof follows from Theorem 3.4.

Remark 6.2. If $q(x) \geq 0$, then Theorem 6.1 generalizes some results given in [31], [2], [3], [43],...

Example 6.3. We consider the boundary value problem

$$L(y) = \sum_{k=1}^{m} \lambda_k e^{u_k x}, \; x \in [a,b],$$
$$l_1(y)(a) = 0, \; l_2(y)(b) = 0.$$

If $\lambda_k > 0$ and $\mu_k > 0$, then this problem satisfies the conditions of Theorem 6.1.

In what follows, we shall use Hadamard linearized technique to study some semilinear problems. Let us consider the following second-order linear differential equations

$$Ly := -y'' + py' + qy = 0, \text{ for } x \in [a,b], \text{ where } p, q \in C[a,b] \tag{6.1}$$

and

$$Ly := -y'' + py' + qy = f(x,y), \text{ for } x \in [a.b], p, q \in C[a,b] \text{ and } f \in C([a,b] \times J), \tag{6.2}$$

with $J \subset \mathbb{R}$ a nondegenerate interval.
If $y \in C^2[a,b]$ is a nontrivial solution of (6.1) such that $y(x) \geq 0$ for every $x \in [a,b]$, then $y(x) > 0$ for every $x \in]a,b[$.

By this well-known property, we have the following lemma.

Lemma 6.4. *We suppose that* $\frac{\partial f}{\partial y} \in C[a,b] \times J)$. *If* $y_1, y_2 \in C^2[a,b]$ *are two solutions of* (6.2) *such that* $y_1(x) \leq y_2(x)$ *for every* $x \in [a,b]$, *then* $y_1(x) < y_2(x)$ *for every* $x \in]a,b[$.

Proof. We have that

$$L(y_2 - y_1)(x) = f(x, y_1(x)) - f(x, y_2(x)) =$$
$$\int_0^1 \frac{\partial f}{\partial y}(x, y_1(x) + t\,(y_2(x) - y_1(x)))\,dt \cdot (y_2 - y_1)(x),$$

i.e., $u := y_2 - y_1$ is a solution of the linear equation

$$Lu(x) - Q(x, y_1(x), y_2(x))u(x) = 0,$$

where $Q(x, y_1(x), y_2(x)) := \int_0^1 \frac{\partial f}{\partial y}(x, y_1(x) + t\,(y_2(x) - y_1(x)))\,dt$. Since $y_1(x) \leq y_2(x)$ for every $x \in [a,b]$, then $y_1(x) < y_2(x)$ for every $x \in]a,b[$. □

By the above lemma, we also have the following theorem.

Theorem 6.5. *We suppose that the following assumptions hold:*

(1) $\frac{\partial f}{\partial y} \in C[a,b] \times J)$;

(2) $\frac{\partial f}{\partial y}(x,t) < 0$ *for every* $x \in [a,b]$ *and* $t \in J$;

(3) the interval $[a,b]$ *is a strong uniqueness interval corresponding to* (L, l_1, l_2), *where* $l_1(y)(a) = y(a)$ *and* $l_2(y)(b) = y(b)$.

Then, the boundary value problem

$$\begin{cases} L(y) := -y'' + p(x)y' + q(x)y = f(x,y), x \in [a,b] \\ y(a) = 0,\; y(b) = 0 \end{cases}$$

has at most a solution.

Proof. Let $y_1, y_2 \in C^2[a,b]$ are two solutions of (6.5) and $u := y_2 - y_1$. Then u is a solution of the linear equation

$$Lu(x) - Q(x, y_1(x), y_2(x))u(x) = 0,$$

where $Q(x, y_1(x), y_2(x))$ was introduced in Lemma 6.4 and has the property that $Q(x, y_1(x), y_2(x))u(x) < 0$ for every $x \in [a,b]$. Thus $q(x) - Q(x, y_1(x), y_2(x))u(x) > 0$ for every $x \in [a,b]$. By (3) and Theorem 2.1 (ii), we get that $u := 0$. □

Another result in the linear case is the following Sturm comparison theorem ([34], [46], [48], [24]).

Let $p, q_1, q_2 \in C[a,b]$ with $q_1(x) < q_2(x)$ for every $x \in [a,b]$. Let y be a nontrivial solution of

$$L(y) := -y'' + p(x)y' + q_1(x)y = 0, x \in [a,b]$$

and z be a notrivial solution of

$$L(z) := -z'' + p(x)z' + q_2(x)z = 0, x \in [a,b].$$

If $z(a) = z(b) = 0$, then there exists $x_0 \in]a, b[$ such that $y(x_0) = 0$. By this result, we immediately obtain the following theorem.

Theorem 6.6. *We consider the boundary value problem (*6.5*). We suppose that the following assumptions hold:*

(1) $\frac{\partial f}{\partial y} \in C[a, b] \times J)$;

(2) $\frac{\partial f}{\partial y}(x, \cdot) : J \to \mathbb{R}$ *is strictly increasing for every* $t \in J$.

*Then, each totally ordered subset of the solution set of (*6.5*) has at most two elements.*

Proof. Let $y_1 \leq y_2 \leq y_3$ three solutions of (6.5). By Lemma 6.4 we get that $y_1 < y_2 < y_3$ for every $x \in]a, b[$. Let $y := y_3 - y_1$ and $z := y_2 - y_1$. Then

$$Ly(x) - Q(x, y_1(x), y_3(x))y(x) = 0,$$

and

$$Lz(x) - Q(x, y_1(x), y_2(x))z(x) = 0,$$

for every $x \in [a, b]$. Since $\frac{\partial f}{\partial y}(x, \cdot)$ is strictly increasing for every $t \in J$ and $y(x) > z(x)$ for every $x \in]a, b[$, by Sturm comparison theorem we get that y must change the sign in $]a, b[$. Since $y(x) > 0$ for every $x \in]a, b[$, this is a contradiction to our initial assumption. The proof is complete. □

Remark 6.7. For similar results to Theorem 6.6 see [44] (pp. 253-254) and the references therein. Another result for the boundary value problem (6.5) can be obtained by Theorem 2.4.

7. Numerical analysis of Bratu type problems

We know that Bratu's problem (1.1) has the exact solution of the form:

$$y(x) = -2 \log \left(\frac{\cosh[(x - \frac{1}{2})\frac{\theta}{2}]}{\cosh \frac{\theta}{4}} \right),$$

where θ is the solution of the equation

$$\theta = \sqrt{2\lambda} \cosh \frac{\theta}{4}. \tag{7.1}$$

To get numerical approximations for the solutions of the equation (7.1) we apply the Newton's method for finding the roots of the function

$$\varphi(\theta) = \theta - \sqrt{2\lambda} \cosh \frac{\theta}{4},$$

defined by

$$\theta_{n+1} = \theta_n - \frac{\varphi(\theta_n)}{\varphi'(\theta_n)}, \tag{7.2}$$

with a starting value $\theta_0 \in [a,b]$ chosen such that in $[a,b]$ equation (7.1) has only one solution and $\varphi'(\theta_0) \neq 0$. It is clear that $\varphi \in C^2(\mathbb{R})$ is a concave function with a maximum in

$$\theta^*(\lambda) = 4arc\sinh\left(\frac{4}{\sqrt{2\lambda}}\right).$$

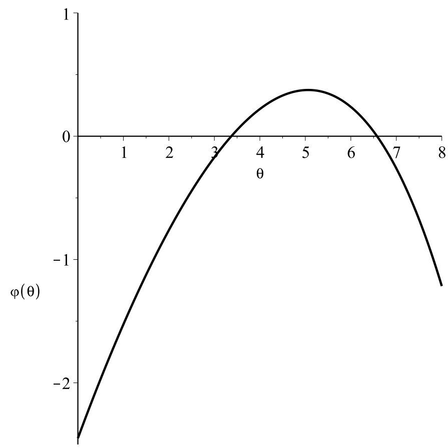


Figure 1. The graph of $\varphi(\theta)$ for $\lambda = 3$.

The existence of the critical value λ_c comes from the condition that, in order to have solutions for the equation (7.1), the maximum value $\varphi(\theta^*(\lambda))$ should be nonnegative. Thus, λ_c is obtained as a solution of

$$\varphi(\theta^*(\lambda)) = 0,$$

and, in the case of Bratu's problem (1.1), we have $\lambda_c \approx 3.513830719$.

If $0 < \lambda < \lambda_c$, then $\varphi(\theta^*(\lambda)) > 0$ and the equation has two solutions $\theta_1(\lambda)$, $\theta_2(\lambda)$. Since $\varphi(0) < 0$ and $\lim\limits_{\theta\to+\infty} \varphi(\theta) = -\infty$ then $\theta_1(\lambda) \in (0, \theta^*(\lambda))$ and $\theta_2(\lambda) \in (\theta^*(\lambda), +\infty)$. In order to get a numerical approximation of $\theta_1(\lambda)$, respectively, of $\theta_2(\lambda)$, we may choose as a starting value $\theta_0 \in (0, \theta^*(\lambda))$, respectively, $\theta_0 \in (\theta^*(\lambda), \theta^*(\lambda) + \varepsilon)$ for some $\varepsilon > 0$.

If $\lambda = \lambda_c$ then $\varphi(\theta^*(\lambda_c)) = 0$, so $\theta^*(\lambda_c)$ is the unique solution of (7.1). This value can be obtained as the limit $\theta_1(\lambda)$ or $\theta_2(\lambda)$ when $\lambda \to \lambda_c$.

In the case of $\lambda = 3$, we have

$$\theta^*(3) = 4\,\text{arcsinh}\left(\frac{2}{3}\sqrt{6}\right) \approx 5.065364187$$

and the following iterations:

	$\theta_0 = 0$	$\theta_0 = 10$
$\theta_1 =$	2.4494897427831780982	8.1438003057516703864
$\theta_2 =$	3.2377069463405279948	7.0734727359775392946
$\vdots$	$\vdots$	$\vdots$
$\theta_{10} =$	3.3735077642858915405	6.5765692592543752601
$\vdots$	$\vdots$	$\vdots$
$\theta_{19} =$	3.3735077642858915405	6.5765692592543752601
$\theta_{20} =$	3.3735077642858915405	6.5765692592543752601

For different values of $\lambda < \lambda_c$, we obtain the following approximating values for $\theta_1(\lambda)$ and $\theta_1(\lambda)$:

λ	$\theta_1(\lambda)$	$\theta_2(\lambda)$
1	1.5171645990507543685	10.938702772122106800
2	2.3575510538774020426	8.5071995707130261296
3	3.3735077642858915407	6.5765692592543752601
3.513	4.7374700066634551382	4.8604846857553034188
3.513830719125	4.7987137042679359281	4.7987154177935504693

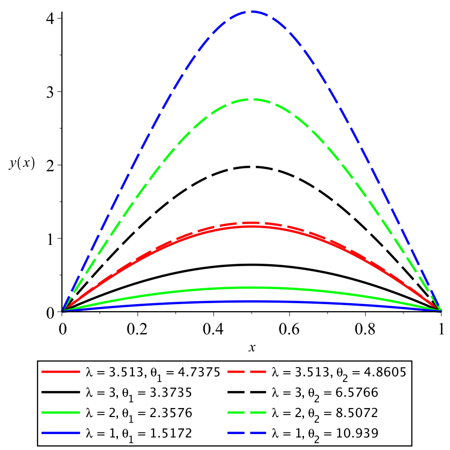


FIGURE 2. The graph of Bratu's problem solutions for different values of λ.

References

[1] Agarwal, R.P., Hodis, S., O'Regan, D., *500 Examples and Problems of Applied Differential Equations*, Springer, 2019.

[2] Amann, H., *Fixed point equations and nonlinear eigenvalue problems in ordered Banach spaces*, SIAM Review, **18**(1976), no. 4, 620-709.

[3] Amann, H., *Supersolutions, monotone iterations and stability*, J. Diff. Eq., **21**(1976), 363-377.

[4] Bailey, P., Shampine, L.F., Waltman, P., *Existence and uniqueness of solution of the second order boundary value problem*, Bull. Amer. Math. Soc., **72**(1966), 96-98.

[5] Bailey, P., Shampine, L.F., Waltman, P., *Nonlinear Two Point Boundary Value Problems*, Acad. Press, New York, 1968.

[6] Bandle, C., *Existence theorems, qualitative results and a priori bounds for a class of nonlinear Dirichlet problems*, Arch. Rational Mech. Anal., **58**(1975), 219-238.

[7] Ben-Romdhane, M., Temimi, H., Baccouch, M., *An iterative finite difference method for approximating the two-branched solution of Bratu's problem*, Appl. Numerical Math., **139**(2019), 62-76.

[8] Bougoffa, L., *Exact solutions of a generalized Bratu equation*, Romanian J. Physics, **62**(2017), article no. 110, 1-5.

[9] Boyd, J.P. *Chebyshev polynomial expantions for simultaneous approximation of two branches of a function with application to the one-dimensional Bratu equation*, Appl. Math. Comput., **142**(2003), 189-200.

[10] Bratu, Gh., *Sur certaines équations integrales nonlinéaire*, C.R. Acad. Sci. Paris, **150**(1910), 896-899.

[11] Bratu, Gh., *Sur l'équation intégrale exponentielle*, C.R. Acad. Sci. Paris, **152**(1911), 1048-1050.

[12] Bratu, Gh., *Sur les équations intégrales nonlinéaires*, Bulletin de la Soc. Math. France, **42**(1914), 113-142.

[13] Buckmire, R. *Application of a Mickens finite-difference scheme to the cylindrical Bratu-Gelfand problem*, Numer. Methods Partial Diff. Equations, **20**(2004), no. 3, 327-337.

[14] Caglar, H., Caglar, N., Özer, M., Anagnostopoulos, A.N., *B-spline method for solving Bratu's problem*, J. Comput. Math., **87**(2010) no. 8, 1885-1891.

[15] Chandrasekhar, S., *An Introduction to the Study of Stellar Structure*, Dover Publ., New York, 1957.

[16] Cohen, N., Toledo Benavides, J.V., *Exact solutions of Bratu and Liouville equations*, Anais do CNMAC, **3**(2010), 750-756.

[17] Feng, X., He, Y., Meng, J., *Application of homotopy perturbation method to the Bratu-type equations*, Topological Meth. in Nonlinear Anal., **31**(2008), 243-252.

[18] Foiaş, C., Gussi, G., Poenaru, V., *Despre problema polilocală la ecuaţii diferenţiale liniare de ordinul al doilea*, Buletin Ştiinţific, Sectiunea de Ştiinţe Matematice şi Fizice, **7**(1955), no. 3, 699-721.

[19] Frank-Kamenetskii, D.A., *Calculation of thermal explosion limits*, Acta Phys., USSR, **10**(1939), 365.

[20] Frank-Kamenetskii, D.A., *Diffusion and Heat Exchange in Chemical Kinetics*, Princeton Univ. Press, 1955.

[21] Fujita, H., *On the nonlinear equations* $\Delta u + e^u = 0$ *and* $\dfrac{\partial v}{\partial t} = \Delta v + e^v$, Bull. Amer. Math. Soc., **75**(1969), 132-135.

[22] Galaktionov, V.A., Vázquez, J.L., *The problem of blow-up in nonlinear parabolic equations*, Discrete and Continuous Dynamical Systems, **8**(2002), no. 2, 399-433.

[23] Gelfand, I.M., *Some problems in the theory of quasilinear equations*, Amer. Math. Soc. Transl., **29**(1963), 295-381 (Usp. Mat. Nauk, **14**(1959), no. 2, 87-158).

[24] Hartman, P., *Ordinary Differential Equations*, Wiley New York, 1964.

[25] Hirsch, M.W., *The dynamical systems approach to differential equations*, Bull. Amer. Math. Soc., **11**(1984), no. 1, 1-64.

[26] Jacobsen, J., Schmitt, K., *The Liouville-Bratu-Gelfand problem for radial operators*, J. Diff. Eq., **184**(2002), 283-298.

[27] Joseph, D., Lundgren, *Quasilinear Dirichlet problems driven by positive source*, Arch. Rational Mech. Anal., **49**(1973), 241-269.

[28] Joseph, D., Sparrow, E.M., *Nonlinear diffusion induced by nonlinear sources*, Quarterly Appl. Math., **28**(1970), no. 3, 327-342.

[29] Kafri, H.Q., Khuri, S.A., *Bratu's problem: A novel approach using fixed point iterations and Green's functions*, Computer Physics Communications, **198**(2016), 97-104.

[30] Keller, H.B., *Existence theory for two point boundary value problems*, Bull. Amer. Math. Soc., **72**(1966), 728-731.

[31] Keller, H.B., Cohen, D.S., *Some positone problems suggested by nonlinear heat generation*, J. Math. Mech., **16**(1967), no. 12, 1361-1376.

[32] Khuri, S.A., Louhichi, I., *A novel Ishikawa-Green's fixed point scheme for the solution of BVPs*, Appl. Math. Letters, **82**(2018), 50-57.

[33] Krasnoselskii, M.A., *Positive Solutions of Operator Equations*, Noordhof, 1964.

[34] Kreith, K. *PDE generalization of the Sturm comparison theorem*, Memoirs Amer. Math. Soc., **48**(1984), no 298, 31-46.

[35] Laetsch, T., *On the number of solutions of boundary value problems with convex nonlinearities*, J. Math. Anal. Appl., **35**(1971), 389-404.

[36] Lebovitz, N., *Oscillation theory and the spectra of eigenvalues*, in: Ordinary Differential Equations, http://people.cs.uchicago.edu/ lebovitz/odes.html

[37] Liouville, J., *Sur l'équation aux différences partielles* $\frac{d^2 \log \lambda}{du dv} \pm \lambda 2a^2 = 0$, J. Math. Pures Appl., **18**(1853), 71-72.

[38] Miller, K.S., Schiffer, M.M., *Monotonic properties of the Green's function*, Proc. Amer. Math. Soc., **3**(1952), 848-956.

[39] Păcurar, M., Rus, I.A., *Some remarks on the notations and terminology in the ordered set theory*, Creat. Math. Inf., **27**(2018), no. 2, 191-195.

[40] Petruşel, A., Rus, I.A., *A class of functional-integral equations with applications to bilocal problem*, in: Topics in Mathematical Analysis and Applications (T.M. Rassias, L. Toth - Eds.), Springer, 2014, 609-631.

[41] Picard, E., *Mémoire sur la théorie des équations aux dérivées partielles et la méthode des approximations successives*, J. Math. Pures Appl., **6**(1890), 145-210.

[42] Picard, E., *Sur l'application des méthodes d'approximations successives à l'étude de certaines équations différentielles ordinaires*, J. Math. Pures Appl., **9**(1893), 217-272.

[43] Picard, E., *Sur certains exemples singuliers d'approximations successives*, Comptes Rendus, **126**(1898), 497-500.

[44] Piccinini, L.C., Stampacchia, G., Vidossich, G., *Ordinary Differential Equations in* $\mathbb{R}^n$, Springer, 1984.

[45] Polyamin, A.D., Zaitsev, V.F., *Handbook of Ordinary Differential Equations: Exact Solutions, Methods and Problems*, CRC Press, Boca Raton, 2018.

[46] Reid, W.T., *Sturmian Theory for Ordinary Differential Equations*, Springer, 1980.

[47] Rus, I.A., *Sur la positivité de la fonction de Green correspondante au problem bilocal*, Glasnik Mat., **5**(1970), 251-257.

[48] Rus, I.A., *Ecuaţii diferenţiale, ecuaţii integrale şi sisteme dinamice*, Transilvania Press Cluj-napoca, 1996.

[49] Rus, I.A., *Some variants of contraction principle, generalization and applications*, Stud. Univ. Babeş-Bolyai Math., **61**(2016), no. 3, 343-358.

[50] Rus, I.A., Petruşel, A., Petruşel, G., *Fixed Point Theory*, Cluj Univ. Press, Cluj-Napoca, 2008.

[51] Syam, M.I., Hamdan, A., *An efficient method for solving Bratu equations*, Appl. Math. Comput., **176**(2006), 704-713.

[52] Wazwaz, A.M., *Adomian decomposition method for a reliable treatment of the Bratu-type equations*, Appl. Math. Comput., **166**(2005), 652-663.

Adrian Petruşel
Babeş-Bolyai University, Faculty of Mathematics and Computer Science
Kogălniceanu Str. no. 1, 400084 Cluj-Napoca, Romania
and
Academy of Romanian Scientists Bucharest, Romania
e-mail: petrusel@math.ubbcluj.ro

Ioan A. Rus
Babeş-Bolyai University, Faculty of Mathematics and Computer Science
Kogălniceanu Str. no. 1, 400084 Cluj-Napoca, Romania
e-mail: iarus@math.ubbcluj.ro

Marcel Adrian Şerban
Babeş-Bolyai University, Faculty of Mathematics and Computer Science
Kogălniceanu Str. no. 1, 400084 Cluj-Napoca, Romania
e-mail: mserban@math.ubbcluj.ro

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 47–53
DOI: 10.24193/subbmath.2021.1.04

On the existence of a periodic solution of the Liénard system

Alexander Ignatyev

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. The Liénard system $\frac{dx}{dt} = y,\ \frac{dy}{dt} = -f(x)y - g(x)$ is considered. Under some assumptions on functions $f(x)$ and $g(x)$, we prove the existence of a periodic solution of this system.

Mathematics Subject Classification (2010): 34C05, 34C07, 34C25, 34C15, 34A26.

Keywords: Liénard system, a periodic solution.

1. Introduction

On the phase plane to periodic solutions of an autonomous system

$$\dot{x} = X(x, y), \quad \dot{y} = Y(x, y)$$

correspond closed trajectories. Such solutions usually describe continuous periodic processes. Periodic solutions are an important class of solutions to ordinary differential equations, since many of the processes described by ordinary differential equations are periodic. A large number of scientific papers are devoted to their study. At his time, Henri Poincaré attached great importance to periodic solutions represented by closed orbits. According to his plans, they were to become a support in the study of all other, non-periodic movements. In a certain sense, periodic solutions are the only type of solutions that can be completely observed in the process of their evolution, since the entire evolution of a periodic solution is determined by the knowledge of this solution over a finite period of time. Periodic solutions are the simplest type of oscillatory solutions.

In 1928, Liénard [7, 8] considered equations of the form

$$\frac{d^2x}{dt^2} + f(x)\frac{dx}{dt} + x = 0, \tag{1.1}$$

where $f(x)$ is a polynomial of even degree. These equations arose as a generalization of the famous van Der Pol equation [12]

$$\frac{d^2x}{dt^2} + \mu(x^2-1)\frac{dx}{dt} + x = 0, \tag{1.2}$$

which studied in detail the case $f(x) = x^2 - 1$. Moreover, the generalization was informal and mathematical, and naturally arose from the nonlinear damping of vibrations in electrical circuits considered by Liénard. Setting $dx/dt = z$, Liénard wrote equation (1.1) in the following form of the system of differential equations of first order

$$\frac{dx}{dt} = z, \quad \frac{dz}{dt} = -x - f(x)z. \tag{1.3}$$

But in his proof of the uniqueness of a periodic solution of equation (1.1), Liénard used other system of differential equations which is equivalent to system (1.3). For this, in system (1.3) he changed the variable $z = y - F(x)$, where

$$F(x) = \int_0^x f(\xi)d\xi, \tag{1.4}$$

and obtained the system

$$\frac{dx}{dt} = y - F(x), \quad \frac{dy}{dt} = -x. \tag{1.5}$$

Equation (1.1) is referred to as a Liénard equation, and both systems of equations (1.3) and (1.5) are called Liénard systems.

Consider the following differential equation

$$\frac{d^2x}{dt^2} + f(x)\frac{dx}{dt} + g(x) = 0, \tag{1.6}$$

which is a generalization of equation (1.1). These equations were obtained by Levinson and Smith [6] in 1942. Equation (1.6) as well as equation (1.1) most of authors call the Liénard equation [1]. The differential equation (1.6) have been studied in many papers [1, 9, 5, 2, 3, 11]. Equation (1.6) one can write in the form of the system of ordinary differential equations

$$\frac{dx}{dt} = y, \quad \frac{dy}{dt} = -f(x)y - g(x). \tag{1.7}$$

This system can model mechanical systems, where $f(x)$ is known as the damping term and $g(x))$ is called the restoring force or stiffness. System (1.7) is also used to model resistor inductor capacitor circuits with nonlinear circuit elements.

In papers [4, 10, 15, 13] the authors obtained conditions, under which system (1.7) or the equivalent system

$$\frac{dx}{dt} = z - F(x), \quad \frac{dz}{dt} = -g(x) \tag{1.8}$$

has a periodic solution.

The aim of this paper is to obtain other sufficient conditions of the existence of a periodic solution of system (1.7).

[1] Some authors call equation (1.6) the generalized Liénard equation.

2. On the existence of periodic solutions of system (1.7)

Let us find the conditions that ensure the existence of periodic solutions of system (1.7). Note that the periodic solution of system (1.7) exists if and only if there is a periodic solution of system (1.8). The following theorem gives sufficient conditions for the existence of periodic solutions of system (1.8).

Theorem 2.1. *Suppose that $F(x)$ is continuously differentiable, $g(x)$ is locally Lipschitz, and besides*

- *$xg(x) > 0$ for $x \neq 0$;*
- *the equation $F(x) = 0$ has three real roots: $x = b_1 > 0, x = b_2 < 0$, and $x = 0$; $F(x) > 0$ for $x \in (b_2, 0) \cup (b_1, +\infty)$; $F(x) < 0$ for $x \in (-\infty, b_2) \cup (0, b_1)$;*
- *$F(x)$ monotonically increases in the intervals $(-\infty, b_2)$ and $(b_1, +\infty)$; $F(x) \to +\infty$ as $x \to +\infty$, $F(x) \to -\infty$ as $x \to -\infty$.*

Then system (1.8) has a nontrivial (nonzero) periodic solution.

Proof. As has been shown in [3, 14], any solution of system (1.8) is a clockwise rotation around the origin, i.e. any solution that starts on the positive semiaxis of ordinate Oz, sequentially passes the first quadrant, then the fourth, third, second, first again, and so on. Consider the trajectory $x(t), z(t)$ of system (1.8) in the plane Oxz starting at the point H with the coordinates $(0, z_H)$ at the zero moment of time t (see Fig. 1).

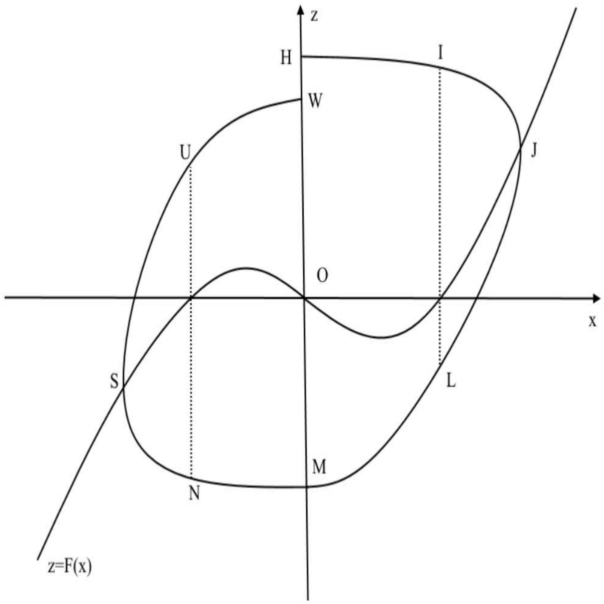


FIGURE 1

Denote by J and S the points of intersection of this trajectory with the curve $z = F(x)$, by I and L the points of intersection of the trajectory with the straight line

$x = b_1$, by U and N the points of intersection of the trajectory with the straight line $x = b_2$, and, finally, by W and M the points of intersection of the trajectory $x(t), z(t)$ with the axis Oz.

Obviously, the solution $x(t), z(t)$ is periodic if and only if the points H and W coincide, i.e. $z_H = z_W$.

Denote

$$G(x) := \int_0^x g(\xi) d\xi.$$

Consider the function

$$v(x, z) = \frac{z^2}{2} + G(x).$$

Its derivative along solutions of system (1.8) is equal

$$\frac{dv(x(t), z(t))}{dt} = -z(t)g(x(t)) + g(x(t))[z(t) - F(x(t))] = -g(x(t))F(x(t)). \qquad (2.1)$$

The change of the function v from point H to point W is equal to

$$\Delta v = v(0, z_W) - v(0, z_H) = \int_0^\tau \frac{dv(x(t), z(t))}{dt} dt = -\int_0^\tau g(x(t))F(x(t))dt \qquad (2.2)$$

where τ is moment of time when the trajectory $x(t), z(t)$ reaches the point W. Assume that z_H is sufficiently large, such that $x_J > b_1$, $x_S < b_2$. Let us show that Δv is a decreasing function of z_H. To do this, we break the trajectory between H and W into 6 pieces, where the first piece is a segment of the trajectory between points H and I, the second piece is a segment of the trajectory between points I and L, the third piece is the segment of the trajectory between the points L and M, the fourth piece is the segment of the trajectory between the points M and N, the fifth piece is a segment of the trajectory between the points N and U, the sixth piece is a segment of the trajectory between the points U and W. So Δv can be represented in the form $\Delta v = \sum_{i=1}^{6} \Delta v_i$ where Δv_i is the change of the function v on i-th piece of the trajectory. On the first, third, fourth and sixth pieces, z can be represented as a function of a variable x, because on these pieces $x(t)$ either monotonically increases or monotonically decreases; hence, the change of variable $dt = \frac{dx}{z - F(x)}$ is quite correct.

On the second and fifth pieces we use the substitution $dt = -\frac{dz}{g(x)}$. We want to argue that Δv is a monotonically decreasing function of z_H. So consider two trajectories starting at $t = 0$ from points $(0, z_H)$ and $(0, z_H + \Delta z_H)$, where $\Delta z_H > 0$. We denote the trajectories of system (1.8), starting at $t = 0$ from the points $(0, z_H)$ and $(0, z_H + \Delta z_H)$ by symbols $T1$ and $T2$ respectively. By virtue of the conditions of the theorem of existence and uniqueness of solutions of system (1.8), trajectories $T1$ and $T2$ have no common points, hence, the trajectory $T2$ is located outside of the trajectory $T1$, i.e. any ray emerging from the origin, first intersects the trajectory $T1$ and then the trajectory $T2$. Let us discover how changes the expression for Δv_i $(i = 1, \dots, 6)$ in the transition from the trajectory $T1$ to the trajectory $T2$.

$$\Delta v_1 = \int_0^{b_1} \frac{g(x)[-F(x)]}{z(x) - F(x)} dx = \int_0^{b_1} \frac{g(x)|F(x)|}{|z(x) - F(x)|} dx.$$

The value for $z(x)$ on the trajectory $T2$ is more then the value for $z(x)$ on $T1$, hence, $\Delta v_1(T2) < \Delta v_1(T1)$. Here and below $\Delta v_i(T2)$ and $\Delta v_i(T1)$ denote the values of Δv_i on trajectories $T2$ and $T1$ respectively.

$$\Delta v_2 = -\int_{z_I}^{z_L} g(x)F(x)\left[-\frac{dz}{g(x)}\right] = -\int_{z_L}^{z_I} F(x(z))dz.$$

Taking into account that on this piece $F(x)$ is positive and monotonically increasing and $x(z)|_{T2} > x(z)|_{T1}$, we obtain that $\Delta v_2(T2) < \Delta v_2(T1)$.

$$\Delta v_3 = \int_{b_1}^{0} \frac{g(x)[-F(x)]}{z(x)-F(x)}dx = \int_{0}^{b_1} \frac{g(x)|F(x)|}{|z(x)-F(x)|}dx$$

In this case we also have $\Delta v_3(T2) < \Delta v_3(T1)$.

$$\Delta v_4 = \int_{0}^{b_2} \frac{[-g(x)]F(x)}{z(x)-F(x)}dx = \int_{b_2}^{0} \frac{[-g(x)]F(x)}{F(x)-z(x)}dx,$$

whence $\Delta v_4(T2) < \Delta v_4(T1)$.

$$\Delta v_5 = -\int_{z_N}^{z_U} g(x)F(x)\left[-\frac{dz}{g(x)}\right] = \int_{z_N}^{z_U} F(x(z))dz.$$

On this piece $F(x)$ is negative. Since $x(z)|_{T2} < x(z)|_{T1}$, then

$$F(x(z))|_{x(z)\in T2} < F(x(z))|_{x(z)\in T1},$$

hence $\Delta v_5(T2) < \Delta v_5(T1)$.

$$\Delta v_6 = -\int_{b_2}^{0} \frac{g(x)F(x)}{z(x)-F(x)}dx = \int_{b_2}^{0} \frac{[-g(x)]F(x)}{z(x)-F(x)}dx.$$

Here $z(x)|_{T2} > z(x)|_{T1}$, therefore $\Delta v_6(T2) < \Delta v_6(T1)$. Thus it has been proved that Δv_i $(i = 1, \ldots, 6)$ decrease if z_H increase, hence Δv also decreases with increasing z_H. Let us show that

$$\lim_{z_H\to+\infty} \Delta v = -\infty.$$

To do this, it is enough to prove that

$$\lim_{z_H\to+\infty} \Delta v_2 = -\infty.$$

We will show that z_I increases indefinitely with unlimited increase of the value z_H. Getting rid of t in system (1.8) and passing to the argument x, we write the differential equation which describes the orbit HIJ:

$$\frac{dz}{dx} = -\frac{g(x)}{z-F(x)}. \tag{2.3}$$

According to the condition of the theorem $F(x) < 0$ for $x \in (0, b_1)$, hence

$$\frac{g(x)}{z-F(x)} < \frac{g(x)}{z} \text{ for } x \in (0, b_1). \tag{2.4}$$

From equation (2.3) and inequality (2.4) it follows

$$-\frac{dz}{dx} < \frac{g(x)}{z} \text{ for } x \in (0, b_1).$$

Separating variables and integrating, we obtain

$$\frac{1}{2}z^2(b_1) - \frac{1}{2}z_H^2 > -\int_0^{b_1} g(x)dx,$$

whence bearing in mind that $z(b_1) = z_I$, we get that $z_I \to +\infty$ if $z_H \to +\infty$.

Let $c \in (b_1, x_J)$. Let us designate the ordinates of the intersection points of the trajectory $T1$ and the line $x = c$ on pieces IJ and JL, respectively z^* and z^{**}. Taking into account that L is the intersection point of the trajectory $T1$ and the line $x = b_1$, we conclude that $z_L < 0$ (see Fig.1). Bearing in mind the continuity of the trajectory $T1$, the value $c \in (b_1, x_J)$ we choose so close to the value of b_1 that $z^{**} < 0$.

Let $z(x)$ be the solution of equation (2.3) such that $z(0) = z_H$. We shall show that $z(c) \to +\infty$ if $z_H \to +\infty$. The inequality $z - F(x) > z - F(c)$ holds on the interval (b_1, c) because the function $F(x)$ monotonically increases on this interval. Hence equation (2.3) yields

$$-\frac{dz}{dx} = \frac{g(x)}{z - F(x)} < \frac{g(x)}{z - F(c)}.$$

Separating variables and integrating, we obtain

$$-\left[\frac{1}{2}z^2 - F(c)z\right]_{z_I}^{z(c)} < \int_{b_1}^{c} g(x)dx,$$

whence (taking into account that $z(c) = z^* > 0$) it follows the inequality

$$z(c) > F(c) + \sqrt{[z_I - F(c)]^2 - 2\int_{b_1}^{c} g(x)dx}.$$

Since $z_I \to +\infty$ if $z_H \to +\infty$, then $z^* = z(c) \to +\infty$ if $z_H \to +\infty$.
Bearing in mind that $F(x)$ increases for $x > b_1$, we have

$$\begin{aligned}\Delta v_2 = -\int_{z_L}^{z_I} F(x(z))dz &< -F(c)(z^* - z^{**}) \\ &< -F(c)\left[F(c) + \sqrt{[z_I - F(c)]^2 - 2\int_{b_1}^{c} g(x)dx}\right].\end{aligned} \tag{2.5}$$

The obtained inequality implies that $\Delta v_2 \to -\infty$ if $z_H \to +\infty$.

If we choose z_H small enough, such that the entire trajectory between points H and W is located in the domain $x \in (b_2, b_1)$, then obviously that $\Delta v > 0$. Taking into account that Δv tends to $-\infty$ when $z_H \to +\infty$, one can conclude that there exists a value $z_H > 0$ such that $\Delta v = 0$. This means that there exists a periodic solution of system (1.8). The proof is comlpete.

Remark 2.2. If additionally to conditions of the theorem, one of the following conditions

- $G(a_1) = G(a_2)$ where a_1 and a_2 are positive and negative roots of equation $f(x) = 0$, and $G(\pm\infty) = +\infty$,
- $f(x)$ is even, $g(x)$ is odd, $G(+\infty) = +\infty$,

is satisfied, then equation (1.7) has a single periodic solution [14].

References

[1] Albarakati, W.A., Lloyd, N.G., Pearson, J.M., *Transformation to Liénard form*, Electronic Journal of Differential Equations, **76**(2000), 1-11.

[2] Bowman, T.T., *Periodic solutions of Liénard systems with symmetries*, Nonlinear Analysis, **2(4)**(1978), 457-464.

[3] Carletti, T., Villari, G., *A note on existence and uniqueness of limit cycles for Liénard systems*, Journal of Mathematical Analysis and Applications, **307**(2005), no. 2, 763-773.

[4] Filippov, A.F., *A sufficient condition for the existence of a stable limit cycle for the equation of the second order*, (in Russian), Sbornik: Mathematics, **30**(1952), no. 1, 171-180.

[5] Ignat'ev, A.O., Kirichenko, V.V., *On necessary conditions for global asymptotic stability of equilibrium for the Liénard equation*, Mathematical Notes, **93**(2013), no. 1-2, 75-82.

[6] Levinson, N., Smith, O., *A general equation for relaxation oscillations*, Duke Mathematical Journal, **9**(1942), no. 2, 382-403.

[7] Liénard, A., *Étude des oscillations entreténues*, Revue Générale de L'électricité, **23**(1928), 901-912.

[8] Liénard, A., *Étude des oscillations entreténues*, Revue Générale de L'électricité, **23**(1928), 946-954.

[9] Lynch, S., Christopher, C.J., *Limit cycles in highly non-linear differential equations*, Journal of Sound and Vibration, **224**(1999), no. 3, 505-517.

[10] Neumann, D.A., Sabbagh, L., *Periodic solutions of Liénard systems*, Journal of Mathematical Analysis and Applications, **62**(1978), no. 1, 148-156.

[11] Odani, K., *Existence of Exactly N Periodic Solutions for Liénard Systems*, Funkcialaj Ekvacioj, **39**(1996), 217-234.

[12] van der Pol, B., *On relaxation-oscillations*, The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science, **2**(1927), 978-992.

[13] Sabatini, M., *On the period function of Liénard systems*, Journal of Differential Equation, **152**(1999), 467-487.

[14] Sabatini, M., Villari, G., *On the uniqueness of limit cycles for Liénard equation: the legacy of G. Sansone*, Matematiche (Catania), **65**(2010), no. 2, 201-214.

[15] Sugie, J., Hara, T., *Non-existence of periodic solutions of the Liénard system*, Journal of Mathematical Analysis and Applications, **159**(1991), no. 1, 224-236.

Alexander Ignatyev
Institute of Applied Mathematics and Mechanics,
R. Luxemburg Street, 74,
Donetsk-83114, Ukraine
e-mail: `aoignat@mail.ru`

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 55–73
DOI: 10.24193/subbmath.2021.1.05

Perturbed eigenvalue problems: an overview

Maria Fărcăşeanu, Andrei Grecu,
Mihai Mihăilescu and Denisa Stancu-Dumitru

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. The study of perturbed eigenvalue problems has been a very active field of investigation throughout the years. In this survey we collect several results in the field.

Mathematics Subject Classification (2010): 35D30, 35D40, 46E30, 49J40, 35A15.

Keywords: Eigenvalue problem, p-Laplace operator, nonlocal (s,p)-Laplace operator, Sobolev space, variational methods.

1. Introduction

This paper is dedicated to prof. Gheorghe Moroşanu on the occasion of his 70th birthday. The topic of our paper fits perfectly with one of prof. Moroşanu 's fields of interests, namely *the study of eigenvalue problems for elliptic operators*, on which he brought a couple of nice contributions which will be recalled in the main body of this article. It is an opportunity and an honour for us to dedicate this work to our professor and friend Gheorghe Moroşanu on the occasion of his 70th birthday.

The goal of this paper is to collect some known results on perturbed eigenvalue problems. We split the discussion in two main parts. More precisely, we will start our survey by presenting results on the classical eigenvalue problem for the p-Laplace operator in both local and nonlocal cases (including a discussion on the limiting case when $p \to \infty$), and we will continue with the case of the perturbed eigenvalue problems of the p-Laplace operator on bounded domains under different boundary conditions or on unbounded domains.

1.1. Notations

Throughout this paper Ω will stand for an open set (bounded or unbounded) of the Euclidean space $\mathbb{R}^N$. We will denote by $\partial\Omega$ the boundary of Ω while ν will stand for the unit outward normal to $\partial\Omega$ and $\frac{\partial u}{\partial \nu}$ will represent the normal derivative of u. The Euclidean norm on $\mathbb{R}^N$ will be denoted by $|\cdot|_N$.

2. Eigenvalue problems for the p-Laplace operator

2.1. The case of the (local) p-Laplace operator

For each real number $p \in (1, \infty)$ and each function $u : \Omega \to \mathbb{R}$, smooth enough, we define *the (local) p-Laplace operator* by

$$\Delta_p u := \operatorname{div}(|\nabla u|_N^{p-2} \nabla u)\,.$$

2.1.1. The case of bounded domains. In this section we will assume that $\Omega \subset \mathbb{R}^N$ $(N \geq 1)$ is a bounded domain with smooth boundary $\partial\Omega$. The *classical eigenvalue problem for the p-Laplace operator* reads as follows

$$-\Delta_p u = \lambda |u|^{p-2} u, \quad \text{in } \Omega\,, \tag{2.1}$$

where $\lambda \in \mathbb{R}$ is a real parameter. This problem was studied under different boundary conditions (see, e.g. Lê [16] for more details), such as

- Dirichlet boundary conditions

$$u = 0, \quad \text{on } \partial\Omega\,, \tag{2.2}$$

- Neumann boundary conditions

$$|\nabla u|_N^{p-2} \frac{\partial u}{\partial \nu} = 0, \quad \text{on } \partial\Omega\,, \tag{2.3}$$

- Robin boundary conditions

$$|\nabla u|_N^{p-2} \frac{\partial u}{\partial \nu} + \alpha |u|^{p-2} u = 0, \quad \text{on } \partial\Omega\,, \tag{2.4}$$

where $\alpha > 0$ is a given real number, etc. In this context, a parameter λ is called an *eigenvalue* of problem (2.1) if the problem possesses a nontrivial (weak) solution u which belongs to a suitable Sobolev space denoted by $W(\Omega)$, where either $W(\Omega) = W_0^{1,p}(\Omega)$, if we are working under the Dirichlet boundary conditions, or $W(\Omega) = W^{1,p}(\Omega)$, if we are working under the Neumann or Robin boundary conditions. More precisely, if we are working under boundary conditions (2.2) or (2.3) then λ is an eigenvalue of problem (2.1) if there exists $u \in W(\Omega) \setminus \{0\}$ such that

$$\int_\Omega |\nabla u|_N^{p-2} \nabla u \nabla \phi \, dx = \lambda \int_\Omega |u|^{p-2} u \phi \, dx, \quad \forall\, \phi \in W(\Omega)\,,$$

while, if we are working under boundary conditions (2.4) then λ is an eigenvalue of problem (2.1) if there exists $u \in W(\Omega) \setminus \{0\}$ such that

$$\int_\Omega |\nabla u|_N^{p-2} \nabla u \nabla \phi \, dx + \alpha \int_{\partial\Omega} |u|^{p-2} u \phi \, d\sigma(x) = \lambda \int_\Omega |u|^{p-2} u \phi \, dx,\ \forall\, \phi \in W(\Omega)\,.$$

A function u as above is called an *eigenfunction* corresponding to the eigenvalue λ.

It is well-known (see, e.g. Lindqvist [18] or Lê [16]) that problem (2.1) (under any of the boundary conditions (2.2), (2.3), or (2.4)) has an increasing and unbounded sequence of nonnegative eigenvalues, say $\{\lambda_k(p; \Omega)\}_{k \geq 1}$, which can be produced using, for instance, the Ljusternik-Schnirelman theory. We recall that for each integer $k \geq 1$

the eigenvalue $\lambda_k(p;\Omega)$, under the boundary conditions (2.2) or (2.3), has the following variational characterisation, (see, e.g. [16]),

$$\lambda_k(p;\Omega) := \inf_{A\in\Sigma_k} \sup_{u\in A} \frac{\displaystyle\int_\Omega |\nabla u|_N^p \, dx}{\displaystyle\int_\Omega |u|^p \, dx}, \tag{2.5}$$

while, under the boundary condition (2.4) its variational characterisation reads as

$$\lambda_k(p;\Omega) := \inf_{A\in\Sigma_k} \sup_{u\in A} \frac{\displaystyle\int_\Omega |\nabla u|_N^p \, dx + \alpha \int_{\partial\Omega} |u|^p \, d\sigma(x)}{\displaystyle\int_\Omega |u|^p \, dx}, \tag{2.6}$$

where

$$\begin{aligned}\Sigma_k \quad := \quad & \{A \subset W(\Omega) \mid A \text{ is symmetric and compact in the} \\ & \text{topology of } W(\Omega), \gamma(A) \geq k\},\end{aligned}$$

and $\gamma(A)$ stands for the Krasnosel'skii genus of A, which is defined as the smallest integer m for which there exists a continuous odd map $f : A \to \mathbb{R}^m \setminus \{0\}$. If no such integer exists, then we set $\gamma(A) = \infty$, while $\gamma(\emptyset) = 0$. Note that in the particular cases when $p = 2$ (and $N \geq 1$), that is the case when the eigenvalue problem (2.1) is linear, or $N = 1$ (and $p \in (1,\infty)$), that is the 1-dimensional case, the sequence $\{\lambda_k(p;\Omega)\}_{k\geq 1}$ describes completely the set of eigenvalues of problem (2.1). However, when $N \geq 2$ and $p \in (1,\infty) \setminus \{2\}$ the existence of other eigenvalues in the interval $(\lambda_2(p;\Omega),\infty)$ different from those given by the sequence $\{\lambda_k(p;\Omega)\}_{k\geq 3}$ remains an open question. Actually, in the latter case it is not known if the set of all eigenvalues of the problem is discrete or not.

In order to simplify the exposition, in the rest of this paper we will use three different notations for the sequences of eigenvalues of problem (2.1) depending on the boundary conditions that will be considered. More precisely, we let $\{\lambda_k^D(p;\Omega)\}_{k\geq 1}$, $\{\lambda_k^N(p;\Omega)\}_{k\geq 1}$ and $\{\lambda_k^R(p;\Omega)\}_{k\geq 1}$ be the sequences of eigenvalues of problem (2.1) under the boundary conditions (2.2), (2.3) and (2.4), respectively.

At this point it is instructive to point out the following simple observations concerning the variational characterisations of the lowest eigenvalues of problem (2.1) under the three different boundary conditions presented above

$$\lambda_1^D(p;\Omega) := \inf_{u\in W_0^{1,p}(\Omega)\setminus\{0\}} \frac{\displaystyle\int_\Omega |\nabla u|_N^p \, dx}{\displaystyle\int_\Omega |u|^p \, dx}, \tag{2.7}$$

$$\lambda_1^N(p;\Omega) := \inf_{u\in W^{1,p}(\Omega)\setminus\{0\}} \frac{\displaystyle\int_\Omega |\nabla u|_N^p \, dx}{\displaystyle\int_\Omega |u|^p \, dx},$$

$$\lambda_1^R(p;\Omega) := \inf_{u\in W^{1,p}(\Omega)\setminus\{0\}} \frac{\displaystyle\int_\Omega |\nabla u|_N^p\,dx + \alpha\int_{\partial\Omega}|u|^p\,d\sigma(x)}{\displaystyle\int_\Omega |u|^p\,dx}\,.$$

All these minimization problems possess minimizers which are corresponding eigenfunctions for the eigenvalues $\lambda_1^D(p;\Omega)$, $\lambda_1^N(p;\Omega)$ and $\lambda_1^R(p;\Omega)$. These minimizers belong to a certain Hölder space $C^{1,\beta}(\Omega)$ (for some $\beta\in(0,1)$) and do not change sign in Ω. On the other hand, the eigenvalues $\lambda_1^D(p;\Omega)$, $\lambda_1^N(p;\Omega)$ and $\lambda_1^R(p;\Omega)$ are simple and isolated. Moreover, we recall that

$$\lambda_1^D(p;\Omega)>0 \quad\text{and}\quad \lambda_1^R(p;\Omega)>0,\quad \forall\, p\in(1,\infty)\,,$$

while

$$\lambda_1^N(p;\Omega)=0,\quad \forall\, p\in(1,\infty)\,.$$

Since the lowest eigenvalue of problem (2.1)+(2.3) vanishes it is important to present the variational characterisation of the second eigenvalue $\lambda_2^N(p;\Omega)$ (that is the first positive eigenvalue of the problem), namely

$$\lambda_2^N(p;\Omega) := \inf_{u\in X_p(\Omega)\setminus\{0\}} \frac{\displaystyle\int_\Omega |\nabla u|_N^p\,dx}{\displaystyle\int_\Omega |u|^p\,dx}\,,$$

where $X_p(\Omega) := \{u\in W^{1,p}(\Omega):\ \int_\Omega |u|^{p-2}u\,dx = 0\}$.

2.1.2. The ∞-eigenvalue problem under the Dirichlet boundary conditions. For each $p\in(1,\infty)$ we can rewrite problem (2.1)+(2.2) as

$$\begin{cases} -\Delta_p u = \lambda|u|^{p-2}u & \text{in }\ \Omega,\\ u=0 & \text{on }\ \partial\Omega\,. \end{cases} \tag{2.8}$$

The asymptotic behavior as $p\to\infty$ of problems (2.8) with $\lambda=\lambda_1^D(p;\Omega)$ has been studied by Fukagai, Ito, & Narukawa [11] and Juutinen, Lindqvist, & Manfredi [15]). A first step in this direction was to show that

$$\lim_{p\to\infty}\sqrt[p]{\lambda_1^D(p;\Omega)} = \left[\max_{x\in\Omega} dist(x,\partial\Omega)\right]^{-1}\,,$$

where $dist(\cdot,\partial\Omega)$ stands for the distance function to the boundary of Ω, (recall that $dist(x,\partial\Omega) := \inf_{y\in\partial\Omega}|x-y|_N$, for all $x\in\Omega$). Next, since the corresponding eigenfunctions of $\lambda_1^D(p;\Omega)$ are, actually, minimizers for the minimization problem (2.7) that do not change sign in Ω, we can let, for each $p\in(1,\infty)$, $u_p>0$ to be an eigenfunction corresponding to the eigenvalue $\lambda_1^D(p;\Omega)$. Juutinen, Lindqvist & Manfredi showed in [15] that there exists a subsequence of $\{u_p\}$ which converges uniformly in Ω to a nontrivial and nonnegative viscosity solution of the limiting problem

$$\begin{cases} \min\left\{|\nabla u|_N - \left[\max\limits_{x\in\Omega} dist(x,\partial\Omega)\right]^{-1} u,\ -\Delta_\infty u\right\} = 0 & \text{in }\Omega,\\ u=0 & \text{on }\partial\Omega, \end{cases} \tag{2.9}$$

where Δ_∞ is the ∞-Laplace operator, which on sufficiently smooth functions $u : \Omega \to \mathbb{R}$ is given by $\Delta_\infty u := \langle D^2 u \nabla u, \nabla u\rangle = \sum_{i,j=1}^N \frac{\partial u}{\partial x_i}\frac{\partial u}{\partial x_j}\frac{\partial^2 u}{\partial x_i \partial x_j}$. Note that $dist(\cdot, \partial\Omega)$ is not always a viscosity solution of (2.9), but, in the particular case when Ω is a ball it turns out that $dist(\cdot, \partial\Omega)$ is the only viscosity solution of (2.9). However, for general domains Ω the convergence of the entire sequence u_p to a unique limit, as $p \to \infty$, is an open question.

2.1.3. The case of unbounded domains. In the first part of this section we will let $\Omega \subseteq \mathbb{R}^N$ ($N \geq 3$) be a general open set (bounded or unbounded) and $V : \Omega \to \mathbb{R}$ be a function which satisfies the hypotheses

$$\begin{cases} V \in L^1_{loc}(\Omega),\ V^+ = V_1 + V_2 \neq 0,\ V_1 \in L^{N/2}(\Omega), \\ \lim\limits_{|x|_N \to \infty} |x|_N^2 V_2(x) = 0,\ \lim\limits_{x \to y} |x-y|_N^2 V_2(x) = 0 \text{ for any } y \in \overline{\Omega}. \end{cases} \tag{2.10}$$

Note that in particular the function V may change sign in Ω.

In [25] Szulkin & Willem analyzed the eigenvalue problem

$$-\Delta u = \lambda V(x) u, \quad u \in \mathcal{D}_0^{1,2}(\Omega)\,, \tag{2.11}$$

where $\mathcal{D}_0^{1,2}(\Omega)$ stands for the closure of $C_0^\infty(\Omega)$ under the L^2-norm of the gradient. Using an elementary argument based on a simple minimization procedure it was proved in [25, Theorems 2.2 & 2.3] the existence of infinitely many eigenvalues of (2.11). A similar result was obtained in the case when instead of the Laplace operator was considered the general p-Laplace operator in equation (2.11) (naturally, in this new case conditions (2.10) were slightly modified in order to be compatible with the new situation).

In the second part of this section we let $\Omega \subset \mathbb{R}^N$ ($N \geq 2$) be a simply connected bounded domain, containing the origin, with C^2 boundary denoted by $\partial\Omega$ and we denote by $\Omega^{\text{ext}} := \mathbb{R}^N \setminus \overline{\Omega}$ the exterior of Ω. Let $K : \Omega^{\text{ext}} \to (0, \infty)$ be a function having the property that $K \in L^\infty(\Omega^{\text{ext}}) \cap L^{N/p}(\Omega^{\text{ext}})$, for some $p \in (1, N)$. Chhetri and Drábek studied in [6] the eigenvalue problem

$$\begin{cases} -\Delta_p u = \lambda K(x) |u|^{p-2} u, & \text{for} \quad x \in \Omega^{\text{ext}}, \\ u(x) = 0, & \text{for} \quad x \in \partial\Omega, \\ u(x) \to 0, & \text{as} \quad |x|_N \to \infty\,. \end{cases} \tag{2.12}$$

In particular, they showed that the lowest eigenvalue of problem (2.12) has the following variational characterization

$$\lambda_1(p; \Omega^{\text{ext}}) := \inf_{u \in C_0^\infty(\Omega^{\text{ext}}) \setminus \{0\}} \frac{\displaystyle\int_{\Omega^{\text{ext}}} |\nabla u|_N^p \mathrm{d}x}{\displaystyle\int_{\Omega^{\text{ext}}} K(x) |u|^p \mathrm{d}x}\,. \tag{2.13}$$

Moreover, $\lambda_1(p; \Omega^{\text{ext}})$ is simple, isolated and its corresponding eigenfunctions have constant sign in Ω^{ext}. In particular, the results from [6] complemented to the case of exterior domains the results obtained on the classical eigenvalue problem of the p-Laplacian on bounded domains subject to the homogeneous Dirichlet boundary conditions (that is problem (2.1)+(2.2), or, equivalently, problem (2.8)).

2.2. The case of the nonlocal p-Laplace operator

Let $\Omega \subset \mathbb{R}^N$ $(N \geq 2)$ be a bounded domain with Lipschitz boundary $\partial\Omega$. For each $p \in (1,\infty)$ and $s \in (0,1)$ we define the nonlocal nonlinear operator

$$(-\Delta_p)^s u(x) := 2 \lim_{\epsilon \searrow 0} \int_{|x-y|_N \geq \epsilon} \frac{|u(x)-u(y)|^{p-2}(u(x)-u(y))}{|x-y|_N^{N+sp}}\, dy, \ \ x \in \mathbb{R}^N . \tag{2.14}$$

Since for $p = 2$ the above definition reduces to the linear fractional Laplacian, $(-\Delta)^s$, we will refer to $(-\Delta_p)^s$ as being a *fractional* (s,p)*-Laplace operator.*

The eigenvalue problem for the fractional (s,p)-Laplacian reads as follows

$$\begin{cases} (-\Delta_p)^s u(x) = \lambda |u(x)|^{p-2}u(x), & \text{for } x \in \Omega, \\ u(x) = 0, & \text{for } x \in \mathbb{R}^N \backslash \Omega . \end{cases} \tag{2.15}$$

Problem (2.15) was extensively studied in the literature in the last decade. Among the results related with this problem we just recall some facts from the paper by Lindgren & Lindqvist [17]. First, in order to explain the notion of eigenvalue for problem (2.15) let us denote by $\widetilde{W}_0^{s,p}(\Omega)$ the fractional Sobolev space where it is natural to seek weak solutions for this problem. Next, for simplicity, for each $p \in (1,\infty)$ and $s \in (0,1)$ we will consider the notation

$$\mathcal{E}_{s,p}(u,v) := \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \frac{|u(x)-u(y)|^{p-2}(u(x)-u(y))(v(x)-v(y))}{|x-y|_N^{N+sp}}\, dx\, dy\, , \tag{2.16}$$

for all $u,\ v \in \widetilde{W}_0^{s,p}(\Omega)$. A real number $\lambda \in \mathbb{R}$ will be called an *eigenvalue* of problem (2.15) if there exists a function $u \in \widetilde{W}_0^{s,p}(\Omega)$ such that

$$\mathcal{E}_{s,p}(u,v) = \lambda \int_\Omega |u(x)|^{p-2}u(x)v(x)\, dx, \quad \forall\, v \in \widetilde{W}_0^{s,p}(\Omega) . \tag{2.17}$$

Further, we define

$$\lambda_1(s,p) := \inf_{u \in \widetilde{W}_0^{s,p}(\Omega)\backslash\{0\}} \frac{\displaystyle\int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \frac{|u(x)-u(y)|^p}{|x-y|_N^{N+sp}}\, dx\, dy}{\displaystyle\int_{\mathbb{R}^N} |u|^p\, dx} . \tag{2.18}$$

It is known that $\lambda_1(s,p)$ is attained at some $u \in \widetilde{W}_0^{s,p}(\Omega) \setminus \{0\}$ (see [17, Theorem 5]), with $\|u\|_{L^p(\mathbb{R}^N)} = \|u\|_{L^p(\Omega)} = 1$ and

$$\frac{\displaystyle\int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \frac{|u(x)-u(y)|^p}{|x-y|_N^{N+sp}}\, dx\, dy}{\displaystyle\int_{\mathbb{R}^N} |u|^p\, dx} = \lambda_1(s,p) .$$

Moreover, it holds true that

$$\mathcal{E}_{s,p}(u,\varphi) = \lambda_1(s,p) \int_{\mathbb{R}^N} |u(x)|^{p-2}u(x)\varphi(x)dx, \quad \forall\, \varphi \in \widetilde{W}_0^{s,p}(\Omega)\, ,$$

which means that $\lambda_1(s,p)$ is an eigenvalue of problem (2.15).

Next, let us recall a result on an eigenvalue problem involving the fractional Laplacian studied on the whole Euclidean space $\mathbb{R}^N$. More precisely, the second author of this survey studied in [12] the eigenvalue problem

$$(-\Delta)^s u(x) = \lambda V(x) u(x), \quad \forall\, x \in \mathbb{R}^N, \tag{2.19}$$

where $s \in (0,1)$ is a given real number, λ is a real parameter and $V : \mathbb{R}^N \to \mathbb{R}$ is a function that may change sign and which satisfies the hypothesis

$$\begin{cases} V \in L^1_{\text{loc}}(\mathbb{R}^N),\ V^+ = V_1 + V_2 \neq 0,\ V_1 \in L^{\frac{N}{2s}}(\mathbb{R}^N) \text{ and} \\[2ex] \lim\limits_{x \to y} |x-y|_N^{2s} V_2(x) = 0, \text{ for all } y \in \mathbb{R}^N \text{ and } \lim\limits_{|x|_N \to \infty} |x|_N^{2s} V_2(x) = 0. \end{cases} \tag{$\widetilde{V}$}$$

It was shown in [12, Theorem 1.3] that under condition ($\widetilde{\mathrm{V}}$) the problem (2.19) has an unbounded, increasing sequence of positive eigenvalues. In particular this result extended to the case of nonlocal operators the result by Szulkin & Willem from [25, Theorems 2.2 & 2.3].

3. Perturbed eigenvalue problems for the p-Laplace operator

In this section we will analyze some perturbations of classical eigenvalue problems. All the perturbed eigenvalue problems are, actually, *nontypical eigenvalue problems* since the differential operators involved in their constructions are inhomogeneous. However, their formulations are similar with those of the typical eigenvalue problems and for that reason we will continue to call the parameter λ involved in these equations an *eigenvalue* if the corresponding problem possesses a nontrivial weak solution.

3.1. The perturbation of the (local) p-Laplace operator

3.1.1. The case of bounded domains. In this section we will assume that $\Omega \subset \mathbb{R}^N$ ($N \geq 1$) is a bounded domain with smooth boundary $\partial\Omega$. Let $p \in (1, \infty)$ be a given real number. We will call a *perturbation of the eigenvalue problem (*2.1*)* a problem of type

$$-\Delta_p u - \Delta_q u = \lambda |u|^{p-2} u, \quad \text{in } \Omega\,, \tag{3.1}$$

where $q \in (1,\infty) \setminus \{p\}$ is a given real number and $\lambda \in \mathbb{R}$ is a real parameter. Our goal will be to determine the set of all parameters λ for which problem (3.1) has nontrivial solutions, under different boundary conditions. This kind of parameters will be called eigenvalues of problem (3.1).

I. *The case of the Dirichlet boundary conditions.* We consider the case when problem (3.1) is investigated subject to the boundary conditions (2.2). More precisely, we consider the problem

$$\begin{cases} -\Delta_p u - \Delta_q u = \lambda |u|^{p-2} u & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega\,. \end{cases} \tag{3.2}$$

For this problem a *weak solution* is a function $u \in W_0^{1,\max\{p,q\}}(\Omega)$ such that

$$\int_\Omega (|\nabla u|_N^{p-2} + |\nabla u|_N^{q-2}) \nabla u \nabla \phi \, dx = \lambda \int_\Omega |u|^{p-2} u \phi \, dx, \quad \forall\, \phi \in W_0^{1,\max\{p,q\}}(\Omega)\,.$$

We will say that λ from the above relation is an *eigenvalue* of problem (3.2) if $u \in W_0^{1,\max\{p,q\}}(\Omega) \setminus \{0\}$. In that case we will refer to u as being an *eigenfunction* corresponding to the eigenvalue λ.

Independently, Tanaka [23] and Bocea and the third author of this paper [5, Theorem 1.1] proved the following result.

Theorem 3.1. *The set of eigenvalues of problem (3.2) is exactly given by the open interval $(\lambda_1^D(p;\Omega), \infty)$. Moreover, for each $\lambda \in (\lambda_1^D(p;\Omega), \infty)$ there exists a nontrivial and nonnegative weak solution for problem (3.2).*

Note the interesting fact that in the case of the perturbed eigenvalue problems under the Dirichlet boundary conditions, such as (3.2), the set of eigenvalues can be entirely described and it is a continuous set. In particular, this is in sharp contrast with the situation which occurs in the case of the Laplace operator when the set of eigenvalues is discrete.

We would like to point out that similar results with those obtained in Theorem 3.1 were obtained by Bhattacharya, Emamizadeh, & Farjudian in [3] and by the first author of this paper in [8, Theorem 1] but for a class of anisotropic differential operators.

Further, let us assume that for each real number $p \in (1,\infty)$ the parameter $q \in (1,\infty) \setminus \{p\}$ which is involved in the construction of problem (3.2) depends on p. In other words we assume that $q : (1,\infty) \to (1,\infty)$ is a function which depends on p, i.e. $q = q(p)$. Furthermore, we assume that $\lim\limits_{p\to\infty} \frac{q(p)}{p} = Q \in (0,\infty) \setminus \{1\}$, and where either $q(p) < p$ if $Q \in (0,1)$ or $q(p) > p$ if $Q \in (1,\infty)$. In [5] the authors investigated the asymptotic behavior of positive solutions of the problems (3.2) as $p \to \infty$. They showed that for any $\Lambda \in [(\max\limits_{x\in\Omega} dist(x,\partial\Omega))^{-1}, \infty)$ and each sequence $\{\lambda_p\}$, with $\lambda_p \in (\lambda_1^D(p;\Omega), \infty)$, such that $\lim_{p\to\infty}(\lambda_p)^{1/p} = \Lambda$ the sequence of positive weak solutions of (3.2) with $\lambda = \lambda_p$ possesses a subsequence which converges to a nontrivial and nonnegative viscosity solution of the limiting problem

$$\begin{cases} \min\Big\{\max\{|\nabla u|_N, |\nabla u|_N^Q\} - \Lambda u, -\Delta_\infty u\Big\} = 0 & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{3.3}$$

On the other hand, it was shown that for all $\Lambda \in (-\infty, (\max\limits_{x\in\Omega} dist(x,\partial\Omega))^{-1})$ there are no nonnegative and nontrivial solutions of problem (3.3). Thus, in comparison to the well-known problem (2.9), the analysis of (3.3) reveals a markedly different situation: while for the original problem a single value of Λ, namely $\left[\max\limits_{x\in\Omega} dist(x,\partial\Omega)\right]^{-1}$, is known for which the corresponding viscosity solution is nonnegative, in the case of problem (3.3) this situation extends to the entire interval $[(\max\limits_{x\in\Omega} dist(x,\partial\Omega))^{-1}, \infty)$ (see [5, Theorem 1.3] for details).

II. *The case of the Neumann boundary conditions.* We consider the case when problem (3.1) is investigated subject to the Neumann-type boundary conditions. More precisely, we consider the problem

$$\begin{cases} -\Delta_p u - \Delta_q u = \lambda |u|^{p-2} u & \text{in } \Omega, \\ (|\nabla u|_N^{p-2} + |\nabla u|_N^{q-2}) \dfrac{\partial u}{\partial \nu} = 0 & \text{on } \partial\Omega \,. \end{cases} \tag{3.4}$$

where p, $q \in (1, \infty)$ and $p \neq q$. For this problem a *weak solution* is a function $u \in W^{1,\max\{p,q\}}(\Omega)$ such that

$$\int_\Omega (|\nabla u|_N^{p-2} + |\nabla u|_N^{q-2}) \nabla u \nabla \phi \, dx = \lambda \int_\Omega |u|^{p-2} u \phi \, dx, \quad \forall\, \phi \in W^{1,\max\{p,q\}}(\Omega) \,.$$

We will say that λ is an *eigenvalue* of problem (3.4) if $u \in W^{1,\max\{p,q\}}(\Omega) \setminus \{0\}$. In that case we will refer to u as being an *eigenfunction* corresponding to the eigenvalue λ.

Problem (3.4) was investigated in the case when $p = 2$ and $q \in (1, \infty) \setminus \{2\}$ by three of the authors of this paper in [19, Theorem 1.1] (for the case $q \in (2, \infty)$) and [9, Theorem 1] (for the case $q \in (1, 2)$) while the case $p \in (2, \infty)$ and $q \in (1, \infty) \setminus \{p\}$ it was analyzed by Moroşanu and the third author of this paper in [21, Theorem 1.1]. We summarise all the results on problem (3.4) in the following theorem.

Theorem 3.2. *Assume that $p \in [2, \infty)$ and $q \in (1, \infty) \setminus \{p\}$. For each such two numbers p and q define*

$$X_{p,q}(\Omega) := \left\{ u \in W^{1,\max\{p,q\}}(\Omega) : \int_\Omega |u|^{p-2} u \, dx = 0 \right\} .$$

Then the set of eigenvalues of problem (3.4) is precisely

$$\{0\} \cup (\mu_1(p, q; \Omega), \infty) \,,$$

where

$$\mu_1(p, q; \Omega) := \inf_{u \in X_{p,q}(\Omega) \setminus \{0\}} \frac{\displaystyle\int_\Omega |\nabla u|_N^p \, dx}{\displaystyle\int_\Omega |u|^p \, dx} \,,$$

is a positive constant.

Note that in the case when $q \in (1, p)$ and $p \geq 2$ we have $\mu_1(p, q; \Omega) = \lambda_2^N(p; \Omega)$ and thus the constant $\mu_1(p, q; \Omega)$ does not depend on q in this case. On the other hand, in the case where $q \in (p, \infty)$ the constant $\mu_1(p, q; \Omega)$ depends on q since in this case $W^{1,\max\{p,q\}}(\Omega) = W^{1,q}(\Omega)$. In that case we can deduce only the fact that $\mu_1(p, q; \Omega) \geq \lambda_2^N(p; \Omega)$.

The conclusion of Theorem 3.2 is interesting if we compare it, for example, with two classical well-known results on similar problems. First, recall the fact that when $q = p = 2$ then problem (3.4) reduces to the eigenvalue problem for the Laplace operator under the homogenous Neumann boundary conditions. In that case we recall the well-known fact that the problem possesses a discrete set of eigenvalues which can be organized as an increasing and unbounded sequence of positive real numbers. On the other hand, if we consider for instance the problem

$$\begin{cases} -\Delta u = \lambda |u|^{q-2}u & \text{in } \Omega, \\ \dfrac{\partial u}{\partial \nu} = 0 & \text{on } \partial\Omega, \end{cases} \tag{3.5}$$

with $q \in (1,\infty) \setminus \{2\}$ then the set of all parameters λ for which problem (3.5) has nontrivial weak solutions is exactly the interval $[0,\infty)$. In that case the set of eigenvalues of problem (3.5) is continuous. The case of problem (3.4) with $p = 2$ and $q \in (1,\infty) \setminus \{2\}$ brings to our attention a new situation when the set of eigenvalues of the problem possesses on the one hand, a continuous part, that is the interval $(\mu_1(2,q;\Omega),\infty)$, and, on the other hand, one more eigenvalue, i.e. $\lambda = 0$, which is isolated.

Finally, we would like to point out three similar results with those obtained in Theorem 3.2. The first result was recently obtained by Abreu & Madeira in [1] in the case when in problem (3.4) we have $p = 2$, $q \in (1,\infty) \setminus \{2\}$ but working under parametric-type boundary conditions instead of the Neumann-type boundary conditions. The other two results are due to Costea & Moroşanu [7] and Barbu & Moroşanu [2] for some Steklov-type eigenvalue problems.

III. *The case of the Robin boundary conditions.* Assume that we are working in an Euclidean space having dimension $N \geq 2$. We consider the case when problem (3.1) is investigated subject to the Robin-type boundary conditions. More precisely, for a given real number $\alpha > 0$ we consider the problem

$$\begin{cases} -\Delta_p u - \Delta_q u = \lambda |u|^{p-2}u & \text{in } \Omega, \\ (|\nabla u|_N^{p-2} + |\nabla u|_N^{q-2})\dfrac{\partial u}{\partial \nu} + \alpha |u|^{p-2}u = 0 & \text{on } \partial\Omega, \end{cases} \tag{3.6}$$

where p, $q \in (1,\infty)$ and $p \neq q$.
For this problem a *weak solution* is a function $u \in W^{1,\max\{p,q\}}(\Omega)$ such that

$$\int_\Omega (|\nabla u|_N^{p-2} + |\nabla u|_N^{q-2})\nabla u \nabla \phi \, dx + \alpha \int_{\partial\Omega} |u|^{p-2}u\phi \, d\sigma(x) = \lambda \int_\Omega |u|^{p-2}u\phi \, dx \,,$$

for all $\phi \in W^{1,\max\{p,q\}}(\Omega)$. We will say that λ is an *eigenvalue* of problem (3.6) if $u \in W^{1,\max\{p,q\}}(\Omega) \setminus \{0\}$. In that case we will refer to u as being an *eigenfunction* corresponding to the eigenvalue λ.

The perturbed eigenvalue problem (3.6) has been investigated by Gyulov & Moroşanu in [14]. In order to recall their result let us define two quantities which play an important role in the analysis of the problem. More precisely, we define

$$\lambda^\star := \alpha \frac{m_{N-1}(\partial\Omega)}{m_N(\Omega)} \,,$$

where $m_{N-1}(\partial\Omega)$ and $m_N(\Omega)$ denote the corresponding $N-1$ and N dimensional Lebesgue measures of the boundary $\partial\Omega$ and the set Ω, respectively, and

$$\nu_1(p,q;\Omega) := \inf_{u\in W^{1,\max\{p,q\}}(\Omega)\setminus\{0\}} \frac{\displaystyle\int_\Omega |\nabla u|_N^p\,dx + \alpha\int_{\partial\Omega}|u|^p\,d\sigma(x)}{\displaystyle\int_\Omega |u|^p\,dx}.$$

By [14, Remark 2] it is clear that $\lambda^\star > \nu_1(p,q;\Omega)$. Moreover, we point out that if $q\in(1,p)$ then $W^{1,\max\{p,q\}}(\Omega) = W^{1,p}(\Omega)$ and, consequently, in that case $\nu_1(p,q;\Omega) = \lambda_1^R(p;\Omega)$. By contrary, if $q\in(p,\infty)$ then $W^{1,\max\{p,q\}}(\Omega) = W^{1,q}(\Omega)$ and, consequently, in that case $\nu_1(p,q;\Omega)\geq\lambda_1^R(p;\Omega)$. The main result on (3.6) is a consequence of Theorems 1-3 from [14].

Theorem 3.3. *For each p, $q\in(1,\infty)$ in the interval $(-\infty,\lambda_1^R(p;\Omega)]$ there is no eigenvalue of problem* (3.6)*. If $q\in(p,\infty)$ then each $\lambda\in(\nu_1(p,q;\Omega),\lambda^\star)$ is an eigenvalue of problem* (3.6)*. If $q\in(1,p)$ then each $\lambda\in(\lambda_1^R(p;\Omega),\lambda^\star)$ is an eigenvalue of problem* (3.6)*.*

The case $\lambda\geq\lambda^\star$ is open.

3.1.2. The case of unbounded domains. In the first part of this section we will let $\Omega\subseteq\mathbb{R}^N$ $(N\geq3)$ be a general open set (bounded or unbounded) and $V:\Omega\to\mathbb{R}$ a function which satisfies the hypothesis (2.10).

Motivated by the results from [25] on the eigenvalue problem (2.11) in [22] the last two authors of this paper studied the set of parameters λ for which the following perturbed eigenvalue problem has nontrivial solutions

$$-\Delta u-\Delta_p u=\lambda V(x)u,\quad u\in\mathcal{D}_0^{1,\Phi_p}(\Omega)\,, \tag{3.7}$$

where $p\in(1,N)\setminus\{2\}$ and $\Phi_p:\mathbb{R}\to\mathbb{R}$ is given by $\Phi_p(t):=\frac{t^2}{2}+\frac{|t|^p}{p}$, the Orlicz-Sobolev type space $\mathcal{D}_0^{1,\Phi_p}(\Omega)$ is obtained as the closure of $C_0^\infty(\Omega)$ under the Luxemburg-type norm

$$\|u\|:=\inf\left\{\mu>0;\ \int_\Omega\Phi_p\left(\frac{|\nabla u(x)|_N}{\mu}\right)\,dx\leq1\right\},$$

(see [22, Section 2] for more details regarding the definition and properties of Φ_p and $\mathcal{D}_0^{1,\Phi_p}(\Omega)$). We recall that in the above framework we say that u is a *weak solution* of equation (3.7) if there exists $u\in\mathcal{D}_0^{1,\Phi_p}(\Omega)\setminus\{0\}$ such that

$$\int_\Omega\nabla u\nabla w\,dx+\int_\Omega|\nabla u|_N^{p-2}\nabla u\nabla w\,dx=\lambda\int_\Omega V(x)uw\,dx,\quad\forall\, w\in\mathcal{D}_0^{1,\Phi_p}(\Omega).$$

The main result on problem (3.7) is formulated in the following theorem.

Theorem 3.4. *Assume condition* (2.10) *is fulfilled. Then the set of parameters λ for which problem* (3.7) *possesses nontrivial solutions is exactly the open interval*

$(\lambda_1, +\infty)$, *where* λ_1 *is given by*

$$\lambda_1 := \inf_{u \in C_0^\infty(\Omega)\setminus\{0\}} \frac{\displaystyle\int_\Omega |\nabla u(x)|_N^2 \, dx}{\displaystyle\int_\Omega V(x) u^2(x) \, dx} . \tag{3.8}$$

Note that by [25, Theorem 2.2] it is obvious that λ_1 defined in (3.8) is achieved in $\mathcal{D}_0^{1,2}(\Omega)$ which is larger than $\mathcal{D}_0^{1,\Phi_p}(\Omega)$ (see [22, Section 2] for details).

In the second part of this section we let $\Omega \subset \mathbb{R}^N$ $(N \geq 2)$ be a simply connected bounded domain, containing the origin, with C^2 boundary denoted by $\partial\Omega$ and we denote by $\Omega^{\text{ext}} := \mathbb{R}^N \setminus \overline{\Omega}$ the exterior of Ω. Let $K : \Omega^{\text{ext}} \to (0, \infty)$ be a function having the property that $K \in L^\infty(\Omega^{\text{ext}}) \cap L^{N/p}(\Omega^{\text{ext}})$ for some $p \in (1, N)$. Let $\lambda_1(p; \Omega^{\text{ext}})$ be the first eigenvalue of problem (2.12) given by relation (2.13). In [13] the second author of this paper investigated a perturbation of problem (2.12) obtained when we perturb the p-Laplacian by a q-Laplacian with $q \neq p$. More precisely, he studied the problem

$$\begin{cases} -\Delta_p u - \Delta_q u = \lambda K(x)|u|^{p-2}u, & \text{for} \quad x \in \Omega^{\text{ext}}, \\ u(x) = 0, & \text{for} \quad x \in \partial\Omega, \\ u(x) \to 0, & \text{as} \quad |x|_N \to \infty, \end{cases} \tag{3.9}$$

where $p, q \in (1, N)$ with $p \neq q$. Note that the natural function space framework for problem (3.9) is given by the Orlicz-Sobolev space $W_0^{1,\Psi_{p,q}}(\Omega^{\text{ext}})$ constructed with the aid of the N-function $\Psi_{p,q} : [0, \infty) \to \mathbb{R}$, given by $\Psi_{p,q}(t) := \frac{t^p}{p} + \frac{t^q}{q}$. In that framework, we say that $u \in W_0^{1,\Psi_{p,q}}(\Omega^{\text{ext}})$ is a weak solution of problem (3.9), if the following relation holds

$$\int_{\Omega^{\text{ext}}} (|\nabla u|_N^{p-2} + |\nabla u|_N^{q-2}) \nabla u \nabla \varphi \mathrm{d}x = \lambda \int_{\Omega^{\text{ext}}} K(x)|u|^{p-2} u \varphi \mathrm{d}x,$$

for all $\varphi \in W_0^{1,\Psi_{p,q}}(\Omega^{\text{ext}})$.

The main result on problem (3.9) is given by the following theorem.

Theorem 3.5. *The set of all parameters λ for which problem (3.9) possesses nontrivial weak solutions is the open interval $(\lambda_1(p; \Omega^{\text{ext}}), \infty)$.*

3.2. The perturbation of the nonlocal (s, p)-Laplace operator

Let $\Omega \subset \mathbb{R}^N$ $(N \geq 2)$ be a bounded domain with Lipschitz boundary $\partial\Omega$. In [10] three of the authors of this paper studied a perturbation of the eigenvalue problem (2.15), namely

$$\begin{cases} (-\Delta_p)^s u(x) + (-\Delta_q)^t u(x) = \lambda |u(x)|^{r-2} u(x), & \text{for } x \in \Omega, \\ u(x) = 0, & \text{for } x \in \mathbb{R}^N \setminus \Omega \,, \end{cases} \tag{3.10}$$

where s, t, p and q are real numbers satisfying the assumption

$$0 < t < s < 1, \quad 1 < p < q < \infty, \quad s - \frac{N}{p} = t - \frac{N}{q} \,, \tag{3.11}$$

$r \in \{p, q\}$ and $\lambda \in \mathbb{R}$ is a parameter. The goal was to determine all the parameters λ for which problem (3.10) possesses nontrivial weak solutions. By a *weak solution of problem (*3.10*)* we understand a function $u \in \widetilde{W}_0^{s,p}(\Omega)$ such that

$$\mathcal{E}_{s,p}(u,v) + \mathcal{E}_{t,q}(u,v) = \lambda \int_\Omega |u(x)|^{r-2}u(x)v(x)\,dx, \quad \forall\, v \in \widetilde{W}_0^{s,p}(\Omega)\,, \tag{3.12}$$

where the quantities $\mathcal{E}_{s,p}(u,v)$ and $\mathcal{E}_{t,q}(u,v)$ are given by relation (2.16).

Define

$$\overline{\lambda}_1 := \begin{cases} \lambda_1(s,p), & \text{if} \quad r = p, \\ \lambda_1(t,q), & \text{if} \quad r = q\,, \end{cases} \tag{3.13}$$

where $\lambda_1(s,p)$ and $\lambda_1(t,q)$ are given by relation (2.18). The main result on problem (3.10) is given by the following theorem (see [10, Theorem 1.1]).

Theorem 3.6. *Assume condition (*3.11*) is fulfilled. Then the set of all real parameters λ for which problem (*3.10*) has at least a nontrivial weak solution is the interval* $(\overline{\lambda}_1, \infty)$*, with $\overline{\lambda}_1$ defined by relation (*3.13*). Moreover, the weak solution could be chosen to be non-negative.*

Next, we recall a result obtained by the second author of this paper in [12] on a perturbation of problem (2.19), namely

$$(-\Delta)^s\, u(x) + (-\Delta_p)^t\, u(x) = \lambda V(x)u(x), \quad \forall\, x \in \mathbb{R}^N, \tag{3.14}$$

under the assumption

$$0 < t < s < 1 \text{ and } s - \frac{N}{2} = t - \frac{N}{p}, \tag{3.15}$$

where λ is a real parameter and $V : \mathbb{R}^N \to [0, \infty)$ is a function satisfying the hypothesis $(\widetilde{\mathrm{V}})$. Note that in the case of problem (3.14) we have $V = V^+$. We will say that $\lambda \in \mathbb{R}$ is an eigenvalue of problem (3.14), if there exists $u \in \mathcal{D}_0^{s,2}(\mathbb{R}^N) \setminus \{0\}$ such that

$$\mathcal{E}_{s,2}(u,\varphi) + \mathcal{E}_{t,p}(u,\varphi) = \lambda \int_{\mathbb{R}^N} V(x)u(x)\varphi(x)\,dx\,, \tag{3.16}$$

for all $\varphi \in \mathcal{D}_0^{s,2}(\mathbb{R}^N)$, where the quantities $\mathcal{E}_{s,2}(u,v)$ and $\mathcal{E}_{t,q}(u,v)$ are given by relation (2.16). Furthermore, u from the above relation will be called an eigenfunction corresponding to the eigenvalue λ.

Define

$$\tilde{\lambda}_1 := \inf_{u \in C_0^\infty(\mathbb{R}^N)\setminus\{0\}} \frac{\displaystyle\int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \frac{|u(x)-u(y)|^2}{|x-y|_N^{N+2s}}\,dxdy}{\displaystyle\int_{\mathbb{R}^N} V(x)u^2\,dx}. \tag{3.17}$$

The main result regarding problem (3.14) is given by the following theorem (see [12, Theorem 1.5]).

Theorem 3.7. *Assume that $V : \mathbb{R}^N \to [0, \infty)$ is a function which satisfies condition $(\widetilde{V})$. Under assumption (*3.15*), the set of eigenvalues of problem (*3.14*) is the open interval* $(\tilde{\lambda}_1, \infty)$*. Moreover, the corresponding eigenfunctions can be chosen to be non-negative.*

Remark. A simple analysis of the proof of Theorem 1.3 from [12] shows that in the case when function V satisfies $V(x) \geq 0$, for all $x \in \mathbb{R}^N$, then $\tilde{\lambda}_1$ defined in relation (3.17) is the smallest eigenvalue of problem (2.19).

3.3. A perturbed eigenvalue problem involving rapidly growing operators in divergence form

Let $\Omega \subset \mathbb{R}^N$ ($N \geq 2$) be a bounded domain with smooth boundary $\partial\Omega$. In this section our goal is to recall some results on the perturbation of the classical eigenvalue problem of the Laplace operator subject to the homogenous Dirichlet boundary conditions (that is problem (2.8) with $p = 2$) with a so-called rapidly growing operator in divergence form (that is $\operatorname{div}(e^{|\nabla u|_N^2 - 1}\nabla u)$). More precisely, we are concerned with the problem

$$\begin{cases} -\operatorname{div}(e^{|\nabla u|_N^2 - 1}\nabla u) - \Delta u = \lambda u & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{3.18}$$

This problem was investigated by Bocea and the third author of this paper in [4]. Using a similar terminology as in the case of the classical eigenvalue problems a real number λ is called an eigenvalue of problem (3.18) if the problem possesses a nontrivial weak solution. Note the fact that the nature of the problem asks for a function space framework involving an Orlicz-Sobolev space, say $X_0 := W_0^{1,\Psi}(\Omega)$ which is constructed with the aid of the N-function $\Psi : [0, \infty) \to \mathbb{R}$, given by $\Psi(t) := e^{t^2} - 1$.

Next, note that the Euler-Lagrange functional associated to the problem (3.18) is $\Lambda : X_0 \to \mathbb{R}$ defined by

$$\Lambda(u) := \frac{1}{2}\int_\Omega \Phi(|\nabla u(x)|_N)\, dx + \frac{e}{2}\int_\Omega |\nabla u(x)|_N^2\, dx - \lambda\frac{e}{2}\int_\Omega |u(x)|^2\, dx.$$

If Λ was smooth on X_0, then one could define an eigenvalue for (3.18) as a real number λ for which there exists a function $u \in X_0 \setminus \{0\}$ such that

$$\int_\Omega e^{|\nabla u|_N^2}\nabla u \nabla v\, dx + e\int_\Omega \nabla u\nabla v\, dx - \lambda e\int_\Omega uv\, dx = 0, \quad \forall\, v \in X_0\,.$$

Unfortunately, in our framework, the functional Λ is not smooth on X_0. However, the functional $g : X_0 \to \mathbb{R}$ defined by

$$g(u) := \frac{e}{2}\int_\Omega |u(x)|^2 dx \tag{3.19}$$

is of class $C^1(X_0, \mathbb{R})$, and we have $\langle g'(u), v\rangle = e\displaystyle\int_\Omega uv\, dx$ for all $u, v \in X_0$. On the other hand, the functional $f : X_0 \to \mathbb{R}$ given by

$$f(u) := \frac{1}{2}\int_\Omega \Phi(|\nabla u(x)|_N) dx + \frac{e}{2}\int_\Omega |\nabla u(x)|_N^2 dx \tag{3.20}$$

is convex, weakly* lower semicontinuous, and coercive but $f \notin C^1(X_0, \mathbb{R})$. To overcome this drawback, we will work with the following reformulation (*à la* Szulkin [24]) of the problem (3.18) as a variational inequality:

$$\begin{cases} f(v) - f(u) - \lambda\langle g^{'}(u), v - u\rangle \geq 0, & \forall\, v \in X_0, \\ u \in X_0. \end{cases} \tag{3.21}$$

A real number λ such that (3.21) has nontrivial solutions $u \in X_0$ is called an *eigenvalue* for the problem (3.21). In this context the main result on problem (3.18) is given by the following theorem (see [4, Theorem 1])

Theorem 3.8. *The set of eigenvalues for problem (*3.18*) is the open interval*

$$\left(\left(1 + \frac{1}{e}\right)\lambda_1^D(2;\Omega), \infty\right),$$

*where $\lambda_1^D(2;\Omega)$ stands for the first eigenvalue of the Laplace operator under the homogenous Dirichlet boundary conditions (see relation (*2.7*) with $p = 2$).*

3.4. The spectrum of the relativistic mean curvature operator

In this section our goal is to characterize the spectrum of the *relativistic mean curvature operator*, i.e.

$$\mathcal{M}u := -\mathrm{div}\left(\frac{\nabla u}{\sqrt{1 - |\nabla u|_N^2}}\right),$$

acting on maps u defined in an open, bounded domain $\Omega \subset \mathbb{R}^N$ $(N \geq 1)$ with smooth boundary $\partial\Omega$, subject to the homogeneous Dirichlet boundary conditions. More precisely, our goal is to analyze the problem

$$\begin{cases} \mathcal{M}u = \lambda u & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{3.22}$$

The starting point in the study of problem (3.22) is to explain the function space framework that will be considered in the sequel. Thus, we note that the structure of the relativistic mean curvature operator asks for a condition of type $|\nabla u(x)|_N \leq 1$ for a.e. $x \in \Omega$. That fact and the homogeneous Dirichlet boundary conditions involved in problem (3.22) imply that a good candidate for the functional space framework would be a subset of

$$W_0^{1,\infty}(\Omega) := \{u \in W^{1,\infty}(\Omega) :\ u = 0, \text{ on } \partial\Omega\},$$

namely

$$K_0 := \{u \in W_0^{1,\infty}(\Omega) :\ |\nabla u(x)|_N \leq 1, \text{ a.e. } x \in \Omega\}.$$

We remark that K_0 is a convex and closed subset of $W^{1,\infty}(\Omega)$ which is the dual of a separable Banach space. This leads to the idea of constructing the Euler-Lagrange functional associated to the relativistic mean curvature operator as $I : W^{1,\infty}(\Omega) \to [0,\infty]$ defined by

$$I(u) := \begin{cases} \displaystyle\int_\Omega F(|\nabla u|_N)\, dx & \text{if } u \in K_0, \\ +\infty & \text{if } u \in W^{1,\infty}(\Omega) \setminus K_0, \end{cases}$$

where $F : [-1,1] \to \mathbb{R}$ is given by $F(t) := 1 - \sqrt{1-t^2}$ for all $t \in [-1,1]$. Then, the Euler-Lagrange functional associated to the problem (3.22) is $J_\lambda : W^{1,\infty}(\Omega) \to \mathbb{R}$ defined by

$$J_\lambda(u) := I(u) - \frac{\lambda}{2}\int_\Omega u^2\, dx, \quad \forall\, u \in W^{1,\infty}(\Omega)\,.$$

We observe that J_λ is the sum of a convex, lower semi-continuous function and a C^1-functional, and, consequently, it has the structure required by *Szulkin's critical point theory* (see [24]). More precisely, the functional J_λ is the sum of the functional $h_\lambda : W^{1,\infty}(\Omega) \to \mathbb{R}$ defined by

$$h_\lambda(u) := -\frac{\lambda}{2}\int_\Omega u^2\, dx\,,$$

which belongs to $C^1(W^{1,\infty}(\Omega), \mathbb{R})$ and has the derivative given by

$$\langle h'_\lambda(u), v\rangle = -\lambda \int_\Omega uv\, dx, \quad \forall\, u,\ v \in W^{1,\infty}(\Omega)\,,$$

with the functional I which is convex and weakly* lower semicontinuous. Then, we will work with a reformulation of problem (3.22) as a *variational inequality*, namely

$$\begin{cases} I(v) - I(u_\lambda) + \langle h'_\lambda(u_\lambda), v - u_\lambda\rangle \geq 0 \quad \text{for all } v \in W^{1,\infty}(\Omega), \\ u_\lambda \in W^{1,\infty}(\Omega)\,. \end{cases} \tag{3.23}$$

or, equivalently,

$$\begin{cases} I(v) - I(u_\lambda) + \langle h'_\lambda(u_\lambda), v - u_\lambda\rangle \geq 0 \quad \text{for all } v \in K_0, \\ u_\lambda \in K_0\,. \end{cases} \tag{3.24}$$

In this context a real number $\lambda \in \mathbb{R}$ is called an *eigenvalue* for problem (3.22) if problem (3.24) has a nontrivial solution $u_\lambda \in K_0$. u_λ will be called an *eigenfunction* corresponding to the eigenvalue λ. According to the terminology from [24], we refer to u_λ as being a *critical point* of functional J_λ.

The main result on problem (3.22) is given by the following theorem (see [20, Theorem 1.1]).

Theorem 3.9. *The set of eigenvalues for problem (*3.22*) is the open interval* $(\lambda_1^D(2;\Omega), \infty)$ *where* $\lambda_1^D(2;\Omega)$ *stands for the principal frequency of the Laplace operator in* Ω *subject to the homogeneous Dirichlet boundary conditions (see relation (*2.7*) with* $p = 2$*). Moreover, for each eigenvalue* λ *we can choose a corresponding eigenfunction* $u_\lambda \in K_0$ *which is nonnegative on* Ω *and minimizes* J_λ.

Note that problem (3.22) can be regarded as a perturbation of the classical eigenvalue problem of the Laplace operator subject to the homogenous Dirichlet boundary conditions (that is problem (2.8) with $p = 2$). Indeed, first note that the function $F : [-1,1] \to \mathbb{R}$, given by $F(t) := 1 - \sqrt{1-t^2}$, for all $t \in [-1,1]$, admits the following extension into power series

$$F(t) = \frac{1}{2}t^2 + \sum_{n\geq 2} a_n t^{2n}, \quad \forall\, t \in [-1,1]\,,$$

where for each integer $n \geq 2$ we let $a_n := \frac{(2n-3)!!}{2^n n!}$. Thus, the above simple remark suggests to us that the differential operator,

$$u \mapsto -\mathrm{div}\left(\frac{\nabla u}{\sqrt{1-|\nabla u|_N^2}}\right),$$

on the left hand side of the PDE in (3.22) can be regarded as being equivalent with the differential operator

$$u \mapsto -\Delta u - \sum_{n\geq 2} a_n \Delta_{2n} u\,,$$

where $\Delta_{2n}u$ stands for the $2n$-Laplacian of u (i.e. $\Delta_{2n}u = \mathrm{div}(|\nabla u|_N^{2n-2}\nabla u)$), for each positive integer n. Thus, problem (3.22) can be reformulated as

$$\begin{cases} -\Delta u - \sum\limits_{n\geq 2} a_n \Delta_{2n} u = \lambda u & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{3.25}$$

Acknowledgements. AG, MF and DS-D have been partially supported by CNCS-UEFISCDI Grant No. PN-III-P1-1.1-TE-2019-0456.

References

[1] Abreu, J., Madeira, G.F., *Generalized eigenvalues of the $(p,2)$-Laplacian under a parametric boundary condition*, Proc. Edinb. Math. Soc., **63**(2020), 287-303.

[2] Barbu, L., Moroşanu, G., *Eigenvalues of the negative (p,q)-Laplacian under a Steklov-like boundary condition*, Complex Var. Elliptic Equ., **64**(2019), 685-700.

[3] Bhattacharya, T., Emamizadeh, B., Farjudian, A., *Existence of continuous eigenvalues for a class of parametric problems involving the $(p,2)$-Laplacian operator*, Acta Appl. Math., **165**(2020), 65-79.

[4] Bocea, M., Mihăilescu, M., *Eigenvalue problems in Orlicz-Sobolev spaces for rapidly growing operators in divergence form*, J. Differential Eqnuations, **256**(2014), 640-657.

[5] Bocea, M., Mihăilescu, M., *Existence of nonnegative viscosity solutions for a class of problems involving the ∞-Laplacian*, Nonlinear Differential Equations and Applications (NoDEA), **23**(2016), Art. 11.

[6] Chhetri, M., Drábek, P., *Principal eigenvalue of p-Laplacian operator in exterior domain*, Results Math., **66**(2014), 461-468.

[7] Costea, N., Moroşanu, G., *Steklov-type eigenvalues of $\Delta_p + \Delta_q$*, Pure Appl. Funct. Anal., **3**(2018), 75-89.

[8] Fărcăşeanu, M., *An eigenvalue problem involving an anisotropic differential operator*, Complex Variables and Elliptic Equations, **62**(2017), 297-306.

[9] Fărcăşeanu, M., Mihăilescu, M., Stancu-Dumitru, D., *On the set of eigenvalues of some PDEs with homogeneous Neumann boundary condition*, Nonlinear Analysis, **116**(2015), 19-25.

[10] Fărcăşeanu, M., Mihăilescu, M., Stancu-Dumitru, D., *Perturbed fractional eigenvalue problems*, Discrete and Continuous Dynamical Systems - Series A, **37**(2017), 6243-6255.

[11] Fukagai, N., Ito, M., Narukawa, K., *Limit as $p \to \infty$ of p-Laplace eigenvalue problems and L^∞-inequality of Poincaré type*, Differential Integral Equations, **12**(1999), 183-206.

[12] Grecu, A., *Fractional eigenvalue problems on $\mathbb{R}^N$*, Electronic Journal of Qualitative Theory of Differential Equations, **26**(2020), 1-17.

[13] Grecu, A., *A perturbed eigenvalue problem in exterior domain*, submitted.

[14] Gyulov, T., Moroşanu, G., *Eigenvalues of $-(\Delta_p + \Delta_q)$ under a Robin-like boundary condition*, Ann. Acad. Rom. Sci. Ser. Math. Appl., **8**(2016), 114-132.

[15] Juutinen, P., Lindqvist, P., Manfredi, J.J., *The ∞-eigenvalue problem*, Arch. Rational Mech. Anal., **148**(1999), 89-105.

[16] Lê, A., *Eigenvalue problems for the p-Laplacian*, Nonlinear Analysis, **64**(2006), 1057-1099.

[17] Lindgren, E., Lindqvist, P., *Fractional eigenvalues*, Calc. Var., **49**(2014), 795-826.

[18] Lindqvist, P., *A nonlinear eigenvalue problem*, In: Ciatti, Paolo (ed.) et al., "Topics in Mathematical Analysis", 175-203, Hackensack, NJ: World Scientific (ISBN 978-981-281-105-9/hbk). Series on Analysis Applications and Computation 3, 2008.

[19] Mihăilescu, M., *An eigenvalue problem possessing a continuous family of eigenvalues plus an isolated eigenvalue*, Communications on Pure and Applied Analysis, **10**(2011), 701-708.

[20] Mihăilescu, M., *The Spectrum of the Mean Curvature Operator*, accepted for publication in: Proceedings of the Royal Society of Edinburgh Section A: Mathematics. (DOI: 10.1017/prm.2020.25).

[21] Mihăilescu, M., Moroşanu, G., *Eigenvalues of $-\Delta_p - \Delta_q$ under Neumann boundary condition*, Canadian Mathematical Bulletin, **59**(2016), 606-616.

[22] Mihăilescu, M., Stancu-Dumitru, D., *A perturbed eigenvalue problem on general domains*, Annals of Functional Analysis, **7**(2016), 529-542.

[23] Tanaka, M., *Generalized eigenvalue problems for (p, q)-Laplacian with indefinite weight*, J. Math. Anal. Appl., **419**(2014), 1181-1192.

[24] Szulkin, A., *Minimax principles for lower semicontinuous functions and applications to nonlinear boundary value problems*, Ann. Inst. H. Poincaré Anal. Non Linéaire, **3**(1986), 77-109.

[25] Szulkin, A., Willem, M., *Eigenvalue problems with indefinite weights*, Studia Mathematica, **135**(1999), 191-201.

Maria Fărcăşeanu
The University of Sydney, School of Mathematics and Statistics,
NSW 2006, Australia
and
University Politehnica of Bucharest,
Research group of the project PN-III-P1-1.1-TE-2019-0456,
060042 Bucharest, Romania
e-mail: `maria.farcaseanu@sydney.edu.au`

Andrei Grecu
University of Craiova, Department of Mathematics,
200585 Craiova, Romania
and
University Politehnica of Bucharest,
Research group of the project PN-III-P1-1.1-TE-2019-0456,
060042 Bucharest, Romania
e-mail: andreigrecu.cv@gmail.com

Mihai Mihăilescu
University of Craiova, Department of Mathematics,
200585 Craiova, Romania
e-mail: mmihailes@yahoo.com

Denisa Stancu-Dumitru
University Politehnica of Bucharest,
Research group of the project PN-III-P1-1.1-TE-2019-0456,
060042 Bucharest, Romania
e-mail: denisa.stancu@yahoo.com

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 75–84
DOI: 10.24193/subbmath.2021.1.06

A class of functionals possessing multiple global minima

Biagio Ricceri

To Professor Gheorghe Moroşanu, with friendship, on his 70th birthday.

Abstract. We get a new multiplicity result for gradient systems. Here is a very particular corollary: Let $\Omega \subset \mathbf{R}^n$ $(n \geq 2)$ be a smooth bounded domain and let $\Phi : \mathbf{R}^2 \to \mathbf{R}$ be a C^1 function, with $\Phi(0,0) = 0$, such that

$$\sup_{(u,v)\in\mathbf{R}^2} \frac{|\Phi_u(u,v)| + |\Phi_v(u,v)|}{1 + |u|^p + |v|^p} < +\infty$$

where $p > 0$, with $p = \frac{2}{n-2}$ when $n > 2$.
Then, for every convex set $S \subseteq L^\infty(\Omega) \times L^\infty(\Omega)$ dense in $L^2(\Omega) \times L^2(\Omega)$, there exists $(\alpha, \beta) \in S$ such that the problem

$$-\Delta u = (\alpha(x)\cos(\Phi(u,v)) - \beta(x)\sin(\Phi(u,v)))\Phi_u(u,v) \ \text{ in } \ \Omega$$

$$-\Delta v = (\alpha(x)\cos(\Phi(u,v)) - \beta(x)\sin(\Phi(u,v)))\Phi_v(u,v) \ \text{ in } \ \Omega$$

$$u = v = 0 \ \text{ on } \ \partial\Omega$$

has at least three weak solutions, two of which are global minima in $H_0^1(\Omega) \times H_0^1(\Omega)$ of the functional

$$(u,v) \to \frac{1}{2}\left(\int_\Omega |\nabla u(x)|^2 dx + \int_\Omega |\nabla v(x)|^2 dx\right)$$

$$-\int_\Omega (\alpha(x)\sin(\Phi(u(x),v(x))) + \beta(x)\cos(\Phi(u(x),v(x))))dx \ .$$

Mathematics Subject Classification (2010): 35J47, 35J50, 49K35.

Keywords: Minimax, multiple global minima, variational methods, semilinear elliptic systems.

1. Introduction

Let S be a topological space. A function $g : S \to \mathbb{R}$ is said to be inf-compact if, for each $r \in \mathbb{R}$, the set $g^{-1}(]-\infty, r])$ is compact.

If Y is a real interval and $f : S \times Y \to \mathbb{R}$ is a function inf-compact and lower semicontinuous in S, and concave in Y, the occurrence of the strict minimax inequality

$$\sup_Y \inf_S f < \inf_S \sup_Y f$$

implies that the interior of the set A of all $y \in Y$ for which $f(\cdot, y)$ has at least two local minima is non-empty. This fact was essentially shown in [4], giving then raise to an enormous number of subsequent applications to the multiplicity of solutions for nonlinear equations of variational nature (see [7] for an account up to 2010).

In [6] (see also [5]), we realized that, under the same assumptions as above, the occurrence of the strict minimax inequality also implies the existence of $\tilde{y} \in Y$ such that the function $f(\cdot, \tilde{y})$ has at least two global minima. It may happen that $\tilde{y}$ is unique and does not belong to the closure of A (see Example 7 of [1]).

In [8] and [12], we extended the result of [6] to the case where Y is an arbitrary convex set in a vector space. We also stress that such an extension is not possible for the result of [4]. We then started to build a network of applications of the results of [8] and [12] which touches several different topics: uniquely remotal sets in normed spaces ([8]); non-expansive operators ([9]); singular points ([10]); Kirchhoff-type problems ([11]); Lagrangian systems of relativistic oscillators ([13]); integral functional of the Calculus of Variations ([14]); non-cooperative gradient systems ([15]); variational inequalities ([16]).

The aim of this paper is to establish a further application within that network.

2. Results

The main abstract result is as follows:

Theorem 2.1. *Let X be a topological space, $(Y, \langle\cdot,\cdot,\rangle)$ a real Hilbert space, $T \subseteq Y$ a convex set dense in Y and $I : X \to \mathbb{R}$, $\varphi : X \to Y$ two functions such that, for each $y \in T$, the function $x \to I(x) + \langle\varphi(x), y\rangle$ is lower semicontinuous and inf-compact. Moreover, assume that there exists a point $x_0 \in X$, with $\varphi(x_0) \neq 0$, such that*
(a) x_0 is a global minimum of both functions I and $\|\varphi(\cdot)\|$;
(b) $\inf_{x\in X}\langle\varphi(x), \varphi(x_0)\rangle < \|\varphi(x_0)\|^2$.

Then, for each convex set $S \subseteq T$ dense in Y, there exists $y^* \in S$ such that the function $x \to I(x) + \langle\varphi(x), y^*\rangle$ has at least two global minima in X.

Proof. In view of (b), we can find $\tilde{x} \in X$ and $r > 0$ such that

$$I(\tilde{x}) + \frac{r}{\|\varphi(x_0)\|}\langle\varphi(\tilde{x}), \varphi(x_0)\rangle < I(x_0) + r\|\varphi(x_0)\| \ . \tag{2.1}$$

Thanks to (a), we have

$$I(x_0)+r\|\varphi(x_0)\|=\inf_{x\in X}(I(x)+r\|\varphi(x)\|)\ . \tag{2.2}$$

The function $y\to\inf_{x\in X}(I(x)+\langle\varphi(x),y\rangle)$ is weakly upper semicontinuous, and so there exists $\tilde{y}\in B_r$ such that

$$\inf_{x\in X}(I(x)+\langle\varphi(x),\tilde{y}\rangle)=\sup_{y\in B_r}\inf_{x\in X}(I(x)+\langle\varphi(x),y\rangle)\ , \tag{2.3}$$

B_r being the closed ball in X, centered at 0, of radius r. We distinguish two cases. First, assume that $\tilde{y}\neq\frac{r\varphi(x_0)}{\|\varphi(x_0\|}$. As a consequence, taking into account that $r\|\varphi(x_0)\|$ is the maximum of the restriction to B_r of the continuous linear functional $\langle\varphi(x_0),\cdot\rangle$ (attained at the point $\frac{r\varphi(x_0)}{\|\varphi(x_0)\|}$ only), we have

$$\inf_{x\in X}(I(x)+\langle\varphi(x),\tilde{y}\rangle)\leq I(x_0)+\langle\varphi(x_0),\tilde{y}\rangle<I(x_0)+r\|\varphi(x_0)\|\ . \tag{2.4}$$

Now, assume that $\tilde{y}=\frac{r\varphi(x_0)}{\|\varphi(x_0\|}$. In this case, due to (2.1), we have

$$\inf_{x\in X}(I(x)+\langle\varphi(x),\tilde{y}\rangle)\leq I(\tilde{x})+\langle\varphi(\tilde{x}),\tilde{y}\rangle=I(\tilde{x})+\frac{r}{\|\varphi(x_0)\|}\langle\varphi(\tilde{x}),\varphi(x_0)\rangle$$
$$<I(x_0)+r\|\varphi(x_0)\|\ . \tag{2.5}$$

Therefore, from (2.2), (2.3), (2.4) and (2.5), it follows that

$$\sup_{y\in B_r}\inf_{x\in X}(I(x)+\langle\varphi(x),y\rangle)<\inf_{x\in X}\sup_{y\in B_r}(I(x)+\langle\varphi(x),y\rangle)\ . \tag{2.6}$$

Now, let $S\subseteq T$ be a convex set dense in Y. By continuity, we clearly have

$$\sup_{y\in B_r\cap S}\langle\varphi(x),y\rangle=\sup_{y\in B_r}\langle\varphi(x),y\rangle$$

for all $x\in X$. Therefore, in view of (2.6), we have

$$\sup_{y\in B_r\cap S}\inf_{x\in X}(I(x)+\langle\varphi(x),y\rangle)\leq\sup_{y\in B_r}\inf_{x\in X}(I(x)+\langle\varphi(x),y\rangle)$$
$$<\inf_{x\in X}\sup_{y\in B_r}(I(x)+\langle\varphi(x),y\rangle)=\inf_{x\in X}\sup_{y\in B_r\cap S}(I(x)+\langle\varphi(x),y\rangle)\ .$$

At this point, the conclusion follows directly applying Theorem 1.1 of [12] to the restriction of the function $(x,y)\to I(x)+\langle\varphi(x),y\rangle$ to $X\times(B_r\cap S)$. □

We now present an application of Theorem 2.1 to elliptic systems.

In the sequel, $\Omega\subseteq\mathbf{R}^n$ $(n\geq 2)$ is a bounded domain with smooth boundary.

We denote by $\mathcal{A}$ the class of all functions $H:\Omega\times\mathbf{R}^2\to\mathbf{R}$ which are measurable in Ω, C^1 in $\mathbf{R}^2$ and satisfy

$$\sup_{(x,u,v)\in\Omega\times\mathbf{R}^2}\frac{|H_u(x,u,v)|+|H_v(x,u,v)|}{1+|u|^p+|v|^p}<+\infty$$

where $p>0$, with $p<\frac{n+2}{n-2}$ when $n>2$.

Given $H \in \mathcal{A}$, we are interested in the problem

$$\begin{aligned} -\Delta u &= H_u(x,u,v) \ \text{ in } \ \Omega \\ -\Delta v &= H_v(x,u,v) \ \text{ in } \ \Omega \\ u = v &= 0 \ \text{ on } \ \partial\Omega \ , \end{aligned}$$

H_u (resp. H_v) denoting the derivative of H with respect to u (resp. v).
As usual, a weak solution of this problem is any $(u,v) \in H^1_0(\Omega) \times H^1_0(\Omega)$ such that

$$\int_\Omega \nabla u(x)\nabla\varphi(x)dx = \int_\Omega H_u(x,u(x),v(x))\varphi(x)dx \ ,$$

$$\int_\Omega \nabla v(x)\nabla\psi(x)dx = \int_\Omega H_v(x,u(x),v(x))\psi(x)dx$$

for all $\varphi, \psi \in H^1_0(\Omega)$.

Define the functional $I_H : H^1_0(\Omega) \times H^1_0(\Omega) \to \mathbf{R}$ by

$$I_H(u,v) = \frac{1}{2}\left(\int_\Omega |\nabla u(x)|^2dx + \int_\Omega |\nabla v(x)|^2dx\right) - \int_\Omega H(x,u(x),v(x))dx$$

for all $(u,v) \in H^1_0(\Omega) \times H^1_0(\Omega)$.

Since $H \in \mathcal{A}$, the functional I_H is C^1 in $H^1_0(\Omega)\times H^1_0(\Omega)$ and its critical points are precisely the weak solutions of the problem. Moreover, due to the Sobolev embedding theorem, the functional $(u,v) \to \int_\Omega H(x,u(x),v(x))$ has a compact derivative and, as a consequence, it is sequentially weakly continuous in $H^1_0(\Omega) \times H^1_0(\Omega)$.

Also, we denote by λ_1 the first eigenvalue of the Dirichlet problem

$$\begin{aligned} -\Delta u &= \lambda u \ \text{ in } \ \Omega \\ u &= 0 \ \text{ on } \ \partial\Omega \ . \end{aligned}$$

Our result is as follows:

Theorem 2.2. *Let $F, G \in \mathcal{A}$, with $p = \frac{2}{n-2}$ when $n > 2$, and let $K \in \mathcal{A}$, with $K(x,0,0) = 0$ for all $x \in \Omega$, satisfy the following conditions:*
(a_1) *one has*

$$\lim_{s^2+t^2\to+\infty} \frac{\sup_{x\in\Omega}(|F(x,s,t)| + |G(x,s,t)|)}{s^2+t^2} = 0 \ ;$$

(a_2) *there is $\eta \in \left]0, \frac{\lambda_1}{2}\right[$ such that*

$$K(x,s,t) \le \eta(s^2+t^2)$$

for all $x \in \Omega$, $s, t \in \mathbf{R}$;
(a_3) *one has*

$$\operatorname{meas}(\{x \in \Omega : 0 < |F(x,0,0)|^2 + |G(x,0,0)|^2\}) > 0 \tag{2.7}$$

and

$$|F(x,0,0)|^2 + |G(x,0,0)|^2 \le |F(x,s,t)|^2 + |G(x,s,t)|^2 \tag{2.8}$$

for all $x \in \Omega$, $s, t \in \mathbf{R}$;

(a_4) *one has*

$$\operatorname{meas}(\{x \in \Omega : \inf_{(s,t)\in\mathbf{R}^2}(F(x,0,0)F(x,s,t) + G(x,0,0)G(x,s,t))$$
$$< |F(x,0,0)|^2 + |G(x,0,0)|^2\}) > 0 \ .$$

Then, for every convex set $S \subseteq L^\infty(\Omega) \times L^\infty(\Omega)$ dense in $L^2(\Omega) \times L^2(\Omega)$, there exists $(\alpha, \beta) \in S$ such that the problem

$$\begin{aligned} -\Delta u &= \alpha(x)F_u(x,u,v) + \beta(x)G_u(x,u,v) + K_u(x,u,v) \ \text{ in } \Omega \\ -\Delta v &= \alpha(x)F_v(x,u,v) + \beta(x)G_v(x,u,v) + K_v(x,u,v) \ \text{ in } \Omega \\ u &= v = 0 \ \text{ on } \partial\Omega \end{aligned}$$

has at least three weak solutions, two of which are global minima in $H^1_0(\Omega) \times H^1_0(\Omega)$ of the functional

$$(u,v) \to \frac{1}{2}\left(\int_\Omega |\nabla u(x)|^2 dx + \int_\Omega |\nabla v(x)|^2 dx\right)$$
$$- \int_\Omega (\alpha(x)F(x,u(x),v(x)) + \beta(x)G(x,u(x),v(x)) + K(x,u(x),v(x)))dx \ .$$

Proof. We are going to apply Theorem 2.1, with the following choices: X is the space $H^1_0(\Omega) \times H^1_0(\Omega)$ endowed with the weak topology induced by the scalar product

$$\langle (u,v),(w,\omega)\rangle_X = \int_\Omega (\nabla u(x)\nabla w(x) + \nabla v(x)\nabla\omega(x))dx \ ;$$

Y is the space $L^2(\Omega) \times L^2(\Omega)$ with the scalar product

$$\langle (f,g),(h,k)\rangle_Y = \int_\Omega (f(x)h(x) + g(x)k(x))dx \ ;$$

T is $L^\infty(\Omega) \times L^\infty(\Omega)$; I is the function defined by

$$I(u,v) = \frac{1}{2}\left(\int_\Omega |\nabla u(x)|^2 dx + \int_\Omega |\nabla v(x)|^2 dx\right) - \int_\Omega K(x,u(x),v(x))dx$$

for all $(u,v) \in X$; φ is the function defined by

$$\varphi(u,v) = (F(\cdot,u(\cdot),v(\cdot)), G(\cdot,u(\cdot),v(\cdot)))$$

for all $(u,v) \in X$; x_0 is the zero of X. Let us show that the assumptions of Theorem 2.1 are satisfied. First, from (2.7) and (2.8) it clearly follows, respectively, that

$$\|\varphi(0,0)\|^2_Y = \int_\Omega (|F(x,0,0)|^2 + |G(x,0,0|^2)dx > 0$$

and that

$$\|\varphi(0,0)\|^2_Y \le \|\varphi(u,v)\|^2_Y$$

for all $(u,v) \in X$. Moreover, from (a_2), thanks to the Poincaré inequality, we get

$$\int_\Omega K(x,u(x),v(x))dx \le \eta \int_\Omega (|u(x)|^2 + |v(x)|^2)dx \le \frac{\eta}{\lambda_1}\int_\Omega (|\nabla u(x)|^2 + |\nabla v(x)|^2)dx \tag{2.9}$$

for all $(u,v)\in X$. In particular, since $K(x,0,0)=0$ in Ω and $\frac{\eta}{\lambda_1}<\frac{1}{2}$, from (2.9) we infer that $(0,0)$ is a global minimum of I in X. So, condition (a) is satisfied. Now, let us verify condition (b). To this end, set

$$P(x,s,t)=F(x,0,0)F(x,s,t)+G(x,0,0)G(x,s,t)-|F(x,0,0)|^2-|G(x,0,0)|^2$$

for all $(x,s,t)\in\Omega\times\mathbf{R}^2$ and

$$D=\left\{x\in\Omega:\inf_{(s,t)\in\mathbf{R}^2}P(x,s,t)<0\right\}\ .$$

By (a_4), D has a positive measure. In view of the Scorza-Dragoni theorem, there exists a compact set $C\subset D$, with positive measure, such that the restriction of P to $C\times\mathbf{R}^2$ is continuous. Fix a point $\tilde{x}\in C$ such that the intersection of C and any ball centered at $\tilde{x}$ has a positive measure. Choose $\tilde{s},\tilde{t}\in\mathbf{R}\setminus\{0\}$ so that $P(\tilde{x},\tilde{s},\tilde{t})<0$. By continuity, there is $r>0$ such that

$$P(x,\tilde{s},\tilde{t})<0$$

for all $x\in C\cap B(\tilde{x},r)$. Set

$$\gamma=\sup_{(x,s,t)\in\Omega\times[-|\tilde{s}|,|\tilde{s}|]\times[-|\tilde{t}|,|\tilde{t}|]}|P(x,t,s)|\ .$$

Since $F,G\in\mathcal{A}$, γ is finite. Now, choose an open set A such that

$$C\cap B(\tilde{x},r)\subset A\subset\Omega$$

and

$$\text{meas}(A\setminus(C\cap B(\tilde{x},r)))<-\frac{\int_{C\cap B(\tilde{x},r)}P(x,\tilde{s},\tilde{t})dx}{\gamma}\ . \tag{2.10}$$

Finally, choose two functions $\tilde{u},\tilde{v}\in H^1_0(\Omega)$ such that

$$\tilde{u}(x)=\tilde{s}\ ,\ \ \tilde{v}(x)=\tilde{t}$$

for all $x\in C\cap B(\tilde{x},r)$,

$$\tilde{u}(x)=\tilde{v}(x)=0$$

for all $x\in\Omega\setminus A$ and

$$|\tilde{u}(x)|\le|\tilde{s}|\ ,\ \ |\tilde{v}(x)|\le|\tilde{t}|$$

for all $x\in\Omega$. Then, taking (2.10) into account, we have

$$\langle\varphi(\tilde{u},\tilde{v}),\varphi(0,0)\rangle_Y-\|\varphi(0,0)\|^2_Y=\int_\Omega P(x,\tilde{u}(x),\tilde{v}(x))dx$$

$$=\int_{C\cap B(\tilde{x},r)}P(x,\tilde{s},\tilde{t})dx+\int_{A\setminus(C\cap B(\tilde{x},r))}P(x,\tilde{u}(x),\tilde{v}(x))dx$$

$$<\int_{C\cap B(\tilde{x},r)}P(x,\tilde{s},\tilde{t})dx+\gamma\text{meas}(A\setminus(C\cap B(\tilde{x},r))<0\ .$$

This shows that (b) is satisfied. Finally, fix $\alpha,\beta\in L^\infty(\Omega)$. Clearly, the function

$$(x,s,t)\to\alpha(x)F(x,s,t)+\beta(x)F(x,s,t)+K(x,s,t)$$

belongs to $\mathcal{A}$, and so the functional

$$(u,v) \to I(u,v) + \langle \varphi(u,v),(\alpha,\beta)\rangle_Y$$

is sequentially weakly lower semicontinuous in X. Let us show that it is coercive. Set

$$\theta = \max\left\{\|\alpha\|_{L^\infty(\Omega)}, \|\beta\|_{L^\infty(\Omega)}\right\}$$

and fix $\epsilon > 0$ so that

$$\epsilon < \frac{1}{\theta}\left(\frac{\lambda_1}{2} - \eta\right) . \tag{2.11}$$

By (a_1), there is $c_\epsilon > 0$ such that

$$|F(x,s,t)| + |G(x,s,t)| \le \epsilon(|s|^2 + |t|^2) + c_\epsilon$$

for all $(x,s,t) \in \Omega \times \mathbf{R}^2$. Then, for each $u, v \in H_0^1(\Omega)$, recalling (2.9), we have

$$\begin{aligned}
&I(u,v) + \langle \varphi(u,v),(\alpha,\beta)\rangle_Y \\
&\ge \left(\frac{1}{2} - \frac{\eta}{\lambda_1}\right)\int_\Omega (|\nabla u(x)|^2 + |\nabla v(x)|^2)dx \\
&\quad - \int_\Omega |\alpha(x)F(x,u(x),v(x)) + \beta(x)G(x,u(x),v(x))|dx \\
&\ge \left(\frac{1}{2} - \frac{\eta}{\lambda_1}\right)\int_\Omega (|\nabla u(x)|^2 + |\nabla v(x)|^2)dx - \theta\epsilon\int_\Omega (|u(x)|^2 + |v(x)|^2)dx - \theta c_\epsilon \mathrm{meas}(\Omega) \\
&\ge \left(\frac{1}{2} - \frac{\eta}{\lambda_1} - \frac{\theta\epsilon}{\lambda_1}\right)\int_\Omega (|\nabla u(x)|^2 + |\nabla v(x)|^2)dx - \theta c_\epsilon \mathrm{meas}(\Omega) .
\end{aligned}$$

Notice that, in view of (2.11), we have $\frac{1}{2} - \frac{\eta}{\lambda_1} - \frac{\theta\epsilon}{\lambda_1} > 0$, and so

$$\lim_{\|(u,v)\|_X \to +\infty} (I(u,v) + \langle \varphi(u,v),(\alpha,\beta)\rangle_Y) = +\infty ,$$

as claimed.
In particular, this also implies that the functional $(u,v) \to I(u,v) + \langle \varphi(u,v),(\alpha,\beta)\rangle_Y$ is weakly lower semicontinuous, by the Eberlein-Smulyan theorem. Thus, the assumptions of Theorem 2.1 are satisfied. Therefore, for each convex set $S \subseteq L^\infty(\Omega) \times L^\infty(\Omega)$ dense in $H_0^1(\Omega) \times H_0^1(\Omega)$, there exists $(\alpha,\beta) \in S$, such that the functional

$$\begin{aligned}
&(u,v) \to \frac{1}{2}\left(\int_\Omega |\nabla u(x)|^2 dx + \int_\Omega |\nabla v(x)|^2 dx\right) \\
&- \int_\Omega (\alpha(x)F(x,u(x),v(x)) + \beta(x)G(x,u(x),v(x)) + K(x,u(x),v(x)))dx
\end{aligned}$$

has at least two global minima in $H_0^1(\Omega) \times H_0^1(\Omega)$. Finally, by Example 38.25 of [17], the same functional satisfies the Palais-Smale condition, and so it admits at least three critical points, in view of Corollary 1 of [3]. The proof is complete. □

Remark 2.3. We are not aware of known results close enough to Theorem 2.2 in order to do a proper comparison. This sentence also applies to the case of single equations, that is to say when F, G, K depend on x and s only. For an account on elliptic systems, we refer to [2].

Among the various corollaries of Theorem 2.2, we wish to stress the following ones:

Corollary 2.4. *Let $K \in \mathcal{A}$, with $K(x,0,0) = 0$ for all $x \in \Omega$, satisfy condition (a_2). Moreover, let $\Phi : \mathbf{R}^2 \to \mathbf{R}$ be a non-constant C^1 function, with $\Phi(0,0) = 0$, belonging to $\mathcal{A}$, with $p = \frac{2}{n-2}$ when $n > 2$.*

Then, for every convex set $S \subseteq L^\infty(\Omega) \times L^\infty(\Omega)$ dense in $L^2(\Omega) \times L^2(\Omega)$, there exists $(\alpha, \beta) \in S$ such that the problem

$$-\Delta u = (\alpha(x)\cos(\Phi(u,v)) - \beta(x)\sin(\Phi(u,v)))\Phi_u(u,v) + K_u(x,u,v) \text{ in } \Omega$$
$$-\Delta v = (\alpha(x)\cos(\Phi(u,v)) - \beta(x)\sin(\Phi(u,v)))\Phi_v(u,v) + K_v(x,u,v) \text{ in } \Omega$$
$$u = v = 0 \text{ on } \partial\Omega$$

has at least three weak solutions, two of which are global minima in $H_0^1(\Omega) \times H_0^1(\Omega)$ of the functional

$$(u,v) \to \frac{1}{2}\left(\int_\Omega |\nabla u(x)|^2 dx + \int_\Omega |\nabla v(x)|^2 dx\right)$$
$$- \int_\Omega (\alpha(x)\sin(\Phi(u(x),v(x))) + \beta(x)\cos(\Phi(u(x),v(x))) + K(x,u(x),v(x)))dx \ .$$

Proof. It suffices to apply Theorem 2.2 to the functions $F, G : \mathbf{R}^2 \to \mathbf{R}$ defined by

$$F(s,t) = \sin(\Phi(s,t)) \ ,$$
$$G(s,t) = \cos(\Phi(s,t))$$

for all $(s,t) \in \mathbf{R}^2$. □

Corollary 2.5. *Let $F, G : \mathbf{R} \to \mathbf{R}$ belong to $\mathcal{A}$, with $p = \frac{2}{n-2}$ when $n > 2$. Moreover, assume that F, G are twice differentiable at 0 and that*

$$\lim_{|s|\to+\infty} \frac{|F(s)| + |G(s)|}{s^2} = 0 \ ,$$
$$0 < |F(0)|^2 + |G(0)|^2 = \inf_{s\in\mathbf{R}} (|F(s)|^2 + |G(s)|^2) \ ,$$
$$F''(0)F(0) + G''(0)G(0) < 0 \ . \tag{2.12}$$

Then, for every convex set $S \subseteq L^\infty(\Omega) \times L^\infty(\Omega)$ dense in $L^2(\Omega) \times L^2(\Omega)$, there exists $(\alpha, \beta) \in S$ such that the problem

$$-\Delta u = \alpha(x)F'(u) + \beta(x)G'(u) \text{ in } \Omega$$
$$u = 0 \text{ on } \partial\Omega$$

has at least three weak solutions, two of which are global minima in $H_0^1(\Omega)$ of the functional

$$u \to \frac{1}{2}\int_\Omega |\nabla u(x)|^2 dx - \int_\Omega (\alpha(x)F(u(x)) + \beta(x)G(u(x)))dx \ .$$

Proof. We apply Theorem 2.2 taking $K = 0$. Since 0 is a global minimum of the function $|F(\cdot)|^2 + |G(\cdot)|^2$, we have

$$F'(0)F(0) + G'(0)G(0) = 0$$

and so, in view of (2.12), 0 is a strict local maximum for the function

$$F(\cdot)F(0) + G(\cdot)G(0).$$

Hence, (a_4) is satisfied and Theorem 2.2 gives the conclusion. □

References

[1] Cabada, A., Iannizzotto, A., *A note on a question of Ricceri*, Appl. Math. Lett., **25**(2012), 215-219.

[2] de Figueiredo, D.G., *Semilinear elliptic systems: existence, multiplicity, symmetry of solutions*, Handbook of Differential Equations: Stationary Partial Differential Equations, Vol. V, 1-48, Handb. Differ. Equ., Elsevier/North-Holland, 2008.

[3] Pucci, P., Serrin, J., *A mountain pass theorem*, J. Differential Equations, **60**(1985), 142-149.

[4] Ricceri, B., *On a three critical points theorem*, Arch. Math., **75**(2000), 220-226.

[5] Ricceri, B., *Well-posedness of constrained minimization problems via saddle-points*, J. Global Optim., **40**(2008), 389-397.

[6] Ricceri, B., *Multiplicity of global minima for parametrized functions*, Rend. Lincei Mat. Appl., **21**(2010), 47-57.

[7] Ricceri, B., *Nonlinear eigenvalue problems*, in "Handbook of Nonconvex Analysis and Applications" D. Y. Gao and D. Motreanu eds., 543-595, International Press, 2010.

[8] Ricceri, B., *A strict minimax inequality criterion and some of its consequences*, Positivity, **16**(2012), 455-470.

[9] Ricceri, B., *A range property related to non-expansive operators*, Mathematika, **60**(2014), 232-236.

[10] Ricceri, B., *Singular points of non-monotone potential operators*, J. Nonlinear Convex Anal., **16**(2015), 1123-1129.

[11] Ricceri, B., *Energy functionals of Kirchhoff-type problems having multiple global minima*, Nonlinear Anal., **15**(2015), 130-136.

[12] Ricceri, B., *On a minimax theorem: an improvement, a new proof and an overview of its applications*, Minimax Theory Appl., **2**(2017), 99-152.

[13] Ricceri, B., *Another multiplicity result for the periodic solutions of certain systems*, Linear Nonlinear Anal., **5**(2019), 371-378.

[14] Ricceri, B., *Miscellaneous applications of certain minimax theorems II*, Acta Math. Vietnam., **45**(2020), 515-524.

[15] Ricceri, B., *An alternative theorem for gradient systems*, Pure Appl. Funct. Anal. (to appear).

[16] Ricceri, B., *A remark on variational inequalities in small balls*, J. Nonlinear Var. Anal., **4**(2020), 21-26.

[17] Zeidler, E., *Nonlinear functional analysis and its applications*, vol. III, Springer-Verlag, 1985.

Biagio Ricceri
Department of Mathematics and Informatics,
University of Catania,
Viale A. Doria 6,
95125 Catania, Italy
e-mail: `ricceri@dmi.unict.it`

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 85–94
DOI: 10.24193/subbmath.2021.1.07

Eigenvalues for anisotropic $p-$Laplacian under a Steklov-like boundary condition

Luminiţa Barbu

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. The eigenvalue problem

$$-\text{div}\,\Big(\frac{1}{p}\nabla_\xi\big(F^p(\nabla u)\big)\Big) = \lambda a(x) \mid u \mid^{q-2} u,$$

with $q \in (1,\infty)$, $p \in \left(\frac{Nq}{N+q-1},\infty\right)$, $p \neq q$, subject to Steklov-like boundary condition,

$$F^{p-1}(\nabla u)\nabla_\xi F(\nabla u)\cdot \nu = \lambda b(x) \mid u \mid^{q-2} u$$

is investigated on a bounded Lipschitz domain $\Omega \subset \mathbb{R}^N$, $N \geq 2$. Here, F stands for a $C^2(\mathbb{R}^N \setminus \{0\})$ norm and $a \in L^\infty(\Omega)$, $b \in L^\infty(\partial\Omega)$ are given nonnegative functions satisfying

$$\int_\Omega a\,dx + \int_{\partial\Omega} b\,d\sigma > 0.$$

Using appropriate variational methods, we are able to prove that the set of eigenvalues of this problem is the interval $[0,\infty)$.

Mathematics Subject Classification (2010): 35J60, 35J92, 35P30.

Keywords: Eigenvalues, anisotropic $p-$Laplacian, Steklov-like boundary condition, Sobolev spaces, variational methods.

1. Introduction

Let F be a norm in $\mathbb{R}^N$, that is a nonnegative, positively homogeneous of degree 1, convex function defined in $\mathbb{R}^N$. Moreover, we assume that $F \in C^2(\mathbb{R}^N \setminus \{0\})$.

Next, let us introduce the so-called anisotropic $p-$Laplacian, defined as follows

$$\mathcal{Q}_p u := \text{div}\,\Big(\frac{1}{p}\nabla_\xi\big(F^p(\nabla u)\big)\Big).$$

When $p = 2$, $\mathcal{Q}_2$ is the anisotropic operator, also known as the Finsler-Laplace operator [6]. We point out that a typical example of F satisfying the above conditions is the l_r-norm

$$F(\xi) := \Big(\sum_{i=1}^{N} \mid \xi_i \mid^r \Big)^{1/r}, \ r > 1,$$

for which the operator $\mathcal{Q}_p$ has the form

$$\Delta_{r,p} u := \text{div } \big(\parallel \nabla u \parallel_r^{p-r} \nabla^r u \big),$$

where

$$\nabla^r u := \left(\Big| \frac{\partial u}{\partial x_1} \Big|^{r-2} \frac{\partial u}{\partial x_1}, \cdots, \Big| \frac{\partial u}{\partial x_N} \Big|^{r-2} \frac{\partial u}{\partial x_N} \right).$$

Note that $\Delta_{r,p}$ is a nonlinear operator unless $p = r = 2$ when it reduces to the usual Laplacian operator. Two important special cases are $r = 2$ and $p \in (1, \infty)$ when $\Delta_{2,p}$ coincides with the usual p-Laplace operator (see [12]) and the case $r = p$, when $\Delta_{p,p}$ is the so-called pseudo p-Laplacian. A physical motivation to study differential equations involving such operators is given by the fact that they appear in well-established models of surface energies in metallurgy, crystallography, crystalline fracture theory, or noise-removal procedures in digital image processing (see for instance, [9], [15], and references therein). Meanwhile, a geometric motivation for the investigation of such operators comes from the fact that such anisotropies appears naturally in the Finsler geometry, such as, for instance, the Minkowski geometry (see the seminal works of P. Finsler [7] and H. Minkowski [13]).

The paper concerns the study of the following Steklov-like eigenvalue problem for $\mathcal{Q}_p$:

$$\begin{cases} -\mathcal{Q}_p u := -\text{div } \Big(\frac{1}{p} \nabla_\xi \big(F^p (\nabla u) \big) = \lambda a(x) \mid u \mid^{q-2} u \ \text{ in } \Omega, \\ F^{p-1}(\nabla u) \nabla_\xi F(\nabla u) \cdot \nu = \lambda b(x) \mid u \mid^{q-2} u \text{ on } \partial\Omega, \end{cases} \tag{1.1}$$

under the following hypotheses

(H_{pq}) $q \in (1, \infty), \ p \in \Big(\frac{Nq}{N+q-1}, \infty \Big), \ p \neq q$;

(H_Ω) $\Omega \subset \mathbb{R}^N, \ N \geq 2$, is a bounded domain with Lipschitz continuous boundary $\partial\Omega$;

(H_{ab}) $a, b \in L^\infty(\Omega)$ are given nonnegative functions satisfying

$$\int_\Omega a \ dx + \int_{\partial\Omega} b \ d\sigma > 0. \tag{1.2}$$

In $(1.1)_2$, ν stands for the outward unit normal to $\partial\Omega$.

The solution u of (1.1) is understood in a weak sense, as an element of the Sobolev space $W^{1,p}(\Omega)$ satisfying equation $(1.1)_1$ in the sense of distributions and boundary condition $(1.1)_2$ in the sense of traces:

Definition 1.1. $\lambda \in \mathbb{R}$ is an eigenvalue of problem (1.1) if there exists $u_\lambda \in W^{1,p} \setminus \{0\}$ such that for all $w \in W^{1,p}(\Omega)$

$$\begin{aligned} &\int_\Omega \Big(F(\nabla u_\lambda)\Big)^{p-1} \nabla_\xi F(\nabla u_\lambda) \cdot \nabla w \; dx \\ &= \lambda \Big(\int_\Omega a \mid u_\lambda \mid^{q-2} u_\lambda w \; dx + \int_{\partial\Omega} b \mid u_\lambda \mid^{q-2} u_\lambda w \; d\sigma \Big). \end{aligned} \tag{1.3}$$

Indeed, according to a Green type formula (see [4], p. 71), $u \in W^{1,p}(\Omega)$ is a solution of (1.1) if and only if it satisfies (1.3).
Our goal is to determine the set of all eigenvalues of problem (1.1). Fortunately we are able to offer a complete description of this set.

The main result of our paper is given by the following theorem

Theorem 1.2. *Assume that* (H_{pq}), (H_Ω) *and* (H_{ab}) *above are fulfilled. Then the set of eigenvalues of problem* (1.1) *is* $[0, \infty)$.

It is worth pointing out that this nice result is due to the fact that operator $\mathcal{Q}_p$ is nonhomogeneous ($p \neq q$). The homogeneous case ($p = q$) is more delicate. For example, if $p = q$ and either $a \equiv 1,\ b \equiv 0$ or $a \equiv 0,\ b \equiv 1$ and F is the usual euclidian norm, then the eigenvalue set of the corresponding (Neumann type) problem is fully known only if $p = q = 2$; otherwise, i.e. if $p = q \in (1, \infty) \setminus \{2\}$, then it is only known that, as a consequence of the Ljusternik-Schnirelman theory, there exists a sequence of positive eigenvalues of problem (1.1) with $Q = -\Delta_p$ (see, e.g., [11]), but this sequence may not constitute the whole eigenvalue set.

Regarding the assumption $p \in \left(\frac{Nq}{N+q-1}, \infty\right)$ we point out that this is directly related to the well-known embeddings $W^{1,p}(\Omega) \hookrightarrow L^q(\Omega)$ which hold in the cases: (1) $1 \le q \le p^* = pN/(N-p)$, if $1 < p < N$; (2) $p \le q < \infty$, if $p = N$; (3) $q = \infty$, if $p > N$. Moreover, these embeddings are compact when $1 \le q < p^*$ in case (1), all q in case (2), and when reinterpreted as $W^{1,p}(\Omega) \hookrightarrow C^1(\overline{\Omega})$ in case (3). We also have trace compactly embeddings $W^{1,p}(\Omega) \hookrightarrow L^q(\partial\Omega)$ for all $1 \le p \le q < p(N-1)/(N-p)$ if $1 \le p < N$, and similarly as before in the other ranges of p (see [1], [3, Section 9.3]).

Also, we restrict ourselves to functions $a \in L^\infty(\Omega),\ b \in L^\infty(\partial\Omega)$ since assuming weaker regularity for these functions leads to similar results without essential changes.

The Dirichlet eigenvalue problem associated with operator $-\mathcal{Q}_p$ for $q = 2$ has been studied in [5]. As far as the problem (1.1) is concerned, a separate analysis is needed since some specific situations have to be addressed, including those related to the trace on $\partial\Omega$ and the fact that the eigenfunctions of our problem belong to the set $\mathcal{C}$ (see Section 2, (2.2) for the definition of $\mathcal{C}$). It is worth pointing out that results concerning the existence and nonexistence of solutions for the case of $p-$Laplacian under Dirichlet boundary conditions and appropriate assumptions on Ω have been obtained by M. Ôtani in the well known paper [14].

2. Preliminary results

Our hypotheses (H_{pq}), (H_Ω), (H_{ab}) will be assumed throughout this paper. Testing equation (1.3) against $w = u_\lambda$ we observe that the eigenvalues of problem (1.1) cannot be negative numbers. It is also obvious that $\lambda_0 = 0$ is an eigenvalue of this problem and the corresponding eigenfunctions are the nonzero constant functions. So any other eigenvalue belongs to $(0, \infty)$.

If we assume that $\lambda > 0$ is an eigenvalue of problem (1.1) and choose $w \equiv 1$ in (1.3) we deduce that every eigenfunction u_λ corresponding to λ satisfies the equation

$$\int_\Omega a \mid u_\lambda \mid^{q-2} u_\lambda \, dx + \int_{\partial\Omega} b \mid u_\lambda \mid^{q-2} u_\lambda \, d\sigma = 0. \tag{2.1}$$

So all eigenfunctions corresponding to positive eigenvalues necessarily belong to the set

$$\mathcal{C} := \Big\{ u \in W^{1,p}(\Omega); \int_\Omega a \mid u \mid^{q-2} u \, dx + \int_{\partial\Omega} b \mid u \mid^{q-2} u \, d\sigma = 0 \Big\}. \tag{2.2}$$

This is a symmetric cone and we can see that $\mathcal{C}$ is a weakly closed subset of $W^{1,p}(\Omega)$. Indeed, let $\big(u_n\big)_n \subset \mathcal{C}$ such that $u_n \rightharpoonup u_0$ in $W^{1,p}(\Omega)$. From assumption (H_{pq}), $W^{1,p}(\Omega) \hookrightarrow L^q(\Omega)$ and $W^{1,p}(\Omega) \hookrightarrow L^q(\partial\Omega)$ compactly, hence there exists a subsequence of $\big(u_n\big)_n$, which is also denoted $\big(u_n\big)_n$, such that

$$u_n \to u_0 \text{ in } L^q(\Omega), \; u_n \to u_0 \text{ in } L^q(\partial\Omega).$$

By Lebesgue's Dominated Convergence Theorem (see also [3, Theorem 4.9]) we obtain $u_0 \in \mathcal{C}$.
In addition, $\mathcal{C}$ has nonzero elements (see [2, Section 2]).
Now let us define the positively homogeneous of order p functional

$$J : W^{1,p}(\Omega) \to \mathbb{R}, \; J(w) := \int_\Omega \big(F(\nabla w)\big)^p \, dx \; \forall \, w \in W. \tag{2.3}$$

Standard arguments can be used in order to deduce that functional J is convex and weakly lower semicontinuous (see, for instance [16, Proposition 25.20]).
Consider the minimization problem

$$\mu := \inf_{w \in \mathcal{C}_1} J(w)\,, \tag{2.4}$$

where

$$\mathcal{C}_1 := \mathcal{C} \cap \Big\{ u \in W^{1,p}(\Omega); \int_\Omega a \mid u \mid^q \; dx + \int_{\partial\Omega} b \mid u \mid^q \; d\sigma = 1 \Big\}.$$

The next result states that J attains its minimal value and this value is positive.

Lemma 2.1. *For each $p > 1$ there exists $u_* \in \mathcal{C}_1$ such that*

$$\mu := J(u_*) = \inf_{w \in \mathcal{C}_1} J(w) > 0.$$

Proof. Let $\big(u_n\big)_n \subset \mathcal{C}_1$ be a minimizing sequence for J, i. e.,

$$J(u_n) \to \inf_{w \in \mathcal{C}_{1q}} J(w) := \mu.$$

We can prove that $\big(u_n\big)_n$ is bounded in $W^{1,p}(\Omega)$. Assume the contrary, that there exists a subsequence of $\big(u_n\big)_n$, again denoted $\big(u_n\big)_n$, such that $\| \, u_n \, \|_{W^{1,p}(\Omega)} \to \infty$ as $n \to \infty$. Define

$$v_n = \frac{u_n}{\| \, u_n \, \|_{W^{1,p}(\Omega)}} \quad \forall \, n \in \mathbb{N}\,.$$

Clearly sequence $\big(v_n\big)_n$ is bounded in $W^{1,p}(\Omega)$ so there exist a $v \in W^{1,p}(\Omega)$ and a subsequence of $\big(v_n\big)_n$, again denoted $\big(v_n\big)_n$, such that

$$v_n \rightharpoonup v \text{ in } W^{1,p}(\Omega).$$

Taking into account assumption (H_{pq}) we obtain that $W^{1,p}(\Omega) \hookrightarrow L^q(\Omega)$ and $W^{1,p}(\Omega) \hookrightarrow L^q(\partial\Omega)$ compactly, therefore, up to a subsequence, we have

$$v_n \to v \text{ in } L^q(\Omega),\ v_n \to v \text{ in } L^q(\partial\Omega).$$

As $\| v_n \|_{W^{1,p}(\Omega)} = 1\ \forall\ n \in \mathbb{N}$ we have $\| v \|_{W^{1,p}(\Omega)} = 1$, and

$$\begin{aligned}\int_\Omega \big(F(\nabla v)\big)^p\ dx &\le \liminf_{n\to\infty} \int_\Omega \big(F(\nabla v_n)\big)^p\ dx \\ &= \liminf_{n\to\infty} \frac{1}{\| u_n \|^p_{W^{1,p}(\Omega)}} J(u_n) = 0,\end{aligned}$$

which shows that v is a constant function. On the other hand, since $\big(v_n\big)_n \subset \mathcal{C}$ and $\mathcal{C}$ is weakly closed in $W^{1,p}(\Omega)$, we infer that $v \in \mathcal{C}$, hence $v \equiv 0$. But this contradicts the fact that $\| v \|_{W^{1,p}(\Omega)} = 1$. Therefore, $\big(u_n\big)_n$ is indeed bounded in $W^{1,p}(\Omega)$, thus, by passing to a subsequence, we can assume that $\big(u_n\big)_n$ converges weakly to a function $u_* \in W^{1,p}(\Omega)$ and

$$u_n \to u_* \text{ in } L^q(\Omega),\ u_n \to u_* \text{ in } L^q(\partial\Omega).$$

By Lebesgue's Dominated Convergence Theorem we obtain $u_* \in \mathcal{C}_1$, so the weak lower semicontinuity of J leads to $\mu = J(u_*)$. In addition, $J(u_*) > 0$. Indeed, assuming by contradiction that $J(u_*) = 0$ would imply that $u_* \equiv Const.$, which is impossible because $u_* \in \mathcal{C}_1$. □

3. Proof of the main result

The following lemma plays a crucial role in the proof of our main theorem

Lemma 3.1. *Assume that* (H_{pq}), (H_Ω) *and* (H_{ab}) *above are fulfilled. Let* $u_* \in W^{1,p}(\Omega)$ *be a minimizer of the functional* J *defined by* (2.3) *on the set*

$$\mathcal{C}_1 := \mathcal{C} \cap \Big\{u \in W^{1,p}(\Omega);\ \int_\Omega a \mid u \mid^q\ dx + \int_{\partial\Omega} b \mid u \mid^q\ d\sigma = 1\Big\}.$$

Then u_* *is an eigenfunction of problem* (1.1) *with eigenvalue* $\mu = \inf\limits_{w\in\mathcal{C}_1} J(w)$.

Proof. Since the constraint $\mathcal{C}_1$ is no more a C^1 manifold if $q < 2$, we can not use a reasoning based on Lagrange Multipliers Rule. In order to avoid this inconvenience let us define the functional

$$\begin{aligned} J_\mu : W^{1,p}(\Omega) \to \mathbb{R},\ J_\mu(u) = &\int_\Omega \big(F(\nabla u)\big)^p\ dx \\ &- \mu\Big(\int_\Omega a \mid u \mid^q\ dx + \int_{\partial\Omega} b \mid u \mid^q\ d\sigma\Big)^{\frac{p}{q}}\ \forall\ u \in W^{1,p}(\Omega). \end{aligned} \tag{3.1}$$

Standard arguments can be used in order to deduce that $J_\mu \in C^1(W^{1,p}(\Omega);\mathbb{R})$, with the derivative given by

$$\begin{aligned}\langle J'_\mu(u), w\rangle = p\int_\Omega \big(F(\nabla u)\big)^{p-1}\nabla_\xi F(\nabla u)\cdot\nabla w\; dx \\ -\mu p\Big(\int_\Omega a\mid u\mid^q\; dx + \int_{\partial\Omega} b\mid u\mid^q\; d\sigma\Big)^{\frac{p}{q}-1} \\ \cdot\Big(\int_\Omega a\mid u\mid^{q-2}uw\; dx + \int_{\partial\Omega} b\mid u\mid^{q-2}uw\; d\sigma\Big)\end{aligned} \tag{3.2}$$

for all $u, w \in W^{1,p}(\Omega)$.

It is obviously that u_* is an eigenfunction of problem (1.1) with eigenvalue μ if and only if u_* is a critical point of J_μ, i. e. $J'_\mu(u_*) = 0$. In order to show this, we fix $v \in \mathrm{Lip}(\Omega)$ arbitrarily. For each $n \in \mathbb{N}^*$ define $f_n : \mathbb{R} \to \mathbb{R}$,

$$f_n(s) = \int_\Omega a\Big|u_* + \frac{1}{n}v + s\Big|^q\; dx + \int_{\partial\Omega} b\Big|u_* + \frac{1}{n}v + s\Big|^q\; d\sigma\ \forall\ s \in \mathbb{R}. \tag{3.3}$$

It is easily seen that f_n is coercive, since we have

$$\begin{aligned}f_n(s) \geq 2^{-q}\mid s\mid^q\big(\parallel a\parallel_{L^\infty(\Omega)}\mid\Omega\mid_N + \parallel b\parallel_{L^\infty(\partial\Omega)}\mid\partial\Omega\mid_{N-1}\big) \\ -\int_\Omega a\Big|u_* + \frac{1}{n}v\Big|^q\; dx - \int_{\partial\Omega} b\Big|u_* + \frac{1}{n}v\Big|^q\; d\sigma,\end{aligned}$$

where $|\cdot|_N$ and $|\cdot|_{N-1}$ denote the Lebesgue measures of the two sets. We have also used the inequality

$$\mid x\mid^q \leq (\mid x + y\mid + \mid y\mid)^q \leq 2^q(\mid x + y\mid^q + \mid y\mid^q)\ \forall\ x, y \in \mathbb{R},\ q > 1.$$

Moreover, function f_n is continuous differentiable on $\mathbb{R}$ (see [8, Theorem 2.27]) and convex (its derivative is an increasing function). Therefore, for all $n \in \mathbb{N}^*$, f_n admits a minimum point s_n, such that $f'_n(s_n) = 0$, that is

$$\begin{aligned}\int_\Omega a\Big|u_* + \frac{1}{n}v + s_n\Big|^{q-2}\Big(u_* + \frac{1}{n}v + s_n\Big)\; dx \\ + \int_{\partial\Omega} b\Big|u_* + \frac{1}{n}v + s_n\Big|^{q-2}\Big(u_* + \frac{1}{n}v + s_n\Big)\; d\sigma = 0.\end{aligned} \tag{3.4}$$

We denote

$$u_n := u_* + 1/n\; v + s_n\ \forall\ n \in \mathbb{N}^*. \tag{3.5}$$

From (3.4) we derive that $\big(u_n\big)_n \subset \mathcal{C}$.

Next, we claim that the sequence $\big(ns_n\big)_n$ is bounded. Arguing by contradiction, let us assume that, up to a sequence, $ns_n \to \infty$ or $ns_n \to -\infty$ as $n \to \infty$. Taking into account that $v \in \mathrm{Lip}(\Omega)$ there exists N_1 large enough such that we have either

$$v(\cdot) + ns_n > 0 \text{ in } \Omega,\ \text{ or } v(\cdot) + ns_n < 0 \text{ in } \Omega\ \forall\ n \geq N_1.$$

Since the function $\gamma \to \mid u^* + \gamma \mid^{q-2} (u^* + \gamma)$ is strictly increasing on $\mathbb{R}$, we get

$$\begin{aligned} 0 &= \int_\Omega a \mid u_n \mid^{q-2} u_n \, dx + \int_{\partial\Omega} b \mid u_n \mid^{q-2} u_n \, d\sigma \\ &> \int_\Omega a \mid u^* \mid^{q-2} u^* \, dx + \int_{\partial\Omega} b \mid u^* \mid^{q-2} u^* \, d\sigma = 0 \ \forall n \geq N_1, \end{aligned} \tag{3.6}$$

if $v(\cdot) + ns_n > 0$ in Ω, or the reverse inequality in the second situation, when

$$v(\cdot) + ns_n < 0 \text{ in } \Omega.$$

In both cases we get a contradiction.

We point out that inequality in relation (3.6) is strictly. Indeed, we note that (1.2) implies that either $|\{x \in \Omega;\ a(x) > 0\}|_N > 0$ or $a = 0$ a.e. in Ω and

$$|\{x \in \partial\Omega;\ b(x) > 0\}|_{N-1} > 0,$$

hence we can not have equality between the two terms containing integrals.

Consequently, $\big(ns_n\big)_n$ should be bounded. This in turn implies there exists $S \in \mathbb{R}$ such that, up to a subsequence, $ns_n \to S$ as $n \to \infty$.

We note that the subsequence of $\big(u_n\big)_n$, denoted $\big(u_n\big)_n$ again, with the property that $\big(ns_n\big)_n$ has the limit S, converges in $W^{1,p}(\Omega)$, more exactly,

$$u_n \to u_* \text{ and } n\big(u_n - u_*\big) \to v + S \text{ in } W^{1,p}(\Omega) \text{ as } n \to \infty. \tag{3.7}$$

We also note that from (3.7), combining with $u_* \not\equiv 0$, there exists N_2 large enough, such that $\big(u_n\big)_n \subset \mathcal{C} \setminus \{0\}\ \forall\ n \geq N_2$. Next, using this subsequence, we are going to construct a minimizing sequence for J_μ restricted to the constraint set $\mathcal{C}_1$. In this respect, we can define

$$t_n := \big(\parallel a^{1/q} u_n \parallel^q_{L^q(\Omega)} + \parallel b^{1/q} u_n \parallel^q_{L^q(\partial\Omega)} \big)^{1/q},\ z_n := \frac{u_n}{t_n}, \tag{3.8}$$

for all n sufficiently large. Obviously, we have

$$\begin{gathered} t_n \to \int_\Omega a \mid u_* \mid^q \ dx + \int_{\partial\Omega} b \mid u_* \mid^q \ d\sigma = 1, \\ \big(z_n\big)_n \subset \mathcal{C}_1,\ z_n \to u_* \text{ in } W^{1,p}(\Omega) \text{ as } n \to \infty. \end{gathered} \tag{3.9}$$

Next, we claim that sequence $\big(n(t_n - 1)\big)_n$ is bounded. In order to proof this, we first show that $\big(n(t_n^{1/q} - 1)\big)_n$ is bounded. To this aim, we define the functional

$$\mathcal{I}_q : W^{1,p}(\Omega) \to \mathbb{R},\ \mathcal{I}_q(u) := \int_\Omega a \mid u \mid^q \ dx + \int_{\partial\Omega} b \mid u \mid^q \ d\sigma \ \forall\ u \in W^{1,p}(\Omega).$$

Under assumption (H_{pq}), it is known that $\mathcal{I}_q \in C^1\big(W^{1,p}(\Omega); \mathbb{R}\big)$ (see, for instance [11]) and for all $u, w \in W^{1,p}(\Omega)$,

$$\langle \mathcal{I}_q'(u), w \rangle = q\Big(\int_\Omega a \mid u \mid^{q-2} uw \ dx + \int_{\partial\Omega} b \mid u \mid^{q-2} uw \ d\sigma \Big). \tag{3.10}$$

Since $\mathcal{I}_q(u_*) = 1$, we note that for all $n \in \mathbb{N}^*$,

$$n\big(t_n^{1/q} - 1\big) = \frac{\mathcal{I}_q(u_n) - \mathcal{I}_q(u_*)}{\frac{1}{n}}. \tag{3.11}$$

Now, taking into account that $\mathcal{I}_q' \in \big(W^{1,p}(\Omega)\big)^*$, we get

$$\begin{aligned}\lim_{n\to\infty} n(t_n^{1/q}-1) &= \lim_{n\to\infty} n\big(\mathcal{I}_q(u_n)-\mathcal{I}_q(u_*)\big)\\ &= \lim_{n\to\infty}\langle \mathcal{I}_q'(u_*), n(u_n-u_*)\rangle + o(n;u_*,v) \\ &= \langle \mathcal{I}_q'(u_*), v+S\rangle = \langle \mathcal{I}_q'(u_*), v\rangle,\end{aligned} \tag{3.12}$$

where $o(n;u_*,v)$ is a notation for the term which tends to zero in the definition of the Fréchet differential of $\mathcal{I}_q$ at u_*, that is $o(n,u_*,v)\to 0$ as $n\to\infty$. Therefore, there exists $K>0$ such that $n\mid t_n^{1/q}-1\mid\leq K$, or equivalently

$$0<1-\frac{K}{n}\leq t_n^{1/q}\leq 1+\frac{K}{n},$$

for all $n\in\mathbb{N}^*$, n large enough, which implies

$$n\left(\left(1-\frac{K}{n}\right)^q-1\right)\leq n(t_n-1)\leq n\left(\left(1+\frac{K}{n}\right)^q-1\right), \tag{3.13}$$

for all n sufficiently large. It is elementary to check that

$$\lim_{x\to 0_+}\frac{(1+Kx)^q-1}{x}=qK,\ \lim_{x\to 0_+}\frac{(1-Kx)^q-1}{x}=-qK.$$

This in combination with (3.13) implies that the sequence $\big(n(t_n-1)\big)_n$ is bounded, thus, by possibly passing to a subsequence, there exists $T\in\mathbb{R}$, such that $n(t_n-1)\to T$ as $n\to\infty$.

Now, it is easy to observe that u_* minimizes functional J_μ over $\mathcal{C}_1$. By using the minimality of u_* and the fact that $\big(z_n\big)_n\subset\mathcal{C}_1$ we obtain that

$$0\leq\lim_{n\to\infty}\frac{J_\mu(z_n)-J_\mu(u_*)}{\frac{1}{n}}. \tag{3.14}$$

Since functional $J_\mu\in C^1(W^{1,p}(\Omega);\mathbb{R})$, we have

$$n\big(J_\mu(z_n)-J_\mu(u_*)\big)=\big(\langle J_\mu'(u_*), n(z_n-u_*)\rangle + o(n;u_*,v), \tag{3.15}$$

with $o(n;u_*,v)\to 0$ as $n\to\infty$. Taking into account (3.5) and (3.8) we can see that

$$n(z_n-u_*)=\frac{1}{t_n}\Big(nu_*\big(1-t_n\big)+v+ns_n\Big)\to -Tu_*+v+S \text{ as } n\to\infty. \tag{3.16}$$

It follows from (3.14)-(3.16) that

$$0\leq\langle J_\mu'(u_*), v+S-Tu_*\rangle. \tag{3.17}$$

From (3.2), Lemma 2.1, and $u_*\in\mathcal{C}_1$, we get that $\langle J_\mu'(u_*),u_*\rangle=0,\ \langle J_\mu'(u_*),S\rangle=0$, hence (3.17) implies

$$0\leq\langle J_\mu'(u_*),v\rangle.$$

A similar reasoning with $-v$ instead of v shows that $0=\langle J_\mu'(u_*),v\rangle$.

The conclusion then follows by exploiting the density of Lipschitz functions in $W^{1,p}(\Omega)$ which is true according to assumption (H_Ω) (see [10, Theorm 3.6]. □

Proof of Theorem 1.2. By Lemma 3.1, there exists an eigenfunction u_* of problem (1.1) corresponding to eigenvalue $\mu = \inf_{w \in \mathcal{C}_1} J(w) > 0$, thus

$$\begin{aligned} &\int_\Omega \Big(F(\nabla u_*)\Big)^{p-1} \nabla_\xi F(\nabla u_*) \cdot \nabla w \; dx \\ &= \mu\Big(\int_\Omega a \mid u_* \mid^{q-2} u_* w \; dx + \int_{\partial\Omega} b \mid u_* \mid^{q-2} u_* w \; d\sigma\Big) \end{aligned} \tag{3.18}$$

for all $w \in W^{1,p}(\Omega)$.
Consider $\lambda > 0$ fixed. Let $\tau > 0$. If we take u_* of the form $u_* = \tau v_*$ in (3.18) and taking into account that F and $\nabla_\xi F$ are positively homogeneous of degree 1 and 0, respectively, we derive

$$\begin{aligned} &\int_\Omega \Big(F(\nabla v_*)\Big)^{p-1} \nabla_\xi F(\nabla v_*) \cdot \nabla w \; dx \\ &= \tau^{q-p}\mu\Big(\int_\Omega a \mid v_* \mid^{q-2} v_* w \; dx + \int_{\partial\Omega} b \mid v_* \mid^{q-2} u_* w \; d\sigma\Big) \end{aligned} \tag{3.19}$$

for all $w \in W^{1,p}(\Omega)$.

Finally, if we choose $\tau = (\lambda/\mu)^{1/(q-p)} > 0$, then $v_* = \tau u_*$ is an eigenfunction of problem (1.1) with eigenvalue λ. As has already been pointed out, $\lambda = 0$ is an eigenvalue of problem (1.1). This conclude the proof.

□

References

[1] Adams, R.A., Fournier, J.J.F., *Sobolev Spaces*, second ed., Pure Appl. Math., **140**, Academic Press, New York-London, 2003.

[2] Barbu, L., Moroşanu, G., *Eigenvalues of the negative (p,q)- Laplacian under a Steklov-like boundary condition*, Complex Var. Elliptic Equ., **64**(2019), no. 4, 685-700.

[3] Brezis, H., *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer, 2011.

[4] Casas, E., Fernández, L.A., *A Green's formula for quasilinear elliptic operators*, J. Math. Anal. Appl., **142**(1989), 62-73.

[5] Fărcăşeanu, M., *Eigenvalues for Finsler p−Laplacian with zero Dirichlet boundary condition*, An. Sti. U. Ovid. Co.-Mat., **24**(2016), no. 1, 231-242.

[6] Ferone, V., Kawohl, B., *Remarks on Finsler-Laplacian*, Proc. Am. Math. Soc., **137** (2008), no. 1, 247-253.

[7] Finsler, P., *Über Kurven und Flachen in Allgemeinen Raumen*, (Dissertation, Gottingen, 1918), Birkhäuser Verlag, Basel, 1951.

[8] Folland, G.B., *Real Analysis: Modern Techniques and Their Applications* (2nd ed.), Pure and Applied Mathematics, John Wiley & Sons, Inc. New York, 1999.

[9] Giga, Y., *Surface Evolution Equations. A Level Set Approach*, Birkhäuser Verlag, Basel, 2006.

[10] Giusti, E., *Direct Methods in the Calculus of Variations*, **7**, World Scientific, Singapore, 2003.

[11] Lê, A., *Eigenvalue problems for p-Laplacian*, Nonlinear Anal. TMA., **64**(2006), 1057-1099.

[12] Lindqvist P., *Notes on p-Laplace equation*, University of Jyväskylä, Lectures notes, 2006.

[13] Minkowski, H., *Allgemeine Lehrsatze uber die konvexen Polyeder*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, 1897, 198-219.

[14] Ôtani, M., *Existence and nonexistence of nontrivial solutions of some nonlinear degenerate elliptic equations*, J. Funct. Anal., **76**(1988), 140-159.

[15] Taylor, J.E., *Crystalline variational problems*, Bulletin of the American Mathematical Society, **84**(1978), 568-588.

[16] Zeidler, E., *Nonlinear Functional Analysis and its Applications: II/B, Nonlinear Monotone Operators*, Springer, New York, 1989.

Luminiţa Barbu
"Ovidius" University,
Faculty of Mathematics and Computer Sciences
124 Mamaia Blvd, 900527 Constanţa, Romania
e-mail: `lbarbu@univ-ovidius.ro`

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 95–103
DOI: 10.24193/subbmath.2021.1.08

Nonstandard Dirichlet problems with competing (p, q)-Laplacian, convection, and convolution

Dumitru Motreanu and Viorica Venera Motreanu

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. The paper focuses on a nonstandard Dirichlet problem driven by the operator $-\Delta_p + \mu\Delta_q$, which is a competing (p, q)-Laplacian with lack of ellipticity if $\mu > 0$, and exhibiting a reaction term in the form of a convection (i.e., it depends on the solution and its gradient) composed with the convolution of the solution with an integrable function. We prove the existence of a generalized solution through a combination of fixed-point approach and approximation. In the case $\mu \leq 0$, we obtain the existence of a weak solution to the respective elliptic problem.

Mathematics Subject Classification (2010): 35J92, 47H30.

Keywords: Competing (p, q)-Laplacian, Dirichlet problem, convection, convolution, generalized solution, weak solution.

1. Introduction

In this paper we consider the following quasilinear problem with homogeneous Dirichlet boundary condition on a bounded domain $\Omega \subset \mathbb{R}^N$ with the boundary $\partial\Omega$,

$$\begin{cases} -\Delta_p u + \mu\Delta_q u = f(x, \rho * u, \nabla(\rho * u)) & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega \end{cases} \tag{1.1}$$

for $1 < q < p < +\infty$, $\mu \in \mathbb{R}$, and $\rho \in L^1(\mathbb{R}^N)$. To ease the exposition we assume $p < N$ mentioning that the complementary case $p \geq N$ can be handled along the same lines.

In order to simplify the notation, for any real number $r > 1$, we set $r' = r/(r-1)$ (the Hölder conjugate of r). In particular, we have $p' = p/(p-1) < q' = q/(q-1)$. In the left-hand side of equation (1.1) there are the negative p-Laplacian

$$-\Delta_p : W_0^{1,p}(\Omega) \to W^{-1,p'}(\Omega)$$

expressed as

$$\langle -\Delta_p u, v\rangle = \int_\Omega |\nabla u(x)|^{p-2}\nabla u(x)\cdot\nabla v(x)\,dx \quad \text{for all } u,v\in W_0^{1,p}(\Omega)$$

and the negative q-Laplacian $-\Delta_q : W_0^{1,q}(\Omega)\to W^{-1,q'}(\Omega)$ expressed as

$$\langle -\Delta_q u, v\rangle = \int_\Omega |\nabla u(x)|^{q-2}\nabla u(x)\cdot\nabla v(x)\,dx \quad \text{for all } u,v\in W_0^{1,q}(\Omega).$$

Hereafter, $|\cdot|$ stands for the Euclidean norm in $\mathbb{R}^N$. Since $1<q<p<+\infty$, it holds the continuous embedding $W_0^{1,p}(\Omega)\hookrightarrow W_0^{1,q}(\Omega)$, so the operator $-\Delta_p+\mu\Delta_q$ is well defined on $W_0^{1,p}(\Omega)$. In the sequel, p^* stands for the Sobolev critical exponent $p^*=Np/(N-p)$ (recall that we assume $p<N$).

The right-hand side of the equation in (1.1) is described by means of a Carathéodory function $f:\Omega\times\mathbb{R}\times\mathbb{R}^N\to\mathbb{R}$ (meaning that $f(\cdot,s,\xi)$ is measurable on Ω for all $(s,\xi)\in\mathbb{R}\times\mathbb{R}^N$ and $f(x,\cdot,\cdot)$ is continuous for a.e. $x\in\Omega$) which is composed with the convolution

$$\rho * u(x) = \int_{\mathbb{R}^N} \rho(x-y)u(y)\,dy \quad \text{for a.e. } x\in\mathbb{R}^N$$

of some $\rho\in L^1(\mathbb{R}^N)$ and $u\in W_0^{1,p}(\Omega)\subset W^{1,p}(\mathbb{R}^N)$. Notice that the convolution $\rho*u$ is well defined.

There are two noticeable aspects related to the right-hand side of the equation in (1.1). The first one is the fact that it exhibits dependence not only with respect to the solution u but also with respect to its gradient ∇u. Such a term is usually called convection and its presence prevents us to make use of variational methods. A systematic study of problems with convection can be found in [4]. A second significant feature related to the right-hand side of the equation in (1.1) is the fact that the convection is composed with a convolution which is nonlocal operator. The study of the problems involving the composition of convection and convolution has been started in [6], specifically for problem (1.1) with $\mu\le 0$. This study incorporates the case where the operator is the p-Laplacian $-\Delta_p$ (for $\mu=0$) and the ordinary (p,q)-Laplacian $-\Delta_p-\Delta_q$ (for $\mu=-1$). The investigation of a (nonsmooth) version of problem (1.1) for an arbitrary $\mu\in\mathbb{R}$, but without convection and convolution, was initiated in [3]. Problem (1.1) with the "competing" (p,q)-Laplacian $-\Delta_p+\Delta_q$ (i.e., in the case where $\mu=1$) and convection but without convolution was addressed in [5].

Let $\lambda_{1,p}>0$ denote the first eigenvalue of the negative p-Laplacian on $W_0^{1,p}(\Omega)$, which is given by the following variational characterization (see, e.g., [7, §9.2]),

$$\lambda_{1,p} = \min\left\{ \frac{\|\nabla u\|^p_{L^p(\Omega,\mathbb{R}^N)}}{\|u\|^p_{L^p(\Omega)}} : u\in W_0^{1,p}(\Omega)\setminus\{0\}\right\}. \tag{1.2}$$

We assume that the following growth condition for $f(x,s,\xi)$ is satisfied.

Assumption 1.1. There holds

$$|f(x,s,\xi)| \leq \sigma(x) + a_1|s|^{p-1} + a_2|\xi|^{p-1} \tag{1.3}$$

for a.e. $x \in \Omega$, all $s \in \mathbb{R}$, and $\xi \in \mathbb{R}^N$, with a function $\sigma \in L^{r'}(\Omega)$ where $r \in [1, p^*)$ and constants $a_1, a_2 \geq 0$ satisfying

$$\|\rho\|^{p-1}_{L^1(\mathbb{R}^N)}(a_1\lambda_{1,p}^{-1} + a_2 N^{p-1}\lambda_{1,p}^{-\frac{1}{p}}) < 1. \tag{1.4}$$

Remark 1.2. The condition (1.4) in Assumption 1.1 can be expressed by saying that the parameter $\rho \in L^1(\mathbb{R}^N)$ in problem (1.1) is small enough with respect to its L^1 norm.

Remark 1.3. (a) If the Carathéodory function f satisfies the growth condition

$$|f(x,s,\xi)| \leq \sigma(x) + a_1|s|^{p-1} + a_2|\xi|^{\beta}$$

as in (1.3) except that the exponent of $|\xi|$ is some $\beta \in [0, p-1)$, then Assumption 1.1 is fulfilled provided that

$$a_1\|\rho\|^{p-1}_{L^1(\mathbb{R}^N)} < \lambda_{1,p}.$$

(b) If f satisfies the stronger growth condition

$$|f(x,s,\xi)| \leq \sigma(x) + a_1|s|^{\alpha} + a_2|\xi|^{\beta}$$

with $\alpha, \beta \in [0, p-1)$, then Assumption 1.1 is fulfilled.

By a *generalized solution* to problem (1.1) we mean any function $u \in W_0^{1,p}(\Omega)$ for which there exists a sequence $\{u_n\}_{n\geq 1}$ in $W_0^{1,p}(\Omega)$ such that

(a) $u_n \rightharpoonup u$ in $W_0^{1,p}(\Omega)$ as $n \to \infty$;
(b) $-\Delta_p u_n + \mu\Delta_q u_n - f(\cdot, \rho * u_n(\cdot), \nabla(\rho * u_n)(\cdot)) \rightharpoonup 0$ in $W^{-1,p'}(\Omega)$ as $n \to \infty$;
(c) $\lim_{n\to\infty} \langle -\Delta_p u_n + \mu\Delta_q u_n, u_n - u\rangle = 0$.

The essential point in our work is that the driving operator $-\Delta_p + \mu\Delta_q$ in problem (1.1) has a fundamentally different behavior depending on whether $\mu \leq 0$ or $\mu > 0$. Indeed, in the latter case, the operator lacks the ellipticity: notice for instance that, for a nonzero $u_0 \in W_0^{1,p}(\Omega)$ and a number $\lambda > 0$, the quantity

$$\langle -\Delta_p(\lambda u_0) + \mu\Delta_q(\lambda u_0), \lambda u_0\rangle = \lambda^p\|\nabla u_0\|^p_{L^p(\Omega,\mathbb{R}^N)} - \lambda^q\mu\|\nabla u_0\|^q_{L^q(\Omega,\mathbb{R}^N)}$$

does not keep a constant sign if $\mu > 0$. It is positive for $\lambda > 0$ sufficiently large and it is negative for $\lambda > 0$ sufficiently small. In view of this, in [3], the operator $-\Delta_p + \mu\Delta_q$ for $\mu > 0$ was called a competing (p,q)-Laplacian. Due to the lack of ellipticity there is no available method to handle problem (1.1) for arbitrary μ. In order to bypass this drawback, the notion of generalized solution was introduced in [3] for a counterpart of problem (1.1) without convolution. Note that, in the case where $\mu \leq 0$, the notions of generalized solution and weak solution coincide (see Lemma 3.3). In Theorem 3.4, we prove the existence of a generalized solution to problem (1.1) for arbitrary μ. Our approach relies on a fixed-point theorem and approximation process. Our treatment of problem (1.1) is unified in the sense that it does not distinguish according to the sign of μ.

2. Preliminaries

In the sequel, the space $W_0^{1,p}(\Omega)$ is considered endowed with the norm $\|\nabla(\cdot)\|_{L^p(\Omega,\mathbb{R}^N)}$.

2.1. Galerkin basis

Due to the density of $C_0^\infty(\Omega)$ in $W_0^{1,p}(\Omega)$, the Banach space $W_0^{1,p}(\Omega)$ with $1 < p < +\infty$ is separable. Therefore, there exists a Galerkin basis of $W_0^{1,p}(\Omega)$, that is a sequence $\{X_n\}_{n\geq 1}$ of vector subspaces of $W_0^{1,p}(\Omega)$ satisfying

(i) $\dim X_n < \infty, \quad \forall n \geq 1$;

(ii) $X_n \subset X_{n+1}, \quad \forall n \geq 1$;

(iii) $\overline{\bigcup_{n\geq 1} X_n} = W_0^{1,p}(\Omega)$.

For the rest of the paper we fix a Galerkin basis $\{X_n\}_{n\geq 1}$ of $W_0^{1,p}(\Omega)$.

2.2. Rellich-Kondrachov theorem

For $1 < p < N$, as known from the Rellich-Kondrachov theorem, the Sobolev space $W_0^{1,p}(\Omega)$ is compactly embedded into $L^\theta(\Omega)$ if $1 \leq \theta < p^*(= \frac{Np}{N-p})$ and continuously embedded if $\theta = p^*$. For every $\theta \in [1, p^*]$ we denote by $S_\theta > 0$ the best constant for this embedding, hence

$$\|u\|_{L^\theta(\Omega)} \leq S_\theta \|\nabla u\|_{L^p(\Omega,\mathbb{R}^N)}, \quad \forall u \in W_0^{1,p}(\Omega). \tag{2.1}$$

For $\theta = p$, we have that $S_p = \lambda_{1,p}^{-\frac{1}{p}}$ (see (1.2)).

2.3. Convolution

For easy reference we list a few useful properties of the convolution $\rho * u$ of $\rho \in L^1(\mathbb{R}^N)$ and $u \in W_0^{1,p}(\Omega)$; we refer to [1, §4.4, §9.1] for details. In order to have well defined the convolution $\rho * u$ of $\rho \in L^1(\mathbb{R}^N)$ with $u \in W_0^{1,p}(\Omega)$, it is convenient to consider the Sobolev space $W_0^{1,p}(\Omega)$ embedded in $W^{1,p}(\mathbb{R}^N)$ by identifying every $u \in W_0^{1,p}(\Omega)$ with its extension equal to zero outside Ω. The convolution $\rho * u$ is defined by

$$\rho * u(x) = \int_{\mathbb{R}^N} \rho(x-y)u(y)\,dy \ \text{ for a.e. } x \in \mathbb{R}^N.$$

The weak partial derivatives of the convolution $\rho * u$ are expressed by

$$\frac{\partial}{\partial x_i}(\rho * u) = \rho * \frac{\partial u}{\partial x_i} \in L^p(\mathbb{R}^N), \quad \forall i = 1, \ldots, N.$$

There hold the estimates

$$\|\rho * u\|_{L^r(\mathbb{R}^N)} \leq \|\rho\|_{L^1(\mathbb{R}^N)} \|u\|_{L^r(\Omega)} \tag{2.2}$$

whenever $r \in [1, p^*]$ and

$$\left\|\rho * \frac{\partial u}{\partial x_i}\right\|_{L^p(\mathbb{R}^N)} \leq \|\rho\|_{L^1(\mathbb{R}^N)} \left\|\frac{\partial u}{\partial x_i}\right\|_{L^p(\Omega)}, \quad \forall i = 1, \ldots, N. \tag{2.3}$$

Using the convexity of the function $t \mapsto t^p$ on $(0, +\infty)$ and (2.3), we derive that

$$\|\nabla(\rho * u)\|^p_{L^p(\mathbb{R}^N,\mathbb{R}^N)} = \int_{\mathbb{R}^N} |\nabla(\rho * u)|^p \, dx = \int_{\mathbb{R}^N} \left(\sum_{i=1}^N \left(\rho * \frac{\partial u}{\partial x_i} \right)^2 \right)^{\frac{p}{2}} dx$$

$$\leq \int_{\mathbb{R}^N} \left(\sum_{i=1}^N \left| \rho * \frac{\partial u}{\partial x_i} \right| \right)^p dx \leq N^{p-1} \sum_{i=1}^N \left\| \rho * \frac{\partial u}{\partial x_i} \right\|^p_{L^p(\mathbb{R}^N)}$$

$$\leq N^{p-1} \|\rho\|^p_{L^1(\mathbb{R}^N)} \sum_{i=1}^N \left\| \frac{\partial u}{\partial x_i} \right\|^p_{L^p(\Omega)} \leq N^p \|\rho\|^p_{L^1(\mathbb{R}^N)} \|\nabla u\|^p_{L^p(\Omega,\mathbb{R}^N)}. \tag{2.4}$$

2.4. Fixed point theorem

An essential tool in our approach will be the following consequence of Brouwer's fixed point theorem (see [8, page 37]).

Lemma 2.1. *Let X be a finite-dimensional space endowed with the norm $\|\cdot\|_X$ and let $A : X \to X^*$ be a continuous mapping. Assume that there is a constant $R > 0$ such that*

$$\langle A(v), v \rangle \geq 0 \textit{ for all } v \in X \textit{ with } \|v\|_X = R.$$

Then there exists $u \in X$ with $\|u\|_X \leq R$ satisfying $A(u) = 0$.

3. Main result

In this section we provide our main result regarding the existence of solutions to problem (1.1).

3.1. Nonlinear operator associated to problem (1.1)

Hereafter we consider the operator $A : W_0^{1,p}(\Omega) \to W^{-1,p'}(\Omega)$ given by

$$\langle A(u), v \rangle = \langle -\Delta_p u + \mu \Delta_q u, v \rangle - \int_\Omega f(x, \rho * u(x), \nabla(\rho * u)(x)) v(x) \, dx \tag{3.1}$$

which arises from problem (1.1).

Lemma 3.1. *Suppose that* (1.3) *in Assumption* 1.1 *is fulfilled. Then, the operator $A : W_0^{1,p}(\Omega) \to W^{-1,p'}(\Omega)$ defined in* (3.1) *is continuous.*

Proof. Relations (2.2) and (2.4) imply that the operator $T : W_0^{1,p}(\Omega) \to L^p(\Omega) \times L^p(\Omega)^N$ given by $T(u) = (\rho * u|_\Omega, \nabla(\rho * u)|_\Omega)$ is linear and continuous. The growth condition in (1.3) allows to apply the Krasnoselskii theorem [2] which implies that the Nemytskii operator

$$N_f : L^p(\Omega) \times L^p(\Omega)^N \to L^{p'}(\Omega), \ (v, w) \mapsto f(\cdot, v(\cdot), w(\cdot))$$

is well defined and continuous. We infer that the operator

$$W_0^{1,p}(\Omega) \to L^{p'}(\Omega), \ u \mapsto f(\cdot, \rho * u(\cdot), \nabla(\rho * u)(\cdot)) \tag{3.2}$$

is continuous as the composition of continuous operators. Note also that $L^{p'}(\Omega)$ is continuously embedded in $W^{-1,p'}(\Omega)$.

The operators $-\Delta_p : W_0^{1,p}(\Omega) \to W^{-1,p'}(\Omega)$ and $-\Delta_q : W_0^{1,q}(\Omega) \to W^{-1,q'}(\Omega)$ are continuous. Since $q < p$ and Ω is bounded, we have that $W_0^{1,p}(\Omega)$ is continuously embedded in $W_0^{1,q}(\Omega)$ and $W^{-1,q'}(\Omega)$ is continuously embedded in $W^{-1,p'}(\Omega)$. Therefore, $-\Delta_p + \mu\Delta_q : W_0^{1,p}(\Omega) \to W^{-1,p'}(\Omega)$ is continuous.

Altogether, this shows that the operator A is continuous. $\square$

3.2. Finite-dimensional approximations

Given a Galerkin basis $\{X_n\}_{n\geq 1}$ of $W_0^{1,p}(\Omega)$, we construct a corresponding sequence of approximate solutions related to problem (1.1).

Proposition 3.2. *Suppose that Assumption* 1.1 *is fulfilled. Then, for every $n \geq 1$, there exists $u_n \in X_n$ such that*

$$\langle -\Delta_p u_n + \mu\Delta_q u_n, v\rangle = \int_\Omega f(x, \rho * u_n(x), \nabla(\rho * u_n)(x))v(x)\,dx \tag{3.3}$$

for all $v \in X_n$. Moreover, the sequence $\{u_n\}_{n\geq 1}$ so obtained is bounded in $W_0^{1,p}(\Omega)$.

Proof. On each finite-dimensional space X_n we consider the mapping $A_n : X_n \to X_n^*$ defined by

$$\langle A_n(u), v\rangle = \langle -\Delta_p u + \mu\Delta_q u, v\rangle - \int_\Omega f(x, \rho * u(x), \nabla(\rho * u)(x))v(x)\,dx$$

for all $u, v \in X_n$. Note that A_n is continuous (see Lemma 3.1). Our goal is to apply Lemma 2.1 to the operator A_n. To this end, we note from (1.3) in Assumption 1.1 and Hölder's inequality that

$$\begin{aligned}
\langle A_n(v), v\rangle &= \int_\Omega (|\nabla v|^p - \mu|\nabla v|^q - f(x, \rho * v, \nabla(\rho * v))v)\,dx \\
&\geq \|\nabla v\|^p_{L^p(\Omega,\mathbb{R}^N)} - \mu|\Omega|^{\frac{p-q}{p}}\|\nabla v\|^q_{L^p(\Omega,\mathbb{R}^N)} - \|\sigma\|_{L^{r'}(\Omega)}\|v\|_{L^r(\Omega)} \\
&\quad -a_1\|\rho * v\|^{p-1}_{L^p(\mathbb{R}^N)}\|v\|_{L^p(\Omega)} - a_2\|\nabla(\rho * v)\|^{p-1}_{L^p(\mathbb{R}^N,\mathbb{R}^N)}\|v\|_{L^p(\Omega)}
\end{aligned}$$

for all $v \in X_n$. Hereafter, we denote by $|\Omega|$ the Lebesgue measure of Ω. Then (2.2), (2.4), and (2.1) lead to the estimate

$$\begin{aligned}
\langle A_n(v), v\rangle &\geq \|\nabla v\|^p_{L^p(\Omega,\mathbb{R}^N)} - \mu|\Omega|^{\frac{p-q}{p}}\|\nabla v\|^q_{L^p(\Omega,\mathbb{R}^N)} \\
&\quad -\|\sigma\|_{L^{r'}(\Omega)}\|v\|_{L^r(\Omega)} - a_1\|\rho\|^{p-1}_{L^1(\mathbb{R}^N)}\|v\|^p_{L^p(\Omega)} \\
&\quad -a_2N^{p-1}\|\rho\|^{p-1}_{L^1(\mathbb{R}^N)}\|\nabla v\|^{p-1}_{L^p(\Omega,\mathbb{R}^N)}\|v\|_{L^p(\Omega)} \\
&\geq \|\nabla v\|^p_{L^p(\Omega,\mathbb{R}^N)} - \mu|\Omega|^{\frac{p-q}{p}}\|\nabla v\|^q_{L^p(\Omega,\mathbb{R}^N)} - S_r\|\sigma\|_{L^{r'}(\Omega)}\|\nabla v\|_{L^p(\Omega,\mathbb{R}^N)} \\
&\quad -(a_1S_p^p\|\rho\|^{p-1}_{L^1(\mathbb{R}^N)} + a_2S_pN^{p-1}\|\rho\|^{p-1}_{L^1(\mathbb{R}^N)})\|\nabla v\|^p_{L^p(\Omega,\mathbb{R}^N)}
\end{aligned} \tag{3.4}$$

for all $v \in X_n$. Taking into account (1.4) (recall that $S_p = \lambda_{1,p}^{-\frac{1}{p}}$) and that $p > q > 1$, the following estimate is true

$$\langle A_n(v), v\rangle \geq 0 \text{ whenever } v \in X_n \text{ with } \|\nabla v\|_{L^p(\Omega,\mathbb{R}^N)} = R$$

provided $R > 0$ is sufficiently large. Then Lemma 2.1 yields the existence of $u_n \in X_n$ satisfying $A_n(u_n) = 0$, that is, (3.3).

It remains to show that the sequence $\{u_n\}_{n\geq 1}$ is bounded in $W_0^{1,p}(\Omega)$. By inserting $v = u_n \in X_n$ in (3.4), we find that

$$\begin{aligned}&\big(1 - \|\rho\|^{p-1}_{L^1(\mathbb{R}^N)}(a_1 S_p^p + a_2 S_p N^{p-1})\big)\|\nabla u_n\|^p_{L^p(\Omega,\mathbb{R}^N)}\\&\leq \mu|\Omega|^{\frac{p-q}{p}}\|\nabla u_n\|^q_{L^p(\Omega,\mathbb{R}^N)} + S_r\|\sigma\|_{L^{r'}(\Omega)}\|\nabla u_n\|_{L^p(\Omega,\mathbb{R}^N)}.\end{aligned}$$

The desired conclusion is readily obtained from assumption (1.4) and the fact that $p > q > 1$. □

3.3. Main result on the existence of a solution to problem (1.1)

First, we show that the notions of generalized solution and weak solution coincide for problem (1.1) in the case where $\mu \leq 0$.

Lemma 3.3. *Suppose that $\mu \leq 0$. For every $u \in W_0^{1,p}(\Omega)$, the following conditions are equivalent:*

(i) *u is a weak solution to problem* (1.1)*, that is, u satisfies*

$$\langle -\Delta_p u + \mu\Delta_q u, v\rangle = \int_\Omega f(x, \rho * u(x), \nabla(\rho * u)(x))v(x)\,dx$$

for all $v \in W_0^{1,p}(\Omega)$;

(ii) *u is a generalized solution to problem* (1.1)*.*

Proof. The implication (i)⇒(ii) is immediate (take $u_n = u$) and actually does not require the condition that $\mu \leq 0$. Conversely, assume that u is a generalized solution to problem (1.1), and let $\{u_n\}_{n\geq 1}$ be a sequence satisfying conditions (a)–(c) of the definition of generalized solution with respect to u. Using the monotonicity of the operator $-\Delta_q$ we note that

$$\begin{aligned}\langle -\Delta_p u_n, u_n - u\rangle &\leq \langle -\Delta_p u_n, u_n - u\rangle - \mu\langle -\Delta_q u_n + \Delta_q u, u_n - u\rangle\\&= \langle -\Delta_p u_n + \mu\Delta_q u_n, u_n - u\rangle - \mu\langle \Delta_q u, u_n - u\rangle.\end{aligned}$$

By (a) and (c), this leads to

$$\limsup_{n\to\infty}\langle -\Delta_p u_n, u_n - u\rangle \leq 0.$$

Then we are able to conclude the strong convergence $u_n \to u$ in $W^{1,p}(\Omega)$ (see, e.g., [7, Proposition 2.72]). By Lemma 3.1, this implies that $A(u_n) \to A(u)$ in $W^{-1,p'}(\Omega)$, where $A : W_0^{1,p}(\Omega) \to W^{-1,p'}(\Omega)$ is the operator defined in (3.1). In view of condition (b) of the definition of generalized solution, this yields $A(u) = 0$, which precisely means that u is a weak solution to problem (1.1). □

We can now state our main result.

Theorem 3.4. *Suppose that Assumption* 1.1 *holds. Then there exists a generalized solution to problem* (1.1)*. In particular, if $\mu \leq 0$, there exists a weak solution to problem* (1.1)*.*

Proof. Consider the sequence $\{u_n\}_{n\geq 1} \subset W_0^{1,p}(\Omega)$ constructed in Proposition 3.2. As asserted therein, this sequence is bounded in $W_0^{1,p}(\Omega)$. In view of the reflexivity of the space $W_0^{1,p}(\Omega)$, we can pass to a subsequence still denoted by $\{u_n\}_{n\geq 1}$ such that

$$u_n \rightharpoonup u \text{ in } W_0^{1,p}(\Omega) \tag{3.5}$$

with some $u \in W_0^{1,p}(\Omega)$. Moreover, since the sequence $\{u_n\}_{n\geq 1}$ is bounded in $W_0^{1,p}(\Omega)$, invoking the continuity of the operator in (3.2), we have that

$$\text{the sequence } \{f(\cdot, \rho * u_n, \nabla(\rho * u_n))\}_{n\geq 1} \text{ is bounded in } L^{p'}(\Omega). \tag{3.6}$$

On the basis of the reflexivity of $W^{-1,p'}(\Omega)$, we can assume that

$$-\Delta_p u_n + \mu\Delta_q u_n - f(\cdot, \rho * u_n, \nabla(\rho * u_n)) \rightharpoonup \eta \text{ in } W^{-1,p'}(\Omega) \tag{3.7}$$

with some $\eta \in W^{-1,p'}(\Omega)$.
Now let $v \in \bigcup_{n\geq 1} X_n$. Fix an integer $m \geq 1$ such that $v \in X_m$. Proposition 3.2 provides that (3.3) holds for all $n \geq m$. Letting $n \to \infty$ in (3.3), by means of (3.7) we get

$$\langle \eta, v\rangle = 0 \text{ for all } v \in \bigcup_{n\geq 1} X_n.$$

By the density of $\bigcup_{n\geq 1} X_n$ in $W_0^{1,p}(\Omega)$ (see (iii) in the definition of Galerkin basis in Section 2.1), it turns out that $\eta = 0$. Therefore, (3.7) renders

$$-\Delta_p u_n + \mu\Delta_q u_n - f(\cdot, \rho * u_n, \nabla(\rho * u_n)) \rightharpoonup 0 \text{ in } W^{-1,p'}(\Omega). \tag{3.8}$$

Next, setting $v = u_n$ in (3.3), we obtain

$$\langle -\Delta_p u_n + \mu\Delta_q u_n, u_n\rangle - \int_\Omega f(x, \rho * u_n, \nabla(\rho * u_n))u_n\, dx = 0 \tag{3.9}$$

for all $n \geq 1$, while (3.8) gives

$$\langle -\Delta_p u_n + \mu\Delta_q u_n, u\rangle - \int_\Omega f(x, \rho * u_n, \nabla(\rho * u_n))u\, dx \to 0 \tag{3.10}$$

as $n \to \infty$. Altogether, (3.9) and (3.10) yield

$$\langle -\Delta_p u_n + \mu\Delta_q u_n, u_n - u\rangle - \int_\Omega f(x, \rho * u_n, \nabla(\rho * u_n))(u_n - u)\, dx \to 0 \tag{3.11}$$

as $n \to \infty$. Moreover, from (3.5), Rellich-Kondrachov compact embedding theorem which ensures that $u_n \to u$ strongly in $L^p(\Omega)$, and (3.6), we derive that

$$\lim_{n\to\infty} \int_\Omega f(x, \rho * u_n, \nabla(\rho * u_n))(u_n - u)\, dx = 0. \tag{3.12}$$

Inserting (3.12) into (3.11) enables us to assert

$$\lim_{n\to\infty} \langle -\Delta_p u_n + \mu\Delta_q u_n, u_n - u\rangle = 0. \tag{3.13}$$

At this point we can notice that (3.5), (3.8), and (3.13) are just the conditions (a), (b), and (c) expressing that $u \in W_0^{1,p}(\Omega)$ is a generalized solution to problem (1.1), which proves the first assertion in the theorem. The last assertion in the theorem is a consequence of Lemma 3.3. $\square$

Acknowledgement. The authors are grateful to Dr. Lucas Fresse for important remarks.

References

[1] Brezis, H., *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Universitext, Springer, New York, 2011.

[2] Krasnoselskii, M.K., *Topological Methods in the Theory of Nonlinear Integral Equations*, Pergamon Press, New York, 1964.

[3] Liu, Z., Livrea, R., Motreanu, D., Zeng, S., *Variational differential inclusions without ellipticity condition*, Electron. J. Qual. Theory Differ. Equ., Paper No. 43 (2020), 17 pp.

[4] Motreanu, D., *Nonlinear Differential Problems with Smooth and Nonsmooth Constraints*, Academic Press, London, 2018.

[5] Motreanu, D., *Quasilinear Dirichlet problems with competing operators and convection*, Open Math., **18**(2020), 1510-1517.

[6] Motreanu D., Motreanu, V.V., *Non-variational elliptic equations involving (p, q)-Laplacian, convection and convolution*, Pure Appl. Funct. Anal., **5**(2020), 1205-1215.

[7] Motreanu, D., Motreanu, V.V., Papageorgiou, N.S., *Topological and Variational Methods with Applications to Nonlinear Boundary Value Problems*, Springer, New York, 2014.

[8] Showalter, R.E., *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, Mathematical Surveys and Monographs, vol. 49, American Mathematical Society, Providence, RI, 1997.

Dumitru Motreanu
Département de Mathématiques,
Université de Perpignan,
66860 Perpignan, France
e-mail: `motreanu@univ-perp.fr`

Viorica Venera Motreanu
Collège Jean Moulin,
14 rue Jean Moulin,
54510 Tomblaine, France
e-mail: `vmotreanu@gmail.com`

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 105–113
DOI: 10.24193/subbmath.2021.1.09

Fredholm and Volterra nonlinear possibilistic integral equations

Sorin G. Gal and Ionuţ T. Iancu

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. In this paper we study the nonlinear functional equations obtained from the classical integral equations of Fredholm and of Volterra of second kind, by replacing there the linear Lebesgue integral with the nonlinear possibilistic integral.

Mathematics Subject Classification (2010): 45B05, 47H10, 28E10, 28A99.

Keywords: Possibility measure, nonlinear possibilistic integral, Fredholm nonlinear possibilistic integral equation, Volterra nonlinear possibilistic integral equation, fixed point theorem, successive approximations.

1. Introduction

It is known that possibility theory is an alternative theory to the probability theory, dealing with certain types of uncertainty and treatment of incomplete information (see, e.g., [2], [4]). In the possibilistic models, all the probabilistic indicators (like expected value, variance, probability measure, integral with respect to a measure, etc) are replaced with suitable possibility indicators. These analogies allow to extend many classical results based on probability theory, to the possibilistic frame. We can mention here the contributions of the first named author to results concerning approximation by possibilistic operators (called also max-product operators), see [1], [3], [7], [9] or to the possibilistic laws of large numbers, see, e.g., [8], and the references therein.

In this paper we continue our researches in this directions, by extending in the frame of possibility theory, results concerning classical integral equations.

In this sense, it is natural and of interest to replace in the classical integral equations, the linear Lebesgue integral by various other kinds of nonlinear integrals. Thus, in the very recent papers [5], [6], the first named author has replaced the linear Lebesgue integral by its non linear extension called Choquet integral and studied the existence of the solutions for the corresponding Fredholm-Choquet and Volterra-Choquet integral equations.

In this spirit of ideas, we study here the nonlinear equations obtained by replacing in the classical Fredholm and Volterra integral equations, the linear Lebesgue integral with the nonlinear possibilistic integral $(Pos)\int$ with respect to a possibility measure P_λ generated by the possibility distribution λ. More exactly, we study the nonlinear possibilistic integral equations

$$\varphi(x) = f(x) + \alpha \cdot (Pos)\int_\Omega K(x,s)\varphi(s)dP_\lambda(s), x \in \Omega, \tag{1.1}$$

with the given data $\alpha \in \mathbb{R}$, $f : \Omega \to \mathbb{R}$, $K : \Omega \times \Omega \to \mathbb{R}$ and the unknown function $\varphi : \Omega \to \mathbb{R}$ in the case of Fredholm type equation, and by

$$\varphi(x) = f(x) + \alpha \cdot (Pos)\int_a^x K(x,s,\varphi(s))dP_\lambda(s), x \in [a,b], \tag{1.2}$$

with the given data $\alpha \in \mathbb{R}$, $f : [a,b] \to \mathbb{R}$, $K : [a,b] \times [a,b] \times \mathbb{R} \to \mathbb{R}$ and the unknown function $\varphi : [a,b] \to \mathbb{R}$, in the case of Volterra type equation.

As we will see, due to the definition of the possibilistic integral, in fact we obtain functional equations which have solutions under some additional conditions. Also, it is worth mentioning that while the classical Fredholm and Volterra integral equations are linear, due to the nonlinearity of the possibilistic integral, obviously that the Fredholm and Volterra possibilistic integral equations are nonlinear.

Section 2 contains some preliminaries on the possibility measures and integrals we will need in the next sections. In Section 3, the existence and construction of the solutions for the Fredholm nonlinear possibilistic integral equation (1.1). Thus, for P_λ belonging to large classes of possibility measures, we show that this functional equation has solutions under some appropriate conditions (similar to those in the classical case) on the given data f, α and K.

Finally, in Section 4 we study the existence of the solutions for the Volterra nonlinear possibilistic integral type equation (1.2).

2. Preliminaries on possibility measures and integrals

Firstly, we summarize some known concepts in possibility theory, which will be used in the next sections. For details, see e.g. [4] or [2].

Definition 2.1. Let Ω be a non-empty set.

(i) A possibilistic (fuzzy) variable X is simply an application $X : \Omega \to \mathbb{R}$.

(ii) A possibility distribution (on Ω), is a function $\lambda : \Omega \to [0, +\infty)$, such that $\sup\{\lambda(s); s \in \Omega\} = V < +\infty$. If $V = 1$, then λ it is called normalized possibility distribution.

(iii) A possibility measure is a mapping $P : \mathcal{P}(\Omega) \to [0, +\infty)$, satisfying the axioms $P(\emptyset) = 0$, $P(\Omega) = 1$ and $P(\bigcup_{i\in I} A_i) = \sup\{P(A_i); i \in I\}$ for all $A_i \in \Omega$, and any I, family of indices.

It is well-known (see e.g. [4]) that any possibility distribution λ on Ω, induces a possibility measure P_λ, given by the formulas $P_\lambda(A) = \sup\{\lambda(s); s \in A\}$, for all $A \subset \Omega$, $A \neq \emptyset$, $P_\lambda(\emptyset) = 0$.

(iv) (see e.g. [2]) The possibilistic integral of $f : \Omega \to \mathbb{R}_+$ on $A \subset \Omega$, with respect to the possibilistic measure P_λ induced by the possibilistic distribution λ, is defined by

$$(Pos)\int_A f(t)dP_\lambda(t) = \sup\{f(t)\cdot\lambda(t); t \in A\}.$$

It is clear that this definition is a particular case of the t-possibilistic integral with respect to a semi-norm t, introduced in [2], by taking there $t(x, y) = x \cdot y$.

(v) The following properties hold : for all $f, g \geq 0$ and $c \in \mathbb{R}, c \geq 0$

$$(Pos)\int_A (f(s)+g(s))dP_\lambda(s) \leq (Pos)\int_A f(s)dP_\lambda(s) + (Pos)\int_A g(s)dP_\lambda(s),$$

$$(Pos)\int_A [cf(s)]dP_\lambda(s) = c\cdot(Pos)\int_A f(s)dP_\lambda(s).$$

3. Fredholm possibilistic integral equations

Let us denote by $B_+(\Omega)$, the Banach space of all positive and bounded functions $f : \Omega \to \mathbb{R}_+$, endowed with the uniform norm, denoted here by $\|\cdot\|$. It is clear that $B_+(\Omega)$ endowed with the metric generated by the uniform norm, is a complete metric space.

Taking into account the definition of the possibilistic integral in Definition 2.1, the Fredholm possibilistic integral equation in (1.1), formally becomes the nonlinear functional equation

$$\varphi(x) = f(x) + \alpha\cdot\sup\{K(x,s)\cdot\varphi(s)\cdot\lambda(s); s \in \Omega\}, x \in \Omega. \tag{3.1}$$

The first main result is the following.

Theorem 3.1. *Let $\Omega \neq \emptyset$ and P_λ be possibility measure induced by the possibility distribution λ on Ω.*

Let us also suppose that

$$0 \leq K(x,s) \leq M < +\infty, \text{ for all } x, s \in \Omega.$$

Then, for any $f \in B_+(\Omega)$ and any $0 < \alpha < \frac{1}{M}$, the Fredholm possibilistic functional equation (3.1) *has a unique solution $\varphi^* \in B_+(\Omega)$.*

Moreover, denoting

$$A(\varphi)(x) = f(x) + \alpha\cdot\sup\{K(x,s)\lambda(s)\cdot\varphi(s); s \in \Omega\} \text{ and } A^n(\varphi_0) = A[A^{n-1}(\varphi_0)],$$

for any arbitrary $\varphi_0 \in B_+(\Omega)$, the following estimate holds

$$\|A^n(\varphi_0) - \varphi^*\| \leq \frac{\alpha\cdot M}{1-\alpha\cdot M}\cdot\|A(\varphi_0)-\varphi_0\|.$$

Proof. For any $x \in \Omega$ fixed and $\varphi \in B_+(\Omega)$, let us denote

$$T(\varphi)(x) = \sup\{K(x,s)\varphi(s)\lambda(s); s \in \Omega\}.$$

By hypothesis we immediately get $T(\varphi) \in B_+(\Omega)$.
This implies that $A(\varphi)(x) = f(x) + \alpha \cdot T(\varphi)(x) \in B_+(\Omega)$, for all $0 \le \alpha < \infty$.

Let $\varphi, \psi \in B_+(\Omega)$. We have $\varphi = \varphi - \psi + \psi \le |\varphi - \psi| + \psi$, which successively implies

$$T(\varphi)(x) \le T(|\varphi - \psi|)(x) + T(\psi)(x),$$

that is

$$T(\varphi)(x) - T(\psi)(x) \le T(|\varphi - \psi|)(x).$$

Writing now $\psi = \psi - \varphi + \varphi \le |\varphi - \psi| + \varphi$ and applying the above reasonings, it follows

$$T(\psi)(x) - T(\varphi)(x) \le T(|\varphi - \psi|)(x),$$

which combined with the above inequality gives

$$|T(\varphi)(x) - T(\psi)(x)| \le T(|\varphi - \psi|)(x),$$

that is

$$\begin{aligned}&|\sup\{K(x,s)\psi(s)\lambda(s); s \in \Omega\} - \sup\{K(x,s)\varphi(s)\lambda(s); s \in \Omega\}|\\ &\le \sup\{|K(x,s)\psi(s)\lambda(s) - K(x,s)\varphi(s)\lambda(s)|; s \in \Omega\}.\end{aligned}$$

Since for all $x \in \Omega$ we have

$$\begin{aligned}|A(\varphi)(x) - A(\psi)(x)| &= \alpha \cdot |T(\varphi)(x) - T(\psi)(x)| \le \alpha \cdot T(|\varphi - \psi|)(x)\\ &\le M \cdot \alpha \cdot \|\varphi - \psi\|,\end{aligned}$$

passing to supremum after $x \in \Omega$, we immediately obtain

$$d(A(\varphi), A(\psi)) := \|A(\varphi) - A(\psi)\| \le M \cdot \alpha \cdot \|\varphi - \psi\| := M \cdot \alpha \cdot d(\varphi, \psi).$$

The hypothesis implies that $d : B_+(\Omega) \times B_+(\Omega) \to \mathbb{R}_+$ is a contraction on the complete metric space $B_+(\Omega)$ endowed with the metric $d(\varphi, \psi) = \|\varphi - \psi\|$, which by the Banach's fixed point theorem implies the desired conclusion. $\square$

Remark 3.2. In general, under the conditions in Theorem 3.1 the sequence of successive approximation cannot be written in the explicit form as in the classical linear Fredholm integral equation (i.e. by using the so called resolvent). However, under some additional hypothesis on the input data f, λ and K, this can be done, exemplified by the following result.

Corollary 3.3. *Let $\Omega = [a,b]$ and P_λ be the possibility measure induced by the possibility distribution λ, supposed to be nondecreasing on Ω.*
Let us suppose that $\alpha > 0$, $f(x) \ge 0$, $K(x,s) \ge 0$, for all $x, s \in [a,b]$, $K(b,b) > 0$,

$$K(\cdot,\cdot) \textit{ is nondecreasing in each variable on } [a,b],$$

$$f \textit{ is nondecreasing on } [a,b].$$

Then, for any $\alpha < \frac{1}{K(b,b)}$, the Fredholm possibilistic functional equation (3.1) has a unique solution $\varphi^* \in B_+[a,b]$, nondecreasing on $[a,b]$.

Moreover, denoting $K_1(x,t)=K(x,t)$ *and by the recurrence formula*

$$K_j(x,t)=(Pos)\int_a^b K_{j-1}(x,s)K(s,t)dP_\lambda(s)$$
$$=\sup\{K_{j-1}(x,s)K(s,t)\cdot\lambda(s); s\in[a,b]\},\ j\in\mathbb{N},\ j\geq 2,$$

for the sequence of successive approximation with φ_0 *positive and nondecreasing on* $[a,b]$*, we have the representation*

$$A^n(\varphi_0)(x)=f(x)+\alpha\cdot(Pos)\int_a^b R_n(x,t;\alpha)f(t)dP_\lambda(t)$$
$$+\alpha^{n+1}\cdot(Pos)\int_a^b K_{n+1}(x,t)\varphi_0(t)dP_\lambda(t)$$
$$=f(x)+\alpha\cdot\sup\{R_n(x,t;\alpha)f(t)\lambda(t);t\in[a,b]\}$$
$$+\alpha^{n+1}\cdot\sup\{K_{n+1}(x,t)\varphi_0(t)\lambda(t);t\in[a,b]\}, \tag{3.2}$$

where

$$R_n(x,t;\alpha)=\sum_{j=1}^{n}\alpha^{j-1}K_j(x,t).$$

Also, for the solution $\varphi^*(x)$ *we have the representation*

$$\varphi^*(x)=f(x)+\alpha\cdot(Pos)\int_a^b R(x,t;\alpha)f(t)dP_\lambda(t)$$
$$=f(x)+\alpha\cdot\sup\{R(x,t;\alpha)f(t)\lambda(t);t\in[a,b]\},\ x\in[a,b], \tag{3.3}$$

with

$$R(x,t;\alpha)=\sum_{j=1}^{\infty}\alpha^{j-1}\cdot K_j(x,t),\ x,t\in[a,b].$$

Proof. Taking $M=K(b,b)$, the hypothesis in Theorem 3.1 are fulfilled, fact which implies that there exists uniquely φ^* satisfying (3.1).

It remains to deal with the sequence of successive approximations.

Let us choose φ_0 be positive and nondecreasing on $[a,b]$ (clearly it follows that φ_0 is bounded too). We get

$$\varphi_1(x)=A(\varphi_0)(x)=f(x)+\alpha\cdot(Pos)\int_a^b K(x,s)\varphi_0(s)dP_\lambda(s)$$
$$=f(x)+\alpha\cdot\sup\{K(x,s)\varphi_0(s)\lambda(s);s\in[a,b]\},$$

which from the hypothesis immediately implies that $\varphi_1(x)\geq 0$, for all $x\in[a,b]$ and φ_1 is nondecreasing (and therefore bounded) on $[a,b]$.

Also, φ_1 is the sum of two positive and both nondecreasing functions on $[a,b]$. Since it is easy to prove that if F and G are both nondecreasing on $[a,b]$ then

$$(Pos)\int_a^b[F(s)+G(s)]dP_\lambda(s)=\sup\{(F(s)+G(s))\lambda(s);s\in[a,b]\}$$
$$=F(b)\lambda(b))+G(b)\lambda(b)=\sup\{F(s)\lambda(s);s\in[a,b]\}+\sup\{G(s)\lambda(s);s\in[a,b]\}$$

$$= (Pos)\int_a^b F(s)dP_\lambda(s) + (Pos)\int_a^b G(s)dP_\lambda(s),$$

we obtain

$$\varphi_2(x) = f(x) + \alpha \cdot (Pos)\int_a^b K(x,t)\cdot\varphi_1(t)dP_\lambda(t)$$
$$= f(x) + \alpha \cdot (Pos)\int_a^b K(x,t)f(t)dP_\lambda(t)$$
$$+\alpha^2 \cdot (Pos)\int_a^b K(x,t)\left[(Pos)\int_a^b K(t,s)\varphi_0(s)dP_\lambda(s)\right]dP_\lambda(t).$$

If, for each fixed $x \in [a,b]$ we denote $F(t,s) = K(x,t)\cdot K(t,s)\cdot\varphi_0(s)$, then $F(t,s) \geq 0$ for all $t, s \in [a,b]$ and $F(t,s)$ is nondecreasing in each variable t and s. Also, since F is bounded on $[a,b]\times[a,b]$, it follows that we can write

$$(Pos)\int_a^b K(x,t)\left[(Pos)\int_a^b K(t,s)\varphi_0(s)dP_\lambda(s)\right]dP_\lambda(t)$$
$$= \sup\{K(x,t)\lambda(t)\cdot\sup\{K(t,s)\varphi_0(s)\lambda(s); s\in[a,b]\}; t\in[a,b]\}$$
$$= \sup\{\sup\{K(x,t)K(t,s)\lambda(t); t\in[a,b]\}\cdot\varphi_0(s)\lambda(s); s\in[a,b]\}$$
$$= \sup\{K_2(x,s)\varphi_0(s)\lambda(s); s\in[a,b]\}$$
$$= (Pos)\int_a^b K_2(x,s)\varphi_0(s)dP_\lambda(s),$$

fact which leads to the formula

$$\varphi_2(x) = f(x) + \alpha\cdot(Pos)\int_a^b K(x,t)\cdot f(t)dP_\lambda(t) + \alpha^2\cdot(Pos)\int_a^b K_2(x,s)\varphi_0(s)dP_\lambda(s).$$

Continuing these kinds of reasonings, step by step we easily get the recurrence formula

$$\varphi_{n+1}(x) = A^n(\varphi_0)(x)$$
$$= f(x) + \sum_{j=1}^{n}\alpha^j\cdot(Pos)\int_a^b K_j(x,t)f(t)dP_\lambda(t)$$
$$+\alpha^{n+1}\cdot(Pos)\int_a^b K_{n+1}(x,t)\varphi_0(t)dP_\lambda(t)$$
$$= f(x) + \alpha\cdot(Pos)\int_a^b\left[\sum_{j=1}^{n}\alpha^{j-1}K_j(x,t)f(t)\right]dP_\lambda(t)$$
$$+\alpha^{n+1}\cdot(Pos)\int_a^b K_{n+1}(x,t)\varphi_0(t)dP_\lambda(t)$$
$$= f(x) + \alpha\cdot\sup\{R_n(x,t;\alpha)f(t)\lambda(t); t\in[a,b]\}$$
$$+\alpha^{n+1}\cdot\sup\{K_{n+1}(x,t)\varphi_0(t)\lambda(t); t\in[a,b]\}.$$

Now, by mathematical induction we easily can prove that

$$0 \leq K_{n+1}(x,t) \leq [K(b,b)]^{n+1},$$

for all $x, t \in [a, b]$ and $n = 0, 1, ...,$. This immediately implies (even uniformly with respect to $x \in [a, b]$)

$$0 \leq \lim_{n\to\infty} \alpha^{n+1} \cdot (Pos) \int_a^b K_{n+1}(x,t)\varphi_0(t) dP_\lambda(t)$$
$$\leq \lim_{n\to\infty} \left([\alpha \cdot K(b,b)]^{n+1} \cdot (Pos) \int_a^b \varphi_0(t) dP_\lambda(t) \right)$$
$$\leq \lim_{n\to\infty} \left([\alpha \cdot K(b,b)]^{n+1} \cdot \|\varphi_0\| \right) = 0$$

and

$$0 \leq \sum_{j=1}^{\infty} \alpha^{j-1} K_j(x,t) f(t)\lambda(t) \leq \sum_{j=1}^{\infty} (\alpha \cdot K(b,b))^{j-1} \cdot [K(b,b)f(b)].$$

Therefore, for each fixed $x \in [a, b]$,

$$R_n(x,t;\alpha) = \sum_{j=1}^{n} \alpha^{j-1} K_j(x,t) f(t)\lambda(t),$$

converges (for $n \to \infty$) to $R(x, t; \alpha) \cdot f(t)\lambda(t)$, uniformly with respect to $t \in [a, b]$.

Applying now the formula

$$|\sup\{R_n(x,t;\alpha)f(t)\lambda(t); t \in [a,b]\} - \sup\{R(x,t;\alpha)f(t)\lambda(t); t \in [a,b]\}|$$
$$\leq \sup\{|R_n(x,t;\alpha)f(t)\lambda(t) - R(x,t;\alpha)f(t)\lambda(t)|; t \in [a,b]\},$$

we immediately arrive to formula (3.3). □

Remark 3.4. It is clear that Corollary 3.3 remains valid if in its statement we replace everywhere the word "nondecreasing" with the word "nonincreasing" and $K(b, b)$ with $K(a, a)$.

4. Volterra possibilistic integral equations

It is known that in the classical case, the Volterra integral equation has solution for any value of the parameter α. Unfortunately, in the case of Volterra possibilistic integral equation, this fact does not hold in general. However, for some appropriate choices of the possibility measure P_λ in equation (1.2), it has unique solution for any value of the parameter α.

Let us make the notation

$$B_+[a,b] = \{f : [a,b] \to \mathbb{R}_+; f \text{ is bounded and positive on } [a,b]\},$$

endowed with the uniform norm $\|\cdot\|$.

For our purpose and taking into account the definition of the possibilistic integral in Definition 2.1, in the Volterra possibilistic integral equation in (1.2) we will consider a family of possibility measures depending on a parameter $\tau > 0$, for which (1.2) formally becomes the nonlinear functional equation

$$\varphi(x) = f(x) + \alpha \cdot \sup\{K(x,s,\varphi(s)) \cdot \lambda_\tau(s); s \in [a,x]\}, x \in [a,b], \tau > 0, \tag{4.1}$$

where λ_τ, $\tau > 0$ is a family of possibility densities defined as in Definition 2.1, (ii).

The main result is the following.

Theorem 4.1. *Let P_{λ_τ} be the possibilistic measure induced by the possibilistic distribution λ_τ on $[a,b]$, $\tau > 0$.*

Let us suppose that

$$K \in B_+([a,b] \times [a,b] \times \mathbb{R}), \tag{4.2}$$

where $B_+([a,b] \times [a,b] \times \mathbb{R})$ denotes the space of all positive and bounded functions $g : [a,b] \times [a,b] \times \mathbb{R} \to \mathbb{R}_+$, there exists $L > 0$ such that

$$|K(x,s,u) - K(x,s,v)| \le L \cdot |u - v|, \text{ for all } x,s \in [a,b], u,v \in \mathbb{R}_+ \tag{4.3}$$

and that

$$\lim_{\tau \to +\infty} \sup\{\lambda_\tau(s); s \in [a,b]\} = 0. \tag{4.4}$$

Then there exists $\tau_0 > 0$ such that for any $f \in B_+[a,b]$, any $\alpha > 0$ and any $\tau > \tau_0$, the Volterra possibilistic integral equation (4.1) has a unique solution $\varphi_\tau \in B_+[a,b]$.

Proof. For any $\varphi \in B_+[a,b]$, let us denote

$$T_\tau(\varphi)(x) = \sup\{K(x,s,\varphi(s))\lambda_\tau(s); s \in [a,x]\}, x \in [a,b], \tau > 0.$$

It is well-defined for any fixed arbitrary $x \in [a,b]$, because from hypothesis on K and $\lambda_\tau(s)$, it easily follows that as function of s, $K(x,s,\varphi(s))$ there exists $M > 0$ such that $|K(x,s,\varphi(s))|\lambda_\tau(s) \le M$, for all $x,s \in [a,b]$. In what follows, we prove that $T_\tau(\varphi) \in B_+[a,b]$. For any fixed $x \in [a,b]$, we immediately get

$$\begin{aligned} T_\tau(\varphi)(x) &= |T_\tau(\varphi)(x)| = |\sup\{K(x,s,\varphi(s))\lambda_\tau(s); s \in [a,x]\}| \\ &\le \sup\{|K(x,s,\varphi(s))\lambda_\tau(s)|; s \in [a,x]\} \le M. \end{aligned}$$

In conclusion $T_\tau(\varphi) \in B_+[a,b]$ and this also implies that $A_\tau(\varphi) = f + \alpha \cdot T_\tau(\varphi) \in B_+[a,b]$.

Therefore, by using the hypothesis (4.3) and (4.4) too, we immediately obtain

$$\begin{gathered} |T_\tau(\varphi)(x) - T_\tau(\psi)(x)| \\ \le \sup\{|K(x,s,\varphi(s)) \cdot \lambda_\tau(s) - K(x,s,\psi(s)) \cdot \lambda_\tau(s)|; s \in [a,x]\} \\ \le L \sup\{|\varphi(s) - \psi(s)|; s \in [a,x]\} \end{gathered}$$

and

$$\begin{aligned} |A_\tau(\varphi)(x) - A_\tau(\psi)(x)| &= \alpha \cdot |T_\tau(\varphi)(x) - T_\tau(\psi)(x)| \\ &\le \alpha \cdot L \sup\{|\varphi(s) - \psi(s)|\lambda_\tau(s); s \in [a,x]\} \\ &\le \alpha \cdot L \cdot \sup\{\lambda_\tau(s); s \in [a,b]\} \cdot \|\varphi - \psi\|, \end{aligned}$$

which immediately implies

$$\begin{gathered} d(A_\tau(\varphi), A_\tau(\psi)) := \|A_\tau(\varphi) - A_\tau(\psi)\| \\ \le \alpha \cdot L \cdot \sup\{\lambda_\tau(s); s \in [a,b]\} \cdot \|\varphi - \psi\|_\tau = \alpha \cdot L \cdot \sup\{\lambda_\tau(s); s \in [a,b]\} \cdot d(\varphi, \psi). \end{gathered}$$

From condition (4.4), there exists τ_0 such that for all $\tau > \tau_0 > 0$ to get

$$\alpha \cdot L \cdot \sup\{\lambda_\tau(s); s \in [a,b]\} < 1,$$

therefore d is a contraction on the complete metric space $B_+[a,b]$ and applying the Banach's fixed point theorem we arrive at the desired conclusion. $\square$

Remark 4.2. An important particular case is when $K(x, s, v) := K(x, s) \cdot v$. In this case, condition (4.2) becomes $K \in B_+([a, b] \times [a, b])$ and it immediately implies the condition (4.3), with $L = \sup\{K(x, s); x, s \in [a, b]\}$.

Remark 4.3. There are very many simple examples of families of possibilistic distributions satisfying condition (4.4) in Theorem 4.1, like, for example,

$$\lambda_\tau(s) = e^{-\tau|s|+1}, \quad \lambda_\tau(s) = \frac{s^2}{\tau}, \quad \lambda_\tau(s) = \frac{|\sin(s)|}{\tau}, \quad \tau > 0$$

and so on.

References

[1] Bede, B., Coroianu, L., Gal, S.G., *Approximation by Max-Product Type Operators*, Springer, New York, 2016.

[2] De Cooman, G., *Possibility theory. I. The measure and integral-theoretic groundwork*, Internat. J. Gen. Systems, **25**(1997), no. 4, 291-323.

[3] Coroianu, L., Gal, S.G., Opris, B., Trifa, S., *Feller' s scheme in approximation by nonlinear possibilistic integral operators*, Numer. Funct. Anal. Optim, **38**(2017), 327-343.

[4] Dubois, D., Prade, H., *Possibility Theory*, Plenum Press, New York, 1988.

[5] Gal, S.G., *Fredholm-Choquet integral equations*, J. Integral Equations Appl., **31**(2019), no. 2, 183-194.

[6] Gal, S.G., *Volterra-Choquet integral equations*, J. Integral Equations Appl., **31**(2019), no. 4, 495-504.

[7] Gal, S.G., *A possibilistic approach of the max-product Bernstein kind operators*, Results Math., **65**(2014), 453-462.

[8] Gal, S.G., *On the laws of large numbers in possibility theory*, Ann. Acad. Rom. Sci. Ser. Math. Appl., **11**(2019), no. 2, 274-284.

[9] Gal, S.G., *Approximation by polynomial possibilistic integral operators*, Ann. Acad. Rom. Sci. Ser. Math. Appl., **12**(2020), no. 1-2, 132-141.

Sorin G. Gal
Department of Mathematics and Computer Science,
University of Oradea,
Universitatii 1, 410087, Oradea, Romania,
and
Academy of Romanian Scientists
Splaiul Independentei No. 54, Bucharest 050094, Romania
e-mail: `galso@uoradea.ro, galsorin23@gmail.com`

Ionuţ T. Iancu
Department of Mathematics and Computer Science,
University of Oradea,
Universitatii 1, 410087, Oradea, Romania
e-mail: `ionutz.tudor.iancu@gmail.com`

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 115–122
DOI: 10.24193/subbmath.2021.1.10

Datko criteria for uniform instability in Banach spaces

Rovana Boruga (Toma) and Mihail Megan

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. The main objective of this paper is to present some necessary and sufficient conditions of Datko type for the uniform exponential and uniform polynomial instability concepts for evolution operators in Banach spaces.

Mathematics Subject Classification (2010): 47D06, 47B01.

Keywords: Evolution operator, uniform instability, Datko criteria.

1. Introduction

The instability behavior of evolution operators is a topic that has witnessed a significant progress in recent years. The importance of the role played by this concept in the theory of dynamical systems is illustrated by the appearance of various papers in this domain for the exponential case ([3], [8], [11], [12], [14]) as well as for the polynomial case ([1], [2], [13]), which appeared due to the fact that in some situation the exponential behavior is too restrictive for the dynamics.

Another direction for the study of the instability behavior is given by M. Megan, A.L. Sasu, B. Sasu in [9] where the authors express the uniform exponential instability of evolution families in terms of Banach function spaces. The property of exponential instability is generalized by M. Megan, C. Stoica [10] for skew-evolution semiflows defined by means of evolution semiflows and evolution cocycles. Recently, P.V. Hai [6] in his paper obtains results from the same perspective of using Banach spaces of sequences for the polynomial instability concept.

In this work we deal with both exponential and polynomial instability behavior for the uniform case of evolution operators in Banach spaces. In this sense, we give some necessary and sufficient conditions due to Datko [5], firstly for the uniform exponential instability concept and then we extend the theory to the polynomial case, our theorems being proved by using different techniques from those known so far.

2. Notations and definitions

We consider X a real or complex Banach space, $\mathcal{B}(X)$ the Banach algebra of all bounded linear operators acting on X and I the identity operator on X. The norms on X and on $\mathcal{B}(X)$ will be denoted by $\|.\|$. We also denote by

$$\Delta = \{(t,s) \in \mathbb{R}_+^2 : t \geq s\}, \qquad T = \{(t,s,t_0) \in \mathbb{R}_+^3 : t \geq s \geq t_0\}.$$

Definition 2.1. An application $U : \Delta \to \mathcal{B}(X)$ is said to be an evolution operator on X if the following relations are satisfied:

(e_1) $U(t,t) = I$ for all $t \geq 0$

(e_2) $U(t,s)U(s,t_0) = U(t,t_0)$ for all $(t,s,t_0) \in T$.

Definition 2.2. An evolution operator $U : \Delta \to \mathcal{B}(X)$ is said to be *strongly measurable* if for all $(s,x) \in \mathbb{R}_+ \times X$, the mapping $t \mapsto \|U(t,s)x\|$ is measurable on $[s,\infty)$.

Definition 2.3. The evolution operator $U : \Delta \to \mathcal{B}(X)$ *has uniform exponential decay (u.e.d.)* if there exist the constants $M \geq 1$ and $\omega > 0$ such that:

$$\|U(s,t_0)x_0\| \leq Me^{\omega(t-s)}\|U(t,t_0)x_0\|, \text{ for all } (t,s,t_0,x_0) \in T \times X.$$

Remark 2.4. The evolution operator $U : \Delta \to \mathcal{B}(X)$ has uniform exponential decay if and only if there exist the constants $M \geq 1$ and $\omega > 0$ such that:

$$\|x\| \leq Me^{\omega(t-s)}\|U(t,s)x\|, \text{ for all } (t,s,x) \in \Delta \times X.$$

Definition 2.5. The evolution operator U is said to be *uniformly exponentially instable (u.e.is.)* if there exist $N \geq 1$ and $\nu > 0$ such that:

$$\|U(s,t_0)x_0\| \leq Ne^{-\nu(t-s)}\|U(t,t_0)x_0\|, \text{ for all } (t,s,t_0,x_0) \in T \times X.$$

Remark 2.6. The evolution operator U is uniformly exponentially instable if and only if there exist $N \geq 1$ and $\nu > 0$ such that:

$$\|x\| \leq Ne^{-\nu(t-s)}\|U(t,s)x\|, \text{ for all } (t,s,x) \in \Delta \times X.$$

Definition 2.7. The evolution operator U has *uniform polynomial decay (u.p.d.)* if there exist the constants $M \geq 1$ and $\omega > 0$ such that:

$$\|U(s,t_0)x_0\| \leq M\left(\frac{t+1}{s+1}\right)^{\omega}\|U(t,t_0)x_0\|, \text{ for all } (t,s,t_0,x_0) \in T \times X.$$

Remark 2.8. The evolution operator U has uniform polynomial decay if and only if there exist the constants $M \geq 1$ and $\omega > 0$ such that:

$$\|x\| \leq M\left(\frac{t+1}{s+1}\right)^{\omega}\|U(t,s)x\|, \text{ for all } (t,s,x) \in \Delta \times X.$$

Definition 2.9. The evolution operator U is said to be *uniformly polynomially instable (u.p.is.)* if there exist $N \geq 1$ and $\nu > 0$ such that:

$$\|U(s,t_0)x_0\| \leq N\left(\frac{s+1}{t+1}\right)^{\nu}\|U(t,t_0)x_0\|, \text{ for all } (t,s,t_0,x_0) \in T \times X.$$

Remark 2.10. The evolution operator U is uniformly polynomially instable if and only if there exist $N \geq 1$ and $\nu > 0$ such that:

$$\|x\| \leq N \left(\frac{s+1}{t+1}\right)^{\nu} \|U(t,s)x\|, \text{ for all } (t,s,x) \in \Delta \times X.$$

Remark 2.11. The connections between the instability concepts and the decay properties mentioned above are given by the following diagram:

$$\begin{array}{ccc} u.e.is. & \Rightarrow & u.p.is. \\ \Downarrow & & \Downarrow \\ u.e.d. & \Leftarrow & u.p.d. \end{array}$$

The converse implications are not true (see [7], [13]).

We define $U_1 : \Delta \to B(X),\ U_1(t,s) = U(e^t - 1, e^s - 1)$ the evolution operator associated to U.

Proposition 2.12. *The evolution operator $U : \Delta \to \mathcal{B}(X)$ has uniform polynomial decay if and only if the evolution operator $U_1 : \Delta \to \mathcal{B}(X)$ has uniform exponential decay.*

Proof. Necessity. We suppose that U has u.p.d. which implies that there exist the constants $M \geq 1$, $\omega > 0$ such that

$$\left(\frac{s+1}{t+1}\right)^{\omega} \|x\| \leq M\|U(t,s)x\|, \text{ for all } (t,s,x) \in \Delta \times X,$$

which implies

$$\left(\frac{e^s}{e^t}\right)^{\omega} \|x\| \leq M\|U(e^t - 1, e^s - 1)x\|,$$

which is equivalent to

$$e^{-\omega(t-s)}\|x\| \leq M\|U(e^t - 1, e^s - 1)x\| = M\|U_1(t,s)x\|, \text{ for all } (t,s,x) \in \Delta \times X.$$

Then U_1 has u.e.d.
Sufficiency. We suppose that U_1 has u.e.d., which implies that there exist $M \geq 1$, $\omega > 0$ such that

$$e^{-\omega(t-s)}\|x\| \leq M\|U_1(t,s)x\| = M\|U(e^t - 1, e^s - 1)x\|, \tag{2.1}$$

for all $(t,s,x) \in \Delta \times X$.
We denote by $e^t - 1 = u,\ e^s - 1 = v$, which implies $t = \ln(1+u), s = \ln(1+v)$. Then, from relation (2.1) we obtain

$$\left(\frac{1+u}{1+v}\right)^{-\omega} \|x\| \leq M\|U(u,v)x\|, \text{ for all } (u,v,x) \in \Delta \times X,$$

which implies that U has u.p.d. $\square$

Proposition 2.13. *The evolution operator $U : \Delta \to \mathcal{B}(X)$ is uniformly polynomially instable if and only if $U_1 : \Delta \to \mathcal{B}(X)$ is uniformly exponentially instable.*

Proof. For the necessity, we suppose that U u.p.is. Then,

$$\begin{aligned} N\|U_1(t,s)x\| &= N\|U(e^t-1, e^s-1)x\| \\ &\geq \left(\frac{e^t}{e^s}\right)^{\nu} \|x\| \\ &= e^{\nu(t-s)}\|x\|, \end{aligned}$$

for all $(t,s,x) \in \Delta \times X$, which implies U_1 is u.e.is.
Conversely, we suppose that U_1 is u.e.is. Then,

$$\begin{aligned} N_1\|U_1(t,s)x\| &= N_1\|U(e^t-1, e^s-1)x\| \\ &= N_1\|U(u,v)x\| \\ &\geq e^{\nu(\ln(1+u)-\ln(1+v))}\|x\| \\ &= e^{\nu \ln \frac{1+u}{1+v}}\|x\| \\ &= \left(\frac{1+u}{1+v}\right)^{\nu} \|x\|, \end{aligned}$$

for all $(t,s,x) \in \Delta \times X$, which implies that U is u.p.is. $\square$

3. The main results

In this section we give some characterization theorems of Datko type for the uniform exponential instability and uniform polynomial instability for evolution operators in Banach spaces.

Theorem 3.1. *Let U be a strongly measurable evolution operator with uniform exponential decay. Then U is uniformly exponentially instable if and only if there exist the constants $D > 1$ and $d \in [0,1)$ such that*

$$\int_s^{\infty} \frac{e^{dt}}{\|U(t,t_0)x_0\|} dt \leq D \frac{e^{ds}}{\|U(s,t_0)x_0\|},$$

for all $(s,t_0,x_0) \in \Delta \times X,\ U(s,t_0)x_0 \neq 0$.

Proof. Necessity. Let $d \in (0,\nu)$. We suppose that U is u.e.is. Then,

$$\begin{aligned} \int_s^{\infty} \frac{e^{dt}}{\|U(t,t_0)x_0\|} dt &\leq N \int_s^{\infty} \frac{e^{dt}e^{-\nu(t-s)}}{\|U(s,t_0)x_0\|} dt = \frac{Ne^{\nu s}}{\|U(s,t_0)x_0\|} \int_s^{\infty} e^{(d-\nu)t} dt \\ &= \frac{N}{\nu - d} \cdot e^{ds}\|U(s,t_0)x_0\| \leq De^{ds}\|U(s,t_0)x_0\|, \end{aligned}$$

where $D = 1 = \dfrac{N}{\nu - d}$.

Sufficiency. Case 1. Let $d \in (0,1)$. For $t \geq s+1$ we obtain

$$\begin{aligned}
\frac{e^{dt}}{\|U(t,t_0)x_0\|} &= \int_{t-1}^{t} \frac{e^{dt}}{\|U(t,t_0)x_0\|} d\tau = \int_{t-1}^{t} \frac{e^{dt}}{\|U(t,\tau)U(\tau,t_0)x_0\|} d\tau \\
&\leq M \int_{t-1}^{t} \frac{e^{dt}e^{\omega(t-\tau)}}{\|U(\tau,t_0)x_0\|} d\tau = M \int_{t-1}^{t} \frac{e^{d\tau} \cdot e^{(d+\omega)(t-\tau)}}{\|U(\tau,t_0)x_0\|} d\tau \\
&\leq Me^{d+\omega} \int_{s}^{\infty} \frac{e^{d\tau}}{\|U(\tau,t_0)x_0\|} d\tau \leq \frac{Ne^{ds}}{\|U(s,t_0)x_0\|},
\end{aligned}$$

which is equivalent to

$$e^{d(t-s)}\|U(s,t_0)x_0\| \leq N\|U(t,t_0)x_0\|, \forall t \geq s+1, s \geq 0. \tag{3.1}$$

Let $t \in [s, s+1]$.

$$\begin{aligned}
e^{d(t-s)}\|U(s,t_0)x_0\| &\leq Me^{(d+\omega)(t-s)}\|U(t,t_0)x_0\| \leq Me^{d+\omega}\|U(t,t_0)x_0\| \\
&\leq N\|U(t,t_0)x_0\|,
\end{aligned}$$

which is equivalent to

$$e^{d(t-s)}\|U(s,t_0)x_0\| \leq N\|U(t,t_0)x_0\|, \forall t \in [s,s+1], s \geq 0. \tag{3.2}$$

From (3.1) and (3.2) we obtain that

$$\|U(s,t_0)x_0\| \leq Ne^{-d(t-s)}\|U(t,t_0)x_0\|, \forall (t,s,x_0) \in \Delta \times X,$$

where $N = 1 + DMe^{d+\omega}$, so U is u.p.is.
Case 2. For $d = 0$, see [3]. □

Theorem 3.2. *Let U be a strongly measurable evolution operator with uniform exponential decay. Then U is uniformly exponentially instable if and only if there exist the constants $D > 1$ and $d \in [0,1)$ such that*

$$\int_{t_0}^{t} \frac{\|U(s,t_0)x_0\|}{e^{ds}} ds \leq D\frac{\|U(t,t_0)x_0\|}{e^{dt}},$$

for all $(t, t_0, x_0) \in \Delta \times X$.

Proof. Case 1. For $d \in (0,1)$ see[12].
Case 2. For $d = 0$ see [14] . □

Theorem 3.3. *Let U be a strongly measurable evolution operator with uniform polynomial decay. Then U is uniformly polynomially instable if and only if there exist $D > 1$ and $d \in [0,1)$ such that*

$$\int_{s}^{\infty} \frac{(t+1)^{d-1}}{\|U(t,t_0)x_0\|} dt \leq \frac{D(s+1)^d}{\|U(s,t_0)x_0\|},$$

for all $(s, t_0, x_0) \in \Delta \times X$, $U(s,t_0)x_0 \neq 0$.

Proof. Case 1. Let $d \in (0,1)$. From Proposition 2.12 we have that U has u.p.d. is equivalent to U_1 has u.e.d. and from Proposition 2.13 we have that U u.p.is. is equivalent to U_1 u.e.is. which means from Theorem 3.1 that there exist $D > 1$ and $d \in [0,1)$ such that

$$\int_s^\infty \frac{e^{dt}}{\|U_1(t,t_0)x_0\|}dt \le \frac{De^{ds}}{\|U_1(s,t_0)x_0\|},$$

that is equivalent to

$$\int_s^\infty \frac{e^{dt}}{\|U(e^t-1,e^{t_0}-1)x_0\|}dt \le \frac{De^{ds}}{\|U(e^s-1,e^{t_0}-1)x_0\|}. \tag{3.3}$$

Using the change of variables $e^t - 1 = u$ and denoting by $v_0 = e^{t_0} - 1,\ u_0 = e^s - 1$, relation (3.3) becomes

$$\int_{u_0}^\infty \frac{e^{d\ln(u+1)}}{\|U(u,v_0)x_0\|}\cdot\frac{du}{u+1} \le \frac{De^{d\ln(u_0+1)}}{\|U(u_0,v_0)x_0\|},$$

that is equivalent to

$$\int_{u_0}^\infty \frac{(u+1)^{d-1}}{\|U(u,v_0)x_0\|}du \le \frac{D(u_0+1)^d}{\|U(u_0,v_0)x_0\|},$$

so the theorem is proved.
Case 2. For $d = 0$ see [4]. □

Theorem 3.4. *Let U be a strongly measurable evolution operator with uniform polynomial decay. Then U is uniformly polynomially instable if and only if there exist $D > 1$ and $d \in [0,1)$ such that*

$$\int_{t_0}^t \frac{\|U(s,t_0)x_0\|}{(s+1)^{d+1}}ds \le \frac{D\|U(t,t_0)x_0\|}{(t+1)^d},$$

for all $(t,t_0,x_0) \in \Delta \times X$.

Proof. Using Proposition 2.12 and Proposition 2.13 we have that U_1 is u.e.is. with u.e.d. and from Theorem 3.1 we obtain that there exist $D > 1$ and $d \in [0,1)$ such that

$$\int_{t_0}^t \frac{\|U_1(s,t_0)x_0\|}{e^{ds}}ds \le \frac{D\|U(t,t_0)x_0\|}{e^{dt}},$$

which is equivalent to

$$\int_{t_0}^t \frac{\|U(e^s-1,e^{t_0}-1)x_0\|}{e^{ds}}ds \le \frac{D\|U(e^t-1,e^{t_0}-1)x_0\|}{e^{dt}}. \tag{3.4}$$

Using the change of variables $e^s - 1 = u$ and denoting by $v = e^t - 1,\ v_0 = e^{t_0} - 1$, relation (3.4) becomes

$$\int_{u_0}^{v} \frac{\|U(u,v_0)x_0\|}{e^{d\ln(u+1)}} \cdot \frac{du}{u+1} \leq \frac{D\|U(v,v_0)x_0\|}{(v+1)^d},$$

that is equivalent to

$$\int_{u_0}^{v} \frac{\|U(u,v_0)x_0\|}{(u+1)^{d+1}} du \leq \frac{D\|U(v,v_0)x_0\|}{(v+1)^d},$$

so the theorem is proved. $\square$

References

[1] Barreira, L., Valls, C., *Polynomial growth rates*, Nonlinear Anal., **71**(2009), 5208-5219.

[2] Bento, A., Silva, C., *Stable manifolds for nonuniform polynomial dichotomies*, J. Funct. Anal., **257**(2009), 122-148.

[3] Biriş, L., *On uniform exponential instability property of evolution operators in Banach spaces*, An. Univ. Vest Timiş. Ser. Mat.-Inform. **47**(2009), no. 1, 3-9.

[4] Boruga, R., Megan, M., *On Some Concepts Regarding the Polynomial Behaviors for Evolution Operators in Banach Spaces*, International Symposium Research and Education in an Innovation Era Conference, Mathematics and Computer Science, 2018, 18-24.

[5] Datko, R., *Uniform Asymptotic Stability of Evolutionary Processes in a Banach Space*, SIAM J. Math. Anal., **3**(1972), 428-445.

[6] Hai, P.V., *Polynomial Stability and Polynomial Instability for Skew-Evolution Semiflows*, Results Math., **74**(2019), no. 4, art. 175.

[7] Megan, M., Boruga (Toma), R., *Barbashin conditions for uniform instability of evolution operators*, (to appear in Stud. Univ. Babeş-Bolyai Math.).

[8] Megan, M., Sasu, A.L., Sasu, B., *The Asymptotic Behavior of Evolution Families*, Mirton Publishing House, 2003.

[9] Megan, M., Sasu, A.L., Sasu, B., *Banach function spaces and exponential instability of evolution families*, Arch. Math. (Brno), **39**(2003), 277-286.

[10] Megan, M., Stoica, C., *Exponential instability of skew-evolution semiflows in Banach spaces*, Stud. Univ. Babeş-Bolyai Math., **53**(2008), no. 1, 17-24.

[11] Mihiţ, C.L., *On uniform h-stability of evolution operators in Banach spaces*, Theory Appl. Math. Comput. Sci., **6**(2016), 19-27.

[12] Popa, I.L., *Nonuniform exponential instability for evolution operators in Banach spaces*, Proc. of the 12^{th} Symposium of Math. and its Appl. "Politehnica" Univ. of Timişoara, November 5-7, 2009.

[13] Rămneanţu, M.L., Megan, M., Ceauşu, T., *Polynomial instability of evolution operators in Banach spaces*, Carpathian J. Math., **29**(2013), no. 1, 77-83.

[14] Stoica, C., Megan, M., *Uniform exponential instability of evolution operators in Banach spaces*, An. Univ. Vest Timiş. Ser. Mat.-Inform., **44**(2006), no. 2, 143-148.

Rovana Boruga (Toma)
West University of Timişoara,
Faculty of Mathematics and Computer Science,
Department of Mathematics,
V. Pârvan Blv. 4,
300223 Timişoara, Romania
e-mail: rovanaboruga@gmail.com

Mihail Megan
Academy of Romanian Scientists,
Independenţei 54,
050094 Bucharest, Romania
and
West University of Timişoara,
Faculty of Mathematics and Computer Science,
Department of Mathematics,
V. Pârvan Blv. 4,
300223 Timişoara, Romania
e-mail: mihail.megan@e-uvt.ro

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 123–126
DOI: 10.24193/subbmath.2021.1.11

The sum theorem for maximal monotone operators in reflexive Banach spaces revisited

Mircea D. Voisei

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. The goal of this note is to present a new shorter proof for the maximal monotonicity of the Minkowski sum of two maximal monotone multi-valued operators defined in a reflexive Banach space under the classical interiority condition involving their domains.

Mathematics Subject Classification (2010): 47H05, 46N10.

Keywords: Maximal monotone operator, Minkowski sum.

1. Preliminaries

Recall the following sum rule for maximal monotone operators:

Theorem 1.1. *(Rockafellar* [5, Theorem 1 (a), p. 76]*) Let* $(X, \|\cdot\|)$ *be a reflexive Banach space with topological dual* X^* *and let* $A,\ B : X \rightrightarrows X^*$ *be multi-valued maximal monotone operators from* X *to* X^**. If* $D(A) \cap \operatorname{int} D(B) \neq \emptyset$ *then* $A + B$ *is maximal monotone.*

Here $D(T) := \{x \in X \mid T(x) \neq \emptyset\}$ is the *domain* of $T : X \rightrightarrows X^*$, "$\operatorname{int} S$" denotes the topological interior of $S \subset X$, and $A + B : X \rightrightarrows X^*$ is the *Minkowski sum* of A and B defined by

$$(A + B)(x) := A(x) + B(x) := \{y + v \mid y \in A(x), v \in B(x)\},$$

for $x \in D(A + B) := D(A) \cap D(B)$.

The proof of [5, Theorem 1, p. 76] relies on the use of the duality mapping J of X and the (Minty's style) characterization of maximal monotone operators defined in reflexive Banach spaces. Similar arguments are used in the presence of an improved qualification constraint in a second proof of Theorem 1.1 (see [2, Corollary 3.5, p. 286]). A third proof of the main theorem involves the exact convolution of some specially constructed functions based on the Fitzpatrick functions of A and B (see [10, Corollary

4, p. 1166]). A different proof of Theorem 1.1 is based on the dual-representability $A+B$ in the presence of the qualification constraint (see [8, Remark 1, p. 276]) and the fact that in a reflexive Banach space dual-representability is equivalent to maximal monotonicity (see e.g. [1, Theorem 3.1, p. 2381]). All the previously mentioned proofs make use of the duality mapping J which is characteristic to a normed space.

Our proof relies on the normal cone, is based on full-range characterizations of maximal monotone operators with bounded domain, and uses the representability of sums of representable operators, but, avoids the use of J or the norm. The following intermediary result, is the main ingredient of our argument.

Theorem 1.2. *Let X be a reflexive Banach space, let $T : X \rightrightarrows X^*$ be maximal monotone, and let $C \subset X$ be closed convex and bounded. If $D(T) \cap \operatorname{int} C \neq \emptyset$ then $T + N_C$ is maximal monotone.*

Here N_C denotes the normal cone to C and is defined by $x^* \in N_C(x)$ if, for every $y \in C$, $\langle y - x, x^* \rangle \leq 0$. Here $\langle \cdot, \cdot \rangle$ denotes the *coupling* or *duality product* of $X \times X^*$ and is defined by

$$c(x, x^*) := \langle x, x^* \rangle := x^*(x), \ x \in X, \ x^* \in X^*.$$

An element $z = (x, x^*) \in X \times X^*$ is *monotonically related* (m.r. for short) *to* T if, for every $(a, a^*) \in \operatorname{Graph} T := \{(u, u^*) \in X \times X^* \mid u \in D(T), \ u^* \in T(u)\}$, $\langle x - a, x^* - a^* \rangle \geq 0$.

Recall that a multi-valued operator $T : X \rightrightarrows X^*$ is

• *monotone* if, for every $x_1^* \in T(x_1)$, $x_2^* \in T(x_2)$, $\langle x_1 - x_2, x_1^* - x_2^* \rangle \geq 0$.

• *maximal monotone* if every m.r. to T element $z = (x, x^*) \in X \times X^*$ belongs to $\operatorname{Graph} T$.

• *representable* if there is a proper convex $s_X \times w^*-$lower semicontinuous $h : X \times X^* \to \mathbb{R} \cup \{+\infty\}$ such that $h \geq c$ and

$$\operatorname{Graph} T = [h = c] := \{(x, x^*) \in X \times X^* \mid h(x, x^*) = \langle x, x^* \rangle\}.$$

Here s_X denotes the strong topology of X and w^* stands for the weak-star topology of X^*.

• *NI* if $\varphi_T \geq c$, where φ_T is the Fitzpatrick function of T which is defined by

$$\varphi_T(x, x^*) := \sup\{\langle x - a, a^* \rangle + \langle a, x^* \rangle \mid (a, a^*) \in \operatorname{Graph} T\}, \ (x, x^*) \in X \times X^*. \quad (1.1)$$

2. Proofs of the main result

Proof of Theorem 1.2. The operator $T + N_C$ is representable, which follows from the facts that T, N_C are maximal monotone thus representable and $D(T) \cap \operatorname{int} C \neq \emptyset$ (see e.g. [6, Corollary 5.6, p. 470] or [7, Theorem 16, p. 818]).

We prove that $R(T + N_C) = X^*$ which implies that $T + N_C$ is of NI–type and so it is maximal monotone (see [6, Theorem 3.4, p. 465] or [8, Theorem 1 (ii), (7)]).

It suffices to prove that $0 \in R(T + N_C)$ otherwise we replace T by $T - x^*$ for an arbitrary $x^* \in X^*$.

Consider $F(x,x^*) := \varphi_T(x,x^*) + g(x,x^*)$, with $g(x,x^*) := \iota_C(x) + \sigma_C(-x^*)$, where $\iota_C(x) = 0$, for $x \in C$; $\iota_C(x) = +\infty$, otherwise, and $\sigma_C(x^*) := \sup_{x\in C}\langle x,x^*\rangle$, $x^* \in X^*$.

Then $F \geq 0$ due to $\varphi_T(x,x^*) \geq \langle x,x^*\rangle$ and $\iota_C(x) + \sigma_C(-x^*) \geq -\langle x,x^*\rangle$ (see f.i. [4]). Hence

$$0 \leq \inf_{X\times X^*} F = -(\varphi_T + g)^*(0,0) = -\min_{(x,x^*)\in X\times X^*}\{\psi_T(x,x^*) + g^*(-x^*,-x)\}, \quad (2.1)$$

because C is bounded, g is $s_X \times s_{X^*}-$continuous on $\operatorname{int} C \times X^*$, and X is reflexive (see f.i. [9, Theorem 2.8.7, p. 126]), where s_{X^*} is the strong topology of X^*. Here "min" denotes an infimum that is attained when finite,

$$\psi_T(x,x^*) = \varphi_T^*(x^*,x), \ (x,x^*) \in X \times X^*, \quad (2.2)$$

the convex conjugation being taken with respect to the dual system

$$(X \times X^*, X^* \times X^{**})$$

and, for every $(x,x^*) \in X \times X^*$, $\psi_T(x,x^*) \geq \langle x,x^*\rangle$ because T is monotone (see e.g. [8, (12)]).

From $g^*(x^*,x) = \iota_C(-x) + \sigma_C(x^*)$, $(x,x^*) \in X \times X^*$ and (2.1) there exists $(\bar{x},\bar{x}^*) \in X \times X^*$ such that $\psi_T(\bar{x},\bar{x}^*) + \iota_C(\bar{x}) + \sigma_C(-\bar{x}^*) \leq 0$ which implies that $\iota_C(\bar{x}) + \sigma_C(-\bar{x}^*) = -\langle \bar{x},\bar{x}^*\rangle$, i.e., $-\bar{x}^* \in N_C(\bar{x})$ and $\psi_T(\bar{x},\bar{x}^*) = \langle \bar{x},\bar{x}^*\rangle$, that is, $\bar{x}^* \in T(\bar{x})$ since T is representable (see [8, Theorem 1, p. 270]).

Therefore $0 \in (T + N_C)(\bar{x},\bar{x}^*)$ and so $0 \in R(T + N_C)$. □

Proof of Theorem 1.1. First we prove that we can assume without loss of generality that $D(B)$ is bounded. Indeed, assume that the result is true for that case.

Let $z = (x,x^*)$ be m.r. to $A + B$. Take $C \subset X$ closed convex and bounded with $x \in \operatorname{int} C$ and $D(A) \cap \operatorname{int} D(B) \cap \operatorname{int} C \neq \emptyset$ e.g. $C := [x_0,x] + U$, where

$$[x_0,x] := \{tx_0 + (1-t)x \mid 0 \leq t \leq 1\}$$

and U is a closed convex bounded neighborhood of 0, and $x_0 \in D(A) \cap \operatorname{int} D(B)$. Note that z is m.r. to $A + B + N_C = A + (B + N_C)$ which is maximal monotone since, according to Theorem 1.2, $B + N_C$ is maximal monotone, $D(B + N_C)$ is bounded, and $x_0 \in D(A) \cap \operatorname{int} D(B + N_C) \neq \emptyset$. Hence $z \in \operatorname{Graph}(A + B + N_C)$ or $x^* \in (A + B)(x)$ because $N_C(x) = \{0\}$. Therefore $A + B$ is maximal monotone.

It remains to prove that, whenever $D(B)$ is bounded, $R(A + B) = X^*$ or sufficiently $0 \in R(A + B)$ (since $A + B$ is representable, see again [6, Corollary 5.6]).

Let

$$F(x,x^*) := \varphi_A(x,x^*) + \varphi_B(x,-x^*), \quad g(x,x^*) := \varphi_B(x,-x^*), \quad (x,x^*) \in X \times X^*.$$

Since A, B are maximal monotone, for every $(x,x^*) \in X \times X^*$,

$$\min\{\varphi_A(x,x^*), \varphi_B(x,x^*)\} \geq \langle x,x^*\rangle$$

which imply $F \geq 0$ and so

$$0 \leq \inf_{X\times X^*} F = -(\varphi_A + g)^*(0,0) = -\min_{(x,x^*)\in X\times X^*}\{\psi_A(x,x) + \psi_B(x,-x^*)\}, \quad (2.3)$$

because $D(B)$ bounded provides $D(B) \times X^* \subset \operatorname{dom} g$, g is $s_X \times s_{X^*}-$continuous on $\operatorname{int} D(B) \times X^*$, and X is reflexive (see again [9, Theorem 2.8.7, p. 126]). More precisely, for every $(x, x^*) \in D(B) \times X^*$ there is $\overline{x}^* \in B(x)$ and so

$$\begin{aligned}\varphi_B(x, x^*) &:= \sup\{\langle x - b, b^*\rangle + \langle b, x^*\rangle \mid (b, b^*) \in \operatorname{Graph} B\}\\ &\le \sup\{\langle x - b, \overline{x}^*\rangle + \langle b, x^*\rangle \mid (b, b^*) \in \operatorname{Graph} B\}\\ &\le \langle x, \overline{x}^*\rangle + \|x^* - \overline{x}^*\| \sup_{b \in D(B)} \|b\| < +\infty.\end{aligned}$$

There exists $(\bar{x}, \bar{x}^*) \in X \times X^*$ such that $\psi_A(\bar{x}, \bar{x}^*) + \psi_B(\bar{x}, -\bar{x}^*) \le 0$ which implies that $\psi_A(\bar{x}, \bar{x}^*) = \langle \bar{x}, \bar{x}^*\rangle$, $\psi_B(\bar{x}, -\bar{x}^*) = -\langle \bar{x}, \bar{x}^*\rangle$, i.e., $\bar{x}^* \in A(\bar{x})$ and $-\bar{x}^* \in B(\bar{x})$ from which $0 \in R(A + B)$. □

Remark 2.1. Theorem 1.2 still holds if we replace the assumption C bounded with $D(T)$ bounded. In this case an alternate proof of Theorem 1.1 can be performed with $A + N_C$ instead of A and a similar argument as in the current proof.

References

[1] Burachik, R.S., Svaiter, B.F., *Maximal monotonicity, conjugation and the duality product*, Proc. Amer. Math. Soc., **131**(2003), no. 3, 2379-2383 (electronic)

[2] Chu, L.-J., *On the sum of monotone operators*, Michigan Math. J., **43**(1996), no. 2, 273-289.

[3] Debrunner, H., Flor, P., *Ein Erweiterungssatz für monotone Mengen*, Arch. Math., **15**(1964), 445-447.

[4] Fitzpatrick, S., *Representing monotone operators by convex functions*, In" "Workshop/Miniconference on Functional Analysis and Optimization" (Canberra, 1988), vol. 20 of Proc. Centre Math. Anal. Austral. Nat. Univ., Austral. Nat. Univ., Canberra, 1988, 59-65.

[5] Rockafellar, R.T., *On the maximality of sums of nonlinear monotone operators*, Trans. Amer. Math. Soc., **149**(1970), 75-88.

[6] Voisei, M.D., *The sum and chain rules for maximal monotone operators*, Set-Valued Anal., **16**(2008), no. 4, 461-476.

[7] Voisei, M.D., *Maximal monotone normal cones in locally convex spaces*, J. Math. Anal. Appl., **476**(2019), no. 2, 811-818.

[8] Voisei, M.D., Zălinescu, C., *Maximal monotonicity criteria for the composition and the sum under weak interiority conditions*, Math. Program., **123**(2010), no. 1, Ser. B, 265-283.

[9] Zălinescu, C., *Convex Analysis in General Vector Spaces*, World Scientific Publishing Co. Inc., River Edge, NJ, 2002.

[10] Zălinescu, C., *A new proof of the maximal monotonicity of the sum using the Fitzpatrick function*, In: Variational analysis and applications, vol. 79 of Nonconvex Optim. Appl., Springer, New York, 2005. 1159-1172.

Mircea D. Voisei
Baltimore, MD-21252, U.S.A.
e-mail: mdvoisei@yahoo.com

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 127–138
DOI: 10.24193/subbmath.2021.1.12

Asymptotic behavior of inexact infinite products of nonexpansive mappings

Simeon Reich and Alexander J. Zaslavski

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary

Abstract. We analyze the asymptotic behavior of inexact infinite products of nonexpansive mappings, which take a nonempty closed subset of a complete metric space into the space, in the case where the errors are sufficiently small.

Mathematics Subject Classification (2010): 47H09, 47H10, 54E50.

Keywords: Complete metric space, fixed point, infinite product, nonexpansive mapping.

1. Introduction

For nearly sixty years now, there has been an intensive research activity regarding the fixed point theory of nonexpansive (that is, 1-Lipschitz) mappings. See, for example, [2, 4, 5, 3, 12, 13, 14, 15, 16, 17, 18, 19, 21, 23, 24, 27, 28, 29, 30, 32, 34, 38, 39, 40, 41] and references cited therein. This activity stems from Banach's classical theorem [1] concerning the existence of a unique fixed point for a strict contraction. It also covers the convergence of (inexact) orbits of a nonexpansive mapping to one of its fixed points. Since that seminal result, numerous developments have taken place in this field including, in particular, studies of feasibility and common fixed point problems, which find important applications in engineering and medical sciences [6, 7, 11, 35, 36, 37, 40, 41]. In this connection, see also the results regarding the asymptotic behavior of infinite products of nonexpansive mappings which have been established in, for instance, [8, 9, 10, 20, 22, 25, 26, 31] and references mentioned therein.

In [30] we collected several results which demonstrate the convergence of inexact iterates of a nonexpansive self-mapping of a complete metric space to one of its fixed points. In the present paper we establish three variants of these results for inexact infinite products of nonexpansive mappings, which take a nonempty closed subset of a complete metric space into the space, in the case where the errors are sufficiently small. Prototypes of these results for inexact orbits of nonexpansive mappings have recently been obtained in [33].

2. Main results

Let (Z,d) be a complete metric space. For each point $z \in Z$ and each positive number M, define

$$B(z,M) := \{y \in Z : \; d(z,y) \le M\}.$$

For every point $z \in Z$ and every nonempty set $D \subset Z$, put

$$d(z,D) := \inf\{d(z,y) : \; y \in D\}.$$

Let $K \subset Z$ be a nonempty closed set, let the mappings $A_j : K \to Z$, $j = 1,2,\dots,$ satisfy

$$d(A_j(z), A_j(y)) \le d(z,y) \text{ for all } z,y \in K \tag{2.1}$$

for each integer $j \ge 1$, and let $\mathcal{R}$ be a nonempty collection of mappings $r : \{1,2,\dots\} \to \{1,2,\dots\}$. Fix a point $\theta \in K$. Assume that a point $z_* \in K$ satisfies

$$A_j(z_*) = z_*, \; j = 1,2,\dots, \tag{2.2}$$

and that the following two properties hold:

(P1) if $r \in \mathcal{R}$ and $q \ge 1$ is an integer, then the mapping $n \to r(n+q)$, $n = 1,2\dots,$ belongs to the collection $\mathcal{R}$;

(P2) for every positive number ε and every positive number M, there exists an integer $n(M,\varepsilon) \ge 1$ such that if $z \in B(\theta, M)$, $r \in \mathcal{R}$ and

$$\prod_{i=1}^{n(M,\varepsilon)} A_{r(i)}(z)$$

exists, then

$$d\left(\prod_{i=1}^{n(M,\varepsilon)} A_{r(i)}(z), z_*\right) \le \varepsilon.$$

Note that property (P2) holds for Banach contractions and for many nonexpansive mappings of contractive type [30]. In [30] we consider a large class of sequences of nonexpansive mappings $\{A_i\}_{i=1}^{\infty}$ and show that a generic (typical) sequence of mappings possesses (P2).

In the present paper we establish the following three theorems.

Theorem 2.1. *Let a pair of positive numbers ε, M be given. Then there exists an integer $n_0 \ge 1$ such that for every $\delta \in (0, \varepsilon/2)$, every natural number $n \ge n_0$, every mapping $r \in \mathcal{R}$ and every sequence $\{z_i\}_{i=0}^{n} \subset K$ satisfying*

$$d(z_0, \theta) \le M,$$

$$d(z_{j+1}, A_{r(j+1)}(z_j)) \le (4n_0)^{-1}\delta, \; j = 0,\dots,n-1,$$

and

$$B(z_j, \delta) \subset K, \; j = 0,\dots,n,$$

the inequality $d(z_j, z_*) \le \varepsilon$ is true for all integers $j = n_0,\dots,n$.

Theorem 2.2. *Let* $r_* \in (0,1)$ *satisfy*

$$B(z_*, r_*) \subset K, \tag{2.3}$$

and let a pair of positive numbers M *and* $\varepsilon \in (0, r_*/2)$ *be given. Then there exists an integer* $n_0 \geq 1$ *such that for every* $\delta \in (0, \varepsilon/2)$*, every natural number* $n \geq n_0$*, every mapping* $r \in \mathcal{R}$ *and every sequence* $\{z_j\}_{j=0}^{n} \subset K$ *satisfying*

$$d(z_0, \theta) \leq M,$$

$$d(z_{j+1}, A_{r(j+1)}(z_j)) \leq (4n_0)^{-1}\delta,\ j = 0, \ldots, n-1,$$

and

$$B(z_j, \delta) \subset K,\ j = 0, \ldots, n_0,$$

the inequality $d(z_j, z_*) \leq \varepsilon$ *is true for all integers* $j = n_0, \ldots, n$.

Note that in Theorem 2.1 the sequence $\{z_j\}_{j=0}^{n} \subset K$ satisfies the inclusion $B(z_j, \delta) \subset K$ for all $j = 0, \ldots, n$, while in Theorem 2.2 the inclusion holds only for $j = 0, \ldots, n_0$. On the other hand, in Theorem 2.2 we assume that z_* is an interior point of K. We do not need this assumption for Theorem 2.1.

Theorem 2.3. *Let* $r_* > 0$,

$$B(z_*, r_*) \subset K, \tag{2.4}$$

$r \in \mathcal{R}$ *and let a sequence* $\{z_j\}_{j=0}^{\infty} \subset K$ *satisfy*

$$\lim_{j\to\infty} d(z_{j+1}, A_{r(j+1)}(z_j)) = 0 \tag{2.5}$$

and have a bounded subsequence $\{z_{j_p}\}_{p=1}^{\infty}$. *Assume that there exists a positive number* Δ *such that*

$$B(z_j, \Delta) \subset K$$

for all sufficiently large natural numbers j. *Then*

$$\lim_{j\to\infty} z_j = z_*.$$

The proofs of these three theorems are presented in Sections 4–6 below. We begin, however, with an auxiliary result which is stated and proved in the next section.

3. An auxiliary result

Proposition 3.1. *Let* $\varepsilon, M > 0$ *be given. Then there exists an integer* $n_0 \geq 1$ *such that for every number* $\delta \in (0, \varepsilon)$*, every mapping* $r \in \mathcal{R}$ *and every sequence* $\{z_j\}_{j=0}^{n_0} \subset K$ *satisfying*

$$d(z_0, \theta) \leq M, \tag{3.1}$$

$$d(z_{j+1}, A_{r(j+1)}(z_j)) \leq (4n_0)^{-1}\delta,\ j = 0, \ldots, n_0 - 1, \tag{3.2}$$

and

$$B(z_j, \delta) \subset K,\ j = 0, \ldots, n_0, \tag{3.3}$$

the inequality $d(z_{n_0}, z_*) \leq \varepsilon$ *holds true.*

Proof. In view of property (P2), there exists an integer $n_0 \geq 1$ such that the following property holds:

(P3) for every point $z \in B(\theta, M)$ and every mapping $r \in \mathcal{R}$ for which $\prod_{j=1}^{n_0} A_{r(j)}(z)$ exists, we have

$$d\left(\prod_{j=1}^{n_0} A_{r(j)}(z), z_*\right) \leq \varepsilon/4.$$

Fix $\delta \in (0, \varepsilon)$ and put

$$\delta_0 := (4n_0)^{-1}\delta. \tag{3.4}$$

Assume that $r \in \mathcal{R}$ and that a sequence of points $\{z_j\}_{j=0}^{n_0} \subset K$ satisfies relations (3.1)–(3.3). Define

$$y_0 := z_0,\ y_1 := A_{r(1)}(z_0). \tag{3.5}$$

By (3.2), (3.4) and (3.5), we have

$$d(y_1, z_1) \leq \delta_0. \tag{3.6}$$

It follows from (3.3), (3.4) and (3.6) that

$$B(z_1, \delta) \subset K,\ y_1 \in K \tag{3.7}$$

and

$$B(y_1, \delta - \delta_0) \subset B(z_1, \delta) \subset K. \tag{3.8}$$

Assume that $1 \leq p < n_0$ is an integer and that a sequence of points $\{y_i\}_{i=0}^{p} \subset K$ satisfies

$$y_0 = z_0, \tag{3.9}$$

$$y_{j+1} = A_{r(j+1)}(y_j),\ j = 0, \dots, p-1, \tag{3.10}$$

and

$$d(y_j, z_j) \leq j\delta_0,\ j = 0, \dots, p. \tag{3.11}$$

(It is clear that by relations (3.5)–(3.8), our assumption is valid for $p = 1$.) We claim that our assumption is true for $p + 1$ too. Indeed, in view of (3.3), (3.4) and (3.11), we have $y_p \in K$ and

$$d(z_p, y_p) \leq p\delta_0. \tag{3.12}$$

Relations (3.4), (3.11) and (3.12) imply that

$$B(y_p, \delta - p\delta_0) \subset B(z_p, \delta) \subset K. \tag{3.13}$$

By (3.13),

$$y_{p+1} = A_{r(p+1)}(y_p) \tag{3.14}$$

is well defined. It now follows from (2.1), (3.2), (3.4), (3.12) and (3.14) that

$$\begin{aligned} d(y_{p+1}, z_{p+1}) &\leq d(y_{p+1}, A_{r(p+1)}(z_p)) + d(A_{r(p+1)}(z_p), z_{p+1}) \\ &\leq d(A_{r(p+1)}(y_p), A_{r(p+1)}(z_p)) + d(A_{r(p+1)}(z_p), z_{p+1}) \\ &\leq d(y_p, z_p) + \delta_0 \leq (p+1)\delta_0. \end{aligned}$$

Thus the assumption made regarding p also holds for $p+1$, as claimed (see (3.9)-(3.11)). This means that we have shown by using induction that our assumption holds for $p = n_0$. Hence there exists a sequence of points $\{y_j\}_{j=0}^{n_0} \subset K$ which satisfies

$$y_0 = z_0, \tag{3.15}$$

$$y_{j+1} = A_{r(j+1)}(y_j),\ j = 0, \dots, n_0 - 1, \tag{3.16}$$

and

$$d(y_j, z_j) \le j\delta_0,\ j = 0, \dots, n_0. \tag{3.17}$$

By (3.15) and (3.16),

$$y_{n_0} = \prod_{j=1}^{n_0} A_{r(j)}(x_0).$$

It follows from property (P3), (3.1) and (3.15)–(3.17) that

$$d(y_{n_0}, z_*) \le \varepsilon/4.$$

When combined with (3.4) and (3.17), this implies that

$$d(z_{n_0}, z_*) \le d(z_{n_0}, y_{n_0}) + d(y_{n_0}, z_*) \le n_0\delta_0 + \varepsilon/4 < \varepsilon.$$

This completes the proof of Proposition 3.1. □

4. Proof of Theorem 2.1

We may assume that

$$M > d(\theta, z_*) + 1 + \varepsilon. \tag{4.1}$$

In view of Proposition 3.1, there exists an integer $n_0 \ge 1$ such that the following property holds:

(P4) for every number $\delta \in (0, \varepsilon/2)$, every mapping $r \in \mathcal{R}$ and every sequence $\{x_j\}_{j=0}^{n_0} \subset K$ satisfying

$$d(x_0, \theta) \le M,$$

$$d(x_{j+1}, A_{r(j+1)}(x_j)) \le (4n_0)^{-1}\delta,\ j = 0, \dots, n_0 - 1,$$

and

$$B(x_j, \delta) \subset K,\ j = 0, \dots, n_0,$$

the inequality

$$d(x_{n_0}, z_*) \le \varepsilon/2$$

holds true.

Fix $\delta \in (0, \varepsilon/2)$. Assume that $n \ge n_0$ is an integer, $r \in \mathcal{R}$ and that a sequence $\{z_j\}_{j=0}^{n} \subset K$ satisfies

$$d(z_0, \theta) \le M, \tag{4.2}$$

$$d(z_{j+1}, A_{r(j+1)}(z_j)) \le (4n_0)^{-1}\delta,\ j = 0, \dots, n - 1, \tag{4.3}$$

and

$$B(z_j, \delta) \subset K,\ j = 0, \dots, n. \tag{4.4}$$

By (P4), the choice of n_0 and relations (4.2)–(4.4), we have

$$d(z_{n_0}, z_*) \le \varepsilon/2. \tag{4.5}$$

In order to complete the proof of the theorem, it suffices to show that

$$d(z_j, z_*) \le \varepsilon,\ j = n_0, \dots, n.$$

To this end, it is sufficient to consider the case where $n > n_0$. Relations (2.1), (2.2), (4.3) and (4.5) imply that

$$\begin{aligned} d(z_{n_0+1}, z_*) &\le d(z_{n_0+1}, A_{r(n_0+1)}(z_{n_0})) + d(A_{r(n_0+1)}(z_{n_0}), z_*) \\ &\le \delta(4n_0)^{-1} + \varepsilon/2. \end{aligned} \tag{4.6}$$

We claim that for each integer $j \in \{n_0+1, \dots, n\}$, we have

$$d(z_j, z_*) \le \varepsilon/2 + \delta(j - n_0)(4n_0)^{-1}. \tag{4.7}$$

By (4.6), relation (4.7) is indeed true for $i = n_0 + 1$.
Assume that $p \in \{n_0+1, \dots, n\} \setminus \{n\}$ and that

$$d(z_p, z_*) \le \varepsilon/2 + \delta(p - n_0)(4n_0)^{-1}. \tag{4.8}$$

Relations (2.1), (2.2), (4.3) and (4.8) imply that

$$\begin{aligned} d(z_{p+1}, z_*) &\le d(z_{p+1}, A_{r(p+1)}(z_p)) + d(A_{r(p+1)}(z_p), z_*) \\ &\le \delta(4n_0)^{-1} + d(z_p, z_*) \\ &\le \varepsilon/2 + \delta(p - n_0)(4n_0)^{-1} + \delta(4n_0)^{-1} \\ &\le \varepsilon/2 + \delta(p + 1 - n_0)(4n_0)^{-1}. \end{aligned}$$

Thus the assumption made regarding p also holds for $p + 1$, as claimed.
This means that we have shown by using induction that (4.7) is indeed true for all integers $j = n_0 + 1, \dots, n$.
Suppose now that there exists an integer $q \in \{n_0, \dots, n\}$ for which

$$d(z_q, z_*) > \varepsilon. \tag{4.9}$$

By (4.5) and (4.9), we have

$$q > n_0.$$

In view of (4.7) and (4.9),

$$\varepsilon < d(z_q, z_*) \le \varepsilon/2 + \delta(q - n_0)(4n_0)^{-1},$$

$$\varepsilon/2 < \delta(q - n_0)(4n_0)^{-1} < (\varepsilon/2)(q - n_0)(4n_0)^{-1},$$

$$q - n_0 > 4n_0$$

and

$$q > 5n_0. \tag{4.10}$$

By (4.5) and (4.10), we may assume without any loss of generality that

$$d(z_j, z_*) \le \varepsilon,\ j = n_0, \dots, q - 1. \tag{4.11}$$

Define

$$x_j := z_{j+q-n_0},\ j = 0, 1, \dots, n_0, \tag{4.12}$$

$$\tilde{r}(j) := r(j + q - n_0),\ j = 1, 2, \dots. \tag{4.13}$$

Property (P1) implies that $\tilde{r} \in \mathcal{R}$. It follows from (4.3), (4.12) and (4.13) that for all integers $j = 0, \ldots, n_0 - 1$, we have

$$d(x_{j+1}, A_{\tilde{r}(j+1)}(x_j)) = d(z_{j+1+q-n_0}, A_{r(j+1+q-n_0)}(z_{j+q-n_0})) \le (4n_0)^{-1}\delta.$$

Property (P4), applied to the sequence x_j, $j = 0, \ldots, n_0$, (4.1), (4.11), (4.12) and (4.14) imply that

$$d(z_*, z_q) = d(z_*, x_{n_0}) \le \varepsilon/2.$$

This, however, contradicts (4.9). The contradiction we have reached completes the proof of Theorem 2.1.

5. Proof of Theorem 2.2

We may assume that

$$M > d(\theta, z_*) + 1 + \varepsilon. \tag{5.1}$$

Recall that $\varepsilon < r_*/2$ (see (2.3)). Proposition 3.1 implies that there exists a natural number n_0 such that property (P4), which was introduced in the previous section, holds. We recall it at this point for the convenience of the reader:

(P4) for every number $\delta \in (0, \varepsilon/2)$, every $r \in \mathcal{R}$ and every sequence of points $\{x_j\}_{j=0}^{n_0} \subset K$ which satisfies

$$d(x_0, \theta) \le M,$$

$$d(x_{j+1}, A_{r(j+1)}(x_j)) \le (4n_0)^{-1}\delta,\ j = 0, \ldots, n_0 - 1,$$

and

$$B(x_j, \delta) \subset K,\ j = 0, \ldots, n_0,$$

the inequality

$$d(x_{n_0}, z_*) \le \varepsilon/2$$

is true.

Let $\delta \in (0, \varepsilon/2)$ be given. Assume that $r \in \mathcal{R}$, $n \ge n_0$ is an integer and that a sequence $\{z_j\}_{j=0}^{n} \subset K$ satisfies

$$d(z_0, \theta) \le M, \tag{5.2}$$

$$d(z_{j+1}, A_{r(j+1)}(z_j)) \le (4n_0)^{-1}\delta,\ j = 0, \ldots, n - 1, \tag{5.3}$$

and

$$B(z_j, \delta) \subset K,\ j = 0, \ldots, n_0. \tag{5.4}$$

Property (P4) and relations (5.2)–(5.4) imply that

$$d(z_{n_0}, z_*) \le \varepsilon/2. \tag{5.5}$$

In order to complete the proof of the theorem, it is sufficient to show that

$$d(z_j, z_*) \le \varepsilon,\ j = n_0, \ldots, n.$$

Suppose to the contrary that these inequalities are not valid. Then there exists a natural number $q \in \{n_0, \ldots, n\}$ for which

$$d(z_q, z_*) > \varepsilon. \tag{5.6}$$

Inequalities (5.5) and (5.6) imply that

$$q > n_0. \tag{5.7}$$

In view of (5.7), we may assume without loss of generality that

$$d(z_j, z_*) \le \varepsilon,\ j = n_0, \dots, q-1. \tag{5.8}$$

Using induction and arguing as in the proof of Theorem 2.1, we can show that for all natural numbers $i = n_0 + 1, \dots, n$, we have

$$d(z_i, z_*) \le \varepsilon/2 + \delta(i - n_0)(4n_0)^{-1}. \tag{5.9}$$

In view of (5.6) and (5.9),

$$\varepsilon < d(z_q, z_*) \le \varepsilon/2 + \delta(q - n_0)(4n_0)^{-1},$$

$$\varepsilon/2 < \delta(q - n_0)(4n_0)^{-1} < (\varepsilon/2)(q - n_0)(4n_0)^{-1},$$

$$q - n_0 > 4n_0$$

and

$$q > 5n_0. \tag{5.10}$$

Put

$$x_j := z_{j+q-n_0},\ j = 0, \dots, n_0, \tag{5.11}$$

$$\tilde{r}(j) := r(j + q - n_0),\ j = 1, 2, \dots. \tag{5.12}$$

It follows from property (P1) that $\tilde{r} \in \mathcal{R}$. In view of (5.3), (5.11) and (5.12), we have, for every integer $j \in \{0, 1, \dots, n_0 - 1\}$,

$$d(x_{j+1}, A_{\tilde{r}(j+1)}(x_j)) = d(z_{j+1+q-n_0}, A_{r(j+1+q-n_0)}(z_{j+q-n_0})) \le (4n_0)^{-1}\delta. \tag{5.13}$$

It follows from property (P4), (5.1), (5.8) and (5.10)–(5.13) that

$$d(z_*, z_q) = d(z_*, x_{n_0}) \le \varepsilon/2.$$

This, however, contradicts (5.6). The contradiction we have reached completes the proof of Theorem 2.2.

6. Proof of Theorem 2.3

We may assume without any loss of generality that

$$B(z_j, \Delta) \subset K,\ j = 0, 1, \dots. \tag{6.1}$$

There exists a number

$$M > d(\theta, z_*) + 1$$

for which

$$d(z_{i_p}, \theta) \le M,\ p = 1, 2, \dots. \tag{6.2}$$

Fix

$$\varepsilon \in (0, r_*/4) \tag{6.3}$$

(see (2.4)). Proposition 3.1 implies that there exists an integer $n_0 \ge 1$ for which property (P4), which was introduced in Section 4, holds. We recall it now for the convenience of the reader:

(P4) for every number $\delta \in (0, \varepsilon/2)$, every mapping $r \in \mathcal{R}$ and every sequence $\{x_j\}_{j=0}^{n_0} \subset K$ which satisfies

$$d(x_0, \theta) \le M,$$

$$d(x_{j+1}, A_{r(j+1)}(x_j)) \le (4n_0)^{-1}\delta,\ j = 0, \dots, n_0 - 1,$$

and

$$B(x_j, \delta) \subset K,\ j = 0, \dots, n_0,$$

the inequality

$$d(x_{n_0}, x_*) \leq \varepsilon/2$$

is true.

Now let

$$\delta := \min\{\varepsilon/4,\ \Delta\}. \tag{6.4}$$

By (6.3) and (6.4),

$$\delta \leq 16^{-1} r_*.$$

Relations (2.5) and (6.2) imply that there exists an integer $p_0 \geq 1$ such that

$$d(z_{i_{p_0}}, \theta) \leq M \tag{6.5}$$

and

$$d(z_{i+1}, A_{r(i+1)}(z_i)) \leq (4n_0)^{-1}\delta \text{ for all integers } i \geq i_{p_0}. \tag{6.6}$$

For all integers $j = 0, \dots, n_0$, define

$$x_j := z_{j+p_0}, \tag{6.7}$$

$$\tilde{r}(i) := r(i + i_{p_0}),\ i = 1, 2, \dots. \tag{6.8}$$

It follows from relations (6.6)–(6.8) that for each integer $j \in \{0, 1, \dots, n_0 - 1\}$, we have

$$d(x_{j+1}, A_{\tilde{r}(j+1)}(x_j)) = d(z_{j+1+i_{p_0}}, A_{r(j+1+i_{p_0})}(z_{j+i_{p_0}})) \leq (4n_0)^{-1}\delta. \tag{6.9}$$

Property (P4) and relations (6.5), (6.7) and (6.9) imply that

$$\varepsilon/2 \geq d(x_{n_0}, z_*) = d(z_{i_{p_0}+n_0}, z_*).$$

In order to complete the proof of the theorem, it is sufficient to show that

$$d(z_j, z_*) \leq \varepsilon \text{ for all integers } j \geq i_{p_0} + n_0.$$

Suppose to the contrary that this is not true. Then there exists a natural number $q > i_{p_0} + n_0$ such that

$$d(z_q, z_*) > \varepsilon. \tag{6.10}$$

We may assume without any loss of generality that

$$d(z_j, z_*) \leq \varepsilon,\ j = i_{p_0} + n_0, \dots, q - 1. \tag{6.11}$$

Using induction and arguing as in the proof of Theorem 2.1, we can show that for each integer $i \geq i_{p_0} + n_0$, we have

$$d(z_i, z_*) \leq \varepsilon/2 + \delta(i - i_{p_0} - n_0)(4n_0)^{-1}. \tag{6.12}$$

Inequalities (6.10) and (6.11) imply that

$$\varepsilon < d(z_q, z_*) \leq \varepsilon/2 + \delta(q - i_{p_0} - n_0)(4n_0)^{-1},$$

$$\varepsilon/2 \leq (\varepsilon/2)(q - i_{p_0} - n_0)(4n_0)^{-1},$$

$$q - i_{p_0} - n_0 \geq 4n_0$$

and

$$q \geq i_{p_0} + 5n_0. \tag{6.13}$$

In view of (6.11) and (6.13), we have

$$d(z_{q-n_0}, z_*) \le \varepsilon. \tag{6.14}$$

By (6.14), we have

$$d(z_{q-n_0}, \theta) \le M. \tag{6.15}$$

For all integers $j = 0, \dots, n_0$, put

$$x_j := z_{q-n_0+j}, \tag{6.16}$$

$$\tilde{r}(j) := r(q - n_0 + j),\ j = 1, 2, \dots. \tag{6.17}$$

It follows from property (P4), (6.1), (6.6), (6.13) and (6.15)–(6.17) that

$$\varepsilon/2 \ge d(z_*, x_{n_0}) = d(z_*, z_q).$$

This, however, contradicts (6.10). The contradiction we have reached completes the proof of Theorem 2.3.

Acknowledgments. The first author was partially supported by the Israel Science Foundation (Grant No. 820/17), by the Fund for the Promotion of Research at the Technion and by the Technion General Research Fund.

References

[1] Banach, S., *Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales*, Fund. Math., **3**(1922), 133-181.

[2] Betiuk-Pilarska, A., Domínguez Benavides, T., *Fixed points for nonexpansive mappings and generalized nonexpansive mappings on Banach lattices*, Pure Appl. Func. Anal., **1**(2016), 343-359.

[3] de Blasi, F.S., Myjak, J., Reich, S., Zaslavski, A.J., *Generic existence and approximation of fixed points for nonexpansive set-valued maps*, Set-Valued Var. Anal., **17**(2009), 97-112.

[4] Bruck, R.E., Kuczumow, T., Reich, S., *Convergence of iterates of asymptotically nonexpansive mappings in Banach spaces with the uniform Opial property*, Colloq. Math., **65**(1993), 169-179.

[5] Butnariu, D., Reich, S., Zaslavski, A. J., *Convergence to fixed points of inexact orbits of Bregman-monotone and of nonexpansive operators in Banach spaces*, Fixed Point Theory Appl., Yokohama Publishers, Yokohama, 2006, 11-32.

[6] Cegielski, A., Reich, S., Zalas, R., *Regular sequences of quasi-nonexpansive operators and their applications*, SIAM J. Optim., **28**(2018), 1508-1532.

[7] Censor, Y., Zaknoon, M., *Algorithms and convergence results of projection methods for inconsistent feasibility problems: a review*, Pure Appl. Func. Anal., **3**(2018), 565-586.

[8] Dye, J.M., Kuczumow, T., Lin, P.-K., Reich, S., *Convergence of unrestricted products of nonexpansive mappings in spaces with the Opial property*, Nonlinear Anal., **26**(1996), 767-773.

[9] Dye, J.M., Reich, S., *Unrestricted iterations of nonexpansive mappings in Hilbert space*, Nonlinear Anal., **18**(1992), 199-207.

[10] Dye, J.M., Reich, S., *Unrestricted itarations of nonexpansive mappings in Banach spaces*, Nonlinear Anal., **19**(1992), 983-992.

[11] Gibali, A., *A new split inverse problem and an application to least intensity feasible solutions*, Pure Appl. Funct. Anal., **2**(2017), 243-258.

[12] Goebel, K., Kirk, W.A., *Topics in Metric Fixed Point Theory*, Cambridge University Press, Cambridge, 1990.

[13] Goebel, K., Kirk, W.A., *Classical Theory of Nonexpansive Mappings*, Handbook of Metric Fixed Point Theory, Kluwer, Dordrecht, 2001, 49-91.

[14] Goebel, K., Reich, S., *Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings*, Marcel Dekker, New York and Basel, 1984.

[15] Granas, A., Dugundji, J., *Fixed Point Theory*, Springer, New York, 2003.

[16] Jachymski, J., *Extensions of the Dugundji-Granas and Nadler's theorems on the continuity of fixed points*, Pure Appl. Funct. Anal., **2**(2017), 657-666.

[17] Kirk, W.A., *Contraction Mappings and Extensions*, Handbook of Metric Fixed Point Theory, Kluwer, Dordrecht, 2001, 1-34.

[18] Kubota, R., Takahashi, W., Takeuchi, Y., *Extensions of Browder's demiclosedness principle and Reich's lemma and their applications*, Pure Appl. Func. Anal., **1**(2016), 63-84.

[19] Plant, A.T., Reich, S., *The asymptotics of nonexpansive iterations*, J. Funct. Anal., **54**(1983), 308-319.

[20] Pustylnik, E., Reich, S., *Infinite products of arbitrary operators and intersections of subspaces in Hilbert space*, J. Approx. Theory, **178**(2014), 91-102.

[21] Pustylnyk, E., Reich, S., Zaslavski, A.J., *Convergence to compact sets of inexact orbits of nonexpansive mappings in Banach and metric spaces*, Fixed Point Theory Appl., **2008**(2008), 1-10.

[22] Pustylnik, E., Reich, S., Zaslavski, A.J., *Inexact infinite products of nonexpansive mappings*, Numer. Funct. Anal. Optim., **30**(2009), 632-645.

[23] Reich, S., *The fixed point property for nonexpansive mappings I*, Amer. Math. Monthly, **83**(1976), 266-268.

[24] Reich, S., *The fixed point property for nonexpansive mappings II*, Amer. Math. Monthly, **87**(1980), 292-294.

[25] Reich, S., *The alternating algorithm of von Neumann in the Hilbert ball*, Dynam. Systems Appl., **2**(1993), 21-25.

[26] Reich, S., Salinas, Z., *Weak convergence of infinite products of operators in Hadamard spaces*, Rend. Circ. Mat. Palermo, **85**(2016), 65-71.

[27] Reich, S., Zaslavski, A.J., *Well-posedness of fixed point problems*, Far East J. Math. Sci., Special Volume (Functional Analysis and Its Applications), Part III (2001), 393-401.

[28] Reich, S., Zaslavski, A.J., *Generic Aspects of Metric Fixed Point Theory*, Handbook of Metric Fixed Point Theory, Kluwer, Dordrecht, 2001, 557-575.

[29] Reich, S., Zaslavski, A.J., *Convergence to attractors under perturbations*, Commun. Math. Anal., **10**(2011), 57-63.

[30] Reich, S., Zaslavski, A.J., *Genericity in Nonlinear Analysis*, Springer, New York, 2014.

[31] Reich, S., Zaslavski, A.J., *Asymptotic behavior of infinite products of nonexpansive mappings in metric spaces*, Z. Anal. Anwend., **33**(2014), 101-117.

[32] Reich, S., Zaslavski, A.J., *Convergence to attractors of nonexpansive set-valued mappings*, Commun. Math. Anal., **22**(2019), 51-60.

[33] Reich, S., Zaslavski, A.J., *Asymptotic behavior of inexact orbits of nonexpansive mappings,* Topol. Methods Nonlinear Anal., accepted for publication.

[34] Rezapour, S., Yao, J.-C., Zakeri, S.H., *A strong convergence theorem for quasi-contractive mappings and inverse strongly monotone mappings*, Pure Appl. Funct. Anal., **5**(2020), 733-745.

[35] Takahashi, W., *The split common fixed point problem and the shrinking projection method for new nonlinear mappings in two Banach spaces*, Pure Appl. Funct. Anal., **2**(2017), 685-699.

[36] Takahashi, W., *A general iterative method for split common fixed point problems in Hilbert spaces and applications*, Pure Appl. Funct. Anal., **3**(2018), 349-369.

[37] Takahashi, W., *A strong convergence theorem for two infinite sequences of nonlinear mappings in Hilbert spaces and applications*, Pure Appl. Funct. Anal., **4**(2019), 629-647.

[38] Takahashi, W., *Weak and strong convergence theorems for two generic generalized non-spreading mappings in Banach spaces*, Pure Appl. Funct. Anal., **5**(2020), 747-767.

[39] Takahashi, W., Wen, C.-F., Yao, J.-C., *A strong convergence theorem by Halpern type iteration for a finite family of generalized demimetric mappings in a Hilbert space*, Pure Appl. Funct. Anal., **4**(2019), 407-426.

[40] Zaslavski, A.J., *Approximate Solutions of Common Fixed Point Problems*, Springer Optim. Appl., Springer, Cham, 2016.

[41] Zaslavski, A.J., *Algorithms for Solving Common Fixed Point Problems*, Springer Optim. Appl., Springer, Cham, 2018.

Simeon Reich
The Technion – Israel Institute of Technology,
Department of Mathematics,
32000 Haifa, Israel
e-mail: `sreich@technion.ac.il`

Alexander J. Zaslavski
The Technion – Israel Institute of Technology,
Department of Mathematics,
32000 Haifa, Israel
e-mail: `ajzasl@technion.ac.il`

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 139–158
DOI: 10.24193/subbmath.2021.1.13

Split equality variational inequality problems for pseudomonotone mappings in Banach spaces

Oganeditse A. Boikanyo and Habtu Zegeye

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. A new algorithm for approximating solutions of the split equality variational inequality problems (SEVIP) for pseudomonotone mappings in the setting of Banach spaces is introduced. Strong convergence of the sequence generated by the proposed algorithm to a solution of the SEVIP is then derived without assuming the Lipschitz continuity of the underlying mappings and without prior knowledge of operator norms of the bounded linear operators involved. In addition, we provide several applications of our method and provide a numerical example to illustrate the convergence of the proposed algorithm. Our results improve, consolidate and complement several results reported in the literature.

Mathematics Subject Classification (2010): 47H09, 47J20, 65K15, 47J05, 90C25.

Keywords: Pseudomonotone mapping, split equality variational inequality problem, strong convergence, variational inequality.

1. Introduction

Let K be a nonempty, closed and convex subset of a real Hilbert space H, and $T : C \to H$ be a nonlinear mapping. The variational inequality problem (VIP), first introduced by Stampacchia [31] and Fichera [19] in 1964, is a problem that consist of finding an element $x \in C$ such that $\langle Tx, y - x\rangle \geq 0$ for all $y \in C$. For a nonlinear mapping $T : C \to H$, we denote the solution of the VIP by $VI(C,T)$ if it is nonempty. It is known that x solves the VIP if and only if x is a fixed point of the map $P_C(I - \lambda T) : C \to C$. Variational inequality problems have been studied extensively by several authors, thanks to their relevance in various applications in areas such as mechanics, physics, engineering, convex programming and control theory. Among these studies, VIPs for continuous and pseudomonotone maps will be of particular interest to us. Let us remember that if T is continuous and pseudomonotone, then $VI(C,T)$ is closed and convex [26]. In [30, 33, 34], the authors studied algorithms for solving uniformly continuous and weakly sequentially continuous pseudomonotone VIPs in Hilbert spaces. The distinctive feature of the algorithms constructed and analyzed in [33, 34] is mainly on the different Armijo-type line search rules used. For further reading on the VIP, particularly iterative methods for finding solutions of VIPs, the interested reader is referred to articles [2, 7, 10, 21–23, 25, 29, 38, 41], and their references.

Let K_1 and K_2 be nonempty, closed and convex subsets of real Hilbert spaces H_1 and H_2, respectively. Also let $A : H_1 \to H_3$ and $B : H_2 \to H_3$ be bounded linear mappings, where H_3 is another real Hilbert space. Consider two nonlinear mappings $T : H_1 \to H_1$ and $S : H_2 \to H_2$. The split equality variational inequality problem (SEVIP) is formulated as a problem of finding:

$$(x, y) \in K_1 \times K_2 \text{ such that } (x, y) \in VI(K_1, T) \times VI(K_2, S) \text{ and } Ax = By. \quad (1.1)$$

The SEVIP is quite general and it includes as special cases, split equality zero point problem (see, [18]), common solutions of the variational inequality problem [12], common zeros of mappings [16], split equality feasibility problem [27], has been studied extensively by many authors and applied to solving many real life problems such as in modelling intensity-modulated radiation therapy treatment planning [8, 9], modelling of inverse problems arising from phase retrieval, and in sensor networks in computerised tomography and data compression [5, 17].

If, in (1.1), we consider $H_2 = H_3$, and $B = I$, the identity mapping on H_2, the SEVIP reduces to the split varitional inequality problem (SVIP) that was recently introduced by Censor *et al.* [10]. The SVIP consists of finding:

$$(x, y) \in K_1 \times K_2 \text{ such that } (x, y) \in VI(K_1, T) \times VI(K_2, S) \text{ and } y = Ax, \quad (1.2)$$

that is, the SVIP constitutes a pair of VIPs, which have to be solved so that the image $y = Ax$, under a given bounded linear operator A of the solution x of the VIP in H_1, is a solution of another VIP in another space H_2. In Moudafi [27], it was noted that the SVIP generalizes the split fixed point problem, split variational inequality problem, split zero point problem and split feasibility problem (see also [3, 4, 6, 13–15, 35, 40], and the references therein). Many of the results cited above were obtained in the setting of real Hilbert spaces. In [11], Censor *et al.* studied an iterative algorithm that approximates a solution of the SVIP for a monotone mapping in Hilbert spaces and proved weak convergence results of the algorithm. In [6], Byrne *et al.* constructed a scheme which approximates the solution of the SVIP for monotone type mappings in Hilbert spaces and proved weak and strong convergence results of the scheme under certain assumptions.

Motivated by the work of Censor *et al.* [11], Byrne *et al.* [6] and Thong *et al* [33], we introduce and study a new algorithm for solving the SEVIP for uniformly continuous and weakly sequentially continuous pseudomonotone mappings in the setting of Banach spaces. Strong convergence of the proposed algorithm is proved under mild assumptions and without prior knowledge of operator norms of bounded linear mappings involved. Some applications of the main results are also provided. A numerical example is given to illustrate the convergence of the proposed algorithm. Our results improve, consolidate and complement several results in the literature.

2. Preliminaries

Let E be a reflexive, strictly convex and smooth Banach space and let C be a nonempty, closed and convex subset of E. Consider the function $\phi : E \times E \to \mathbb{R}$, introduced by Alber [1], defined by

$$\phi(y, x) = ||y||^2 - 2\langle y, Jx\rangle + ||x||^2, \text{ for } x, y \in E, \tag{2.1}$$

where $J : E \to E^*$ is the *normalized duality mapping* defined by

$$Jx := \{x^* \in E^* : \langle x, x^*\rangle = ||x||^2, ||x|| = ||x^*||\}, \forall x \in E.$$

It is known that if E is uniformly smooth, then J is uniformly norm to norm continuous on each bounded subset of E (see, [32]). Furthermore, if E is a reflexive and strictly convex Banach space with a strictly convex dual, then J^{-1} is a duality mapping from E^* into E which satisfies $JJ^{-1} = I_{E^*}$ and $J^{-1}J = I_E$ (see, [32]). The *generalized projection mapping*, introduced by Alber [1], is a mapping $\Pi_C : E \to C$ that assigns an arbitrary point $x \in E$ to the minimizer, $\bar{x}$, of $\phi(., x)$ over C.

Lemma 2.1. [1] *Let C be a nonempty, closed and convex subset of a real reflexive, strictly convex, and smooth Banach space E and let $x \in E$. Then $\phi(y, \Pi_C x) + \phi(\Pi_C x, x) \leq \phi(y, x)$ for all $y \in C$.*

Lemma 2.2. [20] *Let E be a real smooth and uniformly convex Banach space and let (x_n) and (y_n) be two sequences in E. If either (x_n) or (y_n) is bounded and $\phi(x_n, y_n) \to 0$ as $n \to \infty$, then $x_n - y_n \to 0$, as $n \to \infty$.*

Lemma 2.3. [1] *Let C be a convex subset of a real smooth Banach space E. Let $x \in E$. Then $x_0 = \Pi_C x$ if and only if $\langle z - x_0, J_E x - J_E x_0\rangle \leq 0, \forall z \in C$.*

Consider the function $V : E \times E^* \to \mathbb{R}$, studied by Alber [1], defined by

$$V(x, x^*) = \|x\|^2 - 2\langle x, x^*\rangle + \|x^*\|^2, \text{ for all } x \in E \text{ and } x^* \in E^*.$$

Lemma 2.4. [1] *Let E be reflexive, strictly convex and smooth Banach space with E^* as its dual. Then for all $x \in E$ and $x^*, y^* \in E^*$,*

$$V(x, x^*) + 2\langle J_E^{-1} x^* - x, y^*\rangle \leq V(x, x^* + y^*).$$

Lemma 2.5. [28] *If E is a smooth Banach space and $\{t_i\} \in (0, 1)$ with $\sum_{i=1}^N t_i = 1$, then*

$$\phi\Big(z, J_E^{-1}\Big(\sum_{i=1}^N t_i J_E x_i\Big)\Big) \leq \sum_{i=1}^N t_i \phi(z, x_i).$$

Lemma 2.6. [37] *Let (a_n) be a sequence of nonnegative real numbers such that $a_{n+1} \leq (1 - \beta_n)a_n + \beta_n\delta_n$, for all $n \geq 1$, where $(\beta_n) \subset (0, 1)$ and $(\delta_n) \subset R$ satisfying $\sum_{n=1}^{\infty} \beta_n = \infty$, and $\limsup_{n\to\infty} \delta_n \leq 0$. Then, $\lim_{n\to\infty} a_n = 0$.*

Lemma 2.7. [24] *Let (a_n) be a sequence of real numbers such that there exists a subsequence (n_i) of (n) such that $a_{n_i} < a_{n_i+1}$ for all $i \in \mathbb{N}$. Then there exists a nondecreasing sequence $(m_k) \subset \mathbb{N}$ such that $m_k \to \infty$ and $\max\{a_{m_k}, a_k\} \leq a_{m_k+1}$. In fact, $m_k = \max\{j \leq k : a_j < a_{j+1}\}$.*

Lemma 2.8. [39] *Let E be a reflexive and smooth real Banach space. Then, there exists $\alpha > 0$ such that $\langle x - y, J_E x - J_E y\rangle \geq \alpha ||x - y||^2$ for all $x, y \in E$.*

Lemma 2.9. [36] *Let E be a reflexive and smooth real Banach space. Then for each $x, y \in E$, we have $\phi(y, x) \geq \frac{1}{2}||x - y||^2$.*

Lemma 2.10. *Let C be a closed and convex set in a reflexive real Banach space E, h be a real-valued function on E, and K be the set $\{x \in C : h(x) \leq 0\}$. If K is nonempty and h is Lipschitz continuous on C with constant $L > 0$, then*

$$\phi(x, \Pi_K x) \geq \frac{1}{2L^2}(h(x))^2, \text{ for all } x \in C. \tag{2.2}$$

Proof. Clearly (2.2) holds for all $x \in K$. Hence, it suffices to show that (2.2) holds for every $x \in C\backslash K$. Let $x \in C$ but $x \notin K$. Since K is closed, there exists $y \in K$ such that $\phi(x, y) = \phi(x, \Pi_K x)$. It follows from the Lipschitz continuity of h that $|h(x) - h(y)| \leq L\,\|x - y\|$. Since $x \notin K$ and $y \in K$, we have $h(x) > 0$ and $h(y) \leq 0$. Thus, from Lemma 2.9, we have

$$h(x) \leq h(x) - h(y) = |h(x) - h(y)| \leq L||x - y|| \leq L\Big(2\phi(x, \Pi_K x)\Big)^{\frac{1}{2}},$$

and hence the conclusion follows. ☐

Definition 2.11. Let $T : C \to E^*$ be a mapping. Then T is called
(a) sequentially weakly continuous on C if for each sequence $(x_n) \subseteq C$ converging weakly to $x \in C$, the sequence (Ax_n) converges weakly to Ax;
(b) monotone if $\langle x - y, Tx - Ty\rangle \geq 0$ for each $x, y \in C$;
(c) pseudomonotone on C if for all $x, y \in C$,

$$\langle y - x, Tx\rangle \geq 0 \quad \text{implies} \quad \langle y - x, Ty\rangle \geq 0. \tag{2.3}$$

Remark 2.12. In [30], Shehu *et al.* asserted that using the Monte-Carlo approach, it can be shown that the map $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T(x, y) = \left(\left[x^2 + (y - 1)^2\right](1 + y), -x^3 - x(y - 1)^2\right)$$

is pseudomonotone on $\mathbb{R}^2$. The correctness of this method/approach in verifying pseudomonotonicity of an operator is questionable. We claim that there could still be a pair of points, say $(x, y), (u, v) \in \mathbb{R}^2$, such that the implication (2.3) does not hold. Indeed, (2.3) fails to hold for a pair of points $(0, 1)$ and $(-1, 2)$ in $\mathbb{R}^2$, as shown by simple computations below

$$\langle T(0, 1), (-1, 2) - (0, 1)\rangle = 0 \geq 0 \quad \text{and} \quad \langle T(-1, 2), (-1, 2) - (0, 1)\rangle = -4 < 0.$$

Example 2.13. Let the map $S : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$S(x, y) = \left(\left[x^2 + 1 + (y - 1)^2\right](1 + y), -x^3 - x\left[1 + (y - 1)^2\right]\right).$$

Claim 1: S is not monotone. Indeed, for the pair $(1, 0)$ and $(-1, -1)$, we have

$$\langle S(1, 0) - S(-1, -1), (1, 0) - (-1, -1)\rangle = -3 < 0.$$

Claim 2: S is pseudomonotone. To this end, we assume that $\langle S(x, y), (u, v) - (x, y)\rangle \geq 0$ is true for each pair $(x, y), (u, v) \in \mathbb{R}^2$. This means that

$$\left[x^2 + 1 + (y - 1)^2\right](1 + y)(u - x) + \left[-x^3 - x\left[1 + (y - 1)^2\right]\right](v - y) \geq 0$$

which implies that $\left[x^2+1+(y-1)^2\right][u+uy-x-xv] \geq 0$ for all $(x,y),(u,v) \in \mathbb{R}^2$. Therefore, $u(1+y)-x(1+v) \geq 0$ for all $(x,y),(u,v) \in \mathbb{R}^2$. Since $u^2+1+(v-1)^2 > 0$ for any $u,v \in \mathbb{R}$, we have for any $(x,y),(u,v) \in \mathbb{R}^2$,

$$\langle S(u,v),(u,v)-(x,y)\rangle = \left[u^2+1+(v-1)^2\right][u(1+y)-x(1+v)] \geq 0.$$

Lemma 2.14. [26] *Let K be a nonempty, closed and convex subset of a smooth, strictly convex and reflexive real Banach space E. Let A be a continuous pseudomonotone mapping from K into E^*. Then, $VI(K,A)$ is closed and convex, and $p \in VI(K,A)$ if and only if $\langle x-p, Ax\rangle \geq 0,$ for all $x \in K$.*

3. Main results

In the sequel, we shall make use of the following assumptions:
Assumption 1:

(A1) Let E_1 and E_2 be uniformly smooth and uniformly convex real Banach spaces with dual spaces E_1^* and E_2^*, respectively, and let E_3 be a real Banach space with dual space E_3^*.
(A2) Let $A : E_1 \to E_3$ and $B : E_2 \to E_3$ be bounded linear mappings with adjoints $A^* : E_3^* \to E_1^*$ and $B^* : E_3^* \to E_2^*$, respectively.
(A3) Let $C \subseteq E_1$ and $D \subseteq E_2$ be nonempty, closed and convex subsets.
(A4) Let $T : E_1 \to E_1^*$ and $S : E_2 \to E_2^*$ be uniformly continuous pseudomonotone mappings that are sequentially weakly continuous on bounded subset of C and D, respectively.
(A5) Let $\Gamma := \{(p,q) \in C \times D : \langle x-p, Tp\rangle \geq 0, \forall x \in C, \langle y-q, Sq\rangle \geq 0, \forall y \in D, \text{ and } Ap = Bq\} \neq \emptyset$.

Assumption 2:

(B1) Let $\xi = \min\{\xi_1, \xi_2\}$, where ξ_1 and ξ_2 are constants given in Lemma 2.8 associated with J_{E_1} and J_{E_2}, respectively.
(B2) Let $l \in (0,1)$, $\mu > 0$ and $\lambda \in (0, \frac{\xi}{\mu})$.
(B3) Let $(\alpha_n) \subset (0,e] \subset (0,1)$, for some constant $e > 0$, be such that $\lim\limits_{n\to\infty} \alpha_n = 0$ and $\sum_{n=1}^{\infty} \alpha_n = \infty$.
(B4) Let $0 < \gamma \leq \gamma_n \leq \dfrac{\xi \|Ax_n - By_n\|^2}{2[\|A^* J_{E_3}(Ax_n - By_n)\|^2 + \|B^* J_{E_3}(Ax_n - By_n)\|^2]}$
for $n \in \Omega = \{n \in \mathbb{N} : Ax_n - By_n \neq 0\}$, otherwise $\gamma_n = \gamma > 0$.

Now, we introduce our algorithm for the SEVIP.

Algorithm 3.1

For arbitrary $x_0, u \in C$ and $y_0, v \in D$, define an iterative algorithm by

1. **Step 1**. Compute: $u_n = \Pi_C J_{E_1}^{-1}\left[J_{E_1} x_n - \gamma_n A^* J_{E_3}(Ax_n - By_n)\right]$ and $r_1(x_n,u_n) = x_n - u_n$.
   Compute $v_n = \Pi_D J_{E_2}^{-1}\left[J_{E_2} y_n + \gamma_n B^* J_{E_3}(Ax_n - By_n)\right]$ and $s_1(y_n,v_n) = y_n - v_n$.
2. **Step 2**. Compute: $z_n = \Pi_C J_{E_1}^{-1}\left[J_{E_1} u_n - \lambda T u_n\right]$ and $r_2(u_n,z_n) = u_n - z_n$.
   Compute $w_n = \Pi_D J_{E_2}^{-1}\left[J_{E_2} v_n - \lambda S v_n\right]$ and $s_2(v_n,w_n) = v_n - w_n$.

3. **Step 3.** Compute $f_n = u_n - \tau_n r_2(u_n, z_n)$, where $\tau_n = l^{j_n}$ and j_n is the smallest non-negative integer j satisfying

$$\langle r_2(u_n, z_n),\ \ Tu_n - T(u_n - l^j r_2(u_n, z_n))\rangle \leq \mu \left\| r_2(u_n, z_n) \right\|^2 \tag{3.1}$$

and $g_n = v_n - \kappa_n s_2(v_n, w_n)$, where $\kappa_n = l^{h_n}$ and h_n is the smallest non-negative integer h satisfying

$$\langle s_2(v_n, w_n),\ \ Sv_n - S(v_n - l^h s_2(v_n, w_n))\rangle \leq \mu \left\| s_2(v_n, w_n) \right\|^2. \tag{3.2}$$

4. **Step 4.** Compute: $x_{n+1} = J_{E_1}^{-1}\big[\alpha_n J_{E_1} u + (1-\alpha_n) J_{E_1} \Pi_{C_n} u_n\big]$, where

$$C_n := \{x \in C : \langle x - f_n, Tf_n\rangle \leq 0\}, \tag{3.3}$$

and $y_{n+1} = J_{E_2}^{-1}\big[\alpha_n J_{E_2} v + (1-\alpha_n) J_{E_2} \Pi_{D_n} v_n\big]$, where

$$D_n := \{y \in D : \langle y - g_n, Sg_n\rangle \leq 0\}. \tag{3.4}$$

5. **Step 5**. Set $n := n + 1$ and go to **Step 1**.

Lemma 3.1. *Assume that Conditions (A1) - (A5) and (B1) - (B4) are satisfied. Then, the sequences (x_n) and (y_n) generated by Algorithm 3.1 are well defined.*

Proof. It is enough to show that the search rules in (3.1) and (3.2) are well defined, and the sets C_n and D_n are nonempty.
Since $l \in (0, 1)$ and T is continuous on C, it follows that

$$\langle r_2(u_n, z_n),\ \ Tu_n - T(u_n - l^j r_2(u_n, z_n))\rangle \to 0, \ \text{ as } j \to \infty.$$

On the other hand, since $\|r_2(u_n, z_n)\| > 0$, there exists a non-negative integer j_n satisfying inequality (3.1). Similarly, from the continuity of the mapping S on D, there exists a non-negative integer h_n satisfying inequality (3.2).
Furthermore, since $\Gamma \neq \emptyset$, choose $(p, q) \in \Gamma$. Then by Step 3 of the algorithm, $f_n \in C$ and $g_n \in D$ for each $n \geq 0$, and hence by Lemma 2.14, $\langle p - f_n, Tf_n\rangle \leq 0$ and $\langle q - g_n, Sg_n\rangle \leq 0$ for each $n \geq 0$. Hence, $p \in C_n$ and $q \in D_n$ for each $n \geq 0$, showing that $C_n \neq \emptyset$ and $D_n \neq \emptyset$ for each $n \geq 0$. □

Lemma 3.2. *Assume that Conditions (A1) - (A5) and (B1) - (B4) are satisfied. If (u_n), (z_n), (v_n) and (w_n) are sequences generated by Algorithm 3.1, then $\xi\lambda^{-1} \left\| r_2(u_n, z_n) \right\|^2 \leq \langle r_2(u_n, z_n), Tu_n\rangle$ and $\xi\lambda^{-1} \left\| s_2(v_n, w_n) \right\|^2 \leq \langle s_2(v_n, w_n), Sv_n\rangle$.*

Proof. Using Lemma 2.3 and the definition of z_n, we have

$$\langle z - z_n, J_{E_1} u_n - \lambda T u_n - J_{E_1} z_n\rangle \leq 0, \ \forall\, z \in C.$$

In particular, for $z = u_n \in C$, we obtain $\langle u_n - z_n, J_{E_1} u_n - J_{E_1} z_n\rangle \leq \lambda\langle u_n - z_n, Tu_n\rangle$.
Using Lemma 2.8, we obtain

$$\xi \left\| r_2(u_n, z_n) \right\|^2 \leq \xi_1 \left\| u_n - z_n \right\|^2 \leq \lambda\langle u_n - z_n, Tu_n\rangle.$$

The second inequality of the lemma can be proved in a similar way. □

Lemma 3.3. *Assume that Conditions (A1) - (A5) and (B1) - (B4) are met. Let* $(p,q) \in \Gamma$, $F_n(x) = \langle x - f_n, Tf_n \rangle$ *and* $G_n(y) = \langle y - g_n, Sg_n \rangle$. *Then*

(i) $F_n(p) \leq 0$ *and* $F_n(u_n) \geq \tau_n \left(\xi\lambda^{-1} - \mu\right) \|r_2(u_n,z_n)\|^2$, *and*
(ii) $G_n(q) \leq 0$ *and* $G_n(v_n) \geq \kappa_n \left(\xi\lambda^{-1} - \mu\right) \|s_2(v_n,w_n)\|^2$.
In particular, if $r_2(u_n,z_n) \neq 0$ *and* $s_2(v_n,w_n) \neq 0$, *then* $F_n(u_n) > 0$ *and* $G_n(v_n) > 0$, *respectively.*

Proof. (i) Since $(p,q) \in \Gamma$, it follows that $p \in \mathrm{VI}(C,T)$ and $q \in \mathrm{VI}(D,S)$.
By Lemma 2.14, $F_n(p) = \langle p - f_n, Tf_n \rangle \leq 0$ for each $n \geq 0$.
Next, we observe from Step 3 of the algorithm and the definition of F_n that

$$F_n(u_n) = \langle u_n - f_n, Tf_n \rangle = \langle \tau_n r_2(u_n,z_n), Tf_n \rangle = \tau_n \langle r_2(u_n,z_n), Tf_n \rangle.$$

But from the search rule (3.1), $\langle r_2(u_n,z_n), Tu_n - Tf_n \rangle \leq \mu \|r_2(u_n,z_n)\|^2$, which together with Lemma 3.2 imply that

$$\begin{aligned} F_n(u_n) = \tau_n \langle r_2(u_n,z_n), Tf_n \rangle &\geq \tau_n \left[\langle r_2(u_n,z_n), Tu_n \rangle - \mu \|r_2(u_n,z_n)\|^2 \right] \\ &\geq \tau_n \left[\xi\lambda^{-1} \|r_2(u_n,z_n)\|^2 - \mu \|r_2(u_n,z_n)\|^2 \right]. \end{aligned}$$

Obviously, if $r_2(u_n,z_n) \neq 0$, then from Condition (B2), we have $F_n(u_n) > 0$.
(ii) The proof is similar to the proof of part (i) above. □

Lemma 3.4. *Assume that Conditions (A1) - (A5) and (B1) - (B4) hold.*
(a). If there exist $(u_{n_k}) \subset (u_n)$ *and* $(z_{n_k}) \subset (z_n)$ *such that* (u_{n_k}) *converges weakly to* $x \in E_1$ *and* $\tau_{n_k} \|u_{n_k} - z_{n_k}\|^2 \to 0$ *as* $k \to \infty$, *then* $x \in$ *VI*(C,T).
(b). If there exist $(v_{n_i}) \subset (v_n)$ *and* $(w_{n_i}) \subset (w_n)$ *such that* (v_{n_i}) *converges weakly to* $y \in E_2$ *and* $\kappa_{n_i} \|v_{n_i} - w_{n_i}\|^2 \to 0$ *as* $i \to \infty$, *then* $y \in$ *VI*(D,S).

Proof. (a). By considering two possible cases on τ_{n_k}, we first show that

$$\lim_{k\to\infty} \|u_{n_k} - z_{n_k}\| = 0. \tag{3.5}$$

Case I: Assume that $\liminf\limits_{k\to\infty} \tau_{n_k} > 0$.
Then there exists a constant $\tau > 0$ such that $\tau_{n_k} \geq \tau > 0$ for all $k \in \mathbb{N}$. Then

$$\|u_{n_k} - z_{n_k}\|^2 = \tau_{n_k}^{-1} \left[\tau_{n_k} \|u_{n_k} - z_{n_k}\|^2 \right] \leq \tau^{-1} \left[\tau_{n_k} \|u_{n_k} - z_{n_k}\|^2 \right]. \tag{3.6}$$

Therefore, (3.5) follows from (3.6) and the assumption in the lemma.
Case II: Assume that $\liminf\limits_{k\to\infty} \tau_{n_k} = 0$.
In this case, we take a subsequence (n_{k_j}) of (n_k) if necessary, we assume without loss of generality that

$$\lim_{k\to\infty} \tau_{n_k} = 0 \quad \text{and} \quad \lim_{k\to\infty} \|u_{n_k} - z_{n_k}\| = a > 0. \tag{3.7}$$

Let $f_k = \frac{1}{l}\tau_{n_k} z_{n_k} + \left(1 - \frac{1}{l}\tau_{n_k}\right) u_{n_k}$. Using (3.7), we get

$$\lim_{k\to\infty} \|f_k - u_{n_k}\| = \lim_{k\to\infty} l^{-1}\tau_{n_k} \|u_{n_k} - z_{n_k}\| = 0. \tag{3.8}$$

Since T is uniformly continuous on bounded subsets of C, it follows from (3.8) that $\|Tf_k - Tu_{n_k}\| \to 0$ as $k \to \infty$.

From (3.1), we have $\langle Tu_{n_k} - Tf_k, u_{n_k} - z_{n_k}\rangle > \mu \|u_{n_k} - z_{n_k}\|^2$, and it follows that $\|u_{n_k} - z_{n_k}\| \to 0$ as $k \to \infty$. This contradicts (3.7), hence the limit in (3.5) must hold. Finally, we show that $x \in \mathrm{VI}(C, T)$.
Since C is weakly closed, we have $x \in C$. Furthermore, from the fact that J_{E_1} is uniformly continuous on bounded subsets of E_1, we have

$$\lim_{k\to\infty} \|J_{E_1}u_{n_k} - J_{E_1}z_{n_k}\| \quad = \quad 0. \tag{3.9}$$

From Lemma 2.3 and $z_n \in C$, we get $\langle z - z_{n_k}, J_{E_1}u_{n_k} - \lambda Tu_{n_k} - J_{E_1}z_{n_k}\rangle \leq 0$ for all $z \in C$, which implies that

$$\langle z - z_{n_k}, J_{E_1}u_{n_k} - J_{E_1}z_{n_k}\rangle - \lambda\langle u_{n_k} - z_{n_k}, Tu_{n_k}\rangle \leq \lambda\langle z - u_{n_k}, Tu_{n_k}\rangle. \tag{3.10}$$

Taking the limit inferior as $k \to \infty$ and using (3.9), we get

$$\liminf_{k\to\infty}\langle z - u_{n_k}, Tu_{n_k}\rangle \geq 0 \quad \forall\, z \in C. \tag{3.11}$$

Thus the inequality in (3.11) implies that we can choose a decreasing sequence of positive real numbers (δ_k) such that (δ_k) converges to zero as $k \to \infty$, and for each δ_k there exists N_k, the smallest positive integer, such that

$$\langle z - u_{n_j}, Tu_{n_j}\rangle + \delta_k \geq 0 \quad \forall\, j \geq N_k \quad \text{and} \quad \forall\, z \in C. \tag{3.12}$$

Since (δ_k) is decreasing, the sequence (N_k) is increasing.
Note that if there exists $N > 0$ such that $Tu_{N_k} = 0$ for all $k \geq N$, then it can be shown easily that $x \in \mathrm{VI}(C, T)$.
On the other hand, if there exists a subsequence (N_{k_i}) of (N_k), again denoted by (N_k), such that $Tu_{N_k} \neq 0$ for all $k \in \mathbb{N}$, then $\langle a_{N_k}, Tu_{N_k}\rangle = 1$ for each $k \in \mathbb{N}$, where

$$a_{N_k} = \frac{J_{E_1}^{-1}Tu_{N_k}}{\|Tu_{N_k}\|^2}.$$

From (3.12), we deduce that $\langle z + \delta_k a_{N_k} - u_{N_k}, Tu_{N_k}\rangle \geq 0$ for each $k \in \mathbb{N}$ and $z \in C$. Since T is pseudomonotone, it follows that

$$\langle z + \delta_k a_{N_k} - u_{N_k}, T(z + \delta_k a_{N_k})\rangle \geq 0 \quad \forall\, k \in \mathbb{N} \quad \text{and} \quad \forall\, z \in C. \tag{3.13}$$

But by our assumption, (u_{N_k}) converges weakly to $x \in C$. Also T is sequentially weakly continuous on E_1 implies that (Tu_{N_k}) converges weakly to Tx. Moreover, we can suppose that $Tx \neq 0$ (otherwise, x is in $\mathrm{VI}(C, T)$) and so

$$0 \leq \|Tx\| \leq \liminf_{k\to\infty}\|Tu_{N_k}\|.$$

Since $(u_{N_k}) \subset (u_{n_k})$ and (δ_k) converges to zero as $k \to \infty$, we obtain that

$$0 \leq \limsup_{k\to\infty}\|\delta_k a_{N_k}\| = \limsup_{k\to\infty}\left(\frac{\delta_k}{\|Tu_{N_k}\|}\right) \leq \frac{\limsup_{k\to\infty}\delta_k}{\liminf_{k\to\infty}\|Tu_{N_k}\|} \leq \frac{0}{\|Tx\|},$$

and hence $\|\delta_k a_{N_k}\| \to 0$ as $k \to \infty$. Therefore, taking the limit in (3.13) as $k \to \infty$, we get $\langle z - x, Tz\rangle \geq 0$ for all $z \in C$. In view of Lemma 2.14, we conclude that $x \in \mathrm{VI}(C, T)$.
Part (b) of the Lemma can be proved in a similar way. $\square$

Remark 3.5. If in Lemma 3.4, $T : E_1 \to E_1^*$ and $S : E_2 \to E_2^*$ are uniformly continuous and monotone mappings, then for all $z \in C$, we have from (3.10)

$$\begin{aligned}\langle z - z_{n_k}, J_{E_1}u_{n_k} - J_{E_1}z_{n_k}\rangle &+ \lambda\langle z_{n_k} - u_{n_k}, Tu_{n_k}\rangle \leq \lambda\langle z - u_{n_k}, Tu_{n_k} - Tz\rangle \\ &+ \lambda\langle z - u_{n_k}, Tz\rangle \leq \lambda\langle z - u_{n_k}, Tz\rangle.\end{aligned}$$

Taking the limit as $k \to \infty$, we get $0 \leq \langle z - x, Tz\rangle$. It then follows from Lemma 2.14 that $x \in \mathrm{VI}(C,T)$. Similarly, we get $y \in \mathrm{VI}(D,S)$.

Lemma 3.6. *Let (x_n) and (y_n) be sequences generated by Algorithm 3.1. Assume that the Conditions (A1) - (A5) and (B1) - (B4) hold. Then (x_n) and (y_n) are bounded. Hence, (u_n), (v_n), (z_n) and (w_n) are bounded sequences.*

Proof. Let $(p,q) \in \Gamma$. Then $p \in \mathrm{VI}(C,T)$, $q \in \mathrm{VI}(D,S)$ and $Ap = Bq$.
Denote $q_n = J_{E_1}^{-1}\big[J_{E_1}x_n - \gamma_n A^* J_{E_3}(Ax_n - By_n)\big]$. Then $u_n = \Pi_C q_n$ and so from Lemmas 2.1 and 2.4, and the properties of the mapping V, we obtain

$$\begin{aligned}\phi(p,u_n) &\leq \phi\big(p, J_{E_1}^{-1}\big[J_{E_1}x_n - \gamma_n A^* J_{E_3}(Ax_n - By_n)\big]\big) \\ &= V(p, J_{E_1}x_n - \gamma_n A^* J_{E_3}(Ax_n - By_n)) \\ &\leq V(p, J_{E_1}x_n) - 2\langle q_n - p, \gamma_n A^* J_{E_3}(Ax_n - By_n)\rangle \\ &= \phi(p,x_n) - 2\gamma_n\langle q_n - p, A^* J_{E_3}(Ax_n - By_n)\rangle \\ &= \phi(p,x_n) - 2\gamma_n\langle Aq_n - Ap, J_{E_3}(Ax_n - By_n)\rangle. \end{aligned} \tag{3.14}$$

Furthermore, from Lemma 2.5, Lemma 2.1 and (3.14), we have

$$\phi(p,x_{n+1}) \leq \alpha_n\phi(p,u) + (1-\alpha_n)\phi(p,\Pi_{C_n}u_n) \tag{3.15}$$

$$\begin{aligned}&\leq \alpha_n\phi(p,u) + (1-\alpha_n)\phi(p,u_n) \\ &\leq \alpha_n\phi(p,u) + (1-\alpha_n)\left[\phi(p,x_n)\right. \\ &- \left.2\gamma_n\langle Aq_n - Ap, J_{E_3}(Ax_n - By_n)\rangle\right]. \end{aligned} \tag{3.16}$$

Similarly, if we denote $t_n = J_{E_2}^{-1}\big[J_{E_2}y_n + \gamma_n B^* J_{E_3}(Ax_n - By_n)\big]$, then

$$\phi(q,v_n) \leq \phi(q,y_n) + 2\gamma_n\langle Bt_n - Bq, J_{E_3}(Ax_n - By_n)\rangle, \tag{3.17}$$

and therefore, from Lemma 2.5, Lemma 2.1 and (3.17)

$$\phi(q,y_{n+1}) \leq \alpha_n\phi(q,v) + (1-\alpha_n)\phi(q,\Pi_{D_n}v_n) \tag{3.18}$$

$$\begin{aligned}&\leq (1-\alpha_n)\left[\phi(q,y_n) + 2\gamma_n\langle Bt_n - Bq, J_{E_3}(Ax_n - By_n)\rangle\right] \\ &+ \alpha_n\phi(q,v). \end{aligned} \tag{3.19}$$

Denote $\Upsilon = \phi(p,u) + \phi(q,v)$ and $\Theta_n = \phi(p,x_n) + \phi(q,y_n)$. Then adding (3.16) and (3.19), we get

$$\Theta_{n+1} \leq (1-\alpha_n)[\Theta_n - 2\gamma_n\langle Aq_n - Bt_n, J_{E_3}(Ax_n - By_n)\rangle] + \alpha_n\Upsilon. \tag{3.20}$$

Now observe that

$$\begin{aligned}&-\langle Aq_n - Bt_n, J_{E_3}(Ax_n - By_n)\rangle = -\langle Ax_n - By_n, J_{E_3}(Ax_n - By_n)\rangle \\ &-\langle Aq_n - Ax_n, J_{E_3}(Ax_n - By_n)\rangle - \langle By_n - Bt_n, J_{E_3}(Ax_n - By_n)\rangle \\ \leq\ & \|q_n - x_n\|\,\|A^* J_{E_3}(Ax_n - By_n)\| - \|Ax_n - By_n\|^2 \\ +\ & \|y_n - t_n\|\,\|B^* J_{E_3}(Ax_n - By_n)\|. \end{aligned} \tag{3.21}$$

From Lemma 2.8 and the definition of q_n, we obtain

$$\|q_n - x_n\| \leq \frac{1}{\xi_1} \|\gamma_n A^* J_{E_3}(Ax_n - By_n)\| \leq \frac{\gamma_n}{\xi} \|A^* J_{E_3}(Ax_n - By_n)\| . \quad (3.22)$$

Similarly, from Lemma 2.8 and the definition of t_n, we obtain

$$\|y_n - t_n\| \quad \leq \quad \gamma_n \xi^{-1} \|B^* J_{E_3}(Ax_n - By_n)\| . \quad (3.23)$$

Combining (3.21), (3.22) and (3.23), we obtain

$$\begin{aligned} &- \quad 2\gamma_n \langle Aq_n - Bt_n, J_{E_3}(Ax_n - By_n)\rangle \leq -2\gamma_n \|Ax_n - By_n\|^2 \\ &+ \quad 2\gamma_n^2 \xi^{-1} [\|A^* J_{E_3}(Ax_n - By_n)\|^2 + \|B^* J_{E_3}(Ax_n - By_n)\|^2] \\ &\leq \quad -\gamma \|Ax_n - By_n\|^2 \end{aligned} \quad (3.24)$$

for all $n \in \Omega$, where the last inequality follows from Assumption (B4). If $n \notin \Omega$, then $Ax_n - By_n = 0$, and in this case inequality (3.24) follows trivially. Finally, using (3.24) in (3.20), we obtain $\Theta_{n+1} \leq (1 - \alpha_n)\Theta_n + \alpha_n \Upsilon$. By mathematical induction, $\Theta_n \leq \max\{\Theta_0, \Upsilon\}$ for all $n \geq 0$, showing that the sequence $(\phi(p, x_n) + \phi(q, y_n))$ is bounded, which implies that $(\phi(p, x_n))$ and $(\phi(q, y_n))$ are bounded. By the properties of ϕ, we conclude that (x_n) and (y_n) are bounded. Consequently, (u_n), (v_n), (z_n) and (w_n) are bounded. □

Theorem 3.7. *Suppose the Assumptions (A1) - (A5) and (B1) - (B4) hold. Then the sequence $((x_n, y_n))$ generated by Algorithm 3.1 converges strongly to $(x^*, y^*) \in \Gamma$, where $(x^*, y^*) = \prod_\Gamma(u, v)$.*

Proof. Let $(x^*, y^*) \in \Gamma$ be such that $(x^*, y^*) = \prod_\Gamma(u, v)$. Denote

$$\Lambda_n \quad = \quad 2\langle (x_n, y_n) - (x^*, y^*), (J_{E_1}u, J_{E_2}v) - (J_{E_1}x^*, J_{E_2}y^*)\rangle.$$

Then for some $M_1 > 0$, we have

$$\begin{aligned} \Lambda_{n+1} \quad &= \quad 2\left[\langle x_n - x^*, J_{E_1}u - J_{E_1}x^*\rangle + \langle y_n - y^*, J_{E_2}v - J_{E_2}y^*\rangle\right] \\ &+ \quad 2\left[\langle x_{n+1} - x_n, J_{E_1}u - J_{E_1}x^*\rangle + \langle y_{n+1} - y_n, J_{E_2}v - J_{E_2}y^*\rangle\right] \\ &\leq \quad \Lambda_n + M_1\left[\|x_{n+1} - x_n\| + \|y_{n+1} - y_n\|\right] . \end{aligned} \quad (3.25)$$

Since the sequences (u_n) and (z_n) are bounded by Lemma 3.6, the sequence (f_n) is bounded. But T is uniformly continuous implies that there exists $L > 0$ such that $\|Tf_n\| \leq L$ for all $n \geq 0$. We can then deduce that for each $n \geq 0$, the mapping F_n is Lipschitz continuous with Lipschitz constant $L > 0$.
Now, from Lemma 2.1,

$$\phi(x^*, \Pi_{C_n}u_n) \quad \leq \quad \phi(x^*, u_n) - \phi(\Pi_{C_n}u_n, u_n). \quad (3.26)$$

Using Lemmas 2.10 and 3.3, we obtain

$$\phi(\Pi_{C_n}u_n, u_n) \geq L_1 \tau_n^2 \|r_2(u_n, z_n)\|^4 \quad (3.27)$$

for some $L_1 > 0$. Combining (3.26), (3.27) and (3.14), we get

$$\begin{aligned} \phi(x^*, \Pi_{C_n}u_n) \quad &\leq \quad \phi(x^*, x_n) - 2\gamma_n \langle Aq_n - Ax^*, J_{E_3}(Ax_n - By_n)\rangle \\ &- \quad L_1 \tau_n^2 \|u_n - z_n\|^4 . \end{aligned} \quad (3.28)$$

From the definition of x_n, the properties of the map V and Lemma 2.4,

$$\begin{aligned}\phi(x^*, x_{n+1}) &= V(x^*, \alpha_n J_{E_1} u + (1-\alpha_n) J_{E_1} \Pi_{C_n} u_n) \\ &\leq V(x^*, \alpha_n J_{E_1} x^* + (1-\alpha_n) J_{E_1} \Pi_{C_n} u_n) \\ &+ 2\alpha_n \langle x_{n+1} - x^*, J_{E_1} u - J_{E_1} x^* \rangle \\ &= \phi(x^*, J_{E_1}^{-1} [\alpha_n J_{E_1} x^* + (1-\alpha_n) J_{E_1} \Pi_{C_n} u_n]) \\ &+ 2\alpha_n \langle x_{n+1} - x^*, J_{E_1} u - J_{E_1} x^* \rangle. \end{aligned} \tag{3.29}$$

Using Lemma 2.5, (3.29) and (3.28), we obtain for some $\widehat{K}_1 > 0$,

$$\begin{aligned}\phi(x^*, x_{n+1}) &\leq (1-\alpha_n)\phi(x^*, \Pi_{C_n} u_n) + 2\alpha_n \langle x_{n+1} - x^*, J_{E_1} u - J_{E_1} x^* \rangle \\ &\leq (1-\alpha_n)\left[\phi(x^*, x_n) - 2\gamma_n \langle Aq_n - Ax^*, J_{E_3}(Ax_n - By_n)\rangle\right] \\ &+ 2\alpha_n \langle x_{n+1} - x^*, J_{E_1} u - J_{E_1} x^* \rangle - \widehat{K}_1 \tau_n^2 \|u_n - z_n\|^4. \end{aligned} \tag{3.30}$$

Similarly, we deduce that for each $n \geq 0$,

$$\phi(y^*, \Pi_{D_n} v_n) \leq \phi(y^*, v_n) - \phi(\Pi_{D_n} v_n, v_n), \tag{3.31}$$

and also, for some $\widehat{K}_2 > 0$, we derive

$$\begin{aligned}\phi(y^*, y_{n+1}) &\leq (1-\alpha_n)\left[\phi(y^*, y_n) + 2\gamma_n \langle Bt_n - By^*, J_{E_3}(Ax_n - By_n)\rangle\right] \\ &+ 2\alpha_n \langle y_{n+1} - y^*, J_{E_2} v - J_{E_2} y^* \rangle - \widehat{K}_2 \kappa_n^2 \|v_n - w_n\|^4. \end{aligned} \tag{3.32}$$

Denote $\Theta_n^* = \phi(x^*, x_n) + \phi(y^*, y_n)$ and $\Upsilon^* = \phi(x^*, u) + \phi(y^*, v)$. Then combining (3.30), (3.32) and (3.24), we get

$$\Theta_{n+1}^* \leq (1-\alpha_n)\Theta_n^* + \alpha_n \Lambda_{n+1} - L^* \left[\tau_n^2 \|u_n - z_n\|^4 + \kappa_n^2 \|v_n - w_n\|^4\right] \tag{3.33}$$

for some $L^* > 0$. Furthermore, from (3.15), (3.26) and (3.14), we obtain

$$\begin{aligned}\phi(x^*, x_{n+1}) &\leq \alpha_n \phi(x^*, u) + (1-\alpha_n)\left[\phi(x^*, x_n) - \phi(\Pi_{C_n} u_n, u_n)\right] \\ &- 2(1-\alpha_n)\gamma_n \langle Aq_n - Ax^*, J_{E_3}(Ax_n - By_n)\rangle. \end{aligned} \tag{3.34}$$

Similarly, from (3.18), (3.31) and (3.17), we obtain

$$\begin{aligned}\phi(y^*, y_{n+1}) &\leq \alpha_n \phi(y^*, v) + (1-\alpha_n)\left[\phi(y^*, y_n) - \phi(\Pi_{D_n} v_n, v_n)\right] \\ &+ 2(1-\alpha_n)\gamma_n \langle Bt_n - By^*, J_{E_3}(Ax_n - By_n)\rangle. \end{aligned} \tag{3.35}$$

Adding (3.34) and (3.35), and using (3.24), we obtain for some $M > 0$

$$\Theta_{n+1}^* \leq \Theta_n^* + \alpha_n M - \gamma \|Ax_n - By_n\|^2 - \left[\phi(\Pi_{C_n} u_n, u_n) + \phi(\Pi_{D_n} v_n, v_n)\right],$$

which implies that

$$\begin{aligned}\phi(\Pi_{C_n} u_n, u_n) + \phi(\Pi_{D_n} v_n, v_n) + \gamma \|Ax_n - By_n\|^2 &\leq \Theta_n^* - \Theta_{n+1}^* \\ &+ \alpha_n M. \end{aligned} \tag{3.36}$$

Finally, we show that the sequence (Θ_n^*) converges strongly to zero as $n \to \infty$. For this, we consider two possible cases on (Θ_n^*).

Case I. Assume that there exists $n_0 \in \mathbb{N}$ such that the sequence of real numbers (Θ_n^*) is decreasing for all $n \geq n_0$. It then follows that (Θ_n^*) is convergent. Taking the limit in (3.36) as $n \to \infty$, we get

$$\lim_{n\to\infty} \|Ax_n - By_n\| = 0. \tag{3.37}$$

From the definition of q_n, Lemma 2.1, Lemma 2.4 and (3.22), we have

$$\begin{aligned}
\phi(x_n, u_n) &\leq \phi(x_n, J_{E_1}^{-1}[J_{E_1}x_n - \gamma_n A^* J_{E_3}(Ax_n - By_n)]) \\
&= V(x_n, J_{E_1}x_n - \gamma_n A^* J_{E_3}(Ax_n - By_n)) \\
&\leq V(x_n, J_{E_1}x_n) - 2\langle q_n - x_n, \gamma_n A^* J_{E_3}(Ax_n - By_n)\rangle \\
&\leq \phi(x_n, x_n) + 2\gamma_n^2 \xi^{-1} \|A\|^2 \|Ax_n - By_n\|^2 .
\end{aligned}$$

Taking the limit as $n \to \infty$ and noticing (3.37), yield $\phi(x_n, u_n) \to 0$ as $n \to \infty$. Using Lemma 2.2, we obtain

$$\lim_{n\to\infty} \|x_n - u_n\| = 0. \tag{3.38}$$

Similarly, starting with the definition of t_n, we obtain

$$\lim_{n\to\infty} \|y_n - v_n\| = 0. \tag{3.39}$$

Moreover, we also obtain from (3.36) and Lemma 2.2

$$\lim_{n\to\infty} \|\Pi_{C_n} u_n - u_n\| = 0 \quad \text{and} \quad \lim_{n\to\infty} \|\Pi_{D_n} v_n - v_n\| = 0. \tag{3.40}$$

From the definition of x_n and Lemma 2.8,

$$\begin{aligned}
\|x_{n+1} - u_n\| &\leq \xi_1^{-1} \|\alpha_n J_{E_1} u + (1-\alpha_n) J_{E_1} \Pi_{C_n} u_n - J_{E_1} u_n\| \\
&\leq \alpha_n K_1 + \xi_1^{-1} \|J_{E_1} \Pi_{C_n} u_n - J_{E_1} u_n\| ,
\end{aligned} \tag{3.41}$$

for some constant $K_1 > 0$. Since J_{E_1} is norm to norm uniformly continuous on bounded subsets of E_1, we conclude from (3.40) and (3.41) that

$$\lim_{n\to\infty} \|x_{n+1} - u_n\| = 0. \tag{3.42}$$

Therefore, combining (3.38) and (3.42) yield

$$\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0. \tag{3.43}$$

Similarly, one can show that

$$\lim_{n\to\infty} \|y_{n+1} - y_n\| = 0. \tag{3.44}$$

Since (x_n) and (y_n) are bounded by Lemma 3.6, we obtain from (3.33)

$$L^* \left[\tau_n^2 \|u_n - z_n\|^4 + \kappa_n^2 \|v_n - w_n\|^4\right] \leq \Theta_n^* - \Theta_{n+1}^* + \alpha_n \Lambda_{n+1}, \tag{3.45}$$

But the convergence of (Θ_n^*) and Assumption (B3) imply that

$$\lim_{n\to\infty} \tau_n \|u_n - z_n\|^2 = 0 \quad \text{and} \quad \lim_{n\to\infty} \kappa_n \|v_n - w_n\|^2 = 0. \tag{3.46}$$

Now, we deduce from Lemma 3.6 that $((x_n, y_n))$ is a bounded sequence in $C \times D$. Therefore, there exists a subsequence $((x_{n_k}, y_{n_k}))$ of $((x_n, y_n))$ such that $((x_{n_k}, y_{n_k}))$ converges weakly to (x, y) in $E_1 \times E_2$ and

$$\limsup_{n\to\infty} \Lambda_n = \lim_{k\to\infty} \Lambda_{n_k}. \tag{3.47}$$

It then follows that (x_{n_k}) converges weakly to x in E_1 and (y_{n_k}) converges weakly to y in E_2. From (3.38), (u_{n_k}) converges weakly to x in E_1 and from (3.39), (v_{n_k}) converges weakly to y in E_2. Using (3.46) and Lemma 3.4, we conclude that $x \in \mathrm{VI}(C, T)$ and $y \in \mathrm{VI}(D, S)$, respectively. Moreover,

$$\|Ax - By\|^2 \le 2\langle Ax - Ax_{n_k} + By_{n_k} - By, J_{E_3}(Ax - By)\rangle + \|Ax_{n_k} - By_{n_k}\|^2.$$

Since (x_{n_k}) converges weakly to x, it follows that (Ax_{n_k}) converges weakly to Ax. Similarly, (y_{n_k}) converges weakly to y implies that (By_{n_k}) converges weakly to By. Using (3.37), we get $Ax = By$. Consequently, $(x, y) \in \Gamma$.
From (3.25), (3.47), (3.43), (3.44) and Lemma 2.3, we obtain

$$\begin{aligned}
\limsup_{n\to\infty} \Lambda_{n+1} &\le \limsup_{n\to\infty} \Lambda_n + M_1 \limsup_{n\to\infty} [\|x_{n+1} - x_n\| + \|y_{n+1} - y_n\|] \\
&= \lim_{k\to\infty} \Lambda_{n_k} + M_1 \lim_{k\to\infty} [\|x_{n_k+1} - x_{n_k}\| + \|y_{n_k+1} - y_{n_k}\|] \\
&= 2\langle (x, y) - (x^*, y^*), (J_{E_1}u, J_{E_2}v) - (J_{E_1}x^*, J_{E_2}y^*)\rangle \\
&\le 0.
\end{aligned} \tag{3.48}$$

Finally, from (3.33), we have $\Theta^*_{n+1} \le (1-\alpha_n)\Theta^*_n + \alpha_n\Lambda_{n+1}$. Therefore, from (3.48) and Lemma 2.6, we conclude that (Θ^*_n) converges to zero as $n \to \infty$. That is, $\phi(x^*, x_n) \to 0$ and $\phi(y^*, y_n) \to 0$ as $n \to \infty$. Hence by Lemma 2.2, we have (x_n) and (y_n) converges to x^* and y^*, respectively.
Case II. Assume that there exists a subsequence $(\Theta^*_{n_i})$ of (Θ^*_n) such that $\Theta^*_{n_i} < \Theta^*_{n_i+1}$ for all $i \ge 0$. Then in view of Lemma 2.7, we can define a nondecreasing sequence $(m_k) \subset \mathbb{N}$ such that $m_k \to \infty$ as $k \to \infty$ and

$$\Theta^*_{m_k} \le \Theta^*_{m_k+1} \quad \text{and} \quad \Theta^*_k \le \Theta^*_{m_k+1} \tag{3.49}$$

for all $k \in \mathbb{N}$. Following similar steps as in Case I, we derive

$$\limsup_{k\to\infty} \Lambda_{m_k+1} \le 0. \tag{3.50}$$

From (3.33) and (3.49), we obtain $\alpha_{m_k}\Theta^*_{m_k+1} \le \alpha_{m_k}\Lambda_{m_k+1}$, which reduces to $\Theta^*_{m_k+1} \le \Lambda_{m_k+1}$. Taking the limit as $k \to \infty$ and using (3.50), we conclude that $\Theta^*_{m_k+1} \to 0$ as $k \to \infty$. Again from (3.49), it follows that $\Theta^*_k \to 0$ as $k \to \infty$. Therefore, $\phi(x^*, x_k) \to 0$ and $\phi(y^*, y_k) \to 0$ as $k \to \infty$. Hence by Lemma 2.2, we have $x_k \to x^*$ and $y_k \to y^*$ as $k \to \infty$. □

If $u = 0$ and $v = 0$, then Algorithm 3.1 can be used to locate an element of the solution with the minimum norm.

Corollary 3.8. *Let the Assumptions (A1) – (A5) and (B1) – (B4) hold. Then, the sequence $((x_n, y_n))$ generated by Algorithm 3.1 with $u = 0 = v$ converges strongly to the minimum norm point $(x^*, y^*) \in \Gamma$, that is, $(x^*, y^*) = \prod_\Gamma(0, 0)$.*

Corollary 3.9. *Assume that $T : E_1 \to E_1^*$ and $S : E_2 \to E_2^*$ are uniformly continuous and monotone mappings. Let the Assumptions (A1) – (A3), (A5) and (B1) – (B4) be satisfied. Then the sequence $((x_n, y_n))$ generated by Algorithm 3.1 converges strongly to $(x^*, y^*) \in \Gamma$, where $(x^*, y^*) = \prod_\Gamma(u, v)$.*

Proof. The mappings T and S are pseudomonotone, hence by Lemma 3.6, (x_n) and (y_n) are bounded. It then follows from (3.46) and Remark 3.5 that $x \in \mathrm{VI}(C, T)$ and $y \in \mathrm{VI}(D, S)$, where x and y are weak cluster points of (x_n) and (y_n), respectively. The rest of the proof is similar to the proof of Theorem 3.7. $\square$

4. Applications

In this section, we apply our main result to solve the following problems: split equality zero point problem (SEZPP), common solutions of the variational inequality problem, common zeros of pseudomonotone mappings, split variational inequality problem, split zero point problem (SZPP), split equality feasibility problem (SEFP) and split feasibility problem (SFP).

4.1. Split equality zero point problem

If $C = E_1$ and $D = E_2$, then the SEVIP reduces to the SEZPP, which is to find $x \in T^{-1}(0)$ and $y \in S^{-1}(0)$ such that $Ax = By$, where $T^{-1}(0) = \{p \in E_1 : 0 = Tp\}$ and $S^{-1}(0) = \{q \in E_2 : 0 = Sq\}$. Denote the solution of this problem by $\digamma = \{(p, q) \in E_1 \times E_2 : p \in T^{-1}(0), q \in S^{-1}(0) \text{ and } Ap = Bq\}$.

Corollary 4.1. *Assume that $\digamma \neq \emptyset$. Let the Assumptions (A1), (A2), (A4) and (B1) – (B4) be satisfied with $C = E_1$ and $D = E_2$. Then the sequence $((x_n, y_n))$ generated by Algorithm 3.1 converges strongly to $(x^*, y^*) \in \digamma$, where $(x^*, y^*) = \prod_\digamma(u, v)$.*

4.2. Common solutions of the variational inequality problem

Let $E = E_1 = E_2 = E_3$, $A = I$ and $B = I$. In this case, the SEVIP reduces to finding common solutions of two variational inequality problems for pseudomonotone mappings. Denote $\mathcal{F} = \{(p, q) \in C \times D : \langle x - p, Tp\rangle \geq 0, \forall x \in C$ and $\langle y - q, Sq\rangle \geq 0$, $\forall y \in D$ such that $p = q\}$.

Corollary 4.2. *Assume that $\mathcal{F} \neq \emptyset$. Let the Assumptions (A1), (A3), (A4) and (B1) – (B4) be satisfied with $E = E_1 = E_2 = E_3$ and $A = I = B$. Then the sequence $((x_n, y_n))$ generated by Algorithm 3.1 converges strongly to $(x^*, y^*) \in \mathcal{F}$, where $(x^*, y^*) = \prod_{\mathcal{F}}(u, v)$.*

4.3. Common zeros of pseudomonotone mappings

Let $E = E_1 = E_2 = E_3$, $A = I$ and $B = I$. If $C = E$ and $D = E$, then the SEVIP reduces to finding common zeros of pseudomonotone mappings. Denote $\mathcal{F}' = \{(p, q) \in E \times E : p \in T^{-1}(0) \text{ and } q \in S^{-1}(0) \text{ such that } p = q\}$.

Corollary 4.3. *Let the Assumptions (A1), (A4), (B1) – (B4) be satisfied with $C = D = E = E_1 = E_2 = E_3$ and $A = I = B$. If $\mathcal{F}' \neq \emptyset$, then the sequence $((x_n, y_n))$ generated by Algorithm 3.1 converges strongly to $(x^*, y^*) \in \mathcal{F}'$, where $(x^*, y^*) = \prod_{\mathcal{F}'}(u, v)$.*

4.4. Split variational inequality problem

If $E_2 = E_3$ and $B = I$, then the SEVIP reduces to the split variational inequality problem (SVIP) which is to find $x \in VI(C,T)$ and $y \in VI(D,S)$ such that $Ax = y$. Denote $\mathcal{T} = \{(p,q) \in C \times D : p \in VI(C,T), q \in VI(D,S) \text{ and } Ap = q\}$.

Corollary 4.4. *Assume that the Assumptions (A1) – (A4) and (B1) – (B4) hold with $E_3 = E_2$ and $B = I$. If $\mathcal{T} \neq \emptyset$, then the sequence $((x_n, y_n))$ generated by Algorithm 3.1 converges strongly to $(x^*, y^*) \in \mathcal{T}$, where $(x^*, y^*) = \prod_{\mathcal{T}}(u,v)$.*

4.5. Split zero point problem

Let $E = E_2 = E_3$ and $B = I$. If $C = E$ and $D = E$, then the SEVIP reduces to the SZPP which is to find $x \in T^{-1}(0)$ and $y \in S^{-1}(0)$ such that $Ax = y$. Denote $\mathcal{S} = \{(p,q) \in E \times E : p \in T^{-1}(0), q \in S^{-1}(0) \text{ and } Ap = q\}$.

Corollary 4.5. *Assume that the Assumptions (A1), (A2), (A4), (B1) – (B3) and (B4) with $B = I$ hold. If $\mathcal{S} \neq \emptyset$, then the sequence $((x_n, y_n))$ generated by Algorithm 3.1 converges strongly to $(x^*, y^*) \in \mathcal{S}$, where $(x^*, y^*) = \prod_{\mathcal{S}}(u,v)$.*

Remark 4.6. (a). If $E = E_1 = E_2 = E_3$, $S = 0$, $A = 0$ and $B = 0$, then Theorem 3.7 can be used to find solutions of variational inequality problems for uniformly continuous pseudomonotone mappings that are sequentially weakly continuous on bounded subsets of E as well as for uniformly continuous monotone mappings. If in addition, we take $C = E$ and $D = E$, then Corollary 4.1 will approximate zeros of uniformly continuous pseudomonotone mappings that are sequentially weakly continuous on bounded subsets of E and also zeros of uniformly continuous monotone mappings.
(b). In view of Corollary 3.9, and the discussion in this section, one can use the results of this section to find solutions of split equality zero point problem for uniformly continuous monotone mappings, common solutions of the variational inequality problem for uniformly continuous monotone mappings, common zeros of uniformly continuous monotone mappings, split variational inequality problems for uniformly continuous monotone mappings, split zero point problem for monotone mappings.
(c). The special cases of the above results can be obtained by taking $E_1 = H_1$, $E_2 = H_2$ and $E_3 = H_3$ to be real Hilbert spaces.

Note that if $E = H$, a real Hilbert space, then $J_E = I$, the identity mapping on H, and $\Pi_C = P_C$, the metric projection onto C. A well known example of a uniformly continuous, monotone and hence pseudomonotone map is $I - P_C$. Henceforth, $E_1 = H_1$, $E_2 = H_2$ and $E_3 = H_3$ are real Hilbert spaces, $C \subset H_1$ and $D \subset H_2$ are nonempty, closed and convex subsets. Also $A : H_1 \to H_3$ and $B : H_2 \to H_3$ are bounded linear mappings with adjoints A^* and B^*, respectively, $T = I - P_C$ and $S = I - P_D$. Thus, we obtain the following applications in Hilbert spaces.

4.6. Split equality feasibility problem

Replacing T with $I - P_C$ and S with $I - P_D$, then the SEZPP reduces to the SEFP which seeks to find $x \in C$ and $y \in D$ such that $Ax = By$. Denote

$$\Gamma' = \{(p,q) \in C \times D : Ax = By\}.$$

Corollary 4.7. *Assume that $\Gamma' \neq \emptyset$. Let the Assumptions (B2),(B3), (B4) be satisfied. Then the sequence $((x_n, y_n))$ generated by Algorithm 3.1 converges strongly to $(x^*, y^*) \in \Gamma'$, where $(x^*, y^*) = P_{\Gamma'}(u, v)$.*

4.7. Split feasibility problem

Setting $T = I - P_C$, $S = I - P_D$, $H_2 = H_3$ and $B = I$, the identity mapping on H_2, then the SEFP reduces to the SFP which seeks to find $x \in C$ such that $Ax \in D$. This problem can also be expressed as a problem of finding $x \in C$ and $y \in D$ such that $Ax = y$. Denote $\mathcal{S}'' = \{(p, q) \in C \times D : Ap = q\}$.

Corollary 4.8. *Assume that $\mathcal{S}'' \neq \emptyset$. Let the Assumptions (B2), (B3), (B4) with $B = I$ be satisfied. Then the sequence $((x_n, y_n))$ generated by Algorithm 3.1 with $B = I$ converges strongly to $(x^*, y^*) \in \mathcal{S}''$, where $(x^*, y^*) = P_{\mathcal{S}''}(u, v)$.*

Remark 4.9. If we take $u = 0$ and $v = 0$, then one can obtain elements of minimum norm for all the application areas listed in this section.

5. Numerical example

In this section, we give a numerical example to demonstrate that the sequence $(z_n) = ((x_n, y_n))$ generated by Algorithm 3.1 converges to an element $z^* = (x^*, y^*)$ in Γ for different initial values $z_0 = (x_0, y_0)$.

Example 5.1. Let $\|\cdot\|$ be the norm on $\mathbb{R}^2$ induced by the inner product $\langle\cdot,\cdot\rangle$. Define the map $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(x, y) = \left(\frac{3}{2} + \sqrt{x^2 + y^2}\right)(x - 1, y)$. Assume that $\langle T(x, y), (u, v) - (x, y)\rangle \geq 0$ for all $(x, y), (u, v) \in \mathbb{R}^2$. Then

$$\left(\frac{3}{2} + \sqrt{x^2 + y^2}\right)\langle (x - 1, y), (u - x, v - y)\rangle \geq 0,$$

which implies that $\langle (x - 1, y), (u - x, v - y)\rangle \geq 0$ for all $(x, y), (u, v) \in \mathbb{R}^2$. Therefore,

$$\begin{aligned}\langle T(u, v), (u, v) - (x, y)\rangle &= \left(\frac{3}{2} + \sqrt{u^2 + v^2}\right)\langle (x - 1, y), (u - x, v - y)\rangle \\ + &\left(\frac{3}{2} + \sqrt{u^2 + v^2}\right)[\langle (u - 1, v), (u - x, v - y)\rangle - \langle (x - 1, y), (u - x, v - y)\rangle] \\ \geq &\left(\frac{3}{2} + \sqrt{u^2 + v^2}\right)\|(u - x, v - y)\|^2 \geq 0,\end{aligned}$$

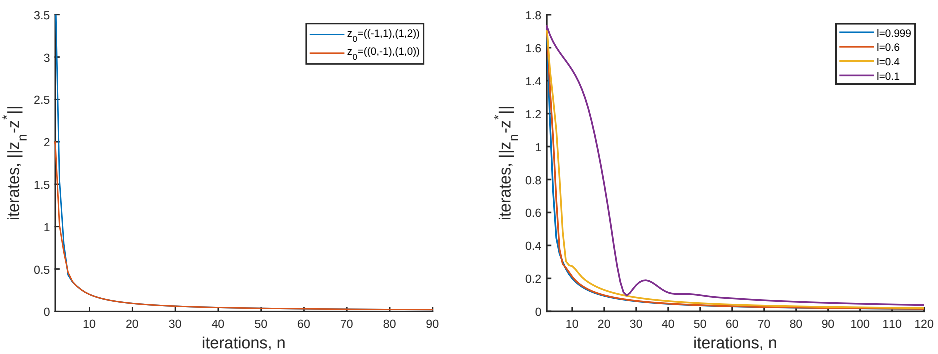
for all $(x, y), (u, v) \in \mathbb{R}^2$, showing that T is pseudomonotone on $\mathbb{R}^2$.

Example 5.2. Let $E_1 = E_2 = E_3 = \mathbb{R}^2$ be equipped with the usual norm. Assume that $A, B : \mathbb{R}^2 \to \mathbb{R}^2$ are given by $A(x, y) = (0, 3y)$ and $B(x, y) = (2x, 0)$ with adjoints $A^*(x, y) = (0, 3y)$ and $B^*(x, y) = (2x, 0)$, respectively.
Let $C = \{(x, y) \in \mathbb{R}^2 : \sqrt{x^2 + y^2} \leq 1\}$ and $D = \{(x, y) \in \mathbb{R}^2 : \sqrt{x^2 + y^2} \leq 3\}$.
Let $S, T : \mathbb{R}^2 \to \mathbb{R}^2$ be pseudomonotone maps defined as in Examples 2.13 and 5.1, respectively.
Then $\langle T(1, 0), (x, y) - (1, 0)\rangle \geq 0$ for all $(x, y) \in C$, $\langle S(0, -1), (x', y') - (0, -1)\rangle \geq 0$

for all $(x', y') \in D$ and $A(1,0) = (0,0) = B(0,-1)$, and so $((1,0),(0,-1)) \in \Gamma \neq \emptyset$. Also, let $\mu = 0.9$, $\lambda = 1$ and let

$$\gamma_n = \begin{cases} \frac{\|Ax_n - By_n\|^2}{8\left[\|A^*(Ax_n - By_n)\|^2 + \|B^*(Ax_n - By_n)\|^2\right]} & \text{if } n \in \Omega \\ \frac{1}{10000} & \text{if } n \notin \Omega. \end{cases}$$

and $\alpha_n = \frac{1}{n+1}$. Thus assumptions (A1) – (A5) and (B1) – (B4) are satisfied. Using MATLAB, we get Figure 1 below which shows that for any choice of initial values the sequence generated by Algorithm 3.1 converges to a solution of the split equality variational inequality problem. The numerical example also shows that the convergence of the sequence is faster if the parameter $l \in (0,1)$ is closer to 1 compared to when it is closer to 0.



(A) Parameter $l = 0.9$ (B) $z_0 = ((0.5, 0.5), (0.5, 0.5))$

FIGURE 1. Convergence of (z_n) to z^*

6. Conclusion

In this paper, we have constructed an algorithm for solving the split equality variational inequality problem in real uniformly smooth and uniformly convex Banach space settings. We also established a strong convergence theorem under the assumption that the associated mappings are uniformly continuous, pseudomonotone and sequentially weakly continuous. The algorithm does not require prior knowledge of operator norms of A and B. We also gave some applications of our results to some problems in Banach spaces. A numerical example was also provided to demonstrate the behavior of the convergence of the proposed algorithm. Our results in this paper extend the results of Censor *et al.* [11], Byrne *et al.* [6] and Thong *et al* [33] to a more general SEVIP in uniformly smooth and uniformly convex Banach spaces more general than Hilbert spaces.

References

[1] Alber, Ya., *Metric and generalized projection operators in Banach spaces: Properties and Applications*, In: "Theory and Applications of Nonlinear Operators of Accretive and Monotone Type" (A. G. Kartsatos Ed.), Lecture Notes in Pure and Appl. Math., Vol. 178, Dekker, New York, 1996, 15-50.

[2] Baiocchi, C., Capelo, A., *Variational and Quasivariational Inequalities; Applications to Free Boundary Problems*, New York, Wiley, 1984.

[3] Boikanyo, O.A., *A strongly convergent algorithm for the split common fixed point problem*, Appl. Math. Comput., **265**(2015), 844-853.

[4] Boikanyo, O.A., Zegeye, H., *The split equality fixed point problem for quasi-pseudo-contractive mappings without prior knowledge of norms*, Numer. Funct. Anal. Optim., **41**(2020), no. 7, 759-777.

[5] Byrne, C., *Iterative oblique projection onto convex sets and the split feasibility problem*, Inverse Problems, **18**(2002), 441-453.

[6] Byrne, C., Censor, Y., Gibali, A., Reich, S., *The split common null point problem*, J. Nonlinear Convex Anal., **13**(2012), 759-775.

[7] Ceng, L.C., Teboulle, M., Yao, J.C., *Weak convergence of an iterative method for pseudomonotone variational inequalities and fixed-point problems*, J. Optim. Theory Appl., **146**(2010), 19-31.

[8] Censor, Y., Bortfeld, T., Martin, B., Trofimov, A., *A unified approach for inversion problems in intensity modulated radiation therapy*, Phys. Med. Biol., **51**(2006), 2353-2365.

[9] Censor Y., Elfving, T., *A multiprojection algorithm using Bregman projections in a product space*, Numer. Algorithms, **8**(1994), 221-239.

[10] Censor, Y., Gibali, A., Reich, S., *The split variational inequality problem*, The Technion-Israel Institute of Technology, Haifa, (2010). arXiv:1009.3780.

[11] Censor, Y., Gibali, A., Reich, S., *Algorithms for the split variational inequality problem*, Numer. Algorithms, **59**(2012), 301-323.

[12] Censor, Y., Gibali, A., Reich, S., Sabach, S., *Common solutions to variational inequalities*, Set-Valued Var. Anal., **20**(2012), 229-247.

[13] Censor, Y., Segal, A., *The split common fixed point problem for directed operators*, J. Convex Anal., **16**(2009), 587-600.

[14] Chang, S.-S., Joseph Lee, H.W., Chan, C.K., Wang L., Qin, L.J., *Split feasibility problem for quasinonexpansive multi-valued mappings and total asymptotically strict pseudo-contractive mappings*, Appl. Math. Comput., **219**(2013), 10416-10424.

[15] Chang, S.-S., Yao, J.-C., Kim, J.J., Yang, L., *Iterative approximation to convex feasibility problems in Banach spaces*, Fixed Point Theory Appl., 2007:19, Article ID 4679.

[16] Chang, S.-S., Yao, J.-C., Wen, C.-F., Yang, L., Qin, L.-J., *Common zero for a finite family of monotone mappings in Hadamard spaces with applications*, Mediterr. J. Math., **15**(2018), Art. no. 160.

[17] Combettes, P.L., *The Convex Feasibility Problem in Image Recovery*, P. Hawkes (Ed.), Advances in Imaging and Electron Physics, vol. 95, Academic Press, New York, 1996, 155-270.

[18] Eslamian, M., Fakhri, A., *Split equality monotone variational inclusions and fixed point problem of set-valued operator*, Acta Univ. Sapientiae. Mathematica, **9**(2017), no. 1, 94-121.

[19] Fichera, G., *Problemi elastostatici con vincoli unilaterali: il problema di Signorini con ambigue condizioni al contorno*, Atti Acad. Naz. Lincei, Mem. Cl. Sci. Fis. Mat. Nat., Sez. I., (1964), no. 7, 91-140.

[20] Kamimura, S., Takahashi, W., *Strong convergence of proximal-type algorithm in a Banach space*, SIAM J. Optim., **13**(2002), 938-945.

[21] Khanh, P.D., Vuong, P.T., *Modified projection method for strongly pseudo-monotone variational inequalities*, J. Global Optim., **58**(2014), 341-350.

[22] Khobotov, E.N., *Modification of the extragradient method for solving variational inequalities and certain optimization problems*, USSR Comput. Math. Math. Phys., **27**(1987), no. 5, 120-127.

[23] Kinderlehrer, D., Stampacchia, G., *An Introduction to Variational Inequalities and Their Applications*, New York: Academic Press, 1980.

[24] Maingé, P.E., *Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization*, Set-Valued Anal., **16**(2008), 899-912.

[25] Marcotte, P., *Applications of Khobotov's algorithm to variational and network equilibrium problems*, Info Syst Oper. Res., **29**(1991), no. 4, 258-270.

[26] Mashreghi, J., Nasri, M., *Forcing strong convergence of Korpelevich's method in Banach spaces with its applications in game theory*, Nonlinear Anal., **72**(2010), 2086-2099.

[27] Moudafi, A., *Split monotone variational inclusions*, J. Optim. Theory Appl., **150**(2011), 275-283.

[28] Phelps, R.P., *Convex Functions, Monotone Operators and Differentiability*, Lecture Notes in Mathematics, vol. 1364, 2nd edn., Springer Verlag, Berlin, 1993.

[29] Shahzad, S., Zegeye, H., *Convergence theorems of common solutions for fixed point, variational inequality and equilibrium problems*, J. Nonlinear Var. Anal., **3**(2019), no. 2, 189-203.

[30] Shehu, Y., Dong, Q.-L., Jiang, D., *Single projection method for pseudo-monotone variational inequality in Hilbert spaces*, Optimization, **68**(2019), 385-409.

[31] Stampacchia, G., *Formes bilinéaires coercitives sur les ensemble convexes*, C.R. Acad. Sci. Paris., **258**(1964), 4413-4416.

[32] Takahashi, W., *Nonlinear Functional Analysis*, Kindikagaku, Tokyo, 1988.

[33] Thong, D.V., Shehu, Y., Iyiola, O.S., *A new iterative method for solving pseudomonotone variational inequalities with non-Lipschitz operators*, Comput. Appl. Math., **39**(2020), no. 2, paper no. 108, 24 pp.

[34] Vuong, P.T., Shehu, Y., *Convergence of an extragradient-type method for variational inequality with applications to optimal control problems*, Numer. Algorithms, **81**(2019), 269-291.

[35] Wang, F., Xu, H.K., *Cyclic algorithms for split feasibility problems in Hilbert spaces*, Nonlinear Anal., **74**(2011), 4105-4111.

[36] Wega, G.B., Zegeye, H., *Convergence results of Forward-Backward method for a zero of the sum of maximally monotone mappings in Banach spaces*, Comp. Appl. Math. 39, Article no. 223 (2020).

[37] Xu, H.K., *Another control condition in an iterative method for nonexpansive mappings*, Bull. Aust. Math. Soc., **65**(2002), 109-113.

[38] Yao Y., Postolache, M., *Iterative methods for pseudo-monotone variational inequalities and fixed-point problems*, J. Optim. Theory Appl., **155**(2012), 273-287.

[39] Zalinescu, C., *Convex Analysis in General Vector Spaces*, World Scientific, River Edge, 2002.

[40] Zegeye, H., *Strong convergence theorems for split equality fixed point problems of η-demimetric in Banach spaces*, DCDIS Ser. B Appl. Algor., **26**(2019), 269-288.

[41] Zegeye, H., Shahzad, N., *Approximating common solution of variational inequality problems for two monotone mappings in Banach spaces*, Optim. Lett., **5**(2011), no. 4, 691-704.

Oganeditse A. Boikanyo
Botswana International University of Science and Technology,
Department of Mathematics and Statistical Sciences,
Private Bag 16, Palapye, Botswana
e-mail: `boikanyoa@gmail.com`

Habtu Zegeye
Botswana International University of Science and Technology,
Department of Mathematics and Statistical Sciences,
Private Bag 16, Palapye, Botswana
e-mail: `habtuzh@yahoo.com`

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 159–177
DOI: 10.24193/subbmath.2021.1.14

A spectral criterion for the existence of the stabilizing solution of a class of Riccati type differential equations with periodic coefficients

Vasile Drăgan and Ioan-Lucian Popa

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. In this paper, we investigate the existence and uniqueness of a stabilizing solution to a periodic backward nonlinear differential equation. This class of nonlinear equations includes as special cases many of the continuous-time Riccati equations arising both in deterministic and stochastic linear quadratic (LQ) type control problems.

Mathematics Subject Classification (2010): 49N10, 93E20.

Keywords: Generalized Riccati equations, stabilizing solution, the periodic case, characteristic multipliers.

1. Introduction

Matrix Riccati differential equations have been extensively studied during the past decades by researchers both in control theory and differential equations area. First, in the work of Kalman [13] it is obtained the existence of global solution for a matrix Riccati differential equation under the controllability and observability conditions. Second, Wonham [20] established these results to the framework of stochastic control considering the so-called Riccati equations of stochastic control. Both results have been related to the so-called LQ optimisation problem.

The above earlier results impose controllability and observability conditions, which are somewhat conservative. More recent works replace these conditions by some weaker ones, as stabilizability and detectability.

In the particular case of constant coefficients, the well-known (standard) algebraic Riccati equation arise naturally and play an essential role. The solutions, whose existence is assured by Popov conditions, properties and applications are studied in many works, see e.g. [8], [15], and [18]. In the study of stochastic case, a class of modified algebraic Riccati equations appear. The result from [23] provides a necessary and sufficient condition for the existence of a mean-square stabilizing solution for this kind of equations. This assures that the corresponding stochastic LQ problem is solvable, i.e. the cost function is minimized and the stochastic closed-loop system is mean-square stable. Also, it was shown in [10] that a mean-square stabilizing solution of a modified algebraic Riccati equation, is unique (if it exists) and coincides with the maximal solution. In [19] it is considered LQ optimal stochastic control problem with random coefficients via the stochastic maximum principle.

Further, if we consider the case of periodic coefficients, the early results that provide the existence of a stabilizing solution for a matrix Riccati differential equation goes to [21]. In addition, the method developed for algebraic Riccati equation has been extended to the Riccati equations for continuous-time linear periodic systems in [17]. More recently, [14] extends the eigenvalue method to the differential Riccati equation in terms of nonlinear eigenvalues and eigenvectors. Various results concerning periodic systems are presented in [1].

The novelty to be pursued in this paper is the use of spectral theory of positive operators to deal with generalized Riccati differential equation (GRDE) in order to obtain necessary and sufficient conditions for the existence and uniqueness of the stabilizing solution. The contribution of the paper can be summarised as follows. First, regarding the set-up of the problem, i.e. the class of nonlinear backward differential equation we first present several special cases. We consider LQ optimal regulator and stochastic LQ optimal regulator. For stochastic framework the cases when the controlled system is affected by multiplicative white noise and of the perturbation model by a Markov process simultaneous, respectively, is considered. This indicates that the considered problem arises in a natural way in optimal control problems and incorporates all these cases. Discussions on backward differential equations and applications can be found in [6] and [22].

Motivated by this goal, we present the definition of stabilizing solution of the class of nonlinear continuous-time backward equation under consideration, see Definition 2.5. Further, we introduce the concept of unobservable characteristic multipliers of a pair formed by a continuous-time linear equation with periodic coefficients and an output, see Definition 3.2. In this framework it is proved the central contribution of the paper, i.e. the existence of the stabilizing solution of the GRDE under consideration, see Theorem 4.3.

Organization of the paper. First, in Subsection 1.1 we introduce useful notations. Section 2 is devoted to the formulation of the problem. It presents our framework, motivated by several special cases of LQ control problems related to our model. Section 3 is dedicated to state and describe some preliminary results, while Section 4 contains the main result of the paper, i.e. the existence of the positive semidefinite and periodic stabilizing solution of the GRDE. Finally, in Section 5, we draw conclusions.

1.1. Notations

The notations used in this work are in general the standard ones. Here we mention some notations less met. If m, n, N are fixed natural numbers, then

$$\mathcal{M}_{nm}^{N} \triangleq \underbrace{\mathbb{R}^{n\times m} \times \cdots \times \mathbb{R}^{n\times m}}_{N\ \ times},$$

$\mathbb{R}^{n\times m}$ being the linear space of the matrices with $n-$rows and $m-$columns. $\mathbf{B} \in \mathcal{M}_{nm}^{N}$ if and only if $\mathbf{B} = (B_1, B_2, \ldots, B_N)$, $B_i \in \mathbb{R}^{n\times m}$. When $n = m$ we shall write $\mathcal{M}_{n}^{N}$ instead of $\mathcal{M}_{nn}^{N}$. If $\mathbf{B} \in \mathcal{M}_{nm}^{N}$, $\mathbf{C} \in \mathcal{M}_{mp}^{N}$ then $\mathbf{BC}$ is defined by

$$\mathbf{BC} = (B_1C_1, B_2C_2, \ldots, B_NC_N) \in \mathcal{M}_{np}^{N}.$$

If $\mathbf{A} = (A_1, A_2, \ldots, A_N) \in \mathcal{M}_{n}^{N}$ is such that $det\ A_i \neq 0$, $1 \leq i \leq N$, then

$$\mathbf{A}^{-1} \triangleq (A_1^{-1}, A_2^{-1}, \ldots, A_N^{-1}) \in \mathcal{M}_{n}^{N}.$$

If $\mathbf{X} = (X_1, X_2, \ldots, X_N) \in \mathcal{M}_{mp}^{N}$ then

$$\mathbf{X}^{T} \triangleq (X_1^{T}, X_2^{T}, \ldots, X_N^{T}) \in \mathcal{M}_{pm}^{N}.$$

Here and in the sequel superscript $()^{T}$ stands for the transpose of a matrix or a vector. $\mathcal{S}_n \subset \mathbb{R}^{n\times n}$ stands for the linear space of symmetric matrices of size $n \times n$ and

$$\mathcal{S}_{n}^{N} \triangleq \underbrace{\mathcal{S}_n \times \mathcal{S}_n \times \cdots \times \mathcal{S}_n}_{N\ \ times}.$$

$\mathcal{S}_{n}^{N}$ has a structure of finite dimensional real Hilbert space induced by the inner product

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{i=1}^{N} Tr\ [X_iY_i], \tag{1.1}$$

for all $\mathbf{X} = (X_1, X_2, \ldots, X_N)$, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)$ from $\mathcal{S}_{n}^{N}$. $Tr\ [\cdot]$ denotes the trace of a matrix. On $\mathcal{S}_{n}^{N}$ one should consider the order relation $\succeq$ induced by the solid, closed, normal, convex cone

$$\mathcal{S}_{n}^{N+} \triangleq \{\mathbf{X} \in \mathcal{S}_{n}^{N} \mid \mathbf{X} = (X_1, X_2, \ldots, X_N) \ \text{ with } \ X_i \geq 0,\ 1 \leq i \leq N\}.$$

Here $X_i \geq 0$ means that X_i is positive semidefinite matrix. If $\mathcal{X}$, $\mathcal{Y}$ are two vector spaces, then $\mathcal{B}(\mathcal{X}, \mathcal{Y})$ denotes the linear space of linear operators $T : \mathcal{X} \to \mathcal{Y}$. If $\mathcal{X} = \mathcal{Y}$ we shall write $\mathcal{B}(\mathcal{X})$ instead of $\mathcal{B}(\mathcal{X}, \mathcal{X})$.

2. A class of nonlinear backward differential equations

2.1. Model description and basic assumptions

On the Hilbert space $\mathcal{S}_{n}^{N}$ we consider backward differential equation

$$-\frac{d}{dt}\mathbf{X}(t) = \mathcal{R}(t, \mathbf{X}(t)), \ \ t \in \mathbb{R}, \tag{2.1}$$

with the unknown function $\mathbf{X}(t) = (X(t,1), X(t,2), \dots, X(t,N)) \in \mathcal{S}_n^N$. In (2.1) the pair $(t, \mathbf{X}) \to \mathcal{R}(t, \mathbf{X}) :\ Dom\ \mathcal{R} \subset \mathbb{R} \times \mathcal{S}_n^N \to \mathcal{S}_n^N$ is described by

$$\mathcal{R}(t,\mathbf{X}) = (\mathcal{R}_1(t,\mathbf{X}), \mathcal{R}_2(t,\mathbf{X}), \dots, \mathcal{R}_N(t,\mathbf{X}))$$

with

$$\begin{aligned}\mathcal{R}_i(t,\mathbf{X}) =& A^T(t,i)X(i) + X(i)A(t,i) + \Pi_1(t)[\mathbf{X}](i)\\ &- (X(i)B(t,i) + \Pi_2(t)[\mathbf{X}](i) + L(t,i))\\ &\cdot (\Pi_3(t)[\mathbf{X}](i) + R(t,i))^{-1}(X(i)B(t,i)\\ &+ \Pi_2(t)[\mathbf{X}](i) + L(t,i))^T + M(t,i), \quad 1 \le i \le N. \end{aligned} \tag{2.2}$$

Here,

$$\mathbf{X} \to \mathbf{\Pi}_1(t)[\mathbf{X}] \triangleq (\Pi_1(t)[\mathbf{X}](1), \dots, \Pi_1(t)[\mathbf{X}](N)) : \mathcal{S}_n^N \to \mathcal{S}_n^N;$$

$$\mathbf{X} \to \mathbf{\Pi}_2(t)[\mathbf{X}] \triangleq (\Pi_2(t)[\mathbf{X}](1), \dots, \Pi_2(t)[\mathbf{X}](N)) : \mathcal{S}_n^N \to \mathcal{M}_{nm}^N;$$

$$\mathbf{X} \to \mathbf{\Pi}_3(t)[\mathbf{X}] \triangleq (\Pi_3(t)[\mathbf{X}](1), \dots, \Pi_3(t)[\mathbf{X}](N)) : \mathcal{S}_n^N \to \mathcal{S}_m^N;$$

are given linear operators.

$$Dom\ \mathcal{R} \triangleq \{(t,\mathbf{X}) \in \mathbb{R} \times \mathcal{S}_n^N \,|\, det\ (\Pi_3(t)[\mathbf{X}](i) + R(t,i)) \neq 0,\ 1 \le i \le N\}.$$

According to the convention of notations from Subsection 1.1, we may rewrite (2.2) in a compact form as:

$$\begin{aligned}\mathcal{R}(t,\mathbf{X}) =& \mathbf{A}^T(t)\mathbf{X} + \mathbf{X}\mathbf{A}(t) + \mathbf{\Pi}_1(t)[\mathbf{X}] - (\mathbf{X}\mathbf{B}(t) + \mathbf{\Pi}_2(t)[\mathbf{X}] + \mathbf{L}(t))\\ &\cdot (\mathbf{\Pi}_3(t)[\mathbf{X}] + \mathbf{R}(t))^{-1}(\mathbf{X}\mathbf{B}(t) + \mathbf{\Pi}_2(t)[\mathbf{X}] + \mathbf{L}(t))^T + \mathbf{M}(t)\end{aligned} \tag{2.3}$$

for all $(t, \mathbf{X}) \in Dom\ \mathcal{R}$, where we have used the following notations:

$$\begin{aligned}&\mathbf{A}(t) = (A(t,1), A(t,2), \dots, A(t,N)) \in \mathcal{M}_n^N,\\ &\mathbf{B}(t) = (B(t,1), B(t,2), \dots, B(t,N)) \in \mathcal{M}_{nm}^N,\end{aligned} \tag{2.4a}$$

$$\begin{aligned}&\mathbf{M}(t) = (M(t,1), M(t,2), \dots, M(t,N)) \in \mathcal{S}_n^N,\\ &\mathbf{L}(t) = (L(t,1), L(t,2), \dots, L(t,N)) \in \mathcal{M}_{nm}^N,\\ &\mathbf{R}(t) = (R(t,1), R(t,2), \dots, R(t,N)) \in \mathcal{S}_m^N.\end{aligned} \tag{2.4b}$$

Based on the operators $\Pi_k(t)[\cdot](i)$ involved in (2.2) we define:

$$\mathbf{\Pi}(t)[\mathbf{X}] = (\Pi(t)[\mathbf{X}](1), \Pi(t)[\mathbf{X}](2), \dots, \Pi(t)[\mathbf{X}](N))$$

as

$$\Pi(t)[\mathbf{X}](i) = \begin{pmatrix} \Pi_1(t)[\mathbf{X}](i) & \Pi_2(t)[\mathbf{X}](i) \\ \Pi_2^T(t)[\mathbf{X}](i) & \Pi_3(t)[\mathbf{X}](i) \end{pmatrix} \in \mathcal{S}_{n+m}^N. \tag{2.5}$$

Hence, for each $t \in \mathbb{R}$, $\mathbf{X} \to \mathbf{\Pi}(t)[\mathbf{X}] : \mathcal{S}_n^N \to \mathcal{S}_{n+m}^N$ is a linear operator.

Remark 2.1. From (2.3) one sees that the operator $\mathcal{R}(\cdot,\cdot)$ is defined by the quadruple $(\mathbf{A}(\cdot), \mathbf{B}(\cdot), \mathbf{\Pi}(\cdot), \mathbf{Q}(\cdot))$ where $\mathbf{A}(\cdot)$, $\mathbf{B}(\cdot)$ are described in (2.4a), the operator valued function $\mathbf{X} \to \mathbf{\Pi}(\cdot)[\mathbf{X}]$ is described in (2.5) and

$$\mathbf{Q}(t) = \begin{pmatrix} \mathbf{M}(t) & \mathbf{L}(t) \\ \mathbf{L}^T(t) & \mathbf{R}(t) \end{pmatrix} \in \mathcal{S}_{n+m}^N. \tag{2.6}$$

Obvious the components $Q(t,i)$ of $\mathbf{Q}(t)$ are

$$Q(t,i) = \begin{pmatrix} M(t,i) & L(t,i) \\ L^T(t,i) & R(t,i) \end{pmatrix} \in \mathcal{S}_{n+m},\ 1 \le i \le N.$$

The developments from this work are done under the following assumption: (**H**).

(a). $\mathbf{A}(\cdot) : \mathbb{R} \to \mathcal{M}_n^N$, $\mathbf{B}(\cdot) : \mathbb{R} \to \mathcal{M}_{nm}^N$, $\mathbf{M}(\cdot) : \mathbb{R} \to \mathcal{S}_n^N$, $\mathbf{L}(\cdot) : \mathbb{R} \to \mathcal{M}_{nm}^N$, $\mathbf{R}(\cdot) : \mathbb{R} \to \mathcal{S}_n^N$ are continuous functions which are periodic of period θ.

(b). $t \to \mathbf{\Pi}(t)[\cdot] : \mathbb{R} \to \mathcal{B}(\mathcal{S}_n^N, \mathcal{S}_{n+m}^N)$ is a continuous operator valued function which is periodic with period θ.

(c). for each $t \in \mathbb{R}$, $\mathbf{X} \to \mathbf{\Pi}(t)[\mathbf{X}] : \mathcal{S}_n^N \to \mathcal{S}_{n+m}^N$ is a positive operator, i.e. $\mathbf{\Pi}(t)[\mathbf{X}] \succeq 0$ whenever $\mathbf{X} \succeq 0$.

(d).

$$R(t,i) > 0 \tag{2.7a}$$

$$M(t,i) - L(t,i)R^{-1}(t,i)L^T(t,i) \ge 0 \tag{2.7b}$$

for all $t \in \mathbb{R}$, $1 \le i \le N$.

2.2. Several relevant special cases of the differential equation (2.1)

In this subsection we display several special cases of the backward differential equation (2.1) arising in a natural way, in some optimal control problems in both deterministic and stochastic framework.

A. The LQ optimal regulator.
We consider the optimal control problem described by the controlled system

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t) \\ x(t_0) &= x_0 \end{aligned} \tag{2.8}$$

the performance criterion

$$J(x_0, u) = \int_{t_0}^{\infty} (x^T(t)M(t)x(t) + 2x^T(t)L(t)u(t) + u^T(t)R(t)u(t))dt \tag{2.9}$$

and the class of admissible controls $u(\cdot) \in L_2([t_0, \infty); \mathbb{R}^m)$. The problem of the optimal regulator requires the finding of the conditions which guarantee the existence of a control $\tilde{u}(\cdot)$ which minimizes the cost function (2.9) along of the trajectories of the controlled system (2.8) determined by the all admissible controls $u(\cdot)$. One shows that if $R(t) > 0$, $M(t) - L(t)R^{-1}(t)L^T(t) \ge 0$, for all $t \ge t_0$, then the optimal control may be computed using a special solution named "stabilizing solution" of the matrix differential equation

$$
\begin{aligned}
-\dot{X}(t) =& A^T(t)X(t) + X(t)A(t) - (X(t)B(t) + L(t))R^{-1}(t) \\
& \cdot (X(t)B(t) + L(t))^T + M(t).
\end{aligned} \tag{2.10}
$$

The problem of the existence of the optimal control which solves the optimization problem described by (2.8)-(2.9) reduces to the problem of the existence of the stabilizing solution of the matrix differential equation (2.10). The differential equation (2.10) may be regarded as a special case of (2.1) when $N = 1$, $\Pi_k(t)[\mathbf{X}] = 0$, $1 \leq k \leq 3$, $\mathbf{X} \in \mathcal{S}_n^1$. Since, for $n = 1$ the differential equation (2.10) reduces to the well known scalar differential equation studied by the mathematician Jacopo Francesco Riccati in the first part of the 18^{th} century, the equation (2.10) was called "matrix Riccati differential equation".

B. The stochastic linear quadratic optimal regulator.

The case when the controlled system is affected by multiplicative white noise perturbations.
In this case, the optimal control problem consists of the finding of a control $\tilde{u}(\cdot)$ which minimizes the cost function

$$
J(x_0, u) = \mathbb{E}\left[\int_0^\infty (x^T(t)M(t)x(t) + 2x^T(t)L(t)u(t) + u^T(t)R(t)u(t))dt\right] \tag{2.11}
$$

along of the trajectories of the controlled system

$$
\begin{aligned}
& dx(t) = (A(t)x(t) + B(t)u(t))dt + (C(t)x(t) + D(t)u(t))dw(t) \\
& x(0) = x_0 \in \mathbb{R}^n
\end{aligned} \tag{2.12}
$$

determined by the admissible controls $u(t)$ from a class of admissible stochastic processes $\mathcal{U}(x_0)$. Here and in the sequel $\mathbb{E}[\cdot]$ stands for the mathematical expectation. In (2.12) $\{w(t)\}_{t\geq 0}$ is an 1-dimensional standard Wiener process defined on a given probability space $(\Omega, \mathcal{F}, \mathcal{P})$ (see [11], [16]). In 1968, to solve this problem W. M. Wonham (see [20]) introduced the following matrix differential equation

$$
\begin{aligned}
-\dot{X}(t) =& A^T(t)X(t) + X(t)A(t) + C^T(t)X(t)C(t) \\
& - (X(t)B(t) + C^T(t)X(t)D(t) + L(t)) \\
& \cdot (D^T(t)X(t)D(t) + R(t))^{-1}(B^T(t)X(t) + D^T(t)X(t)C(t) \\
& + L^T(t)) + M(t).
\end{aligned} \tag{2.13}
$$

Since in the special case $C(t) = 0$, $D(t) = 0$, the differential equation (2.13) reduces to the matrix Riccati differential equation (2.10), it was called "matrix Riccati differential equation of stochastic control".

The problem of the stochastic linear quadratic optimal regulator in the case of the simultaneous presence of multiplicative white noise perturbations and of the perturbation modeled by a Markov process.

In this case, the controlled system is of the form:

$$dx(t) = (A(t,\eta_t)x(t) + B(t,\eta_t)u(t))dt + (C(t,\eta_t)x(t) + D(t,\eta_t)u(t))dw(t) \tag{2.14}$$

and the performance criterion:

$$J(x_0,u) = \mathbb{E}[\int_0^\infty (x^T(t)M(t,\eta_t)x(t) + 2x^T(t)L(t,\eta_t)u(t) + u^T(t)R(t,\eta_t)u(t))dt]. \tag{2.15}$$

Here, $\{w(t)\}_{t\geq 0}$ is 1-dimensional standard Wiener process as before, and $\{\eta_t\}_{t\geq 0}$ is a standard Markov process defined by the same probability space $(\Omega,\mathcal{F},\mathcal{P})$, taking values in the finite set $\mathcal{N} = \{1,2,\ldots,N\}$ and having the transition semigroup $P(t) = e^{Qt}$. The elements q_{ij} of the generator matrix Q have the properties:

$$q_{ij} \geq 0, \text{ if } i \neq j, \qquad \sum_{l=1}^{N} q_{il} = 0, \text{ for all } i,j \in \mathcal{N}.$$

For more details we refer to [2], [3], [7], [11]. It is assumed that the processes $\{w(t)\}_{t\geq 0}$ and $\{\eta_t\}_{t\geq 0}$ are independent stochastic processes.

In the computation of the control which minimizes the cost functional (2.15) along of the trajectories of the controlled system (2.14) one uses the stabilizing solution of the following system of matrix Riccati type differential equations:

$$\begin{aligned}
-\dot{X}(t,i) =& A^T(t,i)X(t,i) + X(t,i)A(t,i) + C^T(t,i)X(t,i)C(t,i)\\
&- (X(t,i)B(t,i) + C^T(t,i)X(t,i)D(t,i) + L(t,i))\\
&\cdot (D^T(t,i)X(t,i)D(t,i) + R(t,i))^{-1}(B^T(t,i)X(t,i)\\
&+ D^T(t,i)X(t,i)C(t,i) + L^T(t,i)) + \sum_{j=1}^{N} q_{ij}X(t,j) + M(t,i).
\end{aligned} \tag{2.16}$$

$1 \leq i \leq N$. One sees that (2.16) is a special case of (2.1) with $A(t,i)$ replaced by $A(t,i) + \frac{1}{2}q_{ii}I_n$,

$$\Pi_1(t)[\mathbf{X}](i) = C^T(t,i)X(i)C(t,i) + \sum_{\substack{j=1\\ j\neq i}}^{N} q_{ij}X(j),$$

$$\Pi_2(t)[\mathbf{X}](i) = C^T(t,i)X(i)D(t,i)$$

$$\Pi_3(t)[\mathbf{X}](i) = D^T(t,i)X(i)D(t,i), \;\; 1 \leq i \leq N,$$

$$\mathbf{X} = (X(1),X(2),\ldots,X(N)) \in \mathcal{S}_n^N.$$

Remark 2.2. From the previous examples, one sees that the differential equation (2.1) contains as special cases different types of Riccati differential equations arising in both deterministic and stochastic framework. That is why, in the sequel we shall call the differential equation (2.1) "generalized Riccati differential equation" (GRDE).

2.3. The stabilizing solution of a GRDE

Let $\mathcal{F}_{ad}$ be the set of all continuous and θ periodic functions $\mathbf{F} : \mathbb{R} \to \mathcal{M}_{mn}^N$. In the developments from this work these functions will be named "admissible feedback gains".

If $\mathbf{\Pi}(\cdot)$ is the operator valued function described in (2.5) and $\mathbf{F}(\cdot)$ is an arbitrary admissible feedback gain, we associate a new operator valued function $\mathbf{\Pi_F}(\cdot)$ defined as follows:

$$\mathbf{\Pi_F}(t)[\mathbf{X}] = (\Pi_\mathbf{F}(t)[\mathbf{X}](1), \Pi_\mathbf{F}(t)[\mathbf{X}](2), \ldots, \Pi_\mathbf{F}(t)[\mathbf{X}](N)) \tag{2.17a}$$

$$\Pi_\mathbf{F}(t)[\mathbf{X}](i) = (I_n \quad F^T(t,i))\Pi(t)[\mathbf{X}](i)(I_n \quad F^T(t,i))^T, \ \ 1 \leq i \leq N. \tag{2.17b}$$

Remark 2.3. **(a).** From (2.5) and (2.17) we infer that $\mathbf{X} \to \mathbf{\Pi_F}(t)[\mathbf{X}] \in \mathcal{B}(\mathcal{S}_n^N)$. Moreover, if the assumption (**H**) **(c)** is fulfilled, then $\mathbf{\Pi_F}(t)[\mathbf{X}] \succeq 0$ whenever $\mathbf{X} \succeq 0$.

(b). In the sequel, $\mathbf{\Pi}^*_\mathbf{F}(t)[\cdot]$ denotes the adjoint operator of $\mathbf{\Pi_F}(t)[\cdot]$ with respect to the inner product (1.1). Thus, in the special case when $\mathbf{\Pi}(t)[\cdot]$ is associated to the Riccati differential equation (2.13) we infer that

$$\mathbf{\Pi}^*_\mathbf{F}(t)[\mathbf{X}] = (C(t) + D(t)F(t))\mathbf{X}(C(t) + D(t)F(t))^T, \ \text{ for all } \ \mathbf{X} \in \mathcal{S}_n^1, \tag{2.18}$$

and in the case when the operator $\mathbf{\Pi}(t)[\cdot]$ is associated to the Riccati differential equation (2.16), one obtains that

$$\begin{aligned}\mathbf{\Pi}^*_\mathbf{F}(t)[\mathbf{X}](i) =& (C(t,i) + D(t,i)F(t,i))X(i)(C(t,i) + D(t,i)F(t,i))^T \\ &+ \sum_{\substack{j=1 \\ j\neq i}}^N q_{ji}X(j), \ \text{ for all } \ \mathbf{X} \in \mathcal{S}_n^N.\end{aligned} \tag{2.19}$$

Employing the triple $(\mathbf{A}(\cdot), \mathbf{B}(\cdot), \mathbf{\Pi}(\cdot))$ described in (2.4a) and (2.5), for each $\mathbf{F} \in \mathcal{F}_{ad}$ we may define the operator valued function $\mathcal{L}_\mathbf{F} : \mathbb{R} \to \mathcal{B}(\mathcal{S}_n^N)$ by

$$\mathcal{L}_\mathbf{F}(t)[\mathbf{X}] = (\mathbf{A}(t) + \mathbf{B}(t)\mathbf{F}(t))\mathbf{X} + \mathbf{X}(\mathbf{A}(t) + \mathbf{B}(t)\mathbf{F}(t))^T + \mathbf{\Pi}^*_\mathbf{F}(t)[\mathbf{X}], \tag{2.20}$$

for all $\mathbf{X} \in \mathcal{S}_n^N$.

Remark 2.4. From (2.18)-(2.20) one sees that in the case when the GRDE (2.1) takes one of the special forms (2.13) or (2.16), the linear operator $\mathcal{L}_\mathbf{F}(t)[\cdot]$ defined in (2.20) recover the well known Lyapunov type operators involved in the definition of the property of the stochastic stabilizability of the systems (2.12) and (2.14), respectively.

Now we are in position to define the notion of "stabilizing solution" of the GRDE (2.1).

Definition 2.5. A solution $\tilde{\mathbf{X}}(\cdot) : \mathbb{R} \to \mathcal{S}_n^N$ of the GRDE (2.1),

$$\tilde{\mathbf{X}}(\cdot) = (\tilde{X}(\cdot,1), \ldots, \tilde{X}(\cdot,N))$$

is named **stabilizing solution** if the linear differential equation on the linear space $\mathcal{S}_n^N$

$$\dot{\mathbf{Y}}(t) = \mathcal{L}_{\tilde{\mathbf{F}}}(t)[\mathbf{Y}(t)] \tag{2.21}$$

is exponentially stable, where $\mathcal{L}_{\tilde{\mathbf{F}}}[\cdot]$ is defined as in (2.20) for $\mathbf{F}(\cdot)$ replaced by

$$\tilde{\mathbf{F}}(\cdot) = (\tilde{F}(\cdot,1), \ldots, \tilde{F}(\cdot,N))$$

described by

$$\tilde{F}(t,i) = -\left(\Pi_3(t)[\tilde{\mathbf{X}}(t)](i) + R(t,i)\right)^{-1}(\tilde{X}(t,i)B(t,i) + \Pi_2(t)[\tilde{\mathbf{X}}(t)](i) \\ + L(t,i))^T. \tag{2.22}$$

Invoking Remark 2.4, we may infer that in the case when the GRDE (2.1) takes one of the special forms (2.10), (2.13) or (2.16), the concept of stabilizing solution introduced in the Definition 2.5, recover the traditional definition of the stabilizing solution of a Riccati differential equation from deterministic / stochastic control. Applying Corollary 5.4.2 and Theorem 5.4.3 from [9] we obtain:

Corollary 2.6. *Under the assumption* $(\mathbf{H})$ *the GRDE (*2.1*) has at most one bounded and stabilizing solution. Moreover, this solution, if it exists is a periodic function with period* θ.

Our aim is to provide a set of necessary and sufficient conditions for the existence of the stabilizing solution of a GRDE of type (2.1).

3. Some auxiliary issues

3.1. Monodromy operators. Characteristic multipliers

In this subsection we recall several definitions and results regarding the linear differential equations with periodic coefficients specialized to the case of linear differential equations on $\mathcal{S}_n^N$ of the form

$$\dot{\mathbf{Y}}(t) = \mathcal{L}_{\mathbf{F}}(t)[\mathbf{Y}(t)] \tag{3.1}$$

when $\mathcal{L}_{\mathbf{F}}(t)[\cdot]$ is associated via (2.20) to the triple $(\mathbf{A}(\cdot), \mathbf{B}(\cdot), \mathbf{\Pi}(\cdot))$ and to an arbitrary $\mathbf{F}(\cdot) \in \mathcal{F}_{ad}$.

For $t, t_0 \in \mathbb{R}$ we denote $\mathbf{T}_{\mathbf{F}}(t, t_0)$ the linear evolution operator on $\mathcal{S}_n^N$ defined by the linear differential equation (3.1) by

$$\mathbf{T}_{\mathbf{F}}(t, t_0)\mathbf{X}_0 = \mathbf{Y}(t; t_0, \mathbf{X}_0)$$

where $\mathbf{Y}(\cdot; t_0, \mathbf{X}_0)$ is the solution of the linear differential equation (3.1) with the initial condition $\mathbf{Y}(t_0; t_0, \mathbf{X}_0) = \mathbf{X}_0$. The main properties of the operator $\mathbf{T}_{\mathbf{F}}(t, t_0)$ involved in the developments of this paper, are summarized in the following lemma.

Lemma 3.1. (a). $\mathbf{T}_{\mathbf{F}}(t, t_0) = \mathcal{I}_{\mathcal{S}_n^N}$, *for all* $t_0 \in \mathbb{R}$, $\mathcal{I}_{\mathcal{S}_n^N}$ *is the identity operator on* $\mathcal{S}_n^N$;
(b). $\mathbf{T}_{\mathbf{F}}(t, t_1)\mathbf{T}_{\mathbf{F}}(t_1, t_0) = \mathbf{T}_{\mathbf{F}}(t, t_0)$, *for all* $t, t_1, t_0 \in \mathbb{R}$;
(c). $\mathbf{T}_{\mathbf{F}}^{-1}(t, t_0) = \mathbf{T}_{\mathbf{F}}(t_0, t)$, *for all* $t, t_0 \in \mathbb{R}$;
(d). *if the assumption* $(\mathbf{H})$ *is fulfilled, then* $\mathbf{T}_{\mathbf{F}}(t + j\theta, t_0 + j\theta) = \mathbf{T}_{\mathbf{F}}(t, t_0)$, *for all* $t, t_0 \in \mathbb{R}$, $j \in \mathbb{Z}$;
(e). *for each* $t \geq t_0 \in \mathbb{R}$, $\mathbf{T}_{\mathbf{F}}(t, t_0) : \mathcal{S}_n^N \to \mathcal{S}_n^N$ *is a positive operator, i.e.* $\mathbf{T}_{\mathbf{F}}(t, t_0)\mathbf{X} \succeq 0$, *for all* $t \geq t_0$ *if* $\mathbf{X} \succeq 0$.

Proof. The proof of these properties is omitted because they are known in a more general framework for linear differential equations in both finite and infinite dimensional case, see e.g. Chapter 3 from [4] for infinite dimensional case and Chapter 1.3. from

[12] for finite dimensional case. Assertion **(e)** may be obtained applying Corollary 2.2.6 from [9]. □

In the rest of the paper we assume that (**H**) is fulfilled. For each $t_0 \in \mathbb{R}$, we set

$$\mathbb{T}_{\mathbf{F}}(t_0) \stackrel{\Delta}{=} \mathbf{T}_{\mathbf{F}}(t_0+\theta, t_0). \tag{3.2}$$

According to the terminology used in connection with the linear differential equations with periodic coefficients, $\mathbb{T}_{\mathbf{F}}(t_0)$ will be named the monodromy operator associated to the linear operator $\mathcal{L}_{\mathbf{F}}(\cdot)$ or, equivalently the monodromy operator associated to the linear differential equation (3.1). From Lemma 3.1 **(d), (e)** and (3.2) one gets that $\mathbb{T}_{\mathbf{F}}(\cdot)$ is an operator valued function periodic of period θ and for each $t_0 \in \mathbb{R}$, $\mathbb{T}_{\mathbf{F}}(t_0)$ is a positive operator on the ordered space $(\mathcal{S}_n^N, \mathcal{S}_n^{N+})$. The elements of the spectrum $\sigma(\mathbb{T}_{\mathbf{F}}(t_0))$ are named characteristic multipliers of the linear differential equation (3.1). If $\lambda \in \sigma(\mathbb{T}_{\mathbf{F}}(t_0))$ then there exists $0 \neq \mathbf{X} \in \mathcal{S}_n^N$ such that

$$\mathbb{T}_{\mathbf{F}}(t_0)\mathbf{X} = \lambda \mathbf{X}. \tag{3.3}$$

By direct calculations based on Lemma 3.1 we obtain that (3.3) is equivalent to

$$\mathbb{T}_{\mathbf{F}}(t_1)\mathbf{T}_{\mathbf{F}}(t_1,t_0)\mathbf{X} = \lambda \mathbf{T}_{\mathbf{F}}(t_1,t_0)\mathbf{X}, \text{ for all } t_1 \in \mathbb{R}. \tag{3.4}$$

From the equivalence of (3.3) and (3.4) we may conclude that the spectrum of the monodromy operator does not depend upon t_0. In our development an important role will be played by the characteristic multipliers from the subset

$$\sigma^+(\mathbb{T}_{\mathbf{F}}(t_0)) = \{\mu \in \sigma(\mathbb{T}_{\mathbf{F}}(t_0)) | \exists 0 \neq \mathbf{Y} \in \mathcal{S}_n^{N+} \text{ such that } \mathbb{T}_{\mathbf{F}}(t_0)\mathbf{Y} = \mu \mathbf{Y}\}.$$

If $\mu \in \sigma^+(\mathbb{T}_{\mathbf{F}}(t_0))$ we denote

$$\mathcal{V}_{\mathbf{F}}(\mu, t_0) = \{\mathbf{Y} \in \mathcal{S}_n^{N+} | \ \mathbf{Y} \neq 0, \ \mathbb{T}_{\mathbf{F}}(t_0)\mathbf{Y} = \mu \mathbf{Y}\}.$$

One sees that $\sigma^+(\mathbb{T}_{\mathbf{F}}(t_0)) \subset [0, \infty)$. In this work, the elements of $\sigma^+(\mathbb{T}_{\mathbf{F}}(t_0))$ will be named distinctive characteristic multipliers. These represent the time-varying counter part of the concept of distinctive eigenvalues introduced in [23] to characterize a subset of the spectrum of a Lyapunov type operators arising in stochastic control.

Let $\rho_{\mathbf{F}}$ be the spectral radius of the monodromy operator. Applying, for example, Theorem 2.6 from [5] in the case of linear operator $\mathbb{T}^*_{\mathbf{F}}(t_0)$ defined on the ordered space $(\mathcal{S}_n^N, \mathcal{S}_n^{N+})$ we obtain that there exists $0 \neq \mathbf{Y} \in \mathcal{S}_n^{N+}$ such that $\mathbb{T}_{\mathbf{F}}(t_0)\mathbf{Y} = \rho_{\mathbf{F}}\mathbf{Y}$, so $\rho_{\mathbf{F}} \in \sigma^+(\mathbb{T}_{\mathbf{F}}(t_0))$, for all $t_0 \in \mathbb{R}$.

3.2. Unobservable characteristic multipliers

Let $\mathbf{C}(\cdot) : \mathbb{R} \to \mathcal{M}_{pn}^N$ be a continuous function which is periodic of period θ.

Definition 3.2. **(a).** We say that the distinctive characteristic multiplier μ is an unobservable characteristic multiplier at t_0 with respect to the pair $(\mathbf{C}(\cdot), \mathcal{L}_{\mathbf{F}}(\cdot))$ if there exists $\mathbf{Y} \in \mathcal{V}_{\mathbf{F}}(\mu, t_0)$ such that

$$\mathbf{C}(t)\mathbf{T}_{\mathbf{F}}(t,t_0)\mathbf{Y} = 0, \ \forall\, t \in [t_0, t_0+\theta]. \tag{3.5}$$

(b). A distinctive characteristic multiplier μ is named unobservable characteristic multiplier for the pair $(\mathbf{C}(\cdot), \mathcal{L}_{\mathbf{F}}(\cdot))$ if it is an unobservable characteristic multiplier at any $t_0 \in \mathbb{R}$.

Remark 3.3. **(a).** Even if the property of unobservability could be defined for any characteristic multiplier we preferred to restrict the definition to the distinctive characteristic multipliers, because, in this framework this property will be involved in the next sections.

(b). From the periodicity property of the functions involved in Definition 3.2, one obtaines that (3.5) holds for any $t \geq t_0$ if it is true for $t \in [t_0, t_0 + \theta]$.

3.3. An useful representation formula of the operator $\mathcal{R}$

Using Lemma 5.1.1 from [9] applied to the operator (2.2), we obtain:

Corollary 3.4. *Let* $\mathbf{F}(\cdot) = (F(\cdot,1), F(\cdot,2), \dots, F(\cdot,N))$ *be an admissible feedback gain. Under the assumption* (**H**) *we have the following representation of the operator* $\mathcal{R}_i(\cdot,\cdot)$*:*

$$\begin{aligned}\mathcal{R}_i(t,\mathbf{X}) =& \mathcal{L}^*_{\mathbf{F}}(t)[\mathbf{X}](i) - (F(t,i) - F^{\mathbf{X}}(t,i))^T(\Pi_3(t)[\mathbf{X}](i) + R(t,i)\\ &\times (F(t,i) - F^{\mathbf{X}}(t,i)) + M(t,i) - L(t,i)R^{-1}(t,i)L^T(t,i)\\ &+ (F(t,i) + R^{-1}(t,i)L^T(t,i))^T R(t,i)(F(t,i) + R^{-1}(t,i)L^T(t,i))\end{aligned} \tag{3.6}$$

for all $1 \leq i \leq N$, $(t,\mathbf{X}) \in$ *Dom* $\mathcal{R}$, *where* $\mathcal{L}^*_{\mathbf{F}}[\mathbf{X}] = (\mathcal{L}^*_{\mathbf{F}}(t)[\mathbf{X}](1), \dots, \mathcal{L}^*_{\mathbf{F}}(t)[\mathbf{X}](N))$ *is the adjoint of the operator* $\mathcal{L}_{\mathbf{F}}(t)[\cdot]$ *associated via (*2.20*) to the triple* $(\mathbf{A}(\cdot), \mathbf{B}(\cdot), \mathbf{\Pi}(\cdot))$ *and to the admissible feedback gain* $\mathbf{F}(\cdot)$*, while,*

$$\begin{aligned}F^{\mathbf{X}}(t,i) \triangleq& -(\Pi_3(t)[\mathbf{X}](i) + R(t,i))^{-1}(X(i)B(t,i) + \Pi_2(t)[\mathbf{X}](i)\\ &+ L(t,i))^T.\end{aligned} \tag{3.7}$$

Employing the convention of notation established in Subsection 1.1, we may rewrite (3.6) and (3.7) in a compact form:

$$\begin{aligned}\mathcal{R}(t,\mathbf{X}) =& \mathcal{L}^*_{\mathbf{F}}(t)[\mathbf{X}] - (\mathbf{F}(t) - \mathbf{F}^{\mathbf{X}}(t))^T(\mathbf{\Pi}_3(t)[\mathbf{X}] + \mathbf{R}(t))\\ &\cdot (\mathbf{F}(t) - \mathbf{F}^{\mathbf{X}}(t)) + \mathbf{M}(t) - \mathbf{L}(t)\mathbf{R}^{-1}(t)\mathbf{L}^T(t)\\ &+ (\mathbf{F}(t) + \mathbf{R}^{-1}(t)\mathbf{L}^T(t))^T\mathbf{R}(t)(\mathbf{F}(t) + \mathbf{R}^{-1}(t)\mathbf{L}^T(t))\end{aligned} \tag{3.8}$$

$$\mathbf{F}^{\mathbf{X}}(t) = -(\mathbf{\Pi}_3(t)[\mathbf{X}] + \mathbf{R}(t))^{-1}(\mathbf{X}\mathbf{B}(t) + \mathbf{\Pi}_2(t)[\mathbf{X}] + \mathbf{L}(t))^T, \tag{3.9}$$

for all $(t,\mathbf{X}) \in$ *Dom* $\mathcal{R}$.

4. The main result

4.1. The statement of the main result

First we recall the definition of the concept of stabilizability of a triple $(\mathbf{A}(\cdot), \mathbf{B}(\cdot), \mathbf{\Pi}(\cdot))$ where $\mathbf{A}(\cdot)$, $\mathbf{B}(\cdot)$ are defined in (2.4a) and $\mathbf{\Pi}(\cdot)$ is described by (2.5). The next definition is an adaptation of the Definition 5.3.2 from [9] to the triples involved here.

Definition 4.1. We say that $(\mathbf{A}(\cdot), \mathbf{B}(\cdot), \mathbf{\Pi}(\cdot))$ is stabilizable if there exists an admissible feedback gain $\mathbf{F}(\cdot)$ with the property that the linear differential equation on $\mathcal{S}_n^N$:

$$\dot{\mathbf{Y}} = \mathcal{L}_{\mathbf{F}}(t)[\mathbf{Y}](t) \tag{4.1}$$

is exponentially stable, $\mathcal{L}_{\mathbf{F}}(t)[\cdot]$ being the linear operator associated via (2.20) to $(\mathbf{A}(\cdot), \mathbf{B}(\cdot), \mathbf{\Pi}(\cdot))$ and to the admissible feedback gain $\mathbf{F}(\cdot)$. The admissible feedback gains for which the linear differential equation (4.1) is exponentially stable will be named stabilizing admissible feedback gains.

Remark 4.2. When the triple $(\mathbf{A}(\cdot), \mathbf{B}(\cdot), \mathbf{\Pi}(\cdot))$ is associated to the Riccati equations (2.10), (2.13) or (2.16), respectively, the concept of stabilizability introduced by Definition 4.1 recover the notions of stabilizability known in the deterministic and/or stochastic control.

The main result of this work is stated in the following theorem:

Theorem 4.3. *Under the assumption* (**H**) *the following are equivalent:*

(i). *the GRDE (*2.1*) has a bounded and stabilizing solution*

$$\tilde{\mathbf{X}}(\cdot) = (\tilde{X}(\cdot, 1), \dots, \tilde{X}(\cdot, N))$$

with $\tilde{X}(t, i) \geq 0$, *for all* $t \in \mathbb{R}$, $1 \leq i \leq N$;

(ii). **(a).** *the triple* $(\mathbf{A}(\cdot), \mathbf{B}(\cdot), \mathbf{\Pi}(\cdot))$ *is stabilizable;*

(b). $\mu = 1$ *is not an unobservable distinctive characteristic multiplier for the pair* $(\tilde{\mathbf{C}}(\cdot), \mathcal{L}_{-\mathbf{R}^{-1}\mathbf{L}^T}(\cdot))$ *where* $\tilde{\mathbf{C}}(\cdot) = (\tilde{C}(\cdot, 1), \dots, \tilde{C}(\cdot, N))$,

$$\tilde{C}(t, i) = (M(t, i) - L(t, i)R^{-1}(t, i)L^T(t, i))^{1/2} \tag{4.2}$$

and $\mathcal{L}_{-\mathbf{R}^{-1}\mathbf{L}^T}(\cdot)$ *is the linear operator of type (*2.20*) associated to the triple* $(\mathbf{A}(\cdot), \mathbf{B}(\cdot), \mathbf{\Pi}(\cdot))$ *and the feedback gains* $F(t, i) = -R^{-1}(t, i)L^T(t, i)$.

4.2. Several intermediate results

For the proof of Theorem 4.3 we need several results which can be interesting by themselves. In this subsection we present their proofs.

Lemma 4.4. *Let* $\mathbf{F}_k(\cdot) = (F_k(\cdot, 1), \dots, F_k(\cdot, N))$, $k = 1, 2$ *be two admissible feedback gains. Let* $\mathcal{L}_{\mathbf{F}_k}(t) : \mathcal{S}_n^N \to \mathcal{S}_n^N$, $k = 1, 2$ *be the linear operators associated to the triple* $(\mathbf{A}(\cdot), \mathbf{B}(\cdot), \mathbf{\Pi}(\cdot))$ *and the feedback gain* $\mathbf{F}_k(\cdot)$ *as in (*2.20*). We denote* $\mathbf{T}_{\mathbf{F}_k}(t, t_0)$ *the linear evolution operator on* $\mathcal{S}_n^N$ *defined by the linear differential equation*

$$\dot{\mathbf{X}}(t) = \mathcal{L}_{\mathbf{F}_k}(t)[\mathbf{X}(\mathbf{t})], \quad k = 1, 2. \tag{4.3}$$

If there exists $\mathbf{Y} \in \mathcal{S}_n^{N+} \setminus \{0\}$ *with the property that*

$$\mathbf{F}_1(t)\mathbf{T}_{\mathbf{F}_1}(t, t_0)\mathbf{Y} = \mathbf{F}_2(t)\mathbf{T}_{\mathbf{F}_1}(t, t_0)\mathbf{Y}, \quad t \in [t_0, t_0 + \theta], \quad t_0 \in \mathbb{R} \tag{4.4}$$

then we have

$$\mathbf{T}_{\mathbf{F}_1}(t, t_0)\mathbf{Y} = \mathbf{T}_{\mathbf{F}_2}(t, t_0)\mathbf{Y}, \quad \text{for all } t \in [t_0, t_0 + \theta]. \tag{4.5}$$

Proof. Let $\mathbf{X}_k(t) = \mathbf{X}_k(t; t_0, \mathbf{Y}) \triangleq \mathbf{T}_{\mathbf{F}_k}(t, t_0)\mathbf{Y}$, $k = 1, 2$. The linear differential equation (4.3) satisfied by $\mathbf{X}_k(\cdot)$ may be rewritten as:

$$\begin{aligned} \dot{\mathbf{X}}_k(t) = &\mathbf{A}(t)\mathbf{X}_k(t) + \mathbf{X}_k(t)\mathbf{A}^T(t) + \mathbf{B}(t)\mathbf{F}_k(t)\mathbf{X}_k(t) \\ &+ (\mathbf{B}(t)\mathbf{F}_k(t)\mathbf{X}_k(t))^T + \mathbf{\Pi}^*_{\mathbf{F}_k}(t)[\mathbf{X}_k(t)], \quad k = 1, 2. \end{aligned} \tag{4.6}$$

Substituting (4.4) in (4.6) written for $k=1$, we obtain that $\mathbf{X}_1(\cdot)$ satisfies the differential equation:

$$\begin{aligned}\dot{\mathbf{X}}_1(t) =&(\mathbf{A}(t)+\mathbf{B}(t)\mathbf{F}_2(t))\mathbf{X}_1(t)+\mathbf{X}_1(t)(\mathbf{A}(t)+\mathbf{B}(t)\mathbf{F}_2(t))^T\\ &+\mathbf{\Pi}^*_{\mathbf{F}_1}(t)[\mathbf{X}_1(t)],\ \ t\in[t_0,t_0+\theta].\end{aligned} \tag{4.7}$$

We show now that if (4.4) holds, then

$$\mathbf{\Pi}^*_{\mathbf{F}_1}(t)[\mathbf{X}_1(t)]=\mathbf{\Pi}^*_{\mathbf{F}_2}(t)[\mathbf{X}_1(t)],\ \ t\in[t_0,t_0+\theta]. \tag{4.8}$$

To this end we set $\mathbf{X}_1(t)=(X_1(t,1),\ldots,X_1(t,N))$. This allows us to write the componentwise version of (4.4) as

$$F_1(t,i)X_1(t,i)=F_2(t,i)X_1(t,i),\ \ t\in[t_0,t_0+\theta],\ \ 1\le i\le N. \tag{4.9}$$

Let $\mathbf{Z}=(Z(1),\ldots,Z(N))\in\mathcal{S}_n^N\setminus\{0\}$ be arbitrary. Using the definition of the adjoint operator with respect to the inner product (1.1) we may write via (2.17) that

$$\begin{aligned}\langle\mathbf{Z},\mathbf{\Pi}^*_{\mathbf{F}_1}(t)[\mathbf{X}_1(t)]\rangle=\langle\mathbf{\Pi}^*_{\mathbf{F}_1}(t)[\mathbf{Z}],\mathbf{X}_1(t)\rangle&=\sum_{i=1}^N Tr[\mathbf{\Pi}^*_{\mathbf{F}_1}(t)[\mathbf{Z}](i)X_1(t,i)]\\ &=\sum_{i=1}^N Tr[\mathbf{\Pi}(t)[\mathbf{Z}](i)\cdot\Psi(X_1(t,i),F_1(t,i))]\end{aligned} \tag{4.10}$$

where

$$\Psi(X_1(t,i),F_1(t,i))\stackrel{\Delta}{=}(I_n\ \ F_1^T(t,i))^TX_1(t,i)(I_n\ \ F_1^T(t,i))\in\mathcal{S}_{n+m}. \tag{4.11}$$

Substituting (4.9) in (4.11) we obtain

$$\Psi(X_1(t,i),F_1(t,i))=\Psi(X_1(t,i),F_2(t,i)). \tag{4.12}$$

Plugging (4.12) in (4.10), we obtain after some calculations based on the properties of the trace operator that

$$\langle\mathbf{Z},\mathbf{\Pi}^*_{\mathbf{F}_1}(t)[\mathbf{X}_1(t)]\rangle=\langle\mathbf{Z},\mathbf{\Pi}^*_{\mathbf{F}_2}(t)[\mathbf{X}_1(t)]\rangle.$$

This equality confirms the validity of (4.8) because $\mathbf{Z}$ is arbitrary in $\mathcal{S}_n^N$. Now, (4.8) allows us to rewrite (4.7) as

$$\begin{aligned}\dot{\mathbf{X}}_1(t) =&(\mathbf{A}(t)+\mathbf{B}(t)\mathbf{F}_2(t))\mathbf{X}_1(t)+\mathbf{X}_1(t)(\mathbf{A}(t)+\mathbf{B}(t)\mathbf{F}_2(t))^T\\ &+\mathbf{\Pi}^*_{\mathbf{F}_2}(t)[\mathbf{X}_1(t)],\ \ t\in[t_0,t_0+\theta].\end{aligned} \tag{4.13}$$

Writing (4.6) for $k=2$ and comparing with (4.13) we remark that $\mathbf{X}_2\stackrel{\Delta}{=}\mathbf{T}_{\mathbf{F}_2}(t,t_0)\mathbf{Y}$ is also a solution of the linear differential equation (4.13). On the other hand $\mathbf{X}_1(t_0)=\mathbf{X}_2(t_0)=\mathbf{Y}$. From the uniqueness of the solution of an initial value problem we deduce that

$$\mathbf{X}_2(t;t_0,\mathbf{Y})=\mathbf{X}_1(t;t_0,\mathbf{Y}),\ \ t\in[t_0,t_0+\theta]$$

which shows that (4.5) is true. Thus the proof is complete. □

Proposition 4.5. *Let $\tilde{\mathbf{X}}(\cdot) = (\tilde{X}(\cdot,1),\dots,\tilde{X}(\cdot,N))$ be a $\theta-$periodic solution of the GRDE (2.1) such that $\tilde{X}(t,i) \geq 0$, for all $t \in \mathbb{R}$, $1 \leq i \leq N$. Let*

$$\tilde{\mathbf{F}}(t) = (\tilde{F}(t,1),\dots,\tilde{F}(t,N))$$

*be the corresponding feedback gain (2.22) associated to the solution $\tilde{\mathbf{X}}(\cdot)$. Under the assumption (**H**), if $\mu \in \sigma^+(\mathbb{T}_{\tilde{\mathbf{F}}}(t_0))$ is such that $\mu \geq 1$ then μ is an unobservable at t_0 distinctive characteristic multiplier for the pair $(\tilde{\mathbf{C}}(\cdot), \mathcal{L}_{-\mathbf{R}^{-1}\mathbf{L}^T}(\cdot))$ where $\tilde{\mathbf{C}}(\cdot)$ is defined as in (4.2).*

Proof. Comparing (2.22) and (3.7) we see that $\tilde{\mathbf{F}}(t) = \mathbf{F}^{\mathbf{X}}(t)$, for all $t \in \mathbb{R}_+$. Using (3.8) with $\mathbf{F}(t) = \tilde{\mathbf{F}}(t)$ we obtain that the equation (2.1) satisfied by $\tilde{\mathbf{X}}(\cdot)$ becomes

$$-\frac{d}{dt}\tilde{\mathbf{X}}(t) = \mathcal{L}^*_{\tilde{\mathbf{F}}}(t)[\tilde{\mathbf{X}}(t)] + \mathbf{W}(t) \tag{4.14}$$

where we denote

$$\begin{aligned}\mathbf{W}(t) =& \mathbf{M}(t) - \mathbf{L}(t)\mathbf{R}^{-1}(t)\mathbf{L}^T(t) + (\tilde{\mathbf{F}}(t) + \mathbf{R}^{-1}(t)\mathbf{L}^T(t))^T \\ & \cdot \mathbf{R}(t)(\tilde{\mathbf{F}}(t) + \mathbf{R}^{-1}(t)\mathbf{L}(t)).\end{aligned} \tag{4.15}$$

From (4.14) we deduce

$$\tilde{\mathbf{X}}(t_0) = \mathbb{T}^*_{\tilde{\mathbf{F}}}(t_0)\tilde{\mathbf{X}}(t_0+\theta) + \int_{t_0}^{t_0+\theta} \mathbf{T}^*_{\tilde{\mathbf{F}}}(t,t_0)\mathbf{W}(t)dt \tag{4.16}$$

with $\mathbb{T}^*_{\tilde{\mathbf{F}}}(t_0)$ and $\mathbf{T}^*_{\tilde{\mathbf{F}}}(t,t_0)$ are the adjoint operators of $\mathbb{T}_{\tilde{\mathbf{F}}}(t_0)$ and $\mathbf{T}_{\tilde{\mathbf{F}}}(t,t_0)$, respectively, with respect to the inner product (1.1). From (2.7) we infer that $\mathbf{W}(t) \succeq 0$, for all $t \in [t_0, t_0+\theta]$. This leads to

$$\Psi \triangleq \int_{t_0}^{t_0+\theta} \mathbf{T}^*_{\tilde{\mathbf{F}}}(t,t_0)\mathbf{W}(t)dt \succeq 0 \tag{4.17}$$

because $\mathbf{T}^*_{\tilde{\mathbf{F}}}(t,t_0)\mathbf{W}(t) \succeq 0$, for all $t \in [t_0, t_0+\theta]$.
On the other hand from $\mu \in \sigma^+(\mathbb{T}_{\tilde{\mathbf{F}}}(t_0))$ we deduce that there exists $\mathbf{Y} \in \mathcal{S}_n^{N+} \setminus \{0\}$ such that

$$\mathbb{T}_{\tilde{\mathbf{F}}}(t_0)\mathbf{Y} = \mu\mathbf{Y}. \tag{4.18}$$

From (1.1), (4.16), (4.17) and (4.18) together with $\tilde{\mathbf{X}}(t_0) = \tilde{\mathbf{X}}(t_0+\theta)$, we get

$$\begin{aligned}0 \leq& \langle\Psi,\mathbf{Y}\rangle = \langle\tilde{\mathbf{X}}(t_0),\mathbf{Y}\rangle - \langle\mathbb{T}^*_{\tilde{\mathbf{F}}}(t_0)\tilde{\mathbf{X}}(t_0+\theta),\mathbf{Y}\rangle \\ &= \langle\tilde{\mathbf{X}}(t_0),\mathbf{Y}\rangle - \langle\tilde{\mathbf{X}}(t_0),\mathbb{T}_{\tilde{\mathbf{F}}}(t_0)\mathbf{Y}\rangle = (1-\mu)\langle\tilde{\mathbf{X}}(t_0),\mathbf{Y}\rangle \leq 0.\end{aligned}$$

Hence, $\langle\Psi,\mathbf{Y}\rangle = 0$, which yields to

$$\int_{t_0}^{t_0+\theta} \langle\mathbf{W}(t), \mathbf{T}_{\tilde{\mathbf{F}}}(t,t_0)\mathbf{Y}\rangle dt = 0. \tag{4.19}$$

From (4.15) and (4.19) we deduce that

$$\int_{t_0}^{t_0+\theta} \langle \mathbf{M}(t) - \mathbf{L}(t)\mathbf{R}^{-1}(t)\mathbf{L}^T(t), \mathbf{T}_{\tilde{\mathbf{F}}}(t,t_0)\mathbf{Y}\rangle dt = 0, \tag{4.20a}$$

$$\int_{t_0}^{t_0+\theta} \langle \mathbf{W}_1(t), \mathbf{T}_{\tilde{\mathbf{F}}}(t,t_0)\mathbf{Y}\rangle dt = 0. \tag{4.20b}$$

where $\mathbf{W}_1(t) = (\tilde{\mathbf{F}}(t) + \mathbf{R}^{-1}(t)\mathbf{L}^T(t))^T\mathbf{R}(t)(\tilde{\mathbf{F}}(t) + \mathbf{R}^{-1}(t)\mathbf{L}^T(t))$.
From (2.7a) and (4.20a) we get

$$\langle \mathbf{M}(t) - \mathbf{L}(t)\mathbf{R}^{-1}(t)\mathbf{L}^T(t), \mathbf{T}_{\tilde{\mathbf{F}}}(t,t_0)\mathbf{Y}\rangle = 0.$$

Further, from (1.1), (4.2) one gets

$$\sum_{i=1}^{N} Tr\left[\tilde{C}(t,i)(\mathbf{T}_{\tilde{\mathbf{F}}}(t,t_0)\mathbf{Y})(i)\tilde{C}^T(t,i)\right] = 0, \text{ for all } t \in [t_0, t_0+\theta].$$

Since $\tilde{C}(t,i)(\mathbf{T}_{\tilde{\mathbf{F}}}(t,t_0)\mathbf{Y})(i)\tilde{C}^T(t,i) \geq 0$ we may conclude that

$$\tilde{\mathbf{C}}(t)\mathbf{T}_{\tilde{\mathbf{F}}}(t,t_0)\mathbf{Y} = 0, \text{ for all } t \in [t_0, t_0+\theta]. \tag{4.21}$$

Proceeding analogously we obtain from (4.20b) that

$$\tilde{\mathbf{F}}(t)\mathbf{T}_{\tilde{\mathbf{F}}}(t,t_0)\mathbf{Y} = -\mathbf{R}^{-1}(t)\mathbf{L}^T(t)\mathbf{T}_{\tilde{\mathbf{F}}}(t,t_0)\mathbf{Y}, \text{ for all } t \in [t_0, t_0+\theta]. \tag{4.22}$$

Applying Lemma 4.4 with $\mathbf{F}_1(t) = \tilde{\mathbf{F}}(t)$ and $\mathbf{F}_2(t) = -\mathbf{R}^{-1}(t)\mathbf{L}^T(t)$ in the case of the equality (4.22) we may infer that

$$\mathbf{T}_{\tilde{\mathbf{F}}}(t,t_0)\mathbf{Y} = \mathbf{T}_{-\mathbf{R}^{-1}\mathbf{L}^T}(t,t_0)\mathbf{Y}, \text{ for all } t \in [t_0, t_0+\theta]. \tag{4.23}$$

From (4.18) together with (4.23) written for $t = t_0 + \theta$, we obtain

$$\mathbb{T}_{-\mathbf{R}^{-1}\mathbf{L}^T}(t_0)\mathbf{Y} = \mu\mathbf{Y}. \tag{4.24}$$

Finally, from (4.21) and (4.23) we obtain that

$$\tilde{\mathbf{C}}(t)\mathbf{T}_{-\mathbf{R}^{-1}\mathbf{L}^T}(t,t_0)\mathbf{Y} = 0, \text{ for all } t \in [t_0, t_0+\theta]. \tag{4.25}$$

From (4.24) and (4.25) we may conclude that μ is a distinctive characteristic multiplier unobservable at instance time t_0 for $(\tilde{\mathbf{C}}(\cdot), \mathcal{L}_{-\mathbf{R}^{-1}\mathbf{L}^T}(\cdot))$ which ends the proof. □

Proposition 4.6. *Under the assumption* $(\mathbf{H})$, *if* $\mu = 1$ *is a distinctive characteristic multiplier unobservable at instance time* t_0 *for the pair* $(\tilde{\mathbf{C}}(\cdot), \mathcal{L}_{-\mathbf{R}^{-1}\mathbf{L}^T}(\cdot))$ *then for any positive semidefinite and* θ*−periodic solution* $\tilde{\mathbf{X}}(\cdot)$ *of the GRDE* (2.1) *we have that* $1 \in \sigma^+(\mathbb{T}_{\tilde{\mathbf{F}}}(t_0))$ *where* $\tilde{\mathbf{F}}(\cdot)$ *is the feedback gain associated via* (3.9) *to the solution* $\tilde{\mathbf{X}}(\cdot)$ *and* $\tilde{\mathbf{C}}(\cdot)$ *is defined in* (4.2).

Proof. If $\mu = 1$ is a distinctive characteristic multiplier unobservable at t_0 for $(\tilde{\mathbf{C}}(\cdot), \mathcal{L}_{-\mathbf{R}^{-1}\mathbf{L}^T}(\cdot))$, there exists $\mathbf{Y} \in \mathcal{S}_n^{N+} \setminus \{0\}$ with the properties

$$\mathbb{T}_{-\mathbf{R}^{-1}\mathbf{L}^T}(t_0)\mathbf{Y} = \mathbf{Y} \tag{4.26}$$

and

$$\tilde{\mathbf{C}}(t)\mathbf{T}_{-\mathbf{R}^{-1}\mathbf{L}^T}(t,t_0)\mathbf{Y} = 0, \quad \text{for all } \; t \in [t_0, t_0+\theta]. \tag{4.27}$$

Let $\tilde{\mathbf{X}}(\cdot)$ be an arbitrary positive semidefinite and $\theta-$periodic solution of GRDE (2.1). Applying Corollary 3.4, taking $\mathbf{F}(t) = -\mathbf{R}^{-1}(t)\mathbf{L}^T(t)$ we may rewrite the equation (2.1) satisfied by $\tilde{\mathbf{X}}(\cdot)$ as:

$$\begin{aligned}-\frac{d}{dt}\tilde{\mathbf{X}}(t) &= \mathcal{L}^*_{-\mathbf{R}^{-1}\mathbf{L}^T}(t_0)[\tilde{\mathbf{X}}(t)] + \mathbf{M}(t) - \mathbf{L}(t)\mathbf{R}^{-1}(t)\mathbf{L}^T(t) \\ &\quad - (\mathbf{R}^{-1}(t)\mathbf{L}^T(t) + \tilde{\mathbf{F}}(t))^T(\mathbf{\Pi}_3(t)[\tilde{\mathbf{X}}] + \mathbf{R}(t))(\mathbf{R}^{-1}(t)\mathbf{L}^T(t) + \tilde{\mathbf{F}}(t)).\end{aligned}$$

This allows us to write the following representation of the solution $\tilde{\mathbf{X}}(\cdot)$:

$$\begin{aligned}\tilde{\mathbf{X}}(t_0) &= \mathbb{T}^*_{-\mathbf{R}^{-1}\mathbf{L}^T}(t_0)\tilde{\mathbf{X}}(t_0+\theta) + \int_{t_0}^{t_0+\theta} \mathbf{T}^*_{-\mathbf{R}^{-1}\mathbf{L}^T}(t,t_0)\tilde{\mathbf{C}}^2(t)dt \\ &\quad - \int_{t_0}^{t_0+\theta} \mathbf{T}^*_{-\mathbf{R}^{-1}\mathbf{L}^T}(t,t_0)\mathbf{\Upsilon}(t)dt \end{aligned} \tag{4.28a}$$

$$\begin{aligned}\mathbf{\Upsilon}(t) &= (\mathbf{R}^{-1}(t)\mathbf{L}^T(t) + \tilde{\mathbf{F}}(t))^T(\mathbf{\Pi}_3(t)[\tilde{\mathbf{X}}(t)] \\ &\quad + \mathbf{R}(t))(\mathbf{R}^{-1}(t)\mathbf{L}^T(t) + \tilde{\mathbf{F}}(t)).\end{aligned} \tag{4.28b}$$

Employing (4.26), (4.27) together with periodicity property of $\tilde{\mathbf{X}}(\cdot)$, we get

$$\langle \mathbb{T}^*_{-\mathbf{R}^{-1}\mathbf{L}^T}(t_0)\tilde{\mathbf{X}}(t_0+\theta) - \tilde{\mathbf{X}}(t_0), \mathbf{Y}\rangle + \left\langle \int_{t_0}^{t_0+\theta} \mathbf{T}^*_{-\mathbf{R}^{-1}\mathbf{L}^T}(t,t_0)\tilde{\mathbf{C}}^2(t)dt, \mathbf{Y} \right\rangle = 0. \tag{4.29}$$

Further, (4.28a) and (4.29) yield

$$\int_{t_0}^{t_0+\theta} \langle \mathbf{T}^*_{-\mathbf{R}^{-1}\mathbf{L}^T}(t,t_0)\mathbf{\Upsilon}(t), \mathbf{Y}\rangle dt = 0. \tag{4.30}$$

From (1.1), (4.28b) we infer that (4.30) is true, if and only if

$$\langle \mathbf{T}^*_{-\mathbf{R}^{-1}\mathbf{L}^T}(t,t_0)\mathbf{\Upsilon}(t), \mathbf{Y}\rangle = 0, \quad \text{for all } \; t \in [t_0, t_0+\theta].$$

This is equivalent to

$$\langle \mathbf{\Upsilon}(t), \mathbf{T}_{-\mathbf{R}^{-1}\mathbf{L}^T}(t,t_0)\mathbf{Y}\rangle = 0, \quad \text{for all } \; t \in [t_0, t_0+\theta]. \tag{4.31}$$

Combining (1.1), (2.7a), (4.28b) we may conclude that (4.31) is true if and only if

$$-\mathbf{R}^{-1}(t)\mathbf{L}^T(t)\mathbf{T}_{-\mathbf{R}^{-1}\mathbf{L}^T}(t,t_0)\mathbf{Y} = \tilde{\mathbf{F}}(t)\mathbf{T}_{-\mathbf{R}^{-1}\mathbf{L}^T}(t,t_0)\mathbf{Y}, \tag{4.32}$$

for all $t \in [t_0, t_0+\theta]$.
Applying Lemma 4.4 taking $\mathbf{F}_1(t) = -\mathbf{R}^{-1}(t)\mathbf{L}^T(t)$ and $\mathbf{F}_2(t) = \tilde{\mathbf{F}}(t)$, we conclude from (4.32) that

$$\mathbf{T}_{-\mathbf{R}^{-1}\mathbf{L}^T}(t,t_0)\mathbf{Y} = \mathbf{T}_{\tilde{\mathbf{F}}}(t,t_0)\mathbf{Y}. \tag{4.33}$$

Finally, (4.26) together with (4.33) written for $t = t_0+\theta$, yield $\mathbb{T}_{\tilde{\mathbf{F}}}(t_0)\mathbf{Y} = \mathbf{Y}$. This shows that $\mu = 1$ lies in $\sigma^+(\mathbb{T}_{\tilde{\mathbf{F}}}(t_0))$ which complete the proof. □

4.3. The proof of the main result

First we prove the implication **(i)** $\Longrightarrow$ **(ii)**. The stabilizability of the triple $(\mathbf{A}(\cdot), \mathbf{B}(\cdot), \mathbf{\Pi}(\cdot))$ is a necessary condition for the existence of the stabilizing solution of the GRDE (2.1) because the feedback gain associated via (2.22) to the stabilizing solution stabilizes this triple in the sense of the Definition 4.1. Let us assume by contrary that **(ii) (b)** from the statement of the Theorem 4.3 is not true. This means that there exists $t_0 \in \mathbb{R}$ such that $\mu = 1$ is a distinctive characteristic multiplier unobservable at t_0 for $(\tilde{\mathbf{C}}(\cdot), \mathcal{L}_{-\mathbf{R}^{-1}\mathbf{L}^T}(\cdot))$. From Proposition 4.6 we deduce that $\mu = 1 \in \sigma^+(\mathbb{T}_{\tilde{\mathbf{F}}}(t_0))$ where $\tilde{\mathbf{F}}(\cdot)$ is the stabilizing feedback gain associated via (2.22) to the solution $\tilde{\mathbf{X}}(\cdot)$ of the GRDE (2.1). Hence $\rho_{\tilde{\mathbf{F}}} \geq 1$, which contradicts the fact that $\tilde{\mathbf{X}}(\cdot)$ is the stabilizing solution. So, we have shown that the implication **(i)** $\Longrightarrow$ **(ii)** holds.

Now we prove **(ii)** $\Longrightarrow$ **(i)**. Applying Theorem 5.3.5 from [9] we deduce that under the assumption **(H)** if the triple $(\mathbf{A}(\cdot), \mathbf{B}(\cdot), \mathbf{\Pi}(\cdot))$ is stabilizable then the GRDE (2.1) has a solution $\tilde{\mathbf{X}}(\cdot)$ which is maximal in the class of positive semidefinite solutions of (2.1). Moreover, $\tilde{\mathbf{X}}(\cdot)$ is a periodic function of period θ. Let $\tilde{\mathbf{F}}(t)$ be the feedback gain associated to the solution $\tilde{\mathbf{X}}(\cdot)$ via (3.9) written for $\mathbf{X}$ replaced by $\tilde{\mathbf{X}}(t)$. We denote $\mathbb{T}_{\tilde{\mathbf{F}}}(t_0)$, $t_0 \in \mathbb{R}$, the monodromy operator defined by the linear differential equation with periodic coefficients of type (3.1) when $\mathbf{F}(t)$ is replaced by $\tilde{\mathbf{F}}(t)$. In the proof of Theorem 5.3.5 from [9] $\mathbb{T}_{\tilde{\mathbf{F}}}(t_0)$ is obtained as the limit of a sequence of linear operators having the spectral radius strictly less than 1. Hence, the spectral radius $\rho_{\tilde{\mathbf{F}}}$ of the monodromy operator $\mathbb{T}_{\tilde{\mathbf{F}}}(t_0)$ satisfies $\rho_{\tilde{\mathbf{F}}} \leq 1$. To show that the maximal solution $\tilde{\mathbf{X}}(\cdot)$ is just the stabilizing solution we have to show that $\rho_{\tilde{\mathbf{F}}} < 1$. Let us assume by contrary that $\rho_{\tilde{\mathbf{F}}} = 1$. In this case, $\mu = 1 \in \sigma^+(\mathbb{T}_{\tilde{\mathbf{F}}}(t_0))$, for all $t_0 \in \mathbb{R}$. Then from Proposition 4.5 it follows that $\mu = 1$ is a distinctive characteristic multiplier unobservable for the pair $(\tilde{\mathbf{C}}(\cdot), \mathcal{L}_{-\mathbf{R}^{-1}\mathbf{L}^T}(\cdot))$ which contradicts **(ii) (b)**. Thus, $\rho_{\tilde{\mathbf{F}}} < 1$ which complete the proof.

5. Conclusions

This paper studies the stabilizing solution for a class of continuous-time backward nonlinear equations. The concept of unobservable characteristic multipliers for a pair adequately chosen is introduced. Based on this spectrum technique, we obtain a necessary and sufficient condition for the existance of the stabilizing solution of the GRDE. The proposed techniques are formulated as the solvability of a linear equation.

References

[1] Bittanti, S., Colaneri, P., *Periodic Systems: Filtering and Control*, Springer-Verlag London, 2009.

[2] Chung, K.L., *Markov Chains with Stationary Transition Probabilities*, Springer, Berlin, Heidelberg, 1967.

[3] Ciucu, G., Tudor C., *Teoria Probabilităţilor şi Aplicaţii*, Editura Ştiinţifică şi Enciclopedică, 1983.

[4] Daleckiĭ, Ju. L., Kreĭn, M.G., *Stability of Solutions of Differential Equations in Banach Space*, Translations of Mathematical Monographs, American Mathematical Society, vol. 43, 1974.

[5] Damm, T., Hinrichsen, D., *Newton's method for concave operators with resolvent positive derivatives in ordered Banach spaces*, Linear Algebra Appl., **363**(2003), 2003.

[6] Delong, L., *Backward Stochastic Differential Equations with Jumps and Their Actuarial and Financial Applications*, Springer-Verlag London, 2013.

[7] Doob, J.L., *Stochastic Processes*, New York, Wiley, 1967.

[8] Drăgan, V., Halanay A., *Stabilization of Linear Systems*, Birkhäuser Basel, 1999.

[9] Drăgan, V., Morozan, T., Stoica, A.-M., *Mathematical Methods in Robust Control of Linear Stochastic Systems*, Springer-Verlag New York, 2013.

[10] Fragoso, M.D., Costa, O.L.V., de Souza, C.E., *A new approach to linearly perturbed Riccati equations arising in stochastic control*, Appl. Math. Optim., **37**(1998), 99-126.

[11] Freedman, A., *Stochastic Differential Equations and Applications*, vol. I, Academic, New York, 1975.

[12] Halanay, A., *Differential Equations: Stability, Oscillations, Time Lags*, Academic Press, vol. 23, 1966.

[13] Kalman, R., *Contributions on the theory of optimal control*, Bult. Soc. Math. Mexicana Segunda Ser., **5**(1960), no. 1, 102-119.

[14] Kawano, Y., Ohtsuka, T., *Nonlinear eigenvalue approach to differential Riccati equations for contraction analysis*, IEEE Trans. Automat. Control, **62**(2017), no. 12, 6497-6504.

[15] Lancaster, P., Rodman, L., *Algebraic Riccati Equations*, Clarendon Press, Oxford, 1995.

[16] Øksendal, B., *Stochastic Differential Equations*, Springer-Verlag Berlin, Heidelberg, 2003.

[17] Pastor, A., Hernández, V., *Differential periodic Riccati equations: existence and uniqueness of nonnegative definite solutions*, Math. Control Signals Systems, **6**(1993), 341-362.

[18] Qiu, L., *On the generalized eigenspace approach for solving Riccati equations*, In: IFAC Proceedings Volumes, 14th IFAC World Congress 1999, Beijing, China, 5-9 July, **32**(1999), no. 2, 2945-2950.

[19] Tang, S., *General linear quadratic optimal stochastic control problems with random coefficients: linear stochastic Hamilton systems and backward stochastic Riccati equations*, SIAM J. Control Optim., **42**(2003), no. 1, 53-75.

[20] Wonham, W.M., *On a matrix Riccati equation of stochastic control*, SIAM J. Control Optim., **6**(1968), no. 4, 681-697.

[21] Yakubovich, V.A., *A linear-quadratic optimization problem and the frequency theorem for nonperiodic systems. I*, Sib. Math. J., **27**(1986), 614-630.

[22] Zhang, J., *Backward Stochastic Differential Equations: From Linear to Fully Nonlinear Theory*, Springer-Verlag New York, vol. 86, 2017.

[23] Zheng, J., Qiu, L., *On the existence of a mean-square stabilizing solution to a continuous-time modified algebraic Riccati equation*, In: 21st International Symposium on Mathematical Theory of Networks and Systems July 7-11, 2014, Groningen, The Netherlands, 2014, 694-700.

Vasile Drăgan
Institute of Mathematics "Simion Stoilow" of the Romanian Academy,
P.O.Box 1-764, RO-014700, Bucharest, Romania
and the Academy of the Romanian Scientists
e-mail: `Vasile.Dragan@imar.ro`

Ioan-Lucian Popa
Department of Computing, Mathematics and Electronics
"1 Decembrie 1918" University of Alba Iulia,
510009 - Alba Iulia, Romania
e-mail: `lucian.popa@uab.ro`

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 179–196
DOI: 10.24193/subbmath.2021.1.15

Kálmán's filtering technique in structural equation modeling

Marianna Bolla and Fatma Abdelkhalek

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. Structural equation modeling finds linear relations between exogenous and endogenous latent and observable random vectors. In this paper, the model equations are considered as a linear dynamical system to which the celebrated R. E. Kálmán's filtering technique is applicable. An artificial intelligence is developed, where the partial least squares algorithm of H. Wold and the block Cholesky decomposition of H. Kiiveri et al. are combined to estimate the parameter matrices from a training sample. Then the filtering technique introduced is capable to predict the latent variable case values along with the prediction error covariance matrices in the test sample. The recursion goes from case to case along the test sample, without having to re-estimate the parameter matrices. The algorithm is illustrated on real life sociological data.

Mathematics Subject Classification (2010): 62H05, 62P25, 68T99.

Keywords: Structural equation modeling, linear dynamical systems, Kálmán's filtering, artificial intelligence, application to social sciences.

1. Introduction

We consider structural equation model (SEM) for two latent random vectors that depend through a linear model on two observable random vectors, respectively (they usually include exogenous and endogenous variables). This kind of models was first investigated by T. Haavelmo [1], who obtained the Nobel Prize for it later. Unlike the traditional factor analysis, where latent variables were introduced and given a meaning based on the factor loadings, here the latent variables are organic parts of the model. The latent variables, e.g., alienation, ambition in [2] or mobility in our example, are given by the experts, and the observed (measurement) variables are indicators of them. In this way, so-called inner and outer relations are stated between the latent variables and between the observable and latent ones, respectively.

The estimation of the parameter matrices of this model was elaborated both in the Gaussian and distribution-free cases, former by K. G. Jöreskog (LISREL) [2], while the other by H. Wold (PLS) [8] in the 1970-1980s. These two approaches are sometimes called covariance-based and component-based SEM. However, we can consider the model equations as a linear dynamical system to which the celebrated R. E. Kálmán's filtering technique [3] is applicable. This technique was developed in the 1960s for time series to make predictions for the hidden state variables of a state space model, and was used in the lunar landing, for instance. We will show how to apply this technique in the more complicated dynamical system, containing two state and two observable equations, describing inner relations between the observable and latent variables, both for the exogenous and endogenous ones. Our contribution is that we connect these two approaches.

The parameter matrices are estimated from a training sample. We combine the first stage of the PLS algorithm of H. Wold to estimate the case values of the latent variables and the method of H. Kiiveri et al. [5] to decompose the inverse of the product moment matrix obtained with the latent case value estimates. At this point, we apply the block Cholesky decomposition. Then the filtering technique to be introduced is capable to make predictions for the endogenous variables based on the exogenous ones, through the latent variables. The driving force is that we propagate the error covariance matrices of the exogenous and endogenous latent variables in a recursion.

The test sample is a succession of observations coming one by one (like a time series or just subsequent observations), and the algorithm predicts their endogenous variables based on their own exogenous ones. Our contribution is that we combine existing methods for parameter estimation, and then apply filtering technique for prediction, so we develop an artificial intelligence. The computational gain is that the parameter matrices need not be estimated for every new-coming case in the test sample, but are estimated only once, in the training sample. The method is distribution-free (just second moments are used in the linear state equations) and applicable to small training, and not necessarily independent test samples.

The organization of the paper is as follows. In Section 2, the most important notions and facts about best linear predictions in Hilbert spaces are introduced. In Section 3, the prediction and propagation of the error covariance matrices are derived in two stages. The main results are summarized in Theorem 3.1 of Section 3.3. Then the proposed algorithm is illustrated on real life sociological data in Section 4. Eventually, the last Section 5 discusses the benefits of the proposed method together with some possible further perspectives.

2. Preliminaries

The following linear dynamical system that resembles the one to which R. E. Kálmán gave a recursive algorithm is considered:

$$\boldsymbol{B}\boldsymbol{\eta} = \boldsymbol{A}\boldsymbol{\xi} + \boldsymbol{\zeta},$$

where $\boldsymbol{\eta}$ is m- and $\boldsymbol{\xi}$ is n-dimensional latent vector, $\boldsymbol{B}$ and $\boldsymbol{A}$ are $m \times m$ and $m \times n$ coefficient matrices, and $\boldsymbol{\zeta}$ is a random vector of residuals of uncorrelated components.

It is also uncorrelated with $\boldsymbol{\xi}$, and $\boldsymbol{B}$ is nonsingular. In the recursive models, $\boldsymbol{B}$ is upper triangular, with 1s along its main diagonal.

Here $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are not observed, but instead, the p-dimensional $\mathbf{Y}$ and the q-dimensional $\mathbf{X}$ are observed such that

$$\mathbf{X} = \boldsymbol{C}\boldsymbol{\xi} + \boldsymbol{\varepsilon}, \quad \mathbf{Y} = \boldsymbol{G}\boldsymbol{\eta} + \boldsymbol{\delta},$$

where $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are vectors of measurement errors in $\mathbf{X}$ and $\mathbf{Y}$, respectively. They are uncorrelated with each other and $\boldsymbol{\zeta}$. Typically, $n \le q$ and $m \le p$.

For the estimation of the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{G}$, and the covariance matrices of the errors, there is the LISREL algorithm of K. G. Jöreskog [2] (assuming multivariate Gaussian distribution of the measurement variables and large sample sizes) and component-wise SEM algorithms (not postulating normality and being able to treat small sample sizes), e.g., [7, 8], at our disposal.

In the first stage of his PLS algorithm, H. Wold [8] gives an iteration to find the case values of the latent variables. He states that this fixed point iteration converges. We use only this first stage to calculate the product moment estimate of the covariance matrix of the latent variables. Then we decompose the inverse of this matrix as $\boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^T$ with $\boldsymbol{L}$ and $\boldsymbol{D}$ having the form

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{B}^T & \boldsymbol{O} \\ -\boldsymbol{A}^T & \boldsymbol{I} \end{pmatrix}, \qquad \boldsymbol{D} = \begin{pmatrix} \boldsymbol{Q}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{F}^{-1} \end{pmatrix}, \tag{2.1}$$

where $\boldsymbol{B}$ is $m \times m$ upper triangular matrix with 1s along its main diagonal, and $\boldsymbol{A}$ is $m \times n$ matrix. The block-diagonal matrix $\boldsymbol{D}$ comprises the inverse of the error covariance matrix $\boldsymbol{Q}$ of $\boldsymbol{\zeta}$ and $\boldsymbol{F}$ of $\boldsymbol{\xi}$, where $\boldsymbol{Q}$ itself is a diagonal matrix. For this purpose we use the block Cholesky decomposition with block sizes $1, \dots, 1, n$ with number[1] m of 1s.

Wold's algorithm also provides the outer relation matrices $\boldsymbol{C}$ and $\boldsymbol{G}$. In this way, we can estimate the parameter matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{G}$ based on a training sample and on adjacency matrices that specify which latent variable is related to which observable one, both among the exogenous and endogenous variables. This is the point where the expert can intervene the system. For a detailed description, see Section 4.

In the heart of the estimation and the forthcoming filtering there is the simultaneous usage of OLS (ordinary least squares) regression, tracing back to the Gauss normal equations. We give a short summary of that.

Now we concentrate on linear estimates in Hilbert spaces that are the best whenever the underlying distribution is multivariate Gaussian, but is also applicable to second order processes.

Lemma 2.1. *Let $\mathbf{Y} \in \mathbb{R}^p$ and $\mathbf{X} \in \mathbb{R}^q$ be random vectors on a joint probability space with existing second moments and zero expectation. Then*
$\mathbb{E}\|\mathbf{Y} - \boldsymbol{A}^T\mathbf{X}\|^2$ *is minimized with*

$$\boldsymbol{A} = [\mathbb{E}\mathbf{X}\mathbf{X}^T]^-[\mathbb{E}\mathbf{X}\mathbf{Y}^T], \tag{2.2}$$

[1]Number of endogenous latent variables in the model.

where $\boldsymbol{A}$ is a $q \times p$ matrix and we use generalized inverse $^-$ if the covariance matrix $\mathbb{E}\mathbf{X}\mathbf{X}^T$ of $\mathbf{X}$ is singular. If it is positive definite, then we get a unique estimate for $\boldsymbol{A}$ with the unique inverse matrix $[\mathbb{E}\mathbf{X}\mathbf{X}^T]^{-1}$.

Note that the notation $^-$ applies to any (not necessarily unique) generalized inverse, whereas $^+$ will be used for the uniquely defined Moore–Penrose generalized inverse, see [6].

Proof. Observe that minimizing

$$\mathbb{E}\|\mathbf{Y} - \boldsymbol{A}^T\mathbf{X}\|^2 = \sum_{i=1}^{p}(Y^i - \mathbf{a}_i^T\mathbf{X})^2$$

with respect to $\boldsymbol{A} = (\mathbf{a}_1 \dots \mathbf{a}_p)$ falls apart into the following p minimization tasks, with respect to the q-dimensional column vectors of $\boldsymbol{A}$:

$$\min_{\mathbf{a}_i}(Y^i - \mathbf{a}_i^T\mathbf{X})^2, \quad i = 1, \dots, p.$$

The solution (e.g., with the help of differentiation) gives the well known *Gauss normal equations* from the classical theory of multivariate regression:

$$[\mathbb{E}\mathbf{X}\mathbf{X}^T]\mathbf{a}_i = [\mathbb{E}\mathbf{X}Y_i], \quad i = 1, \dots, p.$$

Since this system of linear equations is consistent (the vector $\mathbb{E}\mathbf{X}Y_i$ is in the column space of $\mathbb{E}\mathbf{X}\mathbf{X}^T$), it always has a solution in the general form:

$$\mathbf{a}_i = [\mathbb{E}\mathbf{X}\mathbf{X}^T]^-[\mathbb{E}\mathbf{X}Y_i], \quad i = 1, \dots, p.$$

Therefore the matrix $\boldsymbol{A}$ giving the optimum is

$$\boldsymbol{A} = [\mathbb{E}\mathbf{X}\mathbf{X}^T]^-[\mathbb{E}\mathbf{X}\mathbf{Y}^T],$$

that is unique only if $\mathbb{E}\mathbf{X}\mathbf{X}^T$ is invertible (positive definite), otherwise (if $\mathbb{E}\mathbf{X}\mathbf{X}^T$ is singular, positive semidefinite) infinitely many versions of the generalized inverse give infinitely many convenient $\boldsymbol{A}$s. However, these always provide the same optimal linear prediction (projection) for $\mathbf{Y}$ as follows:

$$\mathrm{Proj}_{H(\mathbf{X})}\mathbf{Y} = \hat{\mathbf{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_p \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^T\mathbf{X} \\ \mathbf{a}_2^T\mathbf{X} \\ \vdots \\ \mathbf{a}_p^T\mathbf{X} \end{bmatrix} = \boldsymbol{A}^T\mathbf{X},$$

where $\mathrm{Proj}_{H(\mathbf{X})}$ denotes the projection onto the Hilbert space spanned by the linear combinations of the components of $\mathbf{X}$ (the expectations are zeros and the inner product is the covariance). □

Lemma 2.2. *Let $\mathbf{Y} \in \mathbb{R}^p$ and $\mathbf{X} \in \mathbb{R}^q$ be random vectors on a joint probability space with existing second moments and zero expectation, and let $\mathrm{Proj}_{H(\mathbf{X})}\mathbf{Y}$ denotes the best linear prediction of $\mathbf{Y}$ based on $\mathbf{X}$, as before. Then with any $p \times p$ matrix $\boldsymbol{\Phi}$,*

$$\mathrm{Proj}_{H(\mathbf{X})}(\boldsymbol{\Phi}\mathbf{Y}) = \boldsymbol{\Phi}\mathrm{Proj}_{H(\mathbf{X})}\mathbf{Y}.$$

Proof. We saw that $\mathrm{Proj}_{H(\mathbf{X})}\mathbf{Y} = \boldsymbol{A}^T\mathbf{X}$, where by (2.2), $\boldsymbol{A} = [\mathbb{E}\mathbf{X}\mathbf{X}^T]^-[\mathbb{E}\mathbf{X}\mathbf{Y}^T]$, and we use the generalized inverse $^-$ if the covariance matrix $\mathbb{E}\mathbf{X}\mathbf{X}^T$ of $\mathbf{X}$ is singular. Then

$$\begin{aligned}\mathrm{Proj}_{H(\mathbf{X})}(\boldsymbol{\Phi}\mathbf{Y}) &= \{[\mathbb{E}\mathbf{X}\mathbf{X}^T]^-[\mathbb{E}\mathbf{X}(\boldsymbol{\Phi}\mathbf{Y})^T]\}^T\mathbf{X} = [\mathbb{E}(\boldsymbol{\Phi}\mathbf{Y}\mathbf{X}^T)][\mathbb{E}\mathbf{X}\mathbf{X}^T]^-\mathbf{X}\\ &= \boldsymbol{\Phi}[\mathbb{E}(\mathbf{Y}\mathbf{X}^T)][\mathbb{E}\mathbf{X}\mathbf{X}^T]^-\mathbf{X} = \boldsymbol{\Phi}\mathrm{Proj}_{H(\mathbf{X})}\mathbf{Y}.\end{aligned}$$

□

The above lemma shows that this projection is linear in $\mathbf{Y}$ and it commutes with $\boldsymbol{\Phi}$. In the Gaussian case, obviously, we have that

$$\mathrm{Proj}_{H(\mathbf{X})}(\boldsymbol{\Phi}\mathbf{Y}) = \mathbb{E}(\boldsymbol{\Phi}\mathbf{Y} \mid \mathbf{X}) = \boldsymbol{\Phi}\mathbb{E}(\mathbf{Y} \mid \mathbf{X}) = \boldsymbol{\Phi}\mathrm{Proj}_{H(\mathbf{X})}(\mathbf{Y})$$

by the properties of the conditional expectation.

The above setup is used for simultaneous (in other words, multiple response) regressions when we regress the components of a random vector (target) with all the components of the predictors.

3. The linear dynamical system for the prediction

Discrete time observations $\mathbf{X}_t$, $\mathbf{Y}_t$ arrive, whereas $\boldsymbol{\xi}_t$ and $\boldsymbol{\eta}_t$ are latent state variables corresponding to them. Starting at time 0, for $t = 1, 2, \ldots$, the estimate of $\hat{\boldsymbol{\eta}}_t$ is found, while observing $\mathbf{X}_1, \ldots, \mathbf{X}_t$. Actually, to find $\hat{\boldsymbol{\eta}}_t$, we only need the estimate $\hat{\boldsymbol{\xi}}_t$ and the last observation $\mathbf{X}_t$. Then to find $\hat{\boldsymbol{\xi}}_{t+1}$, the preceding estimate $\hat{\boldsymbol{\eta}}_t$ and the last observation $\mathbf{Y}_t$ are needed. In this way, a recursion is given via the propagation of the error covariance matrices. During the calculations, we use the linearity of the state equations and the predictions, for which we confine ourselves to the second moments of the underlying distributions (second order processes).

The linear dynamical system is

$$\begin{aligned}\boldsymbol{B}\boldsymbol{\eta}_t &= \boldsymbol{A}\boldsymbol{\xi}_t + \boldsymbol{\zeta}_t\\ \boldsymbol{U}\boldsymbol{\xi}_{t+1} &= \boldsymbol{V}\boldsymbol{\eta}_t + \boldsymbol{\gamma}_t\\ \mathbf{X}_t &= \boldsymbol{C}\boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_t\\ \mathbf{Y}_t &= \boldsymbol{G}\boldsymbol{\eta}_t + \boldsymbol{\delta}_t,\end{aligned} \tag{3.1}$$

where $\boldsymbol{A}$ is $m \times n$, $\boldsymbol{B}$ is $m \times m$, $\boldsymbol{V}$ is $n \times m$, $\boldsymbol{U}$ is $n \times n$, $\boldsymbol{C}$ is $q \times n$, and $\boldsymbol{G}$ is $p \times m$ specified matrix; $\boldsymbol{B}$ and $\boldsymbol{U}$ are non-singular (in recursive models they are upper triangular with 1s along their main diagonals). Further, $\boldsymbol{\zeta}_t$ is an orthogonal process with $\mathbb{E}\boldsymbol{\zeta}_t\boldsymbol{\zeta}_s^T = \delta_{st}\boldsymbol{Q}$ with diagonal covariance matrix $\boldsymbol{Q}$; $\boldsymbol{\gamma}_t$ is an orthogonal process with $\mathbb{E}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_s^T = \delta_{st}\boldsymbol{R}$ with diagonal covariance matrix $\boldsymbol{R}$; $\mathbb{E}\boldsymbol{\xi}_s^T\boldsymbol{\zeta}_t = \mathbf{0}$ and $\mathbb{E}\boldsymbol{\eta}_s^T\boldsymbol{\gamma}_t = \mathbf{0}$ for $s \le t$; $\boldsymbol{\varepsilon}_t$ is independent of $\boldsymbol{\xi}_t$, $\boldsymbol{\delta}_t$ is independent of $\boldsymbol{\eta}_t$, they are also independent of each other and of $\boldsymbol{\zeta}_t$ and $\boldsymbol{\gamma}_t$. For simplicity, we assume that all the expectations are zeros.

The $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{U}, \boldsymbol{V}$ matrices are estimated from a training sample. Actually, the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ together with $\boldsymbol{Q}$ and $\boldsymbol{F}$ come from the block Cholesky decomposition (2.1), based on the product-moments of the estimated latent scores of the pairs $\boldsymbol{\xi}_s, \boldsymbol{\eta}_s$, where $s < 0$ is integer from the past (training sample). Likewise, the matrices

$\boldsymbol{U}$ and $\boldsymbol{V}$ together with $\boldsymbol{R}$ and $\boldsymbol{F}^*$ come from the block Cholesky decomposition (3.2) below, based on the product-moments of the estimated latent scores of the shifted pairs $\boldsymbol{\eta}_s, \boldsymbol{\xi}_{s+1}$ ($s < 0$). The inverse of this matrix is decomposed as $\boldsymbol{L}^*\boldsymbol{D}^*\boldsymbol{L}^{*T}$ with $\boldsymbol{L}^*$ and $\boldsymbol{D}^*$ having the form

$$\boldsymbol{L}^* = \begin{pmatrix} \boldsymbol{U}^T & \boldsymbol{O} \\ -\boldsymbol{V}^T & \boldsymbol{I} \end{pmatrix}, \qquad \boldsymbol{D}^* = \begin{pmatrix} \boldsymbol{R}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{F}^{*-1} \end{pmatrix}, \tag{3.2}$$

where recall that, in the recursive model, $\boldsymbol{U}$ is $n \times n$ upper triangular matrix with 1s along its main diagonal, and $\boldsymbol{V}$ is $n \times m$. The block-diagonal matrix $\boldsymbol{D}^*$ comprises the inverse of the error covariance matrix $\boldsymbol{R}$ of $\boldsymbol{\gamma}$ and $\boldsymbol{F}^*$ of $\boldsymbol{\eta}$, where $\boldsymbol{R}$ itself is a diagonal matrix. For this purpose we use the block Cholesky decomposition with block sizes $1, \dots, 1, m$ with number[2] n of 1s.

The matrices $\boldsymbol{C}$ ad $\boldsymbol{G}$ are estimated by the PLS algorithm of H. Wold [8]. The details are given in Section 4.

Now a recursion is introduced for the following problem: starting the observations at time 0 in the test sample, we want to estimate $\boldsymbol{\eta}_t$ based on $\mathbf{X}_t$, and $\boldsymbol{\xi}_{t+1}$ based on $\mathbf{Y}_t$ component-wise, with minimum mean square error. Former observations also play role, but only through the last one and through the propagation of the error covariance matrices. Here $\mathbf{X}_0$ and $\mathbf{Y}_0$ can be taken from the training sample.

3.1. First stage: $\mathbf{X}_t \to \hat{\boldsymbol{\eta}}_t$

For $t \geq 1$, let $H_{t-1}(\mathbf{X}) = \mathrm{Span}\,(\mathbf{X}_0, \dots, \mathbf{X}_{t-1})$ consists of the linear combinations of all the components of $\mathbf{X}_0, \dots, \mathbf{X}_{t-1}$ over a common probability space. They are also in a Hilbert space (L_2 space) with the covariance as inner product. We denote the optimal prediction of $\boldsymbol{\xi}_t$ based on $\mathbf{X}_0, \dots, \mathbf{X}_{t-1}$ by $\hat{\boldsymbol{\xi}}_t$.

If $\mathbf{X}_0, \dots, \mathbf{X}_{t-1}$ are observed, i.e., $H_{t-1}(\mathbf{X})$ is known, then the newly observed (measured) $\mathbf{X}_t$ can be orthogonally decomposed as

$$\mathbf{X}_t = \mathrm{Proj}_{H_{t-1}(\mathbf{X})}\mathbf{X}_t + \tilde{\mathbf{X}}_t = \overline{\mathbf{X}}_t + \tilde{\mathbf{X}}_t, \tag{3.3}$$

where the orthogonal component $\tilde{\mathbf{X}}_t \in I_t(\mathbf{X})$, and $I_t(\mathbf{X})$ is the so-called innovation subspace (actually, the components of $\tilde{\mathbf{X}}_t$ generate $I_t(\mathbf{X})$). Assume that $I_t(\mathbf{X})$ is not the sole $\mathbf{0}$ vector, otherwise observing $\mathbf{X}_t$ does not give any additional information to $H_{t-1}(\mathbf{X})$. If $\{\mathbf{X}_t\}$ is weakly stationary, it means that the process is *regular*.

Equation (3.3) implies the decomposition of the corresponding subspaces like

$$H_t(\mathbf{X}) = H_{t-1}(\mathbf{X}) \oplus I_t(\mathbf{X}), \tag{3.4}$$

that is the analogue of multidimensional Wold decomposition when we make one-step ahead prediction based on finitely many past values. (The Wold decomposition applies to the stationary and infinite past case. Indeed, when $t \to \infty$, i.e., going to the future, we approach this situation in the stationary case).

---

[2]Number of exogenous latent variables in the model.

Assume that we have already found $\hat{\boldsymbol{\xi}}_t$. We shall give a recursion to find $\hat{\boldsymbol{\eta}}_t$ by using the new value of $\mathbf{X}_t$. In view of Equation (3.4), we proceed as follows:

$$\begin{aligned}\boldsymbol{B}\hat{\boldsymbol{\eta}}_t &= \mathrm{Proj}_{H_t(\mathbf{X})}(\boldsymbol{B}\boldsymbol{\eta}_t) = \mathrm{Proj}_{H_{t-1}(\mathbf{X})}(\boldsymbol{B}\boldsymbol{\eta}_t) + \mathrm{Proj}_{I_t(\mathbf{X})}(\boldsymbol{B}\boldsymbol{\eta}_t)\\ &= \boldsymbol{A}\mathrm{Proj}_{H_{t-1}(\mathbf{X})}\boldsymbol{\xi}_t + \mathrm{Proj}_{H_{t-1}(\mathbf{X})}\boldsymbol{\zeta}_t + \boldsymbol{K}_t\tilde{\mathbf{X}}_t\\ &= \boldsymbol{A}\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t\tilde{\mathbf{X}}_t,\end{aligned} \tag{3.5}$$

where we utilized that $\boldsymbol{\zeta}_t \perp H_{t-1}(\mathbf{X})$, Lemma 2.1 and the first state equation of (3.1). We refer to the linearity of the projection, see Lemma 2.2. Since $\mathrm{Proj}_{I_t(\mathbf{X})}\boldsymbol{B}\boldsymbol{\eta}_t$ is the linear combination of the coordinates of the vector $\tilde{\mathbf{X}}_t \in I_t(\mathbf{X})$, its effect can be written as a matrix $\boldsymbol{K}_t$ multiplied with $\tilde{\mathbf{X}}_t$. This $m \times q$ matrix $\boldsymbol{K}_t$ is called *Kálmán gain matrix* after R. E. Kálmán (in fact, this notation was first used in the paper [4] of Kálmán and Bucy).

To specify the matrix $\boldsymbol{K}_t$, we have to write $\tilde{\mathbf{X}}_t$ in terms of $\hat{\boldsymbol{\xi}}_t$ and $\mathbf{X}_t$. For this purpose, let us project both sides of the first observation equation of (3.1), i.e., of $\mathbf{X}_t = \boldsymbol{C}\boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_t$, onto $H_{t-1}(\mathbf{X})$. We get that

$$\overline{\mathbf{X}}_t = \boldsymbol{C}\hat{\boldsymbol{\xi}}_t.$$

Taking the orthogonal decomposition (3.3) of $\mathbf{X}_t$ into consideration yields that

$$\tilde{\mathbf{X}}_t = \mathbf{X}_t - \overline{\mathbf{X}}_t = \mathbf{X}_t - \boldsymbol{C}\hat{\boldsymbol{\xi}}_t. \tag{3.6}$$

We substitute this into the last line of Equation (3.5) and obtain that

$$\boldsymbol{B}\hat{\boldsymbol{\eta}}_t = \boldsymbol{A}\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t\tilde{\mathbf{X}}_t = (\boldsymbol{A} - \boldsymbol{K}_t\boldsymbol{C})\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t\mathbf{X}_t.$$

With the notation

$$\boldsymbol{A}_t^* = \boldsymbol{A} - \boldsymbol{K}_t\boldsymbol{C} \tag{3.7}$$

for the updated transition matrix, we get the new linear dynamics:

$$\boldsymbol{B}\hat{\boldsymbol{\eta}}_t = \boldsymbol{A}_t^*\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t\mathbf{X}_t. \tag{3.8}$$

We also have the alternative expression

$$\boldsymbol{B}\hat{\boldsymbol{\eta}}_t = \boldsymbol{A}\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t\tilde{\mathbf{X}}_t = \boldsymbol{A}\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t(\mathbf{X}_t - \boldsymbol{C}\hat{\boldsymbol{\xi}}_t). \tag{3.9}$$

The estimation error is also governed by the linear dynamical system. This error term has two alternative forms. Using (3.8), the one is

$$\begin{aligned}\boldsymbol{B}\tilde{\boldsymbol{\eta}}_t = \boldsymbol{B}\boldsymbol{\eta}_t - \boldsymbol{B}\hat{\boldsymbol{\eta}}_t &= \boldsymbol{A}\boldsymbol{\xi}_t + \boldsymbol{\zeta}_t - \boldsymbol{A}^*\hat{\boldsymbol{\xi}}_t - \boldsymbol{K}_t\boldsymbol{C}\boldsymbol{\xi}_t - \boldsymbol{K}_t\boldsymbol{\varepsilon}_t\\ &= \boldsymbol{A}_t^*(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_t) + \boldsymbol{\zeta}_t - \boldsymbol{K}_t\boldsymbol{\varepsilon}_t = \boldsymbol{A}_t^*\tilde{\boldsymbol{\xi}}_t + \boldsymbol{\zeta}_t - \boldsymbol{K}_t\boldsymbol{\varepsilon}_t.\end{aligned}$$

Then, using (3.9), the other is

$$\boldsymbol{B}\tilde{\boldsymbol{\eta}}_t = \boldsymbol{A}\boldsymbol{\xi}_t + \boldsymbol{\zeta}_t - \boldsymbol{A}\hat{\boldsymbol{\xi}}_t - \boldsymbol{K}_t(\mathbf{X}_t - \boldsymbol{C}\hat{\boldsymbol{\xi}}_t) = \boldsymbol{A}\tilde{\boldsymbol{\xi}}_t + \boldsymbol{\zeta}_t - \boldsymbol{K}_t(\mathbf{X}_t - \boldsymbol{C}\hat{\boldsymbol{\xi}}_t).$$

From here, we get the following recursion for the covariance matrix

$$\boldsymbol{P}_t = \mathbb{E}\tilde{\boldsymbol{\xi}}_t\tilde{\boldsymbol{\xi}}_t^T \tag{3.10}$$

of the optimal error (of predicting $\boldsymbol{\xi}_t$) and so, of $\boldsymbol{K}_t$:

$$\begin{aligned}\boldsymbol{B}[\mathbb{E}\tilde{\boldsymbol{\eta}}_t\tilde{\boldsymbol{\eta}}_t^T]\boldsymbol{B}^T &= \mathbb{E}[\boldsymbol{B}\tilde{\boldsymbol{\eta}}_t][\boldsymbol{B}\tilde{\boldsymbol{\eta}}_t]^T \\ &= \mathbb{E}[\boldsymbol{A}^*\tilde{\boldsymbol{\xi}}_t + \boldsymbol{\zeta}_t][\boldsymbol{A}\tilde{\boldsymbol{\xi}}_t + \boldsymbol{\zeta}_t - \boldsymbol{K}_t(\mathbf{X}_t - \boldsymbol{C}\hat{\boldsymbol{\xi}}_t)]^T \\ &= \boldsymbol{A}_t^*\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q},\end{aligned} \tag{3.11}$$

where recall that $\boldsymbol{Q} = \mathbb{E}\boldsymbol{\zeta}_t\boldsymbol{\zeta}_t^T$, obtainable by (2.1). We used that $\boldsymbol{\zeta}_t$ is uncorrelated with $\boldsymbol{\xi}_t$ and, therefore, with $\tilde{\boldsymbol{\xi}}_t$ too. We also used that $\mathbf{X}_t - \boldsymbol{C}\hat{\boldsymbol{\xi}}_t$ is in $\boldsymbol{I}_t(\mathbf{X})$, and $\boldsymbol{\zeta}_t$ is uncorrelated with $\boldsymbol{\varepsilon}_t$.

It remains to find an explicit formula for $\boldsymbol{K}_t$, and thus, also for $\boldsymbol{A}_t^*$. Recall that $\boldsymbol{K}_t$ is the matrix of the linear operation $\mathrm{Proj}_{I_t(\mathbf{X})}\boldsymbol{B}\boldsymbol{\eta}_t$, therefore by the projection principle (see Lemma 2.1):

$$\boldsymbol{K}_t = [\mathbb{E}\boldsymbol{B}\boldsymbol{\eta}_t\tilde{\mathbf{X}}_t^T][\mathbb{E}(\tilde{\mathbf{X}}_t\tilde{\mathbf{X}}_t^T]^+,$$

where $^+$ denotes the Moore–Penrose generalized inverse (we use regular inverse if the underlying matrix is invertible).

Now we calculate the matrices in brackets. By the third equation of (3.1), that extends to $\tilde{\mathbf{X}}_t = \boldsymbol{C}\tilde{\boldsymbol{\xi}}_t + \boldsymbol{\varepsilon}_t$ and to their predictions, we get that

$$\mathbb{E}\tilde{\mathbf{X}}_t\tilde{\mathbf{X}}_t^T = \mathbb{E}(\boldsymbol{C}\tilde{\boldsymbol{\xi}}_t + \boldsymbol{\varepsilon}_t)(\boldsymbol{C}\tilde{\boldsymbol{\xi}}_t + \boldsymbol{\varepsilon}_t)^T = \boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E},$$

where $\boldsymbol{E} = \mathbb{E}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T$. $\boldsymbol{E}$ is obtainable by (2.1) in the following way:

$$\mathbb{E}\mathbf{X}_t\mathbf{X}_t^T = \boldsymbol{C}(\mathbb{E}\boldsymbol{\xi}_t\boldsymbol{\xi}_t^T)\boldsymbol{C}^T + \boldsymbol{E} = \boldsymbol{C}\boldsymbol{F}\boldsymbol{C}^T + \boldsymbol{E}.$$

So $\boldsymbol{E}$ is the difference between $\mathbb{E}\mathbf{X}_t\mathbf{X}_t^T$ (estimated as $\hat{\boldsymbol{\Sigma}}_{\mathbf{XX}}$ from the training sample) and $\boldsymbol{C}\boldsymbol{F}\boldsymbol{C}^T$, where $\boldsymbol{F}$ is the inverse of the second diagonal block of $\boldsymbol{D}$ in (2.1).

By the first and third equation of (3.1) and the orthogonality of $\hat{\boldsymbol{\xi}}_t$ and $\tilde{\boldsymbol{\xi}}_t$ we get that

$$\mathbb{E}(\boldsymbol{B}\boldsymbol{\eta}_t\tilde{\mathbf{X}}_t^T) = \boldsymbol{A}\mathbb{E}(\boldsymbol{\xi}_t\tilde{\mathbf{X}}_t^T) = \boldsymbol{A}\mathbb{E}[(\hat{\boldsymbol{\xi}}_t + \tilde{\boldsymbol{\xi}}_t)(\boldsymbol{C}\tilde{\boldsymbol{\xi}}_t)^T] = \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{C}^T. \tag{3.12}$$

Therefore,

$$\boldsymbol{K}_t = \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{C}^T[\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E}]^+ \tag{3.13}$$

with the Moore–Penrose inverse.

With this matrix $\boldsymbol{K}_t$ of Equation (3.13) and using Equation (3.11), we are able to write the error covariance matrix in the form of a symmetric matrix:

$$\begin{aligned}\boldsymbol{B}[\mathbb{E}\tilde{\boldsymbol{\eta}}_t\tilde{\boldsymbol{\eta}}_t^T]\boldsymbol{B}^T &= \boldsymbol{A}^*\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q} = (\boldsymbol{A} - \boldsymbol{K}_t\boldsymbol{C})\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q} \\ &= \left(\boldsymbol{A} - \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{C}^T[\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E}]^+\boldsymbol{C}\right)\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q} \\ &= \boldsymbol{A}\left(\boldsymbol{I} - \boldsymbol{P}_t\boldsymbol{C}^T[\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E}]^+\boldsymbol{C}\right)\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q} \\ &= \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{A}^T - \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{C}^T[\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E}]^+\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q},\end{aligned}$$

so

$$\boldsymbol{B}\boldsymbol{P}_t^*\boldsymbol{B}^T = \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{A}^T - \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{C}^T[\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E}]^+\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q}, \tag{3.14}$$

where $\boldsymbol{P}_t^* = \mathbb{E}(\tilde{\boldsymbol{\eta}}_t\tilde{\boldsymbol{\eta}}_t^T)$ is the covariance matrix of the error when predicting $\boldsymbol{\eta}_t$. In the next stage, we use it to find $\boldsymbol{P}_{t+1}$.

3.2. Second stage: $\mathbf{Y}_t \to \hat{\boldsymbol{\xi}}_{t+1}$

For $t \geq 1$, let $H_{t-1}(\mathbf{Y}) = \mathrm{Span}\,(\mathbf{Y}_0, \dots, \mathbf{Y}_{t-1})$ consists of the linear combinations of all the components of $\mathbf{Y}_0, \dots, \mathbf{Y}_{t-1}$ over a common probability space. We denote the optimal prediction of $\boldsymbol{\eta}_t$ based on $\mathbf{Y}_0, \dots, \mathbf{Y}_{t-1}$ by $\check{\boldsymbol{\eta}}_t$.

If $\mathbf{Y}_0, \dots, \mathbf{Y}_{t-1}$ are observed, i.e., $H_{t-1}(\mathbf{Y})$ is known, then the newly observed (measured) $\mathbf{Y}_t$ can be orthogonally decomposed as

$$\mathbf{Y}_t = \mathrm{Proj}_{H_{t-1}(\mathbf{Y})}\mathbf{Y}_t + \tilde{\mathbf{Y}}_t = \overline{\mathbf{Y}}_t + \tilde{\mathbf{Y}}_t, \tag{3.15}$$

where the orthogonal component $\tilde{\mathbf{Y}}_t \in I_t(\mathbf{Y})$, and $I_t(\mathbf{Y})$ is the innovation subspace (actually, the components of $\tilde{\mathbf{Y}}_t$ generate $I_t(\mathbf{Y})$). Assume that $I_t(\mathbf{Y})$ is not the sole $\mathbf{0}$ vector, otherwise observing $\mathbf{Y}_t$ does not give any additional information to $H_{t-1}(\mathbf{Y})$.

Equation (3.15) implies the decomposition of the corresponding subspaces like

$$H_t(\mathbf{Y}) = H_{t-1}(\mathbf{Y}) \oplus I_t(\mathbf{Y}). \tag{3.16}$$

Assume that we have already found $\check{\boldsymbol{\eta}}_t$. We shall give a recursion to find $\check{\boldsymbol{\xi}}_{t+1}$ by using the new value of $\mathbf{Y}_t$. In view of Equation (3.16):

$$\begin{aligned}
\boldsymbol{U}\check{\boldsymbol{\xi}}_{t+1} &= \mathrm{Proj}_{H_t(\mathbf{Y})}(\boldsymbol{U}\boldsymbol{\xi}_{t+1}) = \mathrm{Proj}_{H_{t-1}(\mathbf{Y})}(\boldsymbol{U}\boldsymbol{\xi}_{t+1}) + \mathrm{Proj}_{I_t(\mathbf{Y})}(\boldsymbol{U}\boldsymbol{\xi}_{t+1}) \\
&= \boldsymbol{V}\mathrm{Proj}_{H_{t-1}(\mathbf{Y})}\boldsymbol{\eta}_t + \mathrm{Proj}_{H_{t-1}(\mathbf{Y})}\gamma_t + \boldsymbol{M}_t\tilde{\mathbf{Y}}_t \\
&= \boldsymbol{V}\check{\boldsymbol{\eta}}_t + \boldsymbol{M}_t\tilde{\mathbf{Y}}_t,
\end{aligned} \tag{3.17}$$

where we utilized that $\gamma_t \perp H_{t-1}(\mathbf{Y})$, Lemma 2.1 and the second state equation of (3.1). Furthermore, we refer to the linearity of the projection, see Lemma 2.2. Since $\mathrm{Proj}_{I_t(\mathbf{Y})}\boldsymbol{U}\boldsymbol{\xi}_{t+1}$ is the linear combination of the coordinates of the vector $\tilde{\mathbf{Y}}_t \in I_t(\mathbf{Y})$, its effect can be written as a matrix $\boldsymbol{M}_t$ multiplied with $\tilde{\mathbf{Y}}_t$. This $n \times p$ matrix $\boldsymbol{M}_t$ is another gain matrix.

To specify the matrix $\boldsymbol{M}_t$, we have to write $\tilde{\mathbf{Y}}_t$ in terms of $\check{\boldsymbol{\eta}}_t$ and $\mathbf{Y}_t$. For this purpose, let us project both sides of the second observation equation of (3.1), i.e., of $\mathbf{Y}_t = \boldsymbol{G}\boldsymbol{\eta}_t + \boldsymbol{\delta}_t$, onto $H_{t-1}(\mathbf{Y})$. We get that

$$\overline{\mathbf{Y}}_t = \boldsymbol{G}\check{\boldsymbol{\eta}}_t.$$

Taking the orthogonal decomposition (3.15) of $\mathbf{Y}_t$ into consideration yields that

$$\tilde{\mathbf{Y}}_t = \mathbf{Y}_t - \overline{\mathbf{Y}}_t = \mathbf{Y}_t - \boldsymbol{G}\check{\boldsymbol{\eta}}_t. \tag{3.18}$$

We substitute this into the last line of Equation (3.17) and obtain that

$$\boldsymbol{U}\check{\boldsymbol{\xi}}_{t+1} = \boldsymbol{V}\check{\boldsymbol{\eta}}_t + \boldsymbol{M}_t\tilde{\mathbf{Y}}_t = (\boldsymbol{V} - \boldsymbol{M}_t\boldsymbol{G})\check{\boldsymbol{\eta}}_t + \boldsymbol{M}_t\mathbf{Y}_t.$$

With the notation

$$\boldsymbol{V}_t^* = \boldsymbol{V} - \boldsymbol{M}_t\boldsymbol{G} \tag{3.19}$$

for the updated transition matrix, we get the new linear dynamics:

$$\boldsymbol{U}\check{\boldsymbol{\xi}}_{t+1} = \boldsymbol{V}_t^*\check{\boldsymbol{\eta}}_t + \boldsymbol{M}_t\mathbf{Y}_t. \tag{3.20}$$

We also have the alternative expression

$$\boldsymbol{U}\check{\boldsymbol{\xi}}_{t+1} = \boldsymbol{V}\check{\boldsymbol{\eta}}_t + \boldsymbol{M}_t\tilde{\mathbf{Y}}_t = \boldsymbol{V}\check{\boldsymbol{\eta}}_t + \boldsymbol{M}_t(\mathbf{Y}_t - \boldsymbol{G}\check{\boldsymbol{\eta}}_t). \tag{3.21}$$

The estimation error is also governed by the linear dynamical system. This error term has two alternative forms. Using (3.20), the one is

$$\begin{aligned}\boldsymbol{U}\breve{\boldsymbol{\xi}}_{t+1} &= \boldsymbol{U}\boldsymbol{\xi}_{t+1} - \boldsymbol{U}\check{\boldsymbol{\xi}}_{t+1} = \boldsymbol{V}\boldsymbol{\eta}_t + \boldsymbol{\gamma}_t - \boldsymbol{V}_t^*\check{\boldsymbol{\eta}}_t - \boldsymbol{M}_t\boldsymbol{G}\boldsymbol{\eta}_t - \boldsymbol{M}_t\boldsymbol{\delta}_t\\ &= \boldsymbol{V}_t^*(\boldsymbol{\eta}_t - \check{\boldsymbol{\eta}}_t) + \boldsymbol{\gamma}_t - \boldsymbol{M}_t\boldsymbol{\delta}_t = \boldsymbol{V}_t^*\breve{\boldsymbol{\eta}}_t + \boldsymbol{\gamma}_t - \boldsymbol{M}_t\boldsymbol{\delta}_t.\end{aligned}$$

Then, using (3.21), the other is

$$\boldsymbol{U}\breve{\boldsymbol{\xi}}_{t+1} = \boldsymbol{V}\boldsymbol{\eta}_t + \boldsymbol{\gamma}_t - \boldsymbol{V}\check{\boldsymbol{\eta}}_t - \boldsymbol{M}_t(\mathbf{Y}_t - \boldsymbol{G}\check{\boldsymbol{\eta}}_t) = \boldsymbol{V}\breve{\boldsymbol{\eta}}_t + \boldsymbol{\gamma}_t - \boldsymbol{M}_t(\mathbf{Y}_t - \boldsymbol{M}_t\check{\boldsymbol{\eta}}_t).$$

From here, we get the following recursion for the covariance matrix

$$\boldsymbol{P}_t^* = \mathbb{E}\breve{\boldsymbol{\eta}}_t\breve{\boldsymbol{\eta}}_t^T \tag{3.22}$$

of the optimal error (of predicting $\boldsymbol{\eta}_t$) and so, of $\boldsymbol{M}_t$:

$$\begin{aligned}\boldsymbol{U}[\mathbb{E}\breve{\boldsymbol{\xi}}_{t+1}\breve{\boldsymbol{\xi}}_{t+1}^T]\boldsymbol{U}^T &= \mathbb{E}[\boldsymbol{U}\breve{\boldsymbol{\xi}}_{t+1}][\boldsymbol{U}\breve{\boldsymbol{\xi}}_{t+1}]^T\\ &= \mathbb{E}[\boldsymbol{V}^*\breve{\boldsymbol{\eta}}_t + \boldsymbol{\gamma}_t - \boldsymbol{M}_t\boldsymbol{\delta}_t][\boldsymbol{V}\breve{\boldsymbol{\eta}}_t + \boldsymbol{\gamma}_t - \boldsymbol{M}_t(\mathbf{Y}_t - \boldsymbol{G}\check{\boldsymbol{\eta}}_t)]^T\\ &= \boldsymbol{V}_t^*\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R},\end{aligned} \tag{3.23}$$

where recall that $\boldsymbol{R} = \mathbb{E}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^T$, obtainable by (3.2), and we used that $\boldsymbol{\gamma}_t$ is uncorrelated with $\boldsymbol{\eta}_t$ and, therefore, with $\breve{\boldsymbol{\eta}}_t$ too; further, $\boldsymbol{V}_t^* = \boldsymbol{V} - \boldsymbol{M}_t\boldsymbol{G}$. We also used that $\mathbf{Y}_t - \boldsymbol{G}\check{\boldsymbol{\eta}}_t$ is in $\boldsymbol{I}_t(\mathbf{Y})$, and that $\boldsymbol{\gamma}_t$ is uncorrelated with $\boldsymbol{\delta}_t$.

Now an explicit formula is found for $\boldsymbol{M}_t$, and thus, also for $\boldsymbol{V}_t^*$. Recall that $\boldsymbol{M}_t$ is the matrix of the linear operation $\mathrm{Proj}_{I_t(\mathbf{Y})}\boldsymbol{U}\boldsymbol{\xi}_{t+1}$, therefore by the projection principle (see Lemma 2.1):

$$\boldsymbol{M}_t = [\mathbb{E}(\boldsymbol{U}\boldsymbol{\xi}_{t+1}\tilde{\mathbf{Y}}_t^T)][\mathbb{E}(\tilde{\mathbf{Y}}_t\tilde{\mathbf{Y}}_t^T]^+,$$

where $^+$ denotes the Moore–Penrose generalized inverse (we use regular inverse if the underlying matrix is invertible). We calculate the matrices in brackets. By the last equation of (3.1), that extends to $\tilde{\mathbf{Y}}_t = \boldsymbol{G}\tilde{\boldsymbol{\eta}}_t + \boldsymbol{\delta}_t$ and to their predictions, we get that

$$\mathbb{E}(\tilde{\mathbf{Y}}_t\tilde{\mathbf{Y}}_t^T) = \mathbb{E}[(\boldsymbol{G}\breve{\boldsymbol{\eta}}_t + \boldsymbol{\delta}_t)(\boldsymbol{G}\breve{\boldsymbol{\eta}}_t + \boldsymbol{\delta}_t)^T = \boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T + \boldsymbol{\Delta},$$

where $\boldsymbol{\Delta} = \mathbb{E}\boldsymbol{\delta}_t\boldsymbol{\delta}_t^T$. $\boldsymbol{\Delta}$ is obtainable by (3.2) in the following way:

$$\mathbb{E}\mathbf{Y}_t\mathbf{Y}_t^T = \boldsymbol{G}(\mathbb{E}\boldsymbol{\eta}_t\boldsymbol{\eta}_t^T)\boldsymbol{G}^T + \Delta = \boldsymbol{G}\boldsymbol{F}^*\boldsymbol{G}^T + \boldsymbol{\Delta}.$$

So $\boldsymbol{\Delta}$ is the difference between $\mathbb{E}\mathbf{Y}_t\mathbf{Y}_t^T$ (estimated as $\hat{\boldsymbol{\Sigma}}_{\mathbf{YY}}$ from the training sample) and $\boldsymbol{G}\boldsymbol{F}^*\boldsymbol{G}^T$, where $\boldsymbol{F}^*$ is the inverse of the second diagonal block of $\boldsymbol{D}^*$ in (3.2).

By the second and fourth equation of (3.1) and the orthogonality of $\check{\boldsymbol{\eta}}_t$ and $\breve{\boldsymbol{\eta}}_t$ we get that

$$\mathbb{E}(\boldsymbol{U}\boldsymbol{\xi}_{t+1}\tilde{\mathbf{Y}}_t^T) = \boldsymbol{V}\mathbb{E}(\boldsymbol{\eta}_t\tilde{\mathbf{Y}}_t^T) = \boldsymbol{V}\mathbb{E}[(\check{\boldsymbol{\eta}}_t + \breve{\boldsymbol{\eta}}_t)(\boldsymbol{G}\breve{\boldsymbol{\eta}}_t)^T] = \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{G}^T. \tag{3.24}$$

Therefore,

$$\boldsymbol{M}_t = \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{G}^T[\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T + \boldsymbol{\Delta}]^+ \tag{3.25}$$

with the Moore–Penrose inverse.

With this matrix $\boldsymbol{M}_t$ of Equation (3.25) and using Equation (3.23), we are able to write the error covariance matrix in the form of a symmetric matrix:

$$\begin{aligned}\boldsymbol{U}[\mathbb{E}\check{\boldsymbol{\xi}}_{t+1}\check{\boldsymbol{\xi}}_{t+1}^T]\boldsymbol{U}^T &= \boldsymbol{V}_t^*\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{\Delta} = (\boldsymbol{V} - \boldsymbol{M}_t\boldsymbol{G})\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R}\\ &= (\boldsymbol{V} - \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{G}^T[\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T + \boldsymbol{\Delta}]^+\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R}\\ &= \boldsymbol{V}(\boldsymbol{I} - \boldsymbol{P}_t^*\boldsymbol{G}^T)[\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T + \boldsymbol{\Delta}]^+\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R}\\ &= \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{V}^T - \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{G}^T[\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T + \boldsymbol{\Delta}]^+\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R},\end{aligned}$$

so

$$\boldsymbol{U}\boldsymbol{P}_{t+1}\boldsymbol{U}^T = \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{V}^T - \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{G}^T[\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T + \boldsymbol{\Delta}]^+\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R}, \tag{3.26}$$

where we assumed that the error covariance matrix of $\tilde{\boldsymbol{\xi}}_t$ and $\check{\boldsymbol{\xi}}_t$, akin to that of $\tilde{\boldsymbol{\eta}}_t$ and $\check{\boldsymbol{\eta}}_t$ is the same. This fact gives rise to a recursion by connecting (3.14) and (3.26).

Finally, with (3.9) and (3.21) we are able to recursively estimate the latent state variables. During the $\boldsymbol{P}_1 \to \boldsymbol{P}_1^* \to \boldsymbol{P}_2 \to \boldsymbol{P}_2^* \dots$ recursion, from $\boldsymbol{P}_t$, we find $\boldsymbol{K}_t$ by (3.13) and $\hat{\boldsymbol{\eta}}_t$ by (3.9). Then, from $\boldsymbol{P}_t^*$, we find $\boldsymbol{M}_t$ by (3.25) and $\check{\boldsymbol{\xi}}_{t+1}$ by (3.21).

As for the relation between $\hat{\boldsymbol{\xi}}_t$ and $\check{\boldsymbol{\xi}}_t$, akin to that between $\hat{\boldsymbol{\eta}}_t$ and $\check{\boldsymbol{\eta}}_t$, we can estimate their cross-covariance matrices from the training sample, and then, linearly predict $\check{\boldsymbol{\eta}}_t$ with $\hat{\boldsymbol{\eta}}_t$ and linearly predict $\check{\boldsymbol{\xi}}_t$ with $\hat{\boldsymbol{\xi}}_t$ by Lemma 2.1 as follows:

$$\check{\boldsymbol{\eta}}_t = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}\mathbf{Y}}\hat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^+\hat{\boldsymbol{\Sigma}}_{\mathbf{YX}}\hat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^+\boldsymbol{\Sigma}_{\mathbf{X}\boldsymbol{\eta}}[\mathbb{E}\hat{\boldsymbol{\eta}}_t\hat{\boldsymbol{\eta}}_t^T]^+\hat{\boldsymbol{\eta}}_t$$

and

$$\hat{\boldsymbol{\xi}}_{t+1} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}\mathbf{X}}\hat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^+\hat{\boldsymbol{\Sigma}}_{\mathbf{XY}}\hat{\boldsymbol{\Sigma}}_{\mathbf{YY}}^+\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}\boldsymbol{\xi}}[\mathbb{E}\check{\boldsymbol{\xi}}_{t+1}\check{\boldsymbol{\xi}}_{t+1}^T]^+\check{\boldsymbol{\xi}}_{t+1}.$$

Here, from Equation (3.11), we conclude that

$$\mathbb{E}\hat{\boldsymbol{\eta}}_t\hat{\boldsymbol{\eta}}_t^T = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}\boldsymbol{\eta}} - \boldsymbol{B}^{-1}[(\boldsymbol{A} - \boldsymbol{K}_t\boldsymbol{C})\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q}](\boldsymbol{B}^{-1})^T.$$

Likewise, from Equation (3.23), we conclude that

$$\mathbb{E}\check{\boldsymbol{\xi}}_t\check{\boldsymbol{\xi}}_t^T = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}\boldsymbol{\xi}} - \boldsymbol{U}^{-1}[(\boldsymbol{V} - \boldsymbol{M}_t\boldsymbol{G})\boldsymbol{P}_t^*\boldsymbol{V}_t^T + \boldsymbol{R}](\boldsymbol{U}^{-1})^T.$$

3.3. The main result

We assume that the system was at rest until time 0. The parameter matrices are estimated from the past, whereas newer and newer estimates for the latent variables are given, as observations arrive at time t ($t = 1, 2, \dots$ up to the end of the experimental time T). Thus, we can summarize the results in the subsequent theorem. It is important that in the derivation of the formulas we used the best linear prediction theory of Hilbert spaces.

Theorem 3.1. *In the linear dynamical system (*3.1*), the optimal estimate $\hat{\boldsymbol{\eta}}_t$ of $\boldsymbol{\eta}_t$ and $\check{\boldsymbol{\xi}}_{t+1}$ of $\boldsymbol{\xi}_{t+1}$ given $\mathbf{X}_1, \dots, \mathbf{X}_t$ and $\mathbf{Y}_1, \dots, \mathbf{Y}_t$ is generated by the new linear dynamical system*

$$\boldsymbol{B}\hat{\boldsymbol{\eta}}_t = \boldsymbol{A}_t^*\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t\mathbf{X}_t$$

and

$$\boldsymbol{U}\check{\boldsymbol{\xi}}_{t+1} = \boldsymbol{V}_t^*\check{\boldsymbol{\eta}}_t + \boldsymbol{M}_t\mathbf{Y}_t.$$

The expected quadratic losses are $tr\boldsymbol{P}_t^*$ and $tr\boldsymbol{P}_{t+1}$, where $\boldsymbol{P}_t^*$ and $\boldsymbol{P}_{t+1}$ are the propagated covariance matrices of the estimation errors. The minimizing matrices and the *one-step ahead predictions* $\hat{\boldsymbol{\eta}}_t$ *and* $\hat{\boldsymbol{\xi}}_{t+1}$ *together with the error covariance matrices* $\boldsymbol{P}_t^*$ *and* $\boldsymbol{P}_{t+1}$ *are uniquely determined by the initial conditions*

$$\hat{\boldsymbol{\xi}}_1 = \mathrm{Proj}_{\mathbf{X}_0}\boldsymbol{\xi}_1, \quad \tilde{\boldsymbol{\xi}}_1 = \boldsymbol{\xi}_1 - \hat{\boldsymbol{\xi}}_1, \quad \boldsymbol{P}_1 = \mathbb{E}\tilde{\boldsymbol{\xi}}_1\tilde{\boldsymbol{\xi}}_1^T$$

and the recursions for $t = 1, 2, \ldots$ *as follows.*

$$\begin{aligned}
\boldsymbol{K}_t &= \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{C}^T[\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E}]^+\\
\hat{\boldsymbol{\eta}}_t &= \boldsymbol{B}^{-1}[\boldsymbol{A}\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t(\mathbf{X}_t - \boldsymbol{C}\hat{\boldsymbol{\xi}}_t)]\\
\hat{\mathbf{Y}}_t &= \boldsymbol{G}\hat{\boldsymbol{\eta}}_t\\
\check{\boldsymbol{\eta}}_t &= \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}\mathbf{X}}\hat{\boldsymbol{\Sigma}}^+_{\mathbf{XX}}\hat{\boldsymbol{\Sigma}}_{\mathbf{XY}}\hat{\boldsymbol{\Sigma}}^+_{\mathbf{YY}}\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}\boldsymbol{\eta}}\left[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta\eta}} - \boldsymbol{B}^{-1}((\boldsymbol{A} - \boldsymbol{K}_t\boldsymbol{C})\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q})(\boldsymbol{B}^{-1})^T\right]^+\hat{\boldsymbol{\eta}}_t\\
\boldsymbol{P}_t^* &= \boldsymbol{B}^{-1}[\boldsymbol{A}\boldsymbol{P}_t\boldsymbol{A}^T - \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{C}^T[\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E}]^+\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q}]\boldsymbol{B}^{-1^T}\\
\boldsymbol{M}_t &= \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{G}^T[\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T + \boldsymbol{\Delta}]^+\\
\check{\boldsymbol{\xi}}_{t+1} &= \boldsymbol{U}^{-1}[\boldsymbol{V}\check{\boldsymbol{\eta}}_t + \boldsymbol{M}_t(\mathbf{Y}_t - \boldsymbol{G}\check{\boldsymbol{\eta}}_t)]\\
\hat{\boldsymbol{\xi}}_{t+1} &= \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}\mathbf{Y}}\hat{\boldsymbol{\Sigma}}^+_{\mathbf{YY}}\hat{\boldsymbol{\Sigma}}_{\mathbf{YX}}\hat{\boldsymbol{\Sigma}}^+_{\mathbf{XX}}\hat{\boldsymbol{\Sigma}}_{\mathbf{X}\boldsymbol{\xi}}\left[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi\xi}} - \boldsymbol{U}^{-1}((\boldsymbol{V} - \boldsymbol{M}_t\boldsymbol{G})\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R})(\boldsymbol{U}^{-1})^T\right]^+\check{\boldsymbol{\xi}}_{t+1}\\
\hat{\mathbf{X}}_{t+1} &= \boldsymbol{C}\hat{\boldsymbol{\xi}}_{t+1}\\
\boldsymbol{P}_{t+1} &= \boldsymbol{U}^{-1}[\boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{V}^T - \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{G}^T[\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T + \boldsymbol{\Delta}]^+\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R}]\boldsymbol{U}^{-1^T},
\end{aligned}$$

where $^+$ *denotes the Moore–Penrose generalized inverse (usual inverse if the matrix is invertible).*

Note that $\hat{\boldsymbol{\xi}}_1 = \mathrm{Proj}_{\mathbf{X}_0}\boldsymbol{\xi}_1 = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}\mathbf{X}}\hat{\boldsymbol{\Sigma}}^+_{\mathbf{XX}}\mathbf{X}_0$, by Lemma 2.1, where the last training sample entry can be chosen for $\mathbf{X}_0$. To initialize $\boldsymbol{P}_1$, the whole training sample can be used: if the L learning sample entries are indexed by ℓ, then $\hat{\boldsymbol{\xi}}_\ell = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}\mathbf{X}}\hat{\boldsymbol{\Sigma}}^+_{\mathbf{XX}}\mathbf{X}_\ell$ and $\tilde{\boldsymbol{\xi}}_\ell = \boldsymbol{\xi}_\ell - \hat{\boldsymbol{\xi}}_\ell$, where $\boldsymbol{\xi}_\ell$ is the ℓth case estimate of $\boldsymbol{\xi}$, based on the forthcoming PLS algorithm in Section 4. Finally, the product moment estimate of $\boldsymbol{P}_1$ is $\frac{1}{L}\sum_{\ell=1}^{L}\tilde{\boldsymbol{\xi}}_\ell\tilde{\boldsymbol{\xi}}_\ell^T$ if the variables have zero expectation (otherwise, the sample means should be subtracted).

Note that the matrices $\mathbf{U}, \mathbf{V}$ are also estimated from the learning sample with the shifted product moments, as discussed in the next section.

4. Application

Using data from three Egyptian villages, we applied our proposed algorithm to examine and predict to what extent parental views affect their daughters' thinking on two empowerment issues. Figure 1 visualizes the hypothesized outer and inner relations. The inner model examines the cause-effect structure between the latent variables (LVs): parental views on girls' participation in decision making and girls' mobility (exogenous: P-DM=ξ^1 and P-Mob=ξ^2) and the daughters' views on the same issues (endogenous: G-DM=η^1 and G-Mob=η^2), respectively. The outer model links LVs and observed variables (OVs) together. Mode A is used to construct all LVs in the model. To clarify, ξ^1 is composed of four independent xs (OVs) that measure parental views on girl's responsibility in making decisions related to marriage, choosing a husband, entering and continuing schooling. ξ^2 is linked to three xs that ask parents whether girls can go alone to places such as market, field, and friends' home. In the same manner, η^1 and η^2 are connected to the same number of OVs, but the dependent ys that reflect the daughters' views on the same indicators. Note, all the OVs are on the same ordinal scale.

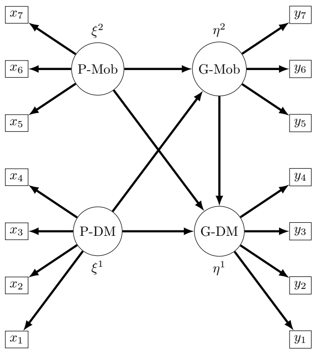


FIGURE 1. Schematic representation of the SEM that considers the effect of parental views on girls' views related to empowerment issues.

A total sample of 349 parents and their daughters are considered. Prior to the analysis, the data were standardized to have zero mean and unit variance. Then, it was divided randomly into a training sample of size 279 and a test sample of the remaining cases. We apply our proposed estimation algorithm on the training sample to obtain the specified parameter matrices ($\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{G}$, $\boldsymbol{Q}$, $\boldsymbol{F}$, $\boldsymbol{U}$, $\boldsymbol{V}$, $\boldsymbol{R}$, and $\boldsymbol{F}^*$). These matrices are used in the derivation of the proposed prediction recursion algorithm. Specifically, the filtering technique in Theorem 3.1 predicts the future values for the test sample observations that come sequentially.

Recall that our integrated estimation algorithm to obtain the model specified matrices combines the first stage of Wold's PLS technique and the block Cholesky decomposition of Kiiveri et al. The detailed explanation follows:

1. The first stage uses stage I of Wold's PLS algorithm, in which the outer relations and the LV case values are obtained from the training sample. It is an iterative process that consists of the following steps:
   i. Initialize the LV scores for each case as the weighted sum of the observed indicators in the block that correspond to each LV:
   $$\boldsymbol{H} = \boldsymbol{N}\boldsymbol{Z},$$
   where $\boldsymbol{H}$ is the exogenous and endogenous LV scores matrix, $\boldsymbol{N}$ is the training sample data matrix of size 279×14, and $\boldsymbol{Z}$ is the 14×4 adjacency matrix of the measurement model. The entries z_{kj} are ones, if the indicator n_{kj} belongs to the block that defines the corresponding LV; and zeros, otherwise. After each step, the LV scores are standardized.
   ii. Update the obtained matrix $\boldsymbol{H}$ with the inner weights
   $$\tilde{\boldsymbol{H}} = \boldsymbol{H}\boldsymbol{W},$$
   where $\boldsymbol{W}$ is the LV inner weights matrix which is computed for each LV to indicate how strong it is connected to the other LVs in the model. There

are three schemes to obtain these weights, for more details, see [7] and [8]. We used the centroid scheme.

iii. Use the obtained LV scores $\tilde{\boldsymbol{H}}$ to estimate the outer relations (loadings/weights). There are two modes of constructing the measurement model:
   - Mode A (reflective), where the arrows point outward from the LV to the OVs, as in our case, see Figure 1. The outer scores are called loadings (λ_{kj}) and estimated by OLS simple linear regression between each observed indicator of the measurement block and the corresponding LV score $\tilde{h}_j$.
   - Mode B (formative), where the arrows point inward from the observed variables to the corresponding LVs. The outer estimated scores called weights (w_{kj}) and are calculated by OLS multiple linear regression in which each LV score $\tilde{h}_j$ is regressed on all the observed indicators of the corresponding block.

   All the outer estimates λ_{kj}s and w_{kj}s are collected in the updated weight matrix $\hat{\boldsymbol{W}}$.

iv. Using the obtained weight matrix $\hat{\boldsymbol{W}}$ to update the LV scores

$$\boldsymbol{H} = \boldsymbol{N}\hat{\boldsymbol{W}},$$

where $\boldsymbol{H}$ contains the LV scores of the last iteration process.

This stage iterates sequentially from Step i. to Step iv. until convergence. At the convergence, the final LV case values $\boldsymbol{H}$ are obtained as well as the estimated outer matrices $\boldsymbol{C}$ and $\boldsymbol{G}$

$$\boldsymbol{C} = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{array} \begin{array}{c} \begin{array}{cc} \xi^1 & \xi^2 \end{array} \\ \begin{pmatrix} .6759 & 0 \\ .7477 & 0 \\ .7739 & 0 \\ .7093 & 0 \\ 0 & .5326 \\ 0 & .8217 \\ 0 & .7356 \end{pmatrix} \end{array}, \quad \boldsymbol{G} = \begin{array}{c} \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{array} \begin{array}{c} \begin{array}{cc} \eta^1 & \eta^2 \end{array} \\ \begin{pmatrix} .6881 & 0 \\ .7408 & 0 \\ .7574 & 0 \\ .7287 & 0 \\ 0 & .4929 \\ 0 & .8132 \\ 0 & .7206 \end{pmatrix} \end{array}.$$

The matrices $\boldsymbol{C}$ and $\boldsymbol{G}$ contain the loadings that link the latent vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with $\mathbf{X}$ and $\mathbf{Y}$, respectively.

2. The second stage runs the block Cholesky decomposition of Kiiveri et al. two times on the obtained LV case values. The first decomposition is applied on the inverse of the product moment of the covariance matrix of the LV final scores $\boldsymbol{\Sigma}_{\boldsymbol{H}}^{-1}$ that is obtained at the convergence of the Wold algorithm, see Equation (2.1). The resulting block matrix $\boldsymbol{L}$ gives the estimated path coefficient matrices of the inner relations,

$$\boldsymbol{B} = \begin{array}{c} \\ \eta^1 \\ \eta^2 \end{array} \begin{array}{c} \begin{array}{cc} \eta^1 & \eta^2 \end{array} \\ \begin{pmatrix} 1 & .107 \\ 0 & 1 \end{pmatrix} \end{array}, \quad \boldsymbol{A} = \begin{array}{c} \\ \eta^1 \\ \eta^2 \end{array} \begin{array}{c} \begin{array}{cc} \xi^1 & \xi^2 \end{array} \\ \begin{pmatrix} .337 & .156 \\ -.068 & .577 \end{pmatrix} \end{array};$$

whereas, the block matrix $\boldsymbol{D}$ yields the error covariance matrices

$$\boldsymbol{Q} = cov(\boldsymbol{\zeta}) = \begin{pmatrix} .869 & 0 \\ 0 & .663 \end{pmatrix}, \quad \boldsymbol{F} = cov(\boldsymbol{\xi}) = \begin{pmatrix} 1 & -.024 \\ -.024 & 1 \end{pmatrix}.$$

The second decomposition is performed on the inverse of the product moments of the shifted LV score pairs $\boldsymbol{\Sigma}^{-1}_{\boldsymbol{H}_{s,s+1}}$, see Equation (3.2). The resulting block matrix $\boldsymbol{L}^*$ contains

$$\boldsymbol{U} = \begin{array}{c} \\ \xi^1 \\ \xi^2 \end{array} \begin{array}{c} \begin{array}{cc} \xi^1 & \xi^2 \end{array} \\ \begin{pmatrix} 1 & .014 \\ 0 & 1 \end{pmatrix} \end{array}, \quad \boldsymbol{V} = \begin{array}{c} \\ \xi^1 \\ \xi^2 \end{array} \begin{array}{c} \begin{array}{cc} \eta^1 & \eta^2 \end{array} \\ \begin{pmatrix} .037 & -.074 \\ .051 & .145 \end{pmatrix} \end{array};$$

while, the matrix $\boldsymbol{D}^*$ gives the error covariance matrices

$$\boldsymbol{R} = cov(\boldsymbol{\gamma}) = \begin{pmatrix} .992 & 0 \\ 0 & .984 \end{pmatrix}, \quad \boldsymbol{F}^* = cov(\boldsymbol{\eta}) = \begin{pmatrix} 1 & -.041 \\ -.041 & 1 \end{pmatrix}.$$

At this point, the specified parameter matrices are obtained from the training sample. The estimated matrices $\boldsymbol{C}$ and $\boldsymbol{G}$ show the loadings of each OV on the corresponding LV. The matrix $\boldsymbol{B}$ displays the extent to which girls' views on mobility affect her views on participating in making decisions; while $\boldsymbol{A}$ shows how parental views on girls' mobility and decision making influence their daughters' opinions on these issues. There is a direct effect of parental views on their daughters' opinions in the same domain, i.e., parents who reported a conservative view on girls' participation in making decisions tend to lead their daughters to think alike. The same scenario is true for mobility, where daughters tend to reproduce their parents' views. On the contrary, the effect is small when we consider parental opinions of one domain on their daughters' views of the other domain.

As new cases come one by one at a time sequence ($t = 1, 2, \dots\ T$), instead of re-running the estimation algorithm, we give a recursion to predict the LV case values for the new observation. To do so, the estimated parameter matrices based on the training sample and the Kálmán filtering technique will be used. Theorem 3.1 discusses the recursion from which the optimal prediction of the latent case values is obtained along with the covariance matrix of the prediction error. Specifically, the prediction of $\hat{\boldsymbol{\eta}}_t$ utilizes the estimated $\hat{\boldsymbol{\xi}}_t$ and the new observation $\mathbf{X}_t$, while the new $\mathbf{Y}_t$ and the obtained $\hat{\boldsymbol{\eta}}_t$ are necessary to find $\hat{\boldsymbol{\xi}}_{t+1}$. To start the recursion, the first propagated matrix $\boldsymbol{P}_1$ ought to be initialized from the training sample. Then, the Kálmán gain matrices $\boldsymbol{K}_t$ and $\boldsymbol{M}_t$ are obtained. The succession of calculations follows the order:

$$\boldsymbol{P}_t \to \boldsymbol{K}_t \to \hat{\boldsymbol{\eta}}_t \to \check{\boldsymbol{\eta}}_t \to \boldsymbol{P}^*_t \to \boldsymbol{M}_t \to \check{\boldsymbol{\xi}}_{t+1} \to \hat{\boldsymbol{\xi}}_{t+1} \to \hat{\mathbf{X}}_{t+1} \to \boldsymbol{P}_{t+1}.$$

For $t = 1$, we show the results of the highlighted matrices of the recursion as they are derived in Section 3.1 and Section 3.2.
First stage: $\mathbf{X}_1 \to \hat{\boldsymbol{\eta}}_1$

$$\boldsymbol{P}_1 = \begin{array}{c} \begin{array}{cc} \xi^1 & \xi^2 \end{array} \\ \begin{pmatrix} .007 & .005 \\ .005 & .003 \end{pmatrix} \end{array},$$

$$\boldsymbol{K}_1 = \begin{array}{c} \begin{array}{ccccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \end{array} \\ \begin{pmatrix} -.018 & -.036 & -.036 & -.022 & .022 & .064 & .051 \\ -.013 & -.025 & -.025 & -.015 & .015 & .045 & .036 \end{pmatrix} \end{array},$$

$$\hat{\boldsymbol{\eta}}_1 = \begin{array}{c} \begin{array}{cc} \eta^1 & \eta^2 \end{array} \\ \begin{pmatrix} .601 & -.377 \end{pmatrix} \end{array}, \quad \boldsymbol{P}_1^* = \begin{array}{c} \begin{array}{cc} \eta^1 & \eta^2 \end{array} \\ \begin{pmatrix} .878 & -.070 \\ -.070 & .664 \end{pmatrix} \end{array}.$$

where $\hat{\boldsymbol{\eta}}_1$ show the predicted case values based on the information $\mathbf{X}_1$ of the new observation; and $\boldsymbol{P}_1^*$ is the covariance matrix of the prediction error of $\hat{\boldsymbol{\eta}}_1$.
Second stage: $\mathbf{Y}_1 \to \hat{\boldsymbol{\xi}}_2$

$$\boldsymbol{M}_1 = \begin{array}{c} \begin{array}{ccccccc} y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 \end{array} \\ \begin{pmatrix} .005 & .016 & .019 & .009 & -.022 & -.042 & -.034 \\ .032 & .011 & .007 & .018 & .041 & .086 & .064 \end{pmatrix} \end{array},$$

$$\hat{\boldsymbol{\xi}}_2 = \begin{array}{c} \begin{array}{cc} \xi^1 & \xi^2 \end{array} \\ \begin{pmatrix} -.595 & -1.460 \end{pmatrix} \end{array}, \quad \boldsymbol{P}_2 = \begin{array}{c} \begin{array}{cc} \xi^1 & \xi^2 \end{array} \\ \begin{pmatrix} .993 & -.014 \\ -.014 & .985 \end{pmatrix} \end{array}.$$

where $\hat{\boldsymbol{\xi}}_2$ presents the estimated case values for the exogenous LVs at $t = 2$ based on the new information $\mathbf{Y}_1$. Then the covariance matrix of the prediction error $\hat{\boldsymbol{\xi}}_2$ is obtained. From this we can calculate the propagation matrix $\boldsymbol{P}_2$ to start the recursion once again at $t = 2$ for the next new observation.

The root mean square error ($RMSE$) statistic measures the prediction errors. For the test sample of size 70 observations, we compared the predicted LV case values that are obtained from the Wold algorithm and the filtering technique, simultaneously. Table 1 shows the values of $RSME$. It indicates that the prediction capability of the filtering technique and that of the Wold algorithm are quite homogeneous. Moreover, the difference (in Frobenius norm) between the error covariance and gain matrices in the tth and $(t+1)$th consecutive steps of the recursion are displayed in Table 2. This shows that these matrices are stabilized after the first few steps.

TABLE 1. RMSE for the predictions of the test sample.

Test Obs. at	Wold Prediction (W)				Filtering Prediction (F)			
Sequence "t"	$\hat{\xi}_t^1$	$\hat{\xi}_t^2$	$\hat{\eta}_t^1$	$\hat{\eta}_t^2$	$\hat{\xi}_t^1$	$\hat{\xi}_t^2$	$\hat{\eta}_t^1$	$\hat{\eta}_t^2$
1	-.44005	-.28108	-.25305	-.46325	-.33149	-.52816	-.32865	-.14788
2	.68377	.28108	.84937	-.46325	.89402	.30331	.44740	-.34648
3	.68377	.28108	.84937	-.46325	.35488	.18442	.46531	-.36156
4	.59157	1.50271	.18313	1.18603	.59467	2.43277	.73569	2.05939
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
70	-.53202	-.28108	-.69213	-.46325	-.47413	-.54895	-.59782	-.31266
$RMSE$:	$\sqrt{\frac{1}{T}\sum_{t=1}^{T}(W(LV_t) - F(LV_t))^2}$				.3075	.4033	.3117	.3229

TABLE 2. Consecutive norm for gain and propagation of predictions.

Sequence	Frobenius Norm $\|\|.\|\|_F$			
"t"	$\boldsymbol{P}_{t+1} - \boldsymbol{P}_t$	$\boldsymbol{K}_{t+1} - \boldsymbol{K}_t$	$\boldsymbol{P}^*_{t+1} - \boldsymbol{P}^*_t$	$\boldsymbol{M}_{t+1} - \boldsymbol{M}_t$
1	-	-	-	-
2	1.391595	.5148828	.008475387	8.145129e-5
3	7.044589e-7	7.491063e-9	1.094965e-10	1.138605e-12
4	9.341999e-15	1.487013e-16	4.388542e-17	1.357636e-17
5	1.734723e-18	0	0	0
6	0	0	0	0
⋮	⋮	⋮	⋮	⋮
70	0	0	0	0

In sum, the numerical results show a good performance of our proposed algorithm. Once the specified matrices are obtained from the training sample, the Kálmán filtering technique yields an optimal prediction for the LV case values along with the error covariance matrices for the test sample.

5. Discussion and Conclusion

It should be emphasized that the PLS method of Wold is applicable to a given sample, where estimates for the endogeneus variables are given through the exogenous latent ones, and the case values of the LVs are also estimated. The algorithm uses a lot of OLS regressions and so, the estimation of the coefficient matrices is time demanding akin to the block Cholesky decomposition we use. This is the case when we have a long time series with small time intervals or data when the observations come frequently in subsequent order. Our point is that for the new observations, we need not to repeat the whole estimation procedure to obtain the model parameters, but instead we can update the latent variable scores with the help of the new observable data, the estimated matrices, and the Kálmán filtering technique.

In this way, an artificial intelligence is developed. The parameter matrices are estimated from a training sample at the beginning, and the latent variable scores are estimated as observable variables arrive one by one from the test sample. Moreover, there is no need for any distribution assumptions and the data are not necessarily independent. It should be noted that in the possession of a stationary time series, the matrix sequences $\boldsymbol{K}_t$ and $\boldsymbol{M}_t$ (as $t \to \infty$) tend to fixed points of an iteration finding the solution of a matrix Riccati equation (see [4]), but this is the topic of a further research.

Acknowledgement. The research reported in this paper which was carried out at the Budapest University of Technology and Economics has been supported by the National Research Development and Innovation Fund based on the charter of bolster issued by the National Research Development and Innovation Office under the auspices of the Ministry for Innovation and Technology; also supported by the National Research, Development and Innovation Fund (TUDFO/51757/2019-ITM, Thematic Excellence Program). The research was suported by the following project too:

EFOP-3.6.2-16-2017-00015, HU-MATHS-IN for deepening the activity of the Hungarian Industrial and Innovation Network.

References

[1] Haavelmo, T., *The statistical implications of a system of simultaneous equations*, Econometrica, **11**(1943), 1-12.

[2] Jöreskog, K.G., *Structural equation models in the social sciences specification, estimation and testing*, In: "Applications of Statistics", (P.R. Krishnaiah, Ed), North-Holland Publishing Co., 1977, 265-287.

[3] Kálmán, R.E., *A new approach to linear filtering and prediction problems*, Trans. ASME J. Basic. Eng., **82D**(1960), 35-45.

[4] Kálmán, R.E., Bucy, R.S., *New results in linear filtering and prediction theory*, Trans. Amer. Soc. Mech. Eng., J. Basic Eng., **83**(1961), 95-108.

[5] Kiiveri, H., Speed, T.P., Carlin, J.B., *Recursive casual models*, J. Aust. Math. Soc., **36**(1984), 30-52.

[6] Rao, C.R., *Linear Statistical Inference and its Applications*, Wiley (1973).

[7] Tenenhaus, M., Esposito Vinzi, V., Chatelinc, Y-L., Lauro, C., *PLS path modeling*, Comput. Statist. Data Anal., **48.1**(2005), 159-205.

[8] Wold, H., *Partial least squares*, In: "Encyclopedia of Statistical Sciences", (Kotz, S., Johnson, N. L., Ed), Wiley, New York, **6**(1985), 581-591.

Marianna Bolla
Budapest University of Technology and Economics,
Institute of Mathematics,
Műegyetem rkp. 3. Budapest 1111, Hungary
e-mail: `marib@math.bme.hu`

Fatma Abdelkhalek
Budapest University of Technology and Economics,
Institute of Mathematics,
Műegyetem rkp. 3. Budapest 1111, Hungary
Also at Faculty of Commerce,
Assiut University, Egypt
e-mail: `fatma@math.bme.hu`

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 197–209
DOI: 10.24193/subbmath.2021.1.16

Permanent solutions for some motions of UCM fluids with power-law dependence of viscosity on the pressure

Constantin Fetecau and Abdul Rauf

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. Steady motion of two types of incompressible Maxwell fluids with power-law dependence of viscosity on the pressure is analytically studied between infinite horizontal parallel plates when the gravity effects are taken into consideration. Simple and exact expressions are established for the permanent components of starting solutions corresponding to two oscillatory motions induced by the lower plate that oscillates in its plane. Such solutions are very important for the experimentalists who want to eliminate the transients from their experiments. The similar solutions for the simple Couette flow of the same fluids, as well as the permanent solutions corresponding to ordinary incompressible Maxwell fluids performing the same motions, are obtained as limiting cases of general solutions. The convergence of starting solutions to their permanent components as well as the influence of physical parameters on the fluid motion is graphically underlined and discussed.

Mathematics Subject Classification (2010): 76A05.

Keywords: Permanent solutions, UCM fluids, Pressure-dependent viscosity.

1. Introduction

Generally, in isothermal processes, the fluid viscosity depends on the flow conditions and particularly on the shear-rate and pressure. The first who remarked that the fluid viscosity could depend on the pressure was Stokes in his seminal work [21]. He also delineated motions in which the viscosity can be considered constant. Later, experiments by Barus [2], Bridgman [3], Griest et al. [9], Bair et al. [1], Prusa et al. [16] and so on certified this dependence. A linear dependence of viscosity on the pressure was proposed by Barus [2] for low to medium pressure difference. At large pressure differences, power or exponential law seems to be more suitable. In any case, exact solutions for steady (permanent) or unsteady motions of non-Newtonian fluids with the pressure-dependent viscosity are lack in the existing literature excepting those of Housiadas [10, 11] corresponding to a linear law of viscosity on the pressure.

In the following, we shall determine simple closed form solutions for the permanent velocity fields corresponding to some unsteady motions of Maxwell fluids with power-law dependence of the viscosity on the pressure. It is worth to point out the fact that high pressure differences appear in many engineering applications such as polymer processing operations [5], fluid film lubrication [22], microfluidics [4], pharmaceutical tablet manufacturing, food processing and geophysics [20]. Techniques for measuring the pressure dependent viscosity and of the pressure-viscosity coefficient can be found in the work of Goubert et al. [8], respectively Park et al. [14]. An extensive literature regarding experimental and theoretical studies on these fluids was provided by Malek and Rajagopal [13].

On the other hand, the gravity effects are important in many flows of the fluids with practical applications. They are stronger in the case of fluid motions in which the gravity acts along the direction in which the pressure varies. The first exact solutions for steady motions of the Newtonian fluids with pressure-dependent viscosity are those of Rajagopal [17, 18] between parallel plates or over an inclined plane due to the gravity. Exact expressions in terms of the Kelvin functions have been established by Prusa [15] for the solutions of modified Stokes problems of the same fluids. General solutions for the same motions of the incompressible Newtonian fluids with power-law or exponential dependence of viscosity on the pressure have been determined by Rajagopal et al. [19] and Fetecau and Vieru [7].

In the present work we provide the first exact and simple expressions for permanent solutions corresponding to two oscillatory motions of incompressible Maxwell fluids with exponential dependence of viscosity on the pressure between infinite horizontal parallel plates. The fluid motion is generated by the lower plate that oscillates in its plane. The similar solutions for the simple Couette flow of the same fluids as well as some known solutions for the Newtonian fluids performing the same motions are obtained as limiting cases of general solutions. Such solutions are important for the experimentalists who want to eliminate the transients from their experiments. In addition, they can be also used as tests to verify different numerical methods that are used to study complex flow problems. Finally, the convergence of starting solutions (numerical solutions) to their permanent components, as well as the effect of physical parameters on the fluid motion, is graphically underlined and discussed.

2. Statement of the problem

Let us consider an incompressible upper-convected Maxwell (UCM) fluid with power-law dependence of viscosity on the pressure at rest between two infinite horizontal parallel plates at the distance d apart. Its constitutive equations, as it results from [12], are

$$\mathbf{T} = -p\mathbf{I} + \mathbf{S}, \quad \mathbf{S} + \lambda\frac{\delta\mathbf{S}}{\delta t} = \eta(p)\mathbf{A}$$
$$\text{with} \;\; tr\mathbf{A} = 0 \;\; \text{and} \;\; \frac{\delta\mathbf{S}}{\delta t} = \frac{d\mathbf{S}}{dt} - \mathbf{LS} - \mathbf{SL}^{\mathrm{T}}, \tag{2.1}$$

where $\mathbf{T}$ is the Cauchy stress tensor, $\mathbf{S}$ the extra-stress tensor, $\mathbf{I}$ is the unit tensor, $\mathbf{A}$ is the first Rivlin-Ericksen tensor, $\mathbf{L}$ is the gradient of the velocity vector $\mathbf{u}$, p is the Lagrange multiplier, λ is the relaxation time and $\eta(p)$ is the fluid viscosity which depends on the pressure.

In the next, we determine exact expressions for the permanent (steady-state or long-time) solutions corresponding to some motions of the incompressible UCM fluids with power-law dependence of viscosity on the pressure of the forms

$$\eta(p) = \mu[\alpha(p - p_0) + 1]^2 \quad \text{or} \quad \eta(p) = \mu[\alpha(p - p_0) + 1]^{1/2}, \tag{2.2}$$

where μ is the fluid viscosity at the reference pressure p_0 and the positive constant α is the dimensional pressure-viscosity coefficient. The gravitational effects will be also taken into consideration. If $\alpha = 0$, $\eta(p) = \mu$ and Eqs. (2.1) correspond to the ordinary UCM fluids. If both α and λ are zero, the governing equations (2.1) correspond to incompressible ordinary Newtonian fluids.

At the moment $t = 0^+$ the lower plate begins to oscillate in its plane according to

$$\mathbf{u} = U\cos(\omega t)\mathbf{i} \quad \text{or} \quad \mathbf{u} = U\sin(\omega t)\mathbf{i}, \tag{2.3}$$

where $\mathbf{i}$ is the unit vector along the x-direction of a suitable Cartesian coordinate system x, y and z while U and ω are the amplitude, respectively the frequency of the oscillations. Due to the shear the fluid is gradually moved and we are looking for a solution of the form [15, 19, 12]

$$\mathbf{u} = u(y,t)\mathbf{i}, \quad p = p(y). \tag{2.4}$$

Using Eqs. (2.4) in (2.1) and introducing the obtained results in the balance of linear momentum, we get the following relevant partial or ordinary differential equations [12]

$$\tau(y,t) + \lambda\frac{\partial\tau(y,t)}{\partial t} = \eta(p)\frac{\partial u(y,t)}{\partial y}, \quad \rho\frac{\partial u(y,t)}{\partial t} = \frac{\partial\tau(y,t)}{\partial y},$$
$$\frac{dp(y)}{dy} + \rho g = 0; \; 0 < y < d, \;\; t > 0, \tag{2.5}$$

where $\tau(y,t)$ is the non-trivial shear stress, ρ is the fluid density and g is the gravitational acceleration. The incompressibility condition is automatically satisfied and Eq. $(2.5)_3$ implies

$$p(y) = \rho g(d - y) + p_0 \quad \text{with} \quad p_0 = p(d). \tag{2.6}$$

Eliminating $\tau(y,t)$ between Eqs. $(2.5)_1$ and $(2.5)_2$ and bearing in mind the expressions of $\eta(p)$ and $p(y)$ from Eqs. (2.2) and (2.6), we find the next governing equations

$$\mu[\alpha\rho g(d-y)+1]^2\frac{\partial^2 u(y,t)}{\partial y^2} - 2\mu\alpha\rho g[\alpha\rho g(d-y)+1]\frac{\partial u(y,t)}{\partial y}$$
$$= \rho\left[\lambda\frac{\partial^2 u(y,t)}{\partial t^2} + \frac{\partial u(y,t)}{\partial t}\right]; \quad 0 < y < d, \;\; t > 0, \tag{2.7}$$

$$\mu\sqrt{\alpha\rho g(d-y)+1}\frac{\partial^2 u(y,t)}{\partial y^2}-\frac{\mu\alpha\rho g}{2\sqrt{\alpha\rho g(d-y)+1}}\frac{\partial u(y,t)}{\partial y}$$
$$=\rho\left[\lambda\frac{\partial^2 u(y,t)}{\partial t^2}+\frac{\partial u(y,t)}{\partial t}\right];\quad 0<y<d,\ t>0, \tag{2.8}$$

for the velocity field $u(y,t)$ corresponding to the motion of the two types of UCM fluids with power-law dependence of viscosity on the pressure between infinite horizontal parallel plates.

The appropriate initial and boundary conditions are given by the relations

$$u(y,0)=0,\quad \left.\frac{\partial u(y,t)}{\partial t}\right|_{t=0}=0 \ \text{ if } \ 0\le y\le d, \tag{2.9}$$

$$u(0,t)=U\cos(\omega t)\ \text{ or }\ u(0,t)=U\sin(\omega t),\ \ u(d,t)=0\ \text{ for }\ t>0. \tag{2.10}$$

Introducing the following non-dimensional variables, functions and parameters

$$y^*=\frac{y}{d},\ \ t^*=\frac{\nu t}{d^2},\ \ u^*=\frac{u}{U},\quad \tau^*=\frac{\tau d}{\mu U},\ \ \omega^*=\frac{d^2}{\nu}\omega,\ \ \alpha^*=\alpha\rho g d, \tag{2.11}$$

where $\nu=\mu/\rho$ is the kinematic viscosity of the fluid and dropping out the star notation, we obtain the following two dimensionless initial and boundary value problems

$$[\alpha(1-y)+1]^2\frac{\partial^2 u(y,t)}{\partial y^2}-2\alpha[\alpha(1-y)+1]\frac{\partial u(y,t)}{\partial y}$$
$$=\mathrm{We}\frac{\partial^2 u(y,t)}{\partial t^2}+\frac{\partial u(y,t)}{\partial t};\quad 0<y<1,\ \ t>0, \tag{2.12}$$

$$u(y,0)=\left.\frac{\partial u(y,t)}{\partial t}\right|_{t=0}=0,\ \ 0\le y\le 1;$$
$$u(0,t)=\cos(\omega t)\ \text{or}\ \sin(\omega t),\ \ u(1,t)=0\ \text{for}\ t>0, \tag{2.13}$$

and

$$\sqrt{\alpha(1-y)+1}\frac{\partial^2 u(y,t)}{\partial y^2}-\frac{\alpha}{2\sqrt{\alpha(1-y)+1}}\frac{\partial u(y,t)}{\partial y}$$
$$=\mathrm{We}\frac{\partial^2 u(y,t)}{\partial t^2}+\frac{\partial u(y,t)}{\partial t};\quad 0<y<1,\ \ t>0, \tag{2.14}$$

$$u(y,0)=\left.\frac{\partial u(y,t)}{\partial t}\right|_{t=0}=0,\ \ 0\le y\le 1;$$
$$u(0,t)=\cos(\omega t)\ \text{or}\ \sin(\omega t),\ \ u(1,t)=0\ \text{for}\ t>0, \tag{2.15}$$

where $\mathrm{We}=\lambda\nu/d^2=\lambda/(d^2/\nu)$ is the Weissenberg number (the ratio of the relaxation time of the fluid and a characteristic time scale).

3. Solution

In the following, in order to avoid confusion, we shall denote by $u_c(y,t)$ and $u_s(y,t)$ the starting solutions of the dimensionless initial and boundary value problems characterized by Eqs. (2.12), (2.13) or (2.14), (2.15). Generally, these solutions can be written as sums of their permanent and transient components, namely

$$u_c(y,t)=u_{cp}(y,t)+u_{ct}(y,t),\quad u_s(y,t)=u_{sp}(y,t)+u_{st}(y,t). \tag{3.1}$$

Some time after the motion initiation, the fluid moves according to starting solutions. After this time when the transients disappear or can be neglected, its movement is characterized by the permanent solutions $u_{cp}(y,t)$ or $u_{sp}(y,t)$. In practice, the experimentalists want to know the required time to reach the permanent state. This is the time after which the fluid flows according to the permanent solutions. In order to determine it, at least the permanent solutions corresponding to a given motion have to be known. This is the reason that, in this section, we shall determine closed form expressions for these solutions only.

3.1. Case $\eta(p) = \mu[\alpha\rho g(d-y)+1]^2$

To determine both permanent solutions in the same time, we use the complex velocity

$$u_p(y,t) = u_{cp}(y,t) + iu_{sp}(y,t), \tag{3.2}$$

where i is the imaginary unit. This velocity has to satisfy the partial differential equation

$$\begin{aligned} &[\alpha(1-y)+1]^2\frac{\partial^2 u_p(y,t)}{\partial y^2} - 2\alpha[\alpha(1-y)+1]\frac{\partial u_p(y,t)}{\partial y} \\ &= \mathrm{We}\frac{\partial^2 u_p(y,t)}{\partial t^2} + \frac{\partial u_p(y,t)}{\partial t}; \quad 0<y<1,\ \ t\in R, \end{aligned} \tag{3.3}$$

with the boundary conditions

$$u_p(0,t) = \mathrm{e}^{i\omega t}, \quad u_p(1,t) = 0; \quad t\in R. \tag{3.4}$$

Making the following suitable change of the spatial variable

$$y = \frac{\alpha+1-\mathrm{e}^r}{\alpha} \quad \text{or equivalently} \quad r = \ln[\alpha(1-y)+1], \tag{3.5}$$

Eq. (3.3) takes the simpler form

$$\begin{aligned} &\alpha^2\left[\frac{\partial^2 u_p(r,t)}{\partial r^2} + \frac{\partial u_p(r,t)}{\partial r}\right] = \mathrm{We}\frac{\partial^2 u_p(r,t)}{\partial t^2} + \frac{\partial u_p(r,t)}{\partial t}; \\ &0<r<a,\ \ t>0, \end{aligned} \tag{3.6}$$

where $a = \ln(\alpha+1)$. The corresponding boundary conditions are

$$u_p(0,t) = 0, \quad u_p(a,t) = \mathrm{e}^{i\omega t}; \quad t\in R. \tag{3.7}$$

Following [12], we are looking for a solution of the form

$$u_p(r,t) = U(r)\mathrm{e}^{i\omega t}; \quad r\in(0,a), \quad t\in R. \tag{3.8}$$

Substituting $u_p(r,t)$ from Eq. (3.8) in (3.6), we find that the unknown function $U(r)$ has to satisfy the following boundary value problem

$$\alpha^2[U''(r)+U'(r)] + \omega(\omega\mathrm{We}-i)U(r) = 0; \quad U(0)=0, \ \ U(a)=1. \tag{3.9}$$

The solution of this boundary value problem is given by the next equality

$$U(r) = \frac{\mathrm{e}^{r_2 r}-\mathrm{e}^{r_1 r}}{\mathrm{e}^{r_2 a}-\mathrm{e}^{r_1 a}}; \quad r_{1,2} = \frac{-1\pm\sqrt{1-4\omega(\omega\mathrm{We}-i)/\alpha^2}}{2}. \tag{3.10}$$

Consequently, the complex velocity field $u_p(y,t)$ is given by

$$u_p(y,t)=\frac{[\alpha(1-y)+1]^{r_2}-[\alpha(1-y)+1]^{r_1}}{(\alpha+1)^{r_2}-(\alpha+1)^{r_1}}e^{i\omega t};\ \ 0<y<1,\ \ t>0, \tag{3.11}$$

and the dimensionless permanent velocities $u_{cp}(y,t)$ and $u_{sp}(y,t)$ have the expressions

$$u_{cp}(y,t)=\mathrm{Re}\left\{\frac{[\alpha(1-y)+1]^{r_2}-[\alpha(1-y)+1]^{r_1}}{(\alpha+1)^{r_2}-(\alpha+1)^{r_1}}\mathrm{e}^{i\omega\, t}\right\}, \tag{3.12}$$

$$u_{sp}(y,t)=\mathrm{Im}\left\{\frac{[\alpha(1-y)+1]^{r_2}-[\alpha(1-y)+1]^{r_1}}{(\alpha+1)^{r_2}-(\alpha+1)^{r_1}}\mathrm{e}^{i\omega\, t}\right\}, \tag{3.13}$$

where Re and Im denote the real, respectively the imaginary part of that which follows.

3.2. Case $\eta(p)=\mu[\alpha(p-p_0)+1]^{1/2}$

In this case the corresponding dimensionless complex velocity $u_p(y,t)$ has to satisfy the following partial differential equation

$$\begin{aligned}&\sqrt{\alpha(1-y)+1}\,\tfrac{\partial^2 u_p(y,t)}{\partial y^2}-\tfrac{\alpha}{2\sqrt{\alpha(1-y)+1}}\tfrac{\partial u_p(y,t)}{\partial y}\\&=\mathrm{We}\tfrac{\partial^2 u_p(y,t)}{\partial t^2}+\tfrac{\partial u_p(y,t)}{\partial t};\quad 0<y<1,\ \ t\in R,\end{aligned} \tag{3.14}$$

with the same boundary conditions (3.4). Making the change of spatial variable

$$y=\frac{\alpha+1-r^2}{\alpha}\ \ \text{ or equivalently }\ \ r=\sqrt{\alpha(1-y)+1}, \tag{3.15}$$

we attain to the next boundary value problem

$$\frac{\alpha^2}{4r}\frac{\partial^2 u_p(r,t)}{\partial r^2}=\mathrm{We}\frac{\partial^2 u_p(r,t)}{\partial t^2}+\frac{\partial u_p(r,t)}{\partial t};\quad 1<r<b,\ \ t>0, \tag{3.16}$$

where the constant $b=\sqrt{\alpha+1}$.

Looking again for a solution of the form (3.8), we find that the corresponding function $U(r)$ has to satisfy the following boundary value problem

$$\alpha^2 U''(r)+4\gamma rU(r)=0;\quad U(1)=0,\ \ U(b)=1, \tag{3.17}$$

where $\gamma=\omega(\omega\mathrm{We}-i)$. The equation (3.17) is an ordinary differential equation of Airy type whose general solution is of the form

$$U(r)=\sqrt{r}\left[C_1J_{1/3}\left(\frac{4r}{3\alpha}\sqrt{\gamma r}\right)+C_2Y_{1/3}\left(\frac{4r}{3\alpha}\sqrt{\gamma r}\right)\right], \tag{3.18}$$

where C_1 and C_2 are arbitrary constants. On the basis of the boundary conditions $(3.17)_2$ and $(3.17)_3$, it immediately results that

$$U(r)=\frac{\sqrt{r}}{\sqrt{b}}\frac{Y_{\frac13}\left(\frac{4\sqrt{\gamma}}{3\alpha}\right)J_{\frac13}\left(\frac{4r}{3\alpha}\sqrt{\gamma r}\right)-J_{\frac13}\left(\frac{4\sqrt{\gamma}}{3\alpha}\right)Y_{\frac13}\left(\frac{4r}{3\alpha}\sqrt{\gamma r}\right)}{Y_{\frac13}\left(\frac{4\sqrt{\gamma}}{3\alpha}\right)J_{\frac13}\left(\frac{4b}{3\alpha}\sqrt{\gamma b}\right)-J_{\frac13}\left(\frac{4\sqrt{\gamma}}{3\alpha}\right)Y_{\frac13}\left(\frac{4b}{3\alpha}\sqrt{\gamma b}\right)}. \tag{3.19}$$

Consequently, the dimensionless velocity fields corresponding to this problem are

$$u_{cp}(y,t)=\frac{\sqrt{r}}{\sqrt{b}}\mathrm{Re}\left\{\frac{Y_{\frac{1}{3}}\left(\frac{4\sqrt{\gamma}}{3\alpha}\right)J_{\frac{1}{3}}\left(\frac{4r}{3\alpha}\sqrt{\gamma r}\right)-J_{\frac{1}{3}}\left(\frac{4\sqrt{\gamma}}{3\alpha}\right)Y_{\frac{1}{3}}\left(\frac{4r}{3\alpha}\sqrt{\gamma r}\right)}{Y_{\frac{1}{3}}\left(\frac{4\sqrt{\gamma}}{3\alpha}\right)J_{\frac{1}{3}}\left(\frac{4b}{3\alpha}\sqrt{\gamma b}\right)-J_{\frac{1}{3}}\left(\frac{4\sqrt{\gamma}}{3\alpha}\right)Y_{\frac{1}{3}}\left(\frac{4b}{3\alpha}\sqrt{\gamma b}\right)}\mathrm{e}^{i\omega t}\right\},\tag{3.20}$$

$$u_{sp}(y,t)=\frac{\sqrt{r}}{\sqrt{b}}\mathrm{Im}\left\{\frac{Y_{\frac{1}{3}}\left(\frac{4\sqrt{\gamma}}{3\alpha}\right)J_{\frac{1}{3}}\left(\frac{4r}{3\alpha}\sqrt{\gamma r}\right)-J_{\frac{1}{3}}\left(\frac{4\sqrt{\gamma}}{3\alpha}\right)Y_{\frac{1}{3}}\left(\frac{4r}{3\alpha}\sqrt{\gamma r}\right)}{Y_{\frac{1}{3}}\left(\frac{4\sqrt{\gamma}}{3\alpha}\right)J_{\frac{1}{3}}\left(\frac{4b}{3\alpha}\sqrt{\gamma b}\right)-J_{\frac{1}{3}}\left(\frac{4\sqrt{\gamma}}{3\alpha}\right)Y_{\frac{1}{3}}\left(\frac{4b}{3\alpha}\sqrt{\gamma b}\right)}\mathrm{e}^{i\omega t}\right\},\tag{3.21}$$

where $r=\sqrt{\alpha(1-y)+1}$. These solutions, as well as those given by Eqs. (3.12) and (3.13), are independent of initial conditions $(2.15)_1$ and $(2.15)_2$ but satisfy the boundary conditions and the corresponding governing equations (2.12), respectively (2.14).

4. Limiting cases

In order to obtain the steady solutions corresponding to the simple Couette flow of the same fluids, as well as the permanent solutions for incompressible ordinary UCM fluids performing the same motions or to recover some known results from the existing literature, we shall consider in this section three special cases.

4.1. Case $\omega\to 0$ (Simple Couette flow)

Taking the limit of the permanent solution $u_{cp}(y,t)$ given by Eq. (3.12) when $\omega\to 0$ and using the asymptotic approximations

$$J_\nu(z)\approx\frac{z^\nu}{2^\nu\Gamma(\nu+1)},\quad Y_\nu(z)\approx-\frac{2^\nu\Gamma(\nu)}{\pi z^\nu}\quad\text{for}\quad \nu>0\quad\text{and}\quad z<<1,\tag{4.1}$$

for $u_{cp}(y,t)$ given by Eq. (3.20), we recover the steady solutions [[6], Eqs. (44)]

$$u_{Cp}(y)=\frac{(\alpha+1)(1-y)}{\alpha(1-y)+1},\quad\text{respectively}\quad u_{Cp}(y)=\frac{\sqrt{\alpha(1-y)+1}-1}{\sqrt{\alpha+1}-1},\tag{4.2}$$

corresponding to the simple Couette flow of the two types of incompressible Newtonian fluids with power-law dependence of viscosity on the pressure. This is not a surprise since the governing equations corresponding to steady motions of incompressible Newtonian or UCM fluids with/without pressure dependent viscosity are identical. It is important to point out the fact that the dimensionless steady solutions given by Eqs. (4.2) can be directly obtained from the corresponding boundary value problems. They correspond to the fluid motion induced by the lower plate that is moving in its plane with the constant velocity U.

4.2. Case $\alpha\to 0$ (Flows of ordinary incompressible UCM fluids)

The governing equation corresponding to unsteady motions of incompressible ordinary UCM fluids between infinite horizontal parallel plates, namely

$$\mathrm{We}\frac{\partial^2 u(y,t)}{\partial t^2}+\frac{\partial u(y,t)}{\partial t}=\frac{\partial^2 u(y,t)}{\partial y^2};\quad 0<y<1,\ \ t>0,\tag{4.3}$$

is obtained making $\alpha \to 0$ in any one of the equations (2.12) or (2.14). Consequently, the velocity fields $u_{Ocp}(y,t)$ and $u_{Osp}(y,t)$ corresponding to unsteady motions of these fluids generated by cosine, respectively sine oscillations of the lower plate can be obtained as limiting cases of the solutions $u_{cp}(y,t)$ and $u_{sp}(y,t)$ given by Eqs. (3.12), (3.13) or (3.20), (3.21). Using the well known asymptotic approximations

$$J_\nu(z) \approx \sqrt{\tfrac{2}{\pi z}} \cos\left[z - \tfrac{(2\nu+1)\pi}{4}\right],$$
$$Y_\nu(z) \approx \sqrt{\tfrac{2}{\pi z}} \sin\left[z - \tfrac{(2\nu+1)\pi}{4}\right] \quad \text{for} \quad z >> 1, \tag{4.4}$$

in Eqs. (3.20) and (3.21), it is easy to show that for small enough values of the dimensionless pressure-viscosity coefficient α

$$u_{cp}(y,t) \approx \frac{\sqrt[8]{\alpha+1}}{\sqrt[8]{\alpha(y-1)+1}} \mathrm{Re}\left\{ \frac{\sin\left\{\frac{4\sqrt{\gamma}}{3\alpha}\left[1-\sqrt[4]{[\alpha(1-y)+1]^3}\right]\right\}}{\sin\left\{\frac{4\sqrt{\gamma}}{3\alpha}\left[1-\sqrt[4]{(\alpha+1)^3}\right]\right\}} \mathrm{e}^{i\omega t}\right\}, \tag{4.5}$$

$$u_{sp}(y,t) \approx \frac{\sqrt[8]{\alpha+1}}{\sqrt[8]{\alpha(y-1)+1}} \mathrm{Im}\left\{ \frac{\sin\left\{\frac{4\sqrt{\gamma}}{3\alpha}\left[1-\sqrt[4]{[\alpha(1-y)+1]^3}\right]\right\}}{\sin\left\{\frac{4\sqrt{\gamma}}{3\alpha}\left[1-\sqrt[4]{(\alpha+1)^3}\right]\right\}} \mathrm{e}^{i\omega t}\right\}. \tag{4.6}$$

Now, using the Maclaurin series expansions for $[1+\alpha(1-y)]^{3/4}$ and $(1+\alpha)^{3/4}$ in Eqs. (4.5) and (4.6) and taking their limits when $\alpha \to 0$, we get the permanent solutions

$$u_{Ocp}(y,t) = \mathrm{Re}\left\{\tfrac{\sin\left[(1-y)\sqrt{\gamma}\right]}{\sin(\sqrt{\gamma})}\mathrm{e}^{i\omega t}\right\},$$
$$u_{Osp}(y,t) = \mathrm{Im}\left\{\tfrac{\sin\left[(1-y)\sqrt{\gamma}\right]}{\sin(\sqrt{\gamma})}\mathrm{e}^{i\omega t}\right\}, \tag{4.7}$$

corresponding to the incompressible ordinary UCM fluids performing the same motions. To the best of our knowledge, the solutions given by Eqs. (4.7) are also new in the literature. As expected, making $\alpha \to 0$ in any one of the equalities (4.2) or $\omega \to 0$ in Eq. $(4.7)_1$, the steady velocity field $u_{NCp}(y) = 1 - y$ corresponding to the simple Couette flow of incompressible ordinary Newtonian fluids is recovered.

4.3. Case We $\to 0$ (Flows of Newtonian fluids with pressure-dependent viscosity)

By now letting We $\to 0$ in all results that have been obtained in the third section, we recover the dimensionless permanent solutions corresponding to the two types of incompressible Newtonian fluids with power-law dependence of viscosity on the pressure performing the same motions. More precisely, the permanent velocity fields $u_{cp}(y,t)$ and $u_{sp}(y,t)$ given by Eqs. (3.12), (3.13) and (3.20), (3.21) take the simplified forms [6]

$$u_{Ncp}(y,t) = \mathrm{Re}\left\{\frac{[\alpha(1-y)+1]^{r_4} - [\alpha(1-y)+1]^{r_3}}{(\alpha+1)^{r_4} - (\alpha+1)^{r_3}}\mathrm{e}^{i\omega t}\right\}, \tag{4.8}$$

$$u_{Nsp}(y,t) = \mathrm{Im}\left\{ \frac{[\alpha(1-y)+1]^{r_4} - [\alpha(1-y)+1]^{r_3}}{(\alpha+1)^{r_4} - (\alpha+1)^{r_3}} \mathrm{e}^{i\omega t} \right\}, \tag{4.9}$$

respectively

$$u_{Ncp}(y,t) =$$

$$\frac{\sqrt{r}}{\sqrt{b}}\mathrm{Re}\left\{ \frac{Y_{\frac{1}{3}}\left(\frac{4\sqrt{-i\omega}}{3\alpha}\right)J_{\frac{1}{3}}\left(\frac{4r}{3\alpha}\sqrt{-i\omega r}\right) - J_{\frac{1}{3}}\left(\frac{4\sqrt{-i\omega}}{3\alpha}\right)Y_{\frac{1}{3}}\left(\frac{4r}{3\alpha}\sqrt{-i\omega r}\right)}{Y_{\frac{1}{3}}\left(\frac{4\sqrt{-i\omega}}{3\alpha}\right)J_{\frac{1}{3}}\left(\frac{4b}{3\alpha}\sqrt{-i\omega b}\right) - J_{\frac{1}{3}}\left(\frac{4\sqrt{-i\omega}}{3\alpha}\right)Y_{\frac{1}{3}}\left(\frac{4b}{3\alpha}\sqrt{-i\omega b}\right)} \mathrm{e}^{i\omega t} \right\}, \tag{4.10}$$

$$u_{Nsp}(y,t) =$$

$$\frac{\sqrt{r}}{\sqrt{b}}\mathrm{Im}\left\{ \frac{Y_{\frac{1}{3}}\left(\frac{4\sqrt{-i\omega}}{3\alpha}\right)J_{\frac{1}{3}}\left(\frac{4r}{3\alpha}\sqrt{-i\omega r}\right) - J_{\frac{1}{3}}\left(\frac{4\sqrt{-i\omega}}{3\alpha}\right)Y_{\frac{1}{3}}\left(\frac{4r}{3\alpha}\sqrt{-i\omega r}\right)}{Y_{\frac{1}{3}}\left(\frac{4\sqrt{-i\omega}}{3\alpha}\right)J_{\frac{1}{3}}\left(\frac{4b}{3\alpha}\sqrt{-i\omega b}\right) - J_{\frac{1}{3}}\left(\frac{4\sqrt{-i\omega}}{3\alpha}\right)Y_{\frac{1}{3}}\left(\frac{4b}{3\alpha}\sqrt{-i\omega b}\right)} \mathrm{e}^{i\omega t} \right\}, \tag{4.11}$$

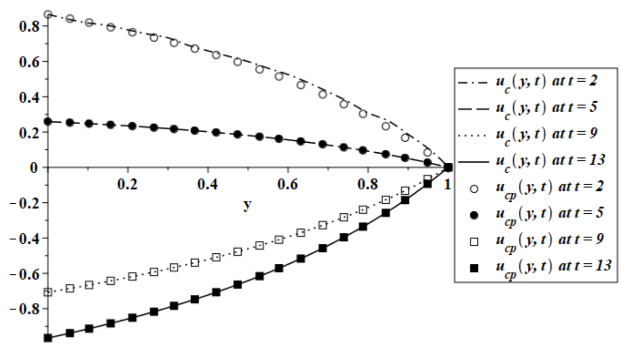
where $r_{3,4} = \frac{-1\pm\sqrt{1+4i\omega/\alpha^2}}{2}$.

5. Numerical results, discussion and conclusions

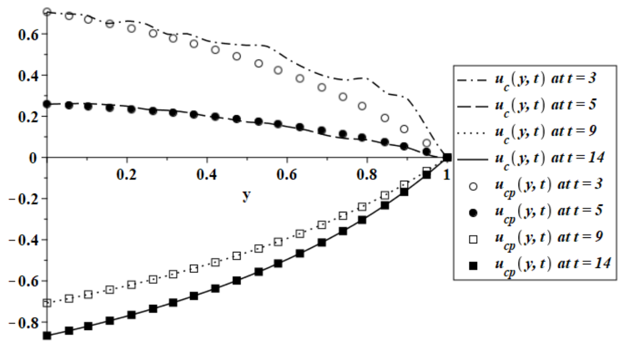
Generally, starting solutions corresponding to oscillatory motions of fluids can be presented as sums of their permanent and transient components. As it is known, the transient solutions tend to zero for increasing values of the time t. Consequently, the transients disappear in time and in practice it is important to know the time after which the fluid flows according to the permanent solution. This is the required time to reach the permanent state. To determine it, for a given motion, at least the permanent solution has to be known. This is the reason that we established here closed form expressions for the permanent solutions corresponding to some oscillatory motions of two types of incompressible UCM fluids with power-law dependence of viscosity on the pressure.

These expressions have been also used to determine the similar solutions $(4.2)_1$ and $(4.2)_2$ for the simple Couette flow of the same fluids and to provide the corresponding solutions from Eqs. (4.8)-(4.11) for ordinary incompressible UCM fluids performing the same motions. In addition, as a proof of their correctness, known solutions for the incompressible Newtonian fluids with power-law dependence of the viscosity on the pressure performing the same motions have been obtained as limiting cases of present solutions.

In order to get some physical insight of results that have been here obtained, as well as to certify their correctness, Figs 1-4 have been prepared for different values of the physical parameters and the time t. Figs. 1 and 2 clearly show that, as expected, the starting solutions $u_c(y,t)$ corresponding to the motion of the two types of incompressible UCM fluids with power-law dependence of viscosity on the pressure induced by cosine oscillations of the lower plate converge to their permanent components $u_{cp}(y,t)$ given by Eqs. (3.12) and (3.20). Furthermore, from these figures it also results that the required time to reach the permanent state diminishes for decreasing values of the Weissenberg number We. Consequently, the permanent state is rather obtained for oscillatory motions of Newtonian fluids as compared to UCM fluids with power-law dependence of viscosity on the pressure.



(A) We = 0.3

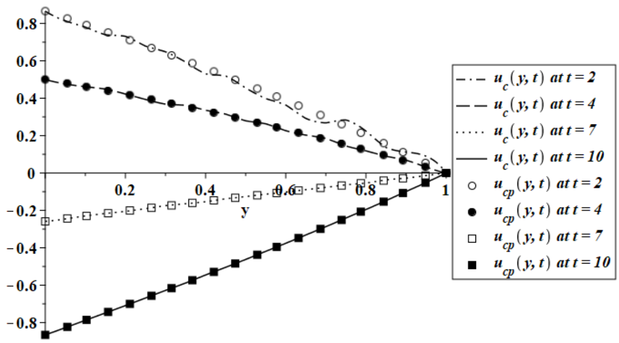


(B) We = 0.7

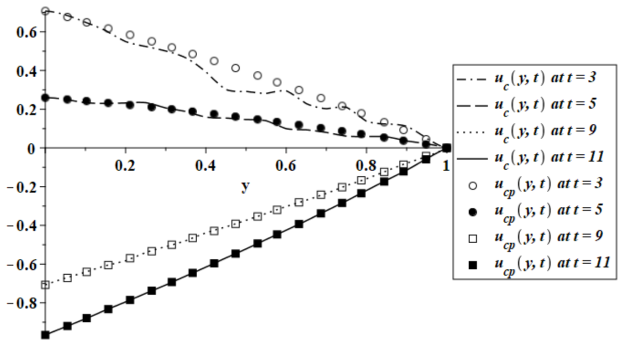
FIGURE (1) Convergence of starting solution $u_c(y,t)$ (numerical solution) to its permanent component $u_{cp}(y,t)$ given by Eq. (3.12) for $\alpha = 0.95$, $\omega = \pi/12$ and two values of We.

In the subsection 4.2 we analytically proved that, if the dimensionless pressure-viscosity coefficient $\alpha \to 0$, the permanent solutions $u_{cp}(y,t)$ and $u_{sp}(y,t)$ given by Eqs. (3.20) and (3.21) tend to the permanent solutions $u_{Ocp}(y,t)$, respectively $u_{Osp}(y,t)$ corresponding to the ordinary incompressible UCM fluids performing the same motions. For completion, as well as for the results validation, Figs. 3 have been prepared to show the convergence of the other permanents solutions $u_{cp}(y,t)$ and $u_{sp}(y,t)$ given by Eqs. (3.12) and (3.13) to $u_{Ocp}(y,t)$ and $u_{Osp}(y,t)$. In all cases, the fluid velocity smoothly decreases from maximum values on the lower plate to the zero value on the stationary wall.

The time variations of the dimensionless mid plane velocity fields $u_{cp}(0.5,t)$ and $u_{sp}(0.5,t)$ given by Eqs. (3.12) and (3.13) have been depicted in Figs. 4 for



(A) We $= 0.3$



(B) We $= 0.7$

FIGURE (2) Convergence of starting solution $u_c(y,t)$ (numerical solution) to its permanent component $u_{cp}(y,t)$ given by Eq. (3.20) for $\alpha = 0.95$, $\omega = \pi/12$ and two values of We.

$\omega = \pi/12$ and We $= 0.2$ and three different values of the pressure-viscosity coefficient α. The oscillatory specific features of the two motions are better underlined and the oscillations' amplitude diminishes for decreasing values of the parameter α. As it was to be expected, the order of magnitude of the oscillations' amplitude is the same for both motions and the phase difference is clearly observed.

References

[1] Bair, S., Jarzynski, J., Winer, W.Q., *The temperature, pressure and time dependence of lubricant viscosity*, Tribol. Int., **34**(2001), 461-468.

[2] Barus, C., *Isothermals, isopiestics and isometrics relative to viscosity*, Am. J. Sci., **45**(1893), 87-96.

[3] Bridgman, P.W., *The Physics of High Pressure*, MacMillan Company, New York, 1931.

[4] Cui, H.H., Silber-Li, Z., Zhu, S.N., *Flow characteristics of liquids in microtubes driven by a high pressure*, Phys. Fluids, **16**(2004), 1803-1810.

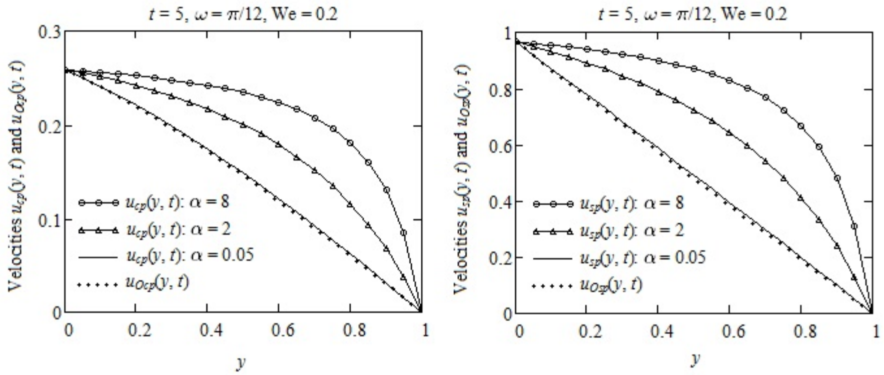


FIGURE (3) Convergence of the permanent solutions $u_{cp}(y,t)$ and $u_{sp}(y,t)$ given by Eqs. (3.12), respectively (3.13) to $u_{Ocp}(y,t)$ and $u_{Osp}(y,t)$.

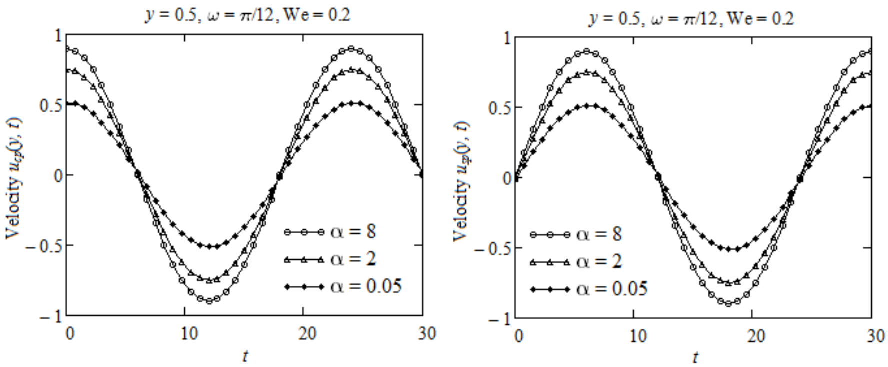


FIGURE (4) Time variation of the mid plane permanent velocities $u_{cp}(0.5,t)$ and $u_{sp}(0.5,t)$ given by Eqs. (3.12), respectively (3.13) for three values of α.

[5] Dealy, J.M., Wang, J., *Melt Rheology and its Applications in the Plastic Industry*, second ed., Springer, Dordrecht, 2013.

[6] Fetecau, C., Agop, M., *Exact solutions for oscillating motions of some fluids with power-law dependence of viscosity on the pressure*, Ann. Acad. Rom. Sci. Ser. Math. Appl., **12**(2020), 295-311.

[7] Fetecau, C., Vieru, D., *Exact solutions for unsteady motion between parallel plates of some fluids with power-law dependence of viscosity on the pressure*, Applications in Engineering Science, **1**(2020), 100003.

[8] Goubert, A., Vermout, J., Moldenaers, P., Gottfert, A., Ernst, B., *Comparison of measurement techniques for evaluating the pressure dependence of the viscosity*, Appl. Rheol., **11**(2001), 26-37.

[9] Griest, E.M. Webb, W., Schiessler, R.W., *Effect of pressure on viscosity of high hydrocarbons and their mixture*, J. Chem. Phys., **29**(1950), 711-720.

[10] Housiadas, K.D., *An exact analytical solution for viscoelastic fluids with pressure-dependent viscosity*, J. Non-Newton. Fluid Mech., **223**(2015), 147-156.

[11] Housiadas, K.D., *Internal viscoelastic flows for fluids with exponential type pressure-dependent viscosity and relaxation time*, J. Rheol., **59**(2015), 769-791.

[12] Karra, S., Prusa, V., Rajagopal, K.R., *On Maxwell fluids with relaxation time and viscosity depending on the pressure*, Int. J. Non Linear Mech., **46**(2011), 819-827.

[13] Malek, J., Rajagopal, K.R., *Mathematical properties of the equations governing the flow of fluids with pressure and shear rate dependent viscosities*, Handbook of Mathematical Fluid Dynamics, In: S. Friedlander, S. Serre (Eds.), Elsevier, Amsterdam, vol. 4, 2006, 407-444.

[14] Park, H.E., Lim, S.T., Laun, H.M., Dealy, J.M., *Measurement of pressure coefficient of melt viscosity: drag flow versus capillary flow*, Rheol. Acta, **47**(2001), 1023-1038.

[15] Prusa, V., *Revisiting Stokes first and second problems for fluids with pressure-dependent viscosities*, Int. J. Eng. Sci., **48**(2010), 2054-2065.

[16] Prusa, V., Srinivasan, S., Rajagopal, K.R., *Role of pressure dependent viscosity in measurements with falling cylinder viscometer*, Int. J. Non Linear Mech., **47**(2012), 743-750.

[17] Rajagopal, K.R., *Couette flows of fluids with pressure dependent viscosity*, Int. J. Appl. Mech. Eng., **9(3)**(2004), 573-585.

[18] Rajagopal, K.R., *A semi-inverse problem of flows of fluids with pressure dependent viscosities*, Inverse Probl. Sci. Eng., **16**(2008), 269-280.

[19] Rajagopal, K.R., Saccomandi G., Vergori, L., *Unsteady flows of fluids with pressure dependent viscosity*, J. Math. Anal. Appl., **404**(2013), 362-372.

[20] Stemmer, K., Hander, H., Hansen, U., *A new method to stimulate convection with strong temperature and pressure-dependent viscosity in a spherical shell: applications to the Earth' mantle*, Phys. Earth. Planet. Inter., **157**(2006), 223-249.

[21] Stokes, G.G., *On the theories of the internal friction of fluids in motion, and motion of elastic solids*, Trans. Camb. Phil. Soc., **8**(1845), 287-305.

[22] Szeri, A.Z., *Fluid Film Lubrication*, Cambridge University Press, Cambridge, 1998.

Constantin Fetecau
"Academy of Romanian Scientists", Section of Mathematics,
54 Splaiul Independentei,
RO-050094, Bucharest, Romania
e-mail: `c_fetecau@yahoo.com`

Abdul Rauf
"Air University Islamabad",
Department of Computer Science and Engineering,
Multan Campus (60000), Pakistan
e-mail: `attari_ab092@yahoo.com`

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 211–222
DOI: 10.24193/subbmath.2021.1.17

An optimization problem for continuous submodular functions

Laszlo Csirmaz

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. Real continuous submodular functions, as a generalization of the corresponding discrete notion to the continuous domain, gained considerable attention recently. The analog notion for entropy functions requires additional properties: a real function defined on the non-negative orthant of $\mathbb{R}^n$ is entropy-like (EL) if it is submodular, takes zero at zero, non-decreasing, and has the Diminishing Returns property. Motivated by problems concerning the Shannon complexity of multipartite secret sharing, a special case of the following general optimization problem is considered: find the minimal cost of those EL functions which satisfy certain constraints. In our special case the cost of an EL function is the maximal value of the n partial derivatives at zero. Another possibility could be the supremum of the function range. The constraints are specified by a smooth bounded surface S cutting off a downward closed subset. An EL function is feasible if at the internal points of S the left and right partial derivatives of the function differ by at least one. A general lower bound for the minimal cost is given in terms of the normals of the surface S. The bound is tight when S is linear. In the two-dimensional case the same bound is tight for convex or concave S. It is shown that the optimal EL function is not necessarily unique. The paper concludes with several open problems.

Mathematics Subject Classification (2010): 90C26, 46N10, 49Q10.

Keywords: Continuous submodular optimization, entropy method, secret sharing.

1. Introduction

Continuous submodularity is a generalization of the discrete notion of submodularity to the continuous domain. It has gained considerable attention recently [2, 4] as efficient convex optimization methods can be extended to find the minimal and maximal value of special multivariable continuous submodular functions over a compact and convex domain. Such optimization algorithms have important applications in many areas of computer science and applied mathematics such as training deep neural networks [5], design of online experiments [6], or budget allocation [12]. For more information see [1].

Interestingly, the same class of continuous submodular functions arises when the continuous version of multipartite secret sharing schemes is considered. In classical secret sharing [3] each participant receives a piece of information – their *shares* – such that a qualified subset of participants can recover the secret from the shares they received, while unqualified subsets – based on their shares only – should have no information on the secret's value at all. In the multipartite case [8, 9] participants are in n disjoint groups, and members in the same group have equal roles. In particular, a qualified subset is described uniquely by the n numbers telling how many members this subset has from each group. The main question in secret sharing is the efficiency – also called complexity – of the scheme, which is typically defined as the worst-case ratio of the size of any of the shares (measured by their Shannon entropy) and the size of the secret. Keeping track of the total entropy of different subsets of shares, traditional entropy inequalities imply a lower bound on the complexity [9, 10] known as the *Shannon-bound*. No general method is known which would effectively determine, or even estimate, the Shannon bound for an arbitrary collection of qualified subsets, and numerical computation is intractable even for moderately sized problems. Investigating the same question in the continuous domain allows applying analytical tools, and results achieved this way might shed light on the discrete case. This paper, based partly on the last section of [7], is an attempt to initiate such a line of research.

No notion from secret sharing or from information theory will be used later as they only serve as motivation for the definitions. The family of real functions corresponding to the (normalized) multipartite entropy will be called *entropy-like* functions and abbreviated as EL. This function family is defined in Section 2; actually it is the family of pointed, increasing, submodular functions with the "Diminishing Returns" property, see [4].

The optimization problem corresponding to finding an optimal multipartite secret sharing scheme is discussed in Section 3. It differs from the well-studied optimization problem for submodular functions [2, 4], where some member of the continuous submodular function family is given, and the task is to find its maximal (minimal) value over a compact, convex set. In our case the optimization problem asks to find an EL function with the smallest cost satisfying certain constraints. Two cost functions are considered. The first one corresponds to the discrete worst case complexity discussed above, and it is the maximal partial derivative of the EL function at the origin. The second possibility is the supremum of the function range; it corresponds to another frequently investigated complexity measure in the discrete case: the total randomness used by the scheme. In Section 3 a general lower bound for the worst case complexity is given as Theorem 3.4. This bound is tight when the constraints are specified by some linear surface.

Section 4 presents results for the bipartite, two-dimensional case. General constructions show that the lower bound of Theorem 3.4 is also tight for strictly convex or strictly concave constraint curves. An alternate construction shows that the optimal

EL function is not necessarily unique. Finally, Section 5 concludes the paper with a list of open problems.

2. Submodular and entropy-like functions

A real function f defined on subsets of a set is *submodular* if

$$f(A) + f(B) \geq f(A \cap B) + f(A \cup B)$$

for arbitrary subsets A and B, see [2] and references therein. The same notion extended to an arbitrary lattice requires

$$f(A) + f(B) \geq f(A \wedge B) + f(A \vee B)$$

for any two lattice members A and B. In particular, the n-variable real function f is *submodular* if it is submodular in the lattice determined by the partial order on $\mathbb{R}^n$ defined by $x \leq y$ if and only if $x_i \leq y_i$ for all coordinates $1 \leq i \leq n$. In this case $x \wedge y = \min(x, y)$ and $x \vee y = \max(x, y)$ where minimization (maximization) is taken coordinatewise, and the submodularity condition rewrites to

$$f(x) + f(y) \geq f(\min(x, y)) + f(\max(x, y)).$$

Entropy-like real functions, also called EL functions, share additional properties with discrete Shannon entropy functions [13], and are defined as follows.

Defnition 2.1. The n-variable real function f is *entropy-like*, or EL function for short, if it satisfies properties (a) – (e) below.

(a) f is defined on the non-negative orthant $\mathbb{R}^n_{\geqslant 0} = \{x \in \mathbb{R}^n : x \geq 0\}$.
(b) f is submodular.
(c) $f(0) = 0$ (f is pointed).
(d) f is non-decreasing: if $0 \leq x \leq y$ then $f(x) \leq f(y)$.
(e) f has the “Diminishing Returns” property [4]. It means that for two points $0 \leq x \leq y$ differing only in their i-th coordinate, increasing that coordinate at x and also at y by the same amount ε, the gain at y is never bigger than the gain at x. Formally, if e_i is the i-th unit vector and $y = x + \lambda e_i$ for some $\lambda > 0$, then for every $\varepsilon > 0$,

$$f(x + \varepsilon e_i) - f(x) \geq f(y + \varepsilon e_i) - f(y). \tag{2.1}$$

The “Diminishing Returns” property models the natural expectation that adding one more unit of some resource contributes more in the case when one has less available amount of that resource.

The left and right partial derivatives of the n-variable function f at $x \in \mathbb{R}^n$ are denoted by $f_i^-(x)$ and $f_i^+(x)$, respectively, and their definition goes as

$$f_i^-(x) = \lim_{\varepsilon \to +0} \frac{f(x) - f(x - \varepsilon e_i)}{\varepsilon}$$

and

$$f_i^+(x) = \lim_{\varepsilon \to +0} \frac{f(x + \varepsilon e_i) - f(x)}{\varepsilon}$$

assuming that the corresponding limits exist. Here e_i is the i-th unit vector.

The following claim summarizes some basic properties of EL functions.

Claim 2.2. *Let f be an n-variable EL function.*

(a) *f is continuous.*

(b) *f is concave along any positive direction: if $0 \le x \le y$ and $0 \le \lambda \le 1$ then*

$$\lambda f(x) + (1-\lambda)f(y) \le f(\lambda x) + (1-\lambda)y).$$

(c) *The Diminishing Returns property* (2.1) *holds for arbitrary pair of points $0 \le x \le y$.*

(d) *f has both left and right partial derivatives at every point of its domain.*

(e) *The partial derivatives are non-negative and non-increasing along any positive direction.*

Proof. (a) It is enough to show that f is continuous along every coordinate. By property (d) it is monotone increasing. The left limit $\lim_{\varepsilon\to+0} f(x-\varepsilon e_i)$ cannot be strictly smaller than the right limit $\lim_{\varepsilon\to+0} f(x+\varepsilon e_i)$ as this would contradict the Diminishing Returns property.

(b) Continuity and the Diminishing Returns property ensures that f is concave along each coordinate. It means that statement (b) is true when points x and y share $n-1$ coordinates. Suppose we have two points sharing i coordinates, and the claim has been established for point pairs sharing $i+1$ or more coordinates. Denote these points by (c, x, a) and (d, y, a) where a stands for the joint i coordinates, x and y are real numbers, and c and d are the remaining tuples. The linear combination $\lambda(c, x, a) + (1-\lambda)(d, y, a)$ is shortened to $(c \between d, x \between y, a)$. Using $(c, x, a) \le (d, y, a)$ and the induction hypothesis for $n-1$ (first line) and for $i+1$ (next two lines) we have

$$\begin{aligned}
\lambda f(c \between d, x, a) + (1-\lambda)f(c \between d, y, a) &\le f(c \between d, x \between y, a),\\
\lambda f(c, x, a) + (1-\lambda)f(d, x, a) &\le f(c \between d, x, a),\\
\lambda f(c, y, a) + (1-\lambda)f(d, y, a) &\le f(c \between d, y, a).
\end{aligned}$$

From here the required inequality

$$\lambda f(c, x, a) + (1-\lambda)f(d, y, a) \le f(c \between d, x \between y, a)$$

follows as the submodularity for the points (c, y, a) and (d, x, a) gives

$$f(c, y, a) + f(d, x, a) \ge f(c, x, a) + f(d, y, a).$$

(c) Similarly to (b) by induction on how many coordinates x and y have in common. Observe that if x and y do not differ at their i-th coordinate then (2.1) is equivalent to submodularity.

(d) This is immediate as f is continuous and non-decreasing.

(e) Non-negativity is clear. Monotonicity: if $x \le y$ then, for example,

$$\begin{aligned}
f_i^+(x) &= \lim_{\varepsilon\to+0} \frac{f(x+\varepsilon e_i) - f(x)}{\varepsilon}\\
&\ge \lim_{\varepsilon\to+0} \frac{f(y+\varepsilon e_i) - f(y)}{\varepsilon} = f_i^+(y),
\end{aligned}$$

where the inequality follows from (c). Other cases are similar. □

The next lemma follows easily from the fact that along each coordinate f is increasing and concave, and is given without proof.

Lemma 2.3. *If $\varepsilon \to +0$, then $f_i^+(x+\varepsilon e_i) \to f_i^+(x)$, and $f_i^+(x-\varepsilon e_i) \to f_i^-(x)$.* □

Remark 2.4. The family of EL functions is closed for non-negative linear combination and truncation: if f_1, f_2 are EL, then so is $\lambda_1 f_1 + \lambda_2 f_2$ for $\lambda_1, \lambda_2 \geq 0$; if f is EL and $M \geq 0$ then $\min(f, M)$ is EL. Consequently

$$f(x) = \min\left(\sum c_i x_i, M\right)$$

is EL for positive c_i and M. Similarly, if f is EL and $a \geq 0$, then $g(x) = f(\min(x, a))$ is EL again. Further examples of EL functions will be given in Section 4.

Remark 2.5. If the sequence f_k of EL functions converge pointwise, then the limit f is also an EL function, moreover

$$f_i^+(x) \leq \liminf_k \,(f_k)_i^+(x) \leq \limsup_k \,(f_k)_i^-(x) \leq f_i^-(x).$$

3. The optimization problem

According to the intuition discussed in Section 1 the value of n-variable EL function f at $x \in \mathbb{R}^n_{\geqslant 0}$ can be considered as the value of the (scaled) entropy of the set of shares assigned to a subset of participants which has members from the i-th group proportional to the i-th coordinate of x. The right derivative $f_i^+(x)$ can be interpreted as the (scaled) entropy increase if one more member from the i-th group joins this subset, and $f_i^-(x)$ as the entropy decrease when one member from the i-th group leaves the subset (defined only if $x_i > 0$). Consequently the share size of a single participant from group i can be identified to $f_i^+(0)$, the i-th right partial derivative of f at zero. Accordingly, the cost function corresponding to the maximal share size is

$$\mathrm{Cost}(f) = \max\{f_1^+(0), f_2^+(0), \ldots, f_n^+(0)\}.$$

While this cost function will be considered in this paper, there are other possibilities. In the discrete cases the total entropy (the amount of randomness needed to generate the whole scheme) is used frequently, this would correspond to the cost function $\sup\{f(x) : x \in \mathrm{Dom}(f)\}$.

In secret sharing the shares of a qualified subset determine the secret, while the same secret is (statistically) independent of the shares of an unqualified subset. We call the point $x \in \mathbb{R}^n_{\geqslant 0}$ *qualified* if the corresponding subset is qualified. When decreasing an unqualified subset it remains unqualified, thus the set of unqualified points are downward closed: if x is unqualified and $0 \leq y \leq x$ then y is unqualified as well. Suppose the unqualified and qualified points are separated by the smooth $(n-1)$-dimensional surface S. Downward closedness means that the normal vectors of S pointing outwards (towards qualified points) have non-negative coordinates. This surface S specifies the secret sharing problem, namely which subsets of the participants are qualified and which are not, and thus the optimization problem as well. The definition below requires slightly stronger properties from such a separating surface excluding certain problematic cases.

Defnition 3.1. An *s-surface* (secret sharing surface) is a smooth $(n-1)$-dimensional surface S in the non-negative orthant $\mathbb{R}^n_{\geqslant 0}$ satisfying the following properties:

(a) S avoids 0,
(b) S is compact, and
(c) for every $x \in S$ the normal vector $\nabla S(x)$ pointing outwards has strictly positive coordinates.

Consider the subset of participants which corresponds to the point $x \in S$ of the s-surface S. If any member from the i-th group leaves this subset, then the subset becomes unqualified – and then the secret must be independent of the joint collection of the associated shares. If any new member from the i-th group joins that subset, it becomes qualified – meaning that the new share collection determines the secret. Thus the difference between the before and after entropy changes, namely $f_i^-(x) - f_i^+(x)$, must cover the entropy of the secret. The entropy of the secret can be taken to be 1 as this changes all values up to a scaling factor only. The following definition summarizes this discussion.

Defnition 3.2. The EL function f is *feasible* for S, or S-feasible, if for every positive $x \in S$ (that is, $x_i > 0$ for all $1 \leq i \leq n$),

$$f_i^-(x) - f_i^+(x) \geq 1 \ (1 \leq i \leq n). \tag{3.1}$$

(Positivity of x is ensures the existence of $f_i^-(x)$.)

Optimization problems considered in this paper are of this form: given the s-surface S, find the minimal cost of the S-feasible functions.

Defnition 3.3. For a given s-surface $S \subseteq \mathbb{R}^n_{\geqslant 0}$ OPT(S) is the optimization problem

$$\begin{cases} \text{minimize:} & \text{Cost}(f) \\ \text{subject to:} & f \text{ is an } S\text{-feasible EL function.} \end{cases}$$

By an abuse of notation, both the problem and its solution – the infimum of the costs of S-feasible functions – will be denoted by OPT(S).

As an example let us consider the case when S is the intersection of the hyperplane

$$c_1x_1 + c_2x_2 + \cdots + c_nx_n = M$$

and the non-negative orthant, here c_i and M are positive constants. Observe that the normal at every $x \in S$ is $\nabla S(x) = (c_1, \ldots, c_n)$. Feasible EL functions will be searched among the one-parameter family

$$f(y) = k \cdot \min\left\{\textstyle\sum c_iy_i, M\right\}$$

with positive k. All of them are EL functions by Remark 2.4. Pick the positive point $x \in S$ and consider $f(x + \varepsilon e_i)$ as a function of ε. It has the constant value $k \cdot M$ for $\varepsilon \geq 0$, and it is linear with slope $k \cdot c_i$ for $\varepsilon \leq 0$. Consequently

$$f_i^-(x) - f_i^+ = k \cdot c_i,$$

which is ≥ 1 if $k \ge 1/\min\{c_i\}$. At zero the partial derivatives of f are $k \cdot c_i$, therefore $\mathrm{Cost}(f) = k \cdot \max\{c_i\}$. The $k = 1/\min\{c_i\}$ choice gives an S-feasible EL function with cost $\max\{c_i\}/\min\{c_i\}$, thus

$$\mathrm{OPT}(S) \le \frac{\max\{c_i\}}{\min\{c_i\}}.$$

According to Theorem 3.4 below the optimal value is actually equal to this amount, as in this case $\nabla S_i(x) = c_i$ for every $x \in S$.

Theorem 3.4. *For every s-surface S, inner point $x \in S$ and $1 \le i, j \le n$ the following inequality holds:*

$$\mathrm{OPT}(S) \ge \frac{\nabla S_j(x)}{\nabla S_i(x)}.$$

Proof. By assumption S behaves linearly on a small neighborhood of x, thus for every small enough positive w there is a unique positive h such that $y = x - we_i + he_j \in S$, and

$$\lim_{w \to +0} \frac{h}{w} = \frac{\nabla S_j(x)}{\nabla S_i(x)}.$$

Let f be any S-feasible EL function, $u = \min(x, y) = x - we_i$ and $v = \max(x, y) = x + he_j$. The following inequalities follow from the facts that f is monotone and concave along each coordinate by Claim 2.2:

$$\begin{aligned} w \cdot f_i^+(u) &\ge f(x) - f(u), \\ h \cdot f_j^+(x) &\ge f(v) - f(x), \\ f(y) - f(u) &\ge h \cdot f_j^-(y), \\ f(v) - f(y) &\ge 0. \end{aligned}$$

Their sum proves the first inequality in the sequence

$$\begin{aligned} w \cdot f_i^+(u) &\ge h\big(f_j^-(y) - f_j^+(x)\big) \\ &\ge h\big(1 + f_j^+(y) - f_j^+(x)\big) \\ &\ge h\big(1 + f_j^+(v) - f_j^+(x)\big). \end{aligned}$$

The second inequality follows from $y \in S$ and that f is an S-feasible function. The third one uses the monotonicity of the derivatives from Claim 2.2 (e). Letting $w \to +0$, $f_i^+(u) \to f_i^-(x)$ and $f_j^+(v) \to f_j^+(x)$ by Lemma 2.3, thus

$$f_i^-(x) \ge \frac{\nabla S_j(x)}{\nabla S_i(x)}.$$

From here the theorem follows as $\mathrm{Cost}(f) \ge f_i^+(0) \ge f_i^-(x)$ by the monotonicity of the derivatives. □

Theorem 3.5. *Suppose* $\mathrm{OPT}(S) < +\infty$ *for an s-surface S. The optimal value is taken by some S-feasible function f, that is,* $\mathrm{Cost}(f) = \mathrm{OPT}(S)$.

Proof. Let $\mathrm{OPT}(S) < M$, and choose the sequence of S-feasible functions f_k such that $\mathrm{Cost}(f_k) < M$ and $\lim_k \mathrm{Cost}(f_k) = \mathrm{OPT}(S)$. Also pick a point $a \in \mathbb{R}^n_{\geqslant 0}$ such that S is contained completely in the box $B = \{x \in \mathbb{R}^n_{\geqslant 0} : x \le a\}$. The functions $g_k(x) = f(\min(x,a))$ are EL by Remark 2.4, and $\mathrm{Cost}(g_k) = \mathrm{Cost}(f_k)$. Each g_k is clearly S-feasible and is bounded by $M\cdot(a_1+\cdots+a_n)$. The sequence $\{g_k\}$ is uniformly equicontinuous as all partial derivatives are bounded by M, thus the Arzelà–Ascoli theorem [11] guarantees a subsequence which converges uniformly on B – and then converges everywhere. Denote this subsequence also by $\{g_k\}$, and let the pointwise limit be g. By Remark 2.5 g is an EL function and $\mathrm{Cost}(g) \le \liminf_k \mathrm{Cost}(g_k) = \mathrm{OPT}(S)$. Also, each g_k is S-feasible, that is, at the points of S the difference between the left and right derivatives is at least 1:

$$(g_k)^-_i(x) - (g_k)^+_i(x) \ge 1, \; x \in S.$$

By Remark 2.5 the same is true for the limit function g. Thus there is an S-feasible function g with $\mathrm{Cost}(g) \le \mathrm{OPT}(S)$, which proves the theorem. □

4. Two-dimensional cases

We have seen that the bound provided by Theorem 3.4 is sharp when S is linear. We show that, at least in the two-dimensional case, it is also sharp when S is strictly convex or strictly concave by constructing matching S-feasible EL functions.

In two dimensions S is a strictly decreasing continuous curve. Write S as $\{(x,\alpha(x)) : 0 \le x \le a\}$, and also as $\{(\beta(y),y) : 0 \le y \le b\}$, see Figure 1.

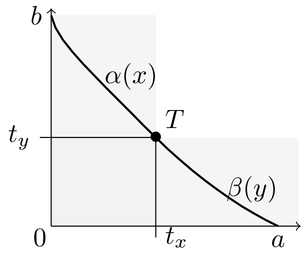


FIGURE 1. The curve S

If S is either convex or concave, then $\nabla S_i(x)/\nabla S_j(x)$ is increasing or decreasing along the curve, thus attains its maximal value at one of the endpoints.

First assume that S is strictly convex. In this case both α and β are convex functions. Let $T = (t_x, t_y)$ be the point on S where the normal is $(1,1)$. On the $[0,t_x]$ interval the derivative $\alpha'(x)$ is ≤ -1, and, similarly, $\beta'(y) \le -1$ on $[0,t_y]$. The function f depicted on Figure 2 is defined as follows.

If both $x \ge t_x$ and $y \ge t_y$ then $f(x,y) = C$, otherwise

$$f(x,y) = \begin{cases} C + \min\{x-\beta(y),0\} & \text{if } x \ge t_x, \\ C + \min\{y-\alpha(x),0\} & \text{if } y \ge t_y, \\ a-\alpha(x)+b-\beta(y) & \text{otherwise,} \end{cases}$$

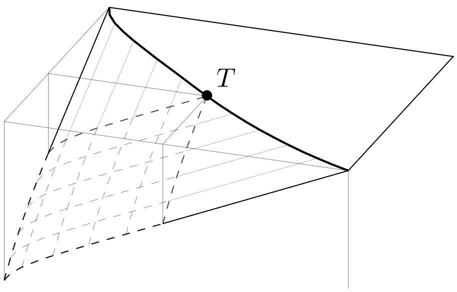


FIGURE 2. Convex case

where $C = a - t_x + b - t_y$. Clearly f has a flat plateau of height C beyond the curve S. It is a routine to check that f is an EL function; one has to use that $-\alpha(x)$ and $-\beta(y)$ are concave functions and have derivative 1 at $x = t_x$ and $y = t_y$, respectively. The left and right partial derivatives of f at $(x, y) \in S$ are $(1, 0)$ and $(-\beta'(y), 0)$ when $x \geq t_x$, and $(-\alpha'(x), 0)$ and $(1, 0)$ when $y \geq t_y$. In all cases the values in the pair differ by at least one, thus f is a feasible S-function. The partial derivatives of f at zero are $-\alpha'(0)$ and $-\beta'(0)$, thus

$$\mathrm{Cost}(f) = \max\{-\alpha'(0), -\beta'(0)\}$$

matching the lower bound of Theorem 3.4.

In the case when no point on S has normal $(1, 1)$ the simpler construction using only the first (or second) line in the definition of the function f works.

A different construction is illustrated on Figure 3 which also meets the lower bound of Theorem 3.4. It also shows that the optimal EL function, if exists, is not necessarily unique. Using the same notation as above,

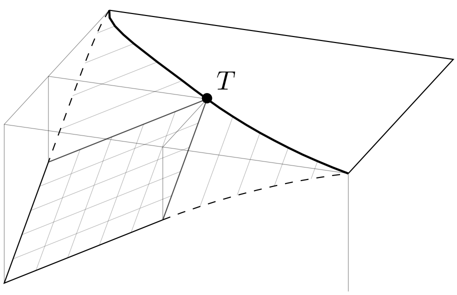


FIGURE 3. Alternate construction for the convex case

the function f is defined analogously: $f(x, y) = C$ if $x \geq t_x$ and $y \geq t_y$, otherwise

$$f(x,y) = \begin{cases} C + \min\{y - \alpha(x), 0\} & \text{if } x \geq t_x, \\ C + \min\{x - \beta(y), 0\} & \text{if } y \geq t_y, \\ x + y & \text{otherwise,} \end{cases}$$

where $C = t_x + t_y$. This is again an EL function, its cost is clearly 1. The difference between the left and right partial derivatives at points of S are $-\alpha'(x)$ and 1 when $x \geq t_x$, and 1 and $-\beta'(y)$ when $y \ge t_y$, thus the difference is at least $k = \min\{-\alpha'(a), -\beta'(b)\}$. Consequently the EL function $k^{-1}f(x,y)$ is feasible for S, and its cost, $1/k$, matches the lower bound in Theorem 3.4.

The third construction, depicted on Figure 4, works for any strictly concave curve S. In this case the plateau is not flat any more. Using the same notations as before, the decreasing functions $\alpha(x)$ and $\beta(y)$ are strictly concave, and $T = (t_x, t_y)$ is the curve point with normal $(1,1)$. The function $f(x,y)$ is defined as $f(x,y) = t_x + t_y$

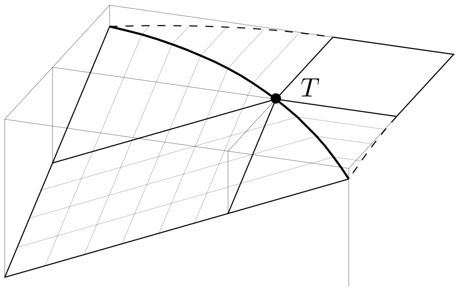


Figure 4. Concave case

if both $x \ge t_x$ and $y \ge t_y$, otherwise

$$f(x,y) = \begin{cases} y + \min\{x, \beta(y)\} & \text{if } x \ge t_x, \\ x + \min\{y, \alpha(x)\} & \text{if } y \ge t_y, \\ x + y & \text{otherwise.} \end{cases}$$

This is an EL function. For example, for a fixed $x \ge t_x$ it is increasing and concave as $y + \beta(y)$ is increasing on the $[0, t_y]$ interval ($\beta'(y) \le -1$ here), and is concave since β is concave. The left and right partial derivatives of f at a point (x,y) of S with $x \ge t_x$ are 1 and 0, and 1 and $1 + \beta'(y)$, respectively. The difference between the corresponding pairs is at least $-\beta'(y) \ge -\beta'(0)$. Choosing the multiplier k such that $k \cdot (-\alpha'(0)) \ge 1$ and $k \cdot (-\beta'(0)) \ge 1$, the EL function $k \cdot f$ will be S-feasible. The minimal such k gives a cost k EL function which again matches the lower bound of Theorem 3.4.

5. Conclusion

A continuous version of the discrete Shannon entropy functions, called *entropy-like*, or EL functions, has been defined in Definition 2.1. They form a natural subclass of multivariate continuous submodular functions which gained considerable attention recently [2]. Interestingly, the same subclass emerged as a crucial one when investigating possible parallelization of traditional submodular optimization algorithms [4].

Motivated by difficult problems in multipartite secret sharing [8], points in the non-negative orthant are flagged as either qualified or unqualified, separated by a *secret sharing surface* S, see Definition 3.1. An EL function is *feasible* for such a surface S if at internal points of S all partial derivatives drop by at least one when passing from left to right. The following optimization problem was considered: for a given s-surface S find the minimal cost of an S-feasible EL function. The first open problem is to prove that this function set is never empty.

Problem 5.1. Prove that for every s-surface S there exists at least one S-feasible function.

The cost of an EL function f is the maximum of its partial derivatives at zero, thus it can be $+\infty$. Definition 3.1 stipulates that for every S-surface there is a positive constant c such that $1/c < \nabla S_i(x) < c$ at each point $x \in S$. The value in Theorem 3.4 bounding the cost of any S-feasible function from below is smaller than c^2, thus it does not exclude the following strengthening of Problem 5.1:

Problem 5.2. Prove that for every s-surface S there is at least one S-feasible function with *finite cost.*

The lower bound on the cost of S-feasible EL functions proved in Theorem 3.4 was shown to be tight for linear s-surfaces, and also for two-dimensional convex and concave s-surfaces.

Problem 5.3. Find an s-surface S for which the bound in Theorem 3.4 is not tight.

As a strenghtening of Problem 5.3 we offer a bold conjecture which might easily turn out to be false.

Problem 5.4. If S is neither convex nor concave, then the bound of Theorem 3.4 is not tight.

Constructions in Section 4 settled the problem of finding the optimal values for two-dimensional convex and concave s-surfaces. It would be interesting to see optimal solutions for convex and concave surfaces in higher dimensions.

Problem 5.5. Determine the optimal costs of convex and concave s-surfaces in dimension > 2.

As mentioned in Section 3, the cost function considered in this paper stems from the *worst case complexity* of general secret sharing schemes. An alternate cost function corresponding to the total entropy would be $\mathrm{Cost}^t(f) = \sup\{\, f(x) : x \in \mathrm{Dom}(f)\,\}$. As an EL function can be truncated, the sup here can be limited to the points of S. The two costs functions are obviously related, but it is not clear how this relationship can be used to connect the corresponding optimization problems.

Problem 5.6. Prove lower bounds, similar to Theorem 3.4, for the optimization problem $\mathrm{OPT}^t(S)$ using the Cost^t function.

By Theorem 3.5, if there is any S-feasible function at all then there is one with minimal cost. The proof relied on the fact that finite cost EL functions have bounded derivatives. For Cost^t this property does not hold anymore.

Problem 5.7. If there is an S-feasible function, then there is one with

$$\mathrm{Cost}^t(f) = \mathrm{OPT}^t(S).$$

Finally, extend the quite meager collection of s-surfaces from Section 4 for which the exact bound is known.

Problem 5.8. Find optimal solutions for additional "interesting" s-surfaces for both cost functions.

Acknowledgment. The work of the author was supported by the GACR grant number 19-045798, and by the Lendület program of the HAS.

References

[1] Bach, F., *Learning with Submodular Functions: A Convex Optimization Perspective*, Foundations and Trends in Machine Learning, 2013.

[2] Bach, F., *Submodular functions: from discrete to continuous domains*, Math. Program., **175**(2019), no. 1-2, 419-459.

[3] Beimel, A., *Secret-sharing schemes: A survey*, In: Proceedings of the Third International Conference on Coding and Cryptology, IWCC'11, Springer-Verlag, Berlin, Heidelberg, 2011, 11-46.

[4] Bian, Y., Buhmann, J.M., Krause, A., *Continuous submodular function maximization*, arXiv preprint arXiv:2006.13474, 2020.

[5] Botton, L., Curtis, F.E., Nocedal, J., *Optimization methods for large-scale machine learning*, Siam Review, **60**(2018), no. 2, 223-311.

[6] Chen, L., Hassani, H., Karbasi, A., *Online continuous submodular maximization*, In: AISTATS, Playa Blanca, Lanzarote, Canary Islands, 2018.

[7] Csirmaz, L., Matús, F., Padró, C., *Bipartite secret sharing and staircases*, Unpublished manuscript, 2020.

[8] Farràs, O., Martí-Farré, J., Padró, C., *Ideal multipartite secret sharing schemes*, J. Cryptology, **25**(2012), no. 3, 434-463.

[9] Farràs, O., Metcalf-Burton, J.R., Padró, C., Vázquez, L., *On the optimization of bipartite secret sharing schemes*, Des. Codes Cryptog., **63**(2012), no. 2, 255-271.

[10] Metcalf-Burton, J.R., *Information rates of minimal non-matroid-related access structures*, CoRR, 2008.

[11] Rudin, W., *Principles of Mathematical Analysis*, McGraw-Hill Book Co., New York, 3rd edition, 1976.

[12] Staib, M., Jegelka, S., *Robust budget allocation via continuous submodular functions*, In: Doina Precup and Yee Whye Teh, editors, Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017, vol. 70 of Proceedings of Machine Learning Research, 3230-3240.

[13] Yeung, R.W., *A First Course in Information Theory*, Information Technology: Transmission, Processing and Storage, Springer US, 2012.

Laszlo Csirmaz
Alfréd Rényi Mathematical Institute, Budapest, Hungary
and
Institute of Information Theory and Automation,
The Czech Academy of Sciences, Pod Vodáreskou věží 4, 182 08 Praha 8, Czech Republic
e-mail: `csirmaz@renyi.hu`

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 223–240
DOI: 10.24193/subbmath.2021.1.18

Properties of Hamiltonian in free final multitime problems

Constantin Udriste and Ionel Tevy

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. In single-time autonomous optimal control problems, the Hamiltonian is constant on optimal evolution. In addition, if the final time is free, the optimal Hamiltonian vanishes on the hole interval of evolution. The purpose of this paper is to extend some of these results to the case of multitime optimal control. The original results include: anti-trace problem, weak and strong multitime maximum principles, multitime-invariant systems and change rate of Hamiltonian, the variational derivative of volume integral, necessary conditions for a free final multitime expressed with the Hamiltonian tensor that replaces the energy-momentum tensor, change of variables in multitime optimal control, conversion of free final multitime problems to problems over fixed interval.

Mathematics Subject Classification (2010): 93C20, 49K20.

Keywords: Weak and strong multitime maximum principles, multitime Hamiltonian, free final multitime.

1. Introduction

The scientific sources for this paper are: necessary conditions for multiple integral problem in the calculus of variations [1], lower semicontinuity of integral functionals [2], time-optimal control of the Bi-Steerable Robot [4], Pontryagin functions for multiple integral control problems [6], multitime maximum principle and multitime dynamic programming [7]- [3].

We give a positive answer to an important question: does the Hamiltonian attached to multitime control problems have properties similar to the Hamiltonian attached to single-time control problems?

Section 2 underlines properties of Hamiltonian in single-time optimal control. Section 3 studies the Hamiltonian in multitime optimal control. Section 4 analyses the strong multitime maximum principle. Section 5 is dedicated to multitime-invariant dynamical systems and change rate of Hamiltonian. The variational derivative of volume integral is analysed in Section 6. The necessary conditions for a free final multitime are given in Section 7. The change of variables in multitime optimal control is described in Section 8. The conversion of free end multitime problems to problems over fixed interval is realized in Section 9. Section 10 contains conclusions.

We tested the theory in relevant applications: multitime control strategies for skilled movements [3], minirobots moving at different partial speeds [16], optimal control of electromagnetic energy [5], multitime optimal control for quantum systems [10] etc.

The basic results are consequences of some properties that deserve to be emphasized: (1) the controlled PDEs used in the paper are completely integrable and this means symmetry conditions, (2) in Section 4 are used the Goursat-Darboux system and Goursat (hyperbolic) PDE, which are totally symmetric, and (3) the dynamical systems analysed in Section 5 are multitime-invariant.

2. Hamiltonian in single-time optimal control

2.1. Maximum principle with algebraic constraints

Single-time optimal control problem. Find

$$\max_u J(u) = \phi(x(t_0), x(t_f)) + \int_{t_0}^{t_f} L(x(t), u(t))\, dt,$$

subject to

$$(i)\ \dot{x}(t) = X(x(t), u(t)),\quad t \in (t_0, t_f),$$

$$(ii)\ u(t) \in U\,,\quad t \in (t_0, t_f),$$

$$(iii)\ \Phi(x(t_0), x(t_f)) \in K\,.$$

We have $x : R \to R^n$, $u : R \to R^q$, $L : R^n \times R^q \to R$, $\phi : R^n \times R^n \to R$, $X : R^n \times R^q \to R^n$, $\Phi : R^n \times R^n \to R^k$, $U \subseteq R^m$, $K \subseteq R^k$. Usually U is bounded and K is compact and convex.

Consider the Hamiltonian

$$H : R^n \times R^n \times R^q \to R,\ H(x, p, u) = L(x, u) + \langle p, X\rangle$$

and the *endpoints Lagrangian*

$$\Psi : R^n \times R^n \times R^k \to R,$$

$$\Psi((x_0, x_f), \psi) = \phi(x(t_0), x(t_f)) + \langle \psi, \Phi(x(t_0), x(t_f))\rangle.$$

Our problem becomes: find $\max \mathcal{J}(u)$, where

$$\mathcal{J}(u) = \Psi((x_0, x_f), \psi) + \int_{t_0}^{t_f} \left(H(x(t), p(t), u(t)) - \langle p(t), X(x(t), u(t))\rangle\right)\, dt\,.$$

Proposition 2.1. *The optimal solution* (x^*, p^*, u^*) *satisfies the conditions*

$$(a)\quad \dot{x}^*(t) = \frac{\partial H}{\partial p}(x^*(t), p^*(t), u^*(t))\,,$$

$$(b)\quad \dot{p}^*(t) = -\frac{\partial H}{\partial x}(x^*(t), p^*(t), u^*(t))\,,$$

$$(c)\quad H(x^*(t), p^*(t), u^*(t)) = \max_{u\in U} H(x^*(t), p^*(t), u)\,,$$

$$(d)\quad p^*(t_f) = \frac{\partial \Psi}{\partial x_f}(x^*(t_0), x^*(t_f), \psi^*)\,,$$

$$(e)\quad p^*(t_0) = -\frac{\partial \Psi}{\partial x_0}(x^*(t_0), x^*(t_f), \psi^*)\,,$$

and ψ^* *is an element of the normal cone to* K *at the point* $(x^*(t_0), x^*(t_f))$ *.*

If the final time is free, we consider it as a new control which maximizes

$$\mathcal{J} = \Psi(x(t_0), x(t_f), \psi) + \int_{t_0}^{t_f} \big(H(x(t), u(t), p(t)) - p_i(t)\dot{x}^i(t)\big)\, dt\,.$$

The necessary condition for extremum, using (d), is

$$0 = \frac{\partial \mathcal{J}}{\partial t}|_{t=t_f} = \frac{\partial \Psi}{\partial x_f^i}\dot{x}^i|_{t=t_f} + \big(H - p_i\dot{x}^i\big)\,|_{t=t_f} = H(t_f)\,.$$

Hence, according with the Proposition 2.1, we have

Proposition 2.2. *Let* x^*, p^*, u^* *be the optimal solution for a free final time autonomous problem. Then* $H^*(t) = 0$ *on the hole interval* $t_0 \leq t \leq t_f$*.*

2.2. The Hamiltonian as a first integral

For any kind of single-time autonomous problem with bounded control, the following statement is true:

Proposition 2.3. *Let* x^*, p^*, u^* *be the optimal solution and*

$$H^*(t) = H(x^*(t), p^*(t), u^*(t)) \tag{2.1}$$

the pull-back of Hamiltonian on this solution. Then $H^*(t) =$ *constant.*

Proof. According with maximum principle, in any interval of continuity, for each τ and σ, we have

$$H^*(\tau) - H^*(\sigma) \geq H(x^*(\tau), p^*(\tau), u^*(\sigma)) - H^*(\sigma)\,.$$

Then, for $\tau > \sigma$, by the state and costate equations:

$$\lim_{\tau\downarrow\sigma} \frac{H^*(\tau) - H^*(\sigma)}{\tau - \sigma} \geq \lim_{\tau\downarrow\sigma} \frac{H(x^*(\tau), p^*(\tau), u^*(\sigma)) - H^*(\sigma)}{\tau - \sigma}$$

$$= \frac{\partial H}{\partial x}(x^*(\sigma), p^*(\sigma), u^*(\sigma))\,\dot{x}^*(\sigma) + \frac{\partial H}{\partial p}(x^*(\sigma), p^*(\sigma), u^*(\sigma))\,\dot{p}^*(\sigma) = 0\,.$$

Taking $\tau < \sigma$, we obtain in a similar way the opposite inequality. Hence $\dot{H}^*(\sigma) = 0$. The result follows.
With a bit completions at a point of discontinuity we have $H^*(t) \equiv ct$. □

2.3. Conversion to problems over a fixed interval

For an optimal problem with free end time T, consider the change of variable $\tau = t/T$. Then the final time in the new variable is 1.

The functions of time expressed in this new variable become $\widetilde{x}(\tau) = x(\tau T)$, $\widetilde{u}(\tau) = u(\tau T)$, the evolution is $T\,X(x,u)$ and the running cost is $T\,L(x,u)$.

Viewing T as a new state variable for the new problem, with $\dot{T} = 0$ and costate q, the new Hamiltonian will be $\mathcal{H} = T\,H$. The optimality condition gives us

$$\dot{q} = \frac{\partial \mathcal{H}}{\partial T} = H\ ,\ \ q(0) = q(1) = 0\,.$$

Then

$$0 = q(1) - q(0) = \int_0^1 H^*(\tau)\,d\tau$$

and, according with the Proposition 2.2, we have $H^*(\tau) = 0$.

3. Hamiltonian in multitime optimal control

Generally, a multitime optimal control problem [7]-[3] is formulated in the following way: find

$$\max_u\ Q(u(\cdot)) = \int_{\Omega_{0t_0}} L(x(t),u(t))\omega + g(x(t_0))$$

subject to

$$\frac{\partial x^i}{\partial t^\alpha}(t) = X^i_\alpha(x(t),u(t)),\quad t \in \Omega_{0t_0} \subset R^m_+,$$

a controlled PDEs evolution system which is completely integrable (m-flow), where $\Omega_{0t_0} \subset \mathbb{R}^m_+$ is the parallelepiped determined by the diagonal opposite points 0 and t_0, $x(t) = (x^1(t),...,x^n(t)),\ t = (t^1,...,t^m) \in \Omega_{0t_0}$ is the state vector, $u(t) \in U,\ t \in \Omega_{0t_0}$ is the control vector, the C^1 function $L(x,u)$ is the *running cost*, $\omega = dt^1 \wedge ... \wedge dt^m$ is the volume element, and g is a C^1 function that defines the *terminal cost.*
The multitime maximum principle [7]- [3] involves the Hamiltonian $H = L + p^\alpha_i X^i_\alpha$, the initial and adjoint PDEs

$$\frac{\partial x^i}{\partial t^\alpha} = \frac{\partial H}{\partial p^\alpha_i},\ \frac{\partial p^\alpha_i}{\partial t^\alpha} = -\frac{\partial H}{\partial x^i}$$

and the condition $\max_u H$. Since the adjoint PDEs have too many solutions, we attach an anti-trace problem which involves the Hamiltonian tensor field $H^\alpha_\beta = \frac{1}{m}\delta^\alpha_\beta L + p^\alpha_i X^i_\beta$, the initial and adjoint (completely integrable) PDEs

$$\frac{\partial x^i}{\partial t^\alpha} = \frac{\partial H}{\partial p^\alpha_i},\ \frac{\partial p^\alpha_i}{\partial t^\beta} = -\frac{\partial H^\alpha_\beta}{\partial x^i}$$

and the condition $\max_u H$.
Anti-trace property: *Any solution of the anti-trace problem is solution of multitime maximum principle.*

Remark 3.1. The complete integrability condition for the adjoint PDEs is essential. Generally, we can write the anti-trace adjoint PDEs in the Pfaff form

$$dp_i^\alpha = -\frac{\partial H_\gamma^\alpha}{\partial x^i} dt^\gamma.$$

Then, let us consider $\omega = dt^1 \wedge ... \wedge dt^m$ and $\omega_\alpha = \frac{\partial}{\partial t^\alpha} \rfloor \omega$. The $(m-1)$-forms $p_i^\alpha \, \omega_\alpha$ have the exterior differentials $d(p_i^\alpha \, \omega_\alpha) = dp_i^\alpha \wedge \omega_\alpha$. This suggests to use

$$dp_i^\alpha \wedge \omega_\alpha = -\frac{\partial H_\gamma^\alpha}{\partial x^i} dt^\gamma \wedge \omega_\alpha$$

and the identities $dt^\beta \wedge \omega_\alpha = \delta_\alpha^\beta \, \omega$. We find the divergence PDEs system

$$\frac{\partial p_i^\alpha}{\partial t^\alpha} = -\frac{\partial H}{\partial x^i}$$

if and only if $\omega \neq 0$ on the solutions (complete integrability conditions).

For any type of autonomous multitime problem with bounded control, let x^*, p^*, u^* be the optimal solution and we denote $H^*(t) = H(x^*(t), p^*(t), u^*(t))$, where $t = (t^1, ..., t^m)$.

Theorem 3.2. *Suppose we have an autonomous multitime optimal problem (multitime-invariant dynamics and Lagrangian), with bounded control. If the Lagrangian L is independent on $x = (x^i)$ and the optimal solution x^*, p^*, u^* fulfills the anti-trace PDEs, then H^* is constant on the optimal m-sheets.*

Proof. According to the multitime maximum principle for any fixed σ in any m-interval of continuity, $\tau \in \mathbb{R}^m$ and $\varepsilon \in \mathbb{R}$ we have

$$H^*(\sigma + \varepsilon\tau) - H^*(\sigma) \geq H(x^*(\sigma + \varepsilon\tau), p^*(\sigma + \varepsilon\tau), u^*(\sigma)) - H^*(\sigma)\,.$$

Then, for $\varepsilon > 0$,

$$\lim_{\varepsilon \downarrow 0} \frac{H^*(\sigma + \varepsilon\tau) - H^*(\sigma)}{\varepsilon} \geq \lim_{\varepsilon \downarrow 0} \frac{H(x^*(\sigma + \varepsilon\tau), p^*(\sigma + \varepsilon\tau), u^*(\sigma)) - H^*(\sigma)}{\varepsilon}$$

$$= \left[\frac{\partial H}{\partial x^i}(x^*(\sigma), p^*(\sigma), u^*(\sigma))\, \frac{\partial x^{*i}}{\partial t^\gamma}(\sigma) + \frac{\partial H}{\partial p_i^\alpha}(x^*(\sigma), p^*(\sigma), u^*(\sigma))\, \frac{\partial p_i^{*\alpha}}{\partial t^\gamma}(\sigma)\right] \tau^\gamma\,.$$

By hypotheses (anti-trace property of multitime maximum principle), at $x^*(\sigma)$, $p^*(\sigma)$, $u^*(\sigma)$, we have

$$H_\beta^\alpha = \frac{1}{m}\delta_\beta^\alpha L + p_i^\alpha X_\beta^i,\; H = L + p_i^\alpha X_\alpha^i.$$

$$\frac{\partial x^i}{\partial t^\alpha} = \frac{\partial H}{\partial p_i^\alpha},\; \frac{\partial p_i^\alpha}{\partial t^\gamma} = -\frac{\partial H_\gamma^\alpha}{\partial x^i}.$$

It follows

$$\frac{\partial H}{\partial t^\gamma} = \frac{\partial H}{\partial x^i}\frac{\partial x^i}{\partial t^\gamma} + \frac{\partial H}{\partial p_i^\alpha}\frac{\partial p_i^\alpha}{\partial t^\gamma} = \frac{\partial L}{\partial x^i}\frac{\partial x^i}{\partial t^\gamma} + \frac{\partial p_i^\alpha}{\partial t^\gamma} X_\alpha^i + p_i^\alpha \frac{\partial X_\alpha^i}{\partial x^j}\frac{\partial x^j}{\partial t^\gamma}$$

$$= \frac{\partial L}{\partial x^i} X_\gamma^i \left(1 - \frac{1}{m}\right) + p_i^\alpha [X_\gamma, X_\alpha]^i = 0,$$

($[X_\gamma, X_\alpha] = 0$ means the complete integrability condition), i.e., H is a first integral of the anti-trace PDEs.

Hence

$$\lim_{\varepsilon\downarrow 0} \frac{H^*(\sigma+\varepsilon\tau) - H^*(\sigma)}{\varepsilon} \geq 0\,.$$

Taking $\varepsilon < 0$, we obtain in a similar way the opposite inequality; the derivative of H^* at σ in any direction τ vanishes. The result follows.

With some additions for points of discontinuity, it follows $H^*(t) \equiv ct$. □

4. Strong multitime maximum principle

This Section discusses the differences between two kind of evolution systems involved into multitime optimal control problems: (i) a full completely integrable PDEs system and (ii) a hyperbolic (diagonal) PDEs system which is completely integrable via an m-order hyperbolic PDE.

4.1. Full PDEs evolution system, no running cost

This case involves the PDEs evolution system (m-flow)

$$(PDE_f) \qquad \frac{\partial x^i}{\partial t^\alpha}(t) = X^i_\alpha(x(t), u(t)), \quad t \in \Omega_{0t_0} \subset \mathbb{R}^m_+$$

with $i = 1, ..., n$, $\alpha = 1, ..., m$, and a terminal cost.

The *Hamiltonian* and the *Hamiltonian tensor* are respectively

$$H(x,p,u) = p^\alpha_i X^i_\alpha(x,u)\,, \;\; H^\alpha_\beta(x,p,u) = p^\alpha_i X^i_\beta(x,u)\,,$$

p^α_i being the costate variables. The (PDE_f) can be written $\frac{\partial x^i}{\partial t^\alpha} = \frac{\partial H}{\partial p^\alpha_i}$. According to the *strong multitime maximum principle*, we can built an optimal costate function via the adjoint equations

$$(ADJ\,1,\,2) \qquad \frac{\partial p^\alpha_i}{\partial t^\alpha} = -\frac{\partial H}{\partial x^i}\,, \;\; \frac{\partial p^\alpha_i}{\partial t^\beta} = -\frac{\partial H^\alpha_\beta}{\partial x^i}\,.$$

4.2. Missing equations in PDEs evolution system, no running cost

Let us suppose that a PDEs evolution system does not contain all equations previously indexed by $i = 1, ..., n$, $\alpha = 1, ..., m$. An example could be a diagonal system (hyperbolic system, Goursat-Darboux system)

$$\frac{\partial x^\alpha}{\partial t^\alpha}(t) = X^\alpha_\alpha(x(t), u(t)), \;\; \alpha = 1, ..., m \;\; \text{(no sum)}.$$

To include these kinds of PDEs in the set of all first order normal PDEs, let us use an *indicator (characteristic)* function χ which, generally, is a function defined on a set $\mathcal{A}$ that indicates membership of an element in a subset A of $\mathcal{A}$, having the value 1 for all elements of A and the value 0 for all elements of $\mathcal{A}$ not in A. In our case, $\chi = 1$, if the equation with indices i and α appears in the initial evolution system and $\chi = 0$, if not. So the (PDE_f) can be written

$$(PDE_m) \qquad \chi\frac{\partial x^i}{\partial t^\alpha}(t) = \chi X^i_\alpha(x(t), u(t)), \quad t \in \Omega_{0t_0} \subset \mathbb{R}^m_+$$

with $i = 1, ..., n$, $\alpha = 1, ..., m$.

Then the *Hamiltonian* and the *Hamiltonian tensor* are respectively

$$H(x, p, u) = p_i^\alpha \ \chi X_\alpha^i(x, u)\,, \ \ H_\beta^\alpha(x, p, u) = p_i^\alpha \ \chi X_\beta^i(x, u)\,,$$

p_i^α being the costate variables.

According to the *strong multitime maximum principle*, we can built an optimal costate function via the adjoint equations

$$(ADJ\,1,\,2) \qquad \frac{\partial \chi p_i^\alpha}{\partial t^\alpha} = -\frac{\partial H}{\partial x^i}\,, \ \ \frac{\partial \chi p_i^\alpha}{\partial t^\beta} = -\frac{\partial H_\beta^\alpha}{\partial x^i}\,.$$

Remark 4.1. In the case of missing equations in PDEs evolution system, we work only with the "active" equations. The formalism of characteristic function is doing this.

4.3. Free endpoint problem with running cost

Let us consider that the cost functional include a running cost, i.e.,

$$(Q) \qquad Q(u(\cdot)) = \int_{\Omega_{0t_0}} X^0(x(t), u(t))\omega + g(x(t_0)),$$

where $x(t) = (x^1(t), ..., x^n(t))$ is the state vector, Ω_{0t_0} is the parallelepiped determined by the diagonal opposite points 0 and t_0, the *running cost* $X^0(x, u)$ is a C^1 function, and g is a C^1 function associated to the *terminal cost.* Suppose the controlled PDEs evolution system

$$\frac{\partial x^i}{\partial t^\alpha}(t) = X_\alpha^i(x(t), u(t)), \quad t \in \Omega_{0t_0} \subset \mathbb{R}_+^m$$

is full.

Adding new variables. Introducing new variables $x^{n+1}, ..., x^{n+m}$, and new costates $p_{n+\alpha}^\alpha(\cdot)$, we convert the theory to the foregoing case. The new state variables are constrained by the diagonal PDEs system (hyperbolic system, Goursat-Darboux system)

$$\frac{\partial x^{n+\alpha}}{\partial t^\alpha} = x_\alpha^{n+\alpha} = x^{n+\alpha+1}, \alpha = \overline{1, m-1}\,, \ \ \frac{\partial x^{n+m}}{\partial t^m} = X^0(x^1, ..., x^n, u),$$

equivalent to the Goursat (hyperbolic) PDE

$$\frac{\partial^m x^{n+1}}{\partial t^1 ... \partial t^m} = X^0(x^1, ..., x^n, u)\,;$$

denote also, for convenience, $x^{n+m+1} = X^0(x^1, ..., x^n, u)$.

We introduce a costate matrix $\bar{p}(\cdot) = (p_i^\alpha(\cdot)) \oplus (p_{n+\alpha}^\alpha(\cdot))$. For the new equations and new costates, $p_{n+\beta}^\alpha(\cdot)$, the values of the indicator χ are summarized by δ_β^α.

The *control Hamiltonian* is

$$H(\bar{x}, \bar{p}, u) = p_i^\alpha X_\alpha^i(x, u) + \sum_{\alpha=1}^m p_{n+\alpha}^\alpha x^{n+\alpha+1}$$

and the *control Hamiltonian tensor field* H_β^α must have the form

$$H_\beta^\alpha(\bar{x}, \bar{p}, u) = \begin{cases} p_i^\alpha X_\beta^i(x, u) & \text{if} \quad \alpha \neq \beta \\ p_i^\alpha X_\alpha^i(x, u) + p_{n+\alpha}^\alpha x^{n+\alpha+1} & \text{if} \quad \alpha = \beta \quad \text{(no sum upon } \alpha). \end{cases}$$

To understand that the matrix H^α_β is the anti-trace of H, we need to have in mind the diagonal matrices operations.

According to the *strong multitime maximum principle*, we can built a costate function $\bar{p}^*(\cdot) = ({p^*}^\alpha_i(\cdot)) \oplus ({p^*}^\alpha_{n+\alpha}(\cdot))$ satisfying

$$(EDP1) \qquad x^i_\alpha = \frac{\partial H}{\partial p^\alpha_i} \text{ or } x^i_\alpha = \frac{\partial H^\alpha_\beta}{\partial p^\beta_i}\,,\ \text{(no sum)},\ \alpha,\beta = \overline{1,m}\ \ i = \overline{1,n}\,,$$

$$(EDP2) \qquad x^{n+\alpha}_\alpha = \frac{\partial H}{\partial p^\alpha_{n+\alpha}} \text{ or } x^{n+\alpha}_\alpha = \frac{\partial H^\alpha_\alpha}{\partial p^\alpha_{n+\alpha}}\,,\ \alpha = \overline{1,m}\ \text{(no sum)},$$

$$(ADJ1) \qquad \frac{\partial \bar{p}^\alpha_i}{\partial t^\alpha} = -\frac{\partial H}{\partial \bar{x}^i}\,,\ i = \overline{1,n}\,;\ \frac{\partial p^\alpha_{n+\alpha}}{\partial t^\alpha} = -\frac{\partial H}{\partial x^{n+\alpha}}\,, \alpha = \overline{1,m}\,\text{(no sum)}$$

$$(ADJ2) \qquad \frac{\partial p^\alpha_i}{\partial t^\beta} = -\frac{\partial H^\alpha_\beta}{\partial x^i}\,,\ i = \overline{1,n}\,;\ \frac{\partial p^\alpha_{n+\alpha}}{\partial t^\alpha} = -\frac{\partial H^\alpha_\alpha}{\partial x^{n+\alpha}}\,, \alpha = \overline{1,m}\ \text{(no sum)}.$$

All these PDEs systems are completely integrable.

5. Multitime-invariant dynamical systems and change rate of Hamiltonian

Let us refer to open-end-multitime optimization problem. In the conditions of Section 3, we have $H^* = constant$ as an alternative scalar necessary condition for optimality.

Let us consider $\omega = dt^1 \wedge \ldots \wedge dt^m$ and $\omega_\alpha = \frac{\partial}{\partial t^\alpha} \rfloor \omega$. Since the final multitime t_f is free to vary, we rewrite the functional

$$J = \int_{\partial\Omega_{0t_f}} v^\alpha \omega_\alpha + \int_{\Omega_{0t_f}} \left(H - p^\alpha_i \frac{\partial x^i}{\partial t^\alpha} \right) \omega$$
$$= \int_{\Omega_{0t_f}} Div\, v + \int_{\Omega_{0t_f}} \left(H - p^\alpha_i \frac{\partial x^i}{\partial t^\alpha} \right) \omega,$$

where $v(x(t)) = (v^\alpha(x(t)))$ is the generating vector field, and $\frac{\partial v^\alpha}{\partial x^i}(t_f) = p^\alpha_i(t_f)$. Now t_f is an additional control variable for maximizing J. Consequently the cost sensitivity via the total mixed operator $D_{t^1\ldots t^m}$, to final multitime t_f, should be zero, i.e.,

$$0 = D_{t^1\ldots t^m} J|_{t=t_f} = \frac{\partial v^\alpha}{\partial x^i}\frac{\partial x^i}{\partial t^\alpha}|_{t=t_f} + \left(H - p^\alpha_i \frac{\partial x^i}{\partial t^\alpha} \right)|_{t=t_f} = H(t_f).$$

Consequently, $H^* = 0$ in the closed interval $0 \leq t \leq t_f$, i.e., in the hyperrectangle Ω_{0t_f}.

Lemma 5.1. *Let ϕ be a terminal cost, ψ be an algebraic condition for the terminal point, both of class C^m, and v be a generating C^1 vector field related by the PDE $D_{t^1\ldots t^m}(\phi + \nu\psi) = Div\, v$, where ν is a constant Lagrange multiplier.*

(i) Given v, there exists $\phi + \nu\psi$; (ii) given $\phi + \nu\psi$, there exists v.

Proof. (i) Consequence of the formula $\phi(x(t)) + \nu\psi(x(t)) = \int_{\Omega_{0t}} Div\, v\, \omega.$

(ii) Explicitly $v^1 = \frac{1}{m} D_{t^2 \ldots t^m}(\phi + \nu\psi), ..., v^m = \frac{1}{m} D_{t^1 \ldots t^{m-1}}(\phi + \nu\psi).$ □

This Lemma shows that the cost functional can be written also as

$$J = \phi(x(t_f)) + \nu\psi(x(t_f)) + \int_{\Omega_{0t_f}} \left(H - p_i^\alpha \frac{\partial x^i}{\partial t^\alpha} \right) \omega.$$

6. The variational derivative of volume integral

For a domain that evolves with the velocity v from Ω_0 to Ω_ϵ and for the function

$$I(\varepsilon) = \int_{\Omega_\varepsilon} f(t, \varepsilon)\omega,$$

we have

$$\frac{dI}{d\varepsilon}(\varepsilon = 0) = \int_{\Omega_0} \left(\frac{\partial f}{\partial \varepsilon} + div\,(f\,v) \right) \omega\,.$$

If we want that the hyperrectangle $\Omega_0 = [0, T]$ to become $\Omega_\varepsilon = [0, T + \varepsilon\,\delta t]$, it should take the transformation $t \to t + \epsilon v(t)$, where for example, $v = (v^\alpha)$, with $v^\alpha = \frac{t^\alpha}{T^\alpha}\,\delta t^\alpha$ (no summation). Then, for the function

$$\varepsilon \to I(x(\cdot) + \varepsilon h(\cdot); T + \varepsilon\delta t) = \int_{\Omega_\varepsilon} L(x(t) + \varepsilon h(t), x_\alpha(t) + \varepsilon h_\alpha(t))\omega,$$

we find

$$\frac{dI}{d\varepsilon}(\varepsilon = 0) = \int_{\Omega_0} \left[\frac{\partial L}{\partial x^i} - \frac{\partial}{\partial t^\alpha} \left(\frac{\partial L}{\partial x^i_\alpha} \right) \right] h^i\,\omega$$

$$+ \int_{\Omega_0} \left[\frac{\partial}{\partial t^\alpha} \left(\frac{\partial L}{\partial x^i_\alpha}\, h^i \right) + div(L(x(t), x_\gamma(t))\,v) \right] \omega$$

$$= \int_{\Omega_0} \left[\frac{\partial L}{\partial x^i} - \frac{\partial}{\partial t^\alpha} \left(\frac{\partial L}{\partial x^i_\alpha} \right) \right] h^i\,\omega$$

$$+ \int_{\partial\Omega_0} \delta_{\alpha\beta} \left(\frac{\partial L}{\partial x^i_\alpha}\, h^i\ + L(x(t), x_\gamma(t)) \frac{t^\alpha}{T^\alpha}\,\delta t^\alpha \right) n^\beta\, d\sigma\,,$$

with the transversality tensor

$$\mathcal{T}^\alpha = \frac{\partial L}{\partial x^i_\alpha}\, h^i\ + L(x(t), x_\gamma(t)) \frac{t^\alpha}{T^\alpha}\,\delta t^\alpha\,.$$

In this way we have consistency because on the initial faces the integrand is 0, since $h^i = 0$, $t^\alpha = 0$ and canonical normals, and on the final faces $t^\alpha = T^\alpha$. Moreover, using the vector v, we find the connection between h and δt on the faces, from their relationship to the final multitime T.

On the faces, h is related to δt. Indeed

$$x(t + \varepsilon v) + \varepsilon h(t + \varepsilon v) = \phi(t + \varepsilon v),$$

whence, differentiating at $\varepsilon = 0$, we find

$$\frac{\partial x}{\partial t^\gamma}\, v^\gamma + h = \frac{\partial \phi}{\partial t^\gamma}\, v^\gamma \quad \text{or} \quad h^i = \left(\frac{\partial \phi^i}{\partial t^\gamma} - \frac{\partial x^i}{\partial t^\gamma}\right) v^\gamma\,.$$

The transversality vector becomes

$$\mathcal{T}^\alpha = \frac{\partial L}{\partial x^i_\alpha}\left(\frac{\partial \phi^i}{\partial t^\gamma} - \frac{\partial x^i}{\partial t^\gamma}\right) v^\gamma \; + L(x(t), x_\gamma(t))\, v^\gamma$$

$$= \left(\frac{\partial L}{\partial x^i_\alpha}\,\frac{\partial \phi^i}{\partial t^\gamma} - \frac{\partial L}{\partial x^i_\alpha}\,\frac{\partial x^i}{\partial t^\gamma} + \delta^\alpha_\gamma L(x(t), x_\gamma(t))\right) v^\gamma$$

$$= \left(\frac{\partial L}{\partial x^i_\alpha}\,\frac{\partial \phi^i}{\partial t^\gamma} - T^\alpha_\gamma\right) v^\gamma\,,$$

where T^α_γ is the energy-momentum tensor.

Case of boundary integral

We can write

$$0 = \int_{\partial\Omega_0} \delta_{\alpha\beta}\left(\frac{\partial L}{\partial x^i_\alpha}\,\frac{\partial \phi^i}{\partial t^\gamma} - T^\alpha_\gamma\right) v^\gamma\; n^\beta\, d\sigma,$$

where

$$v^\gamma|_{t^\gamma = T^\gamma} = \delta t^\gamma\,, \quad v^\gamma|_{t^\gamma = 0} = 0$$

($2m$ faces, m terms of summation). Hence

$$0 = \delta t^\gamma\, \sum_{\alpha=1}^{m} \int_{F_\alpha}\left(\frac{\partial L}{\partial x^i_\alpha}\,\frac{\partial \phi^i}{\partial t^\gamma} - T^\alpha_\gamma\right) d\sigma.$$

Since δt^γ is arbitrary, it follows

$$0 = \sum_{\alpha=1}^{m} \int_{F_\alpha}\left(\frac{\partial L}{\partial x^i_\alpha}\,\frac{\partial \phi^i}{\partial t^\gamma} - T^\alpha_\gamma\right) d\sigma.$$

Here we have an algebraic system of m equations with m unknowns $T^1, ..., T^m$.

Case of multiple integral

Consequently

$$0 = \int_{\Omega_0} \frac{\partial}{\partial t^\alpha}\left(\left(\frac{\partial L}{\partial x^i_\alpha}\,\frac{\partial \phi^i}{\partial t^\gamma} - T^\alpha_\gamma\right) v^\gamma\right) \omega.$$

7. Necessary conditions for a free final multitime

Let us look for optimization of the functional

$$I(x(\cdot); t_f) = \int_{\Omega_{0t_f}} L(x(t), x_\gamma(t))\ \omega,$$

where the final multitime t_f is free to vary.

Setting the final time free means that we want to use the final time as yet another parameter for optimization. Let us return back to the calculus of variations, having in mind that understanding of the boundary conditions is crucial.

The key idea: derive the necessary conditions with the end point t_f of Ω_{0t_f} on a sheet prescribed by the function $\phi : \Omega_{0t_f} \to \mathbb{R}^n$, $t \to \phi(t)$. This trick is, that the stretching or shrinking of the hyperrectangle Ω_{0t_f} is done by perturbing the stationary value of the final multitime, denoted t_f, with the same ϵ as we use to perturb the functions $x(t)$:

$$I(x(\cdot) + \epsilon h(\cdot); t_f + \epsilon\delta t_f) = \int_{\Omega_{0t_f+\epsilon\delta t_f}} L(x(t) + \epsilon h(t), x_\gamma(t) + \epsilon h_\gamma(t))\omega.$$

Using the differentiation of a multiple integral with a parameter, we impose the necessary condition

$$0 = \frac{d}{d\epsilon} I(\epsilon)|_{\epsilon=0} = \int_{\Omega_{0t_f}} \left(\frac{\partial L}{\partial x^i} - D_\gamma \frac{\partial L}{\partial x^i_\gamma}\right) h^i\ \omega$$

$$+ \int_{\partial\Omega_{0t_f}} \delta_{\alpha\beta} \left(\frac{\partial L}{\partial x^i_\alpha}\, h^i + L\, \delta t^\alpha_f\right) n^\beta d\sigma.$$

Via Euler-Lagrange equations, it remains the only condition

$$0 = \int_{\partial\Omega_{0t_f}} \delta_{\alpha\beta} \left(\frac{\partial L}{\partial x^i_\alpha}\, h^i + L\, \delta t^\alpha_f\right) n^\beta d\sigma,$$

where

$$\delta t^\alpha_f|_{t^\beta=0} = 0,\ \forall\alpha, \beta = 1, ..., m.$$

The surface integral represents the flux of the vector field

$$\mathcal{T}^\alpha = \frac{\partial L}{\partial x^i_\alpha}\, h^i + L\, \delta t^\alpha_f$$

through the surface $\partial\Omega_{0t_f}$.
Obviously $h(t)$ and δt are related since $x(t)$ is requested to lie in the sheet $\phi(t)$ on the end faces $t^\alpha = t^\alpha_f$, i.e.,

$$x(t + \epsilon\delta t) + \epsilon h(t + \epsilon\delta t) = \phi(t + \epsilon\delta t), \text{ on the end faces of } \Omega_{0t_f}.$$

Differentiating with respect to ϵ and evaluating at $\epsilon = 0$, we find

$$\frac{\partial x}{\partial t^\alpha}\ \delta t^\alpha + h = \frac{\partial \phi}{\partial t^\alpha}\ \delta t^\alpha.$$

Computing $h(t)$, and replacing in $T^\alpha(t)$, we get the *transversality vector*

$$\mathcal{T}^\alpha(t) = \frac{\partial L}{\partial x^i_\alpha}(x(t), x_\gamma(t)) \left(\frac{\partial \phi^i}{\partial t^\beta}(t) - \frac{\partial x^i}{\partial t^\beta}(t) \right) \delta t^\beta + L(x(t), x_\gamma(t)\delta^\alpha_\beta \delta t^\beta$$
$$= \left(\frac{\partial L}{\partial x^i_\alpha}(x(t), x_\gamma(t)) \frac{\partial \phi^i}{\partial t^\beta}(t) - T^\alpha_\beta(t) \right) \delta t^\beta,$$

where T^α_β is the *energy-momentum tensor*. Since δt^β is arbitrary, and the normal vector field n^α of each face of Ω_{0t_f} belongs to the set of canonical orthonormal versors and their opposites in $\mathbb{R}^m$, the transversality relation can be written as

$$\frac{\partial L}{\partial x^i_\alpha}(x(t), x_\gamma(t)) \frac{\partial \phi^i}{\partial t^\beta}(t) - T^\alpha_\beta(t) = 0, \; t \in \text{union of end faces.}$$

The energy-momentum tensor $T^\alpha_\beta = p^\alpha_i x^i_\beta - L\, \delta^\alpha_\beta$ can be changed into Hamiltonian tensor $H^\alpha_\beta = p^\alpha_i x^i_\beta - \frac{1}{m}\, L\, \delta^\alpha_\beta$ by scaling the partial velocities. The trace of the Hamiltonian tensor is $H = p^\alpha_i x^i_\alpha - L$. It follows that for a free-final-multitime and fixed-final-state scenario, in which $\phi(t) = c$, $c \in \mathbb{R}^n$, the transversality condition simplifies to

$$H^\alpha_\beta(t) = 0 \Longrightarrow H(t) = 0, \; t \in \text{union of end faces.}$$

Consequently, $H^* = 0$ in the interval $0 \leq t \leq t_f$, i.e., in the hyperrectangle Ω_{0t_f}.

Remark 7.1. (i) The transition from the multitime calculus of variations to the multitime optimal control, especially when it comes to the definition of Hamiltonian, is somewhat tricky.

(ii) The classical Reynolds' transport theorem is:

$$\frac{d}{d\epsilon} \int_{\Omega(\epsilon)} f(x, \epsilon) dV = \int_{\Omega(\epsilon)} \frac{\partial}{\partial \epsilon} f(x, \epsilon) dV + \int_{\partial\Omega(\epsilon)} (v^b \cdot n) f(x, \epsilon) dA,$$

where $n(x, \epsilon)$ is the outward-pointing unit-normal, x is a point in the region and is the variable of integration, and dV, dA are volume and surface elements at x, and $v^b(x, \epsilon)$ is the velocity of the area element - so not necessarily the flow velocity.

8. Change of variables in multitime optimal control

In order to transform the control conditions to other coordinates and, over all, to converse a free end multitime problem to a fixed end one, we must use the transformation of the independent variables as $t = w(\tau)$, i. e. $t^\alpha = w^\alpha(\tau^1, ..., \tau^m)\,, \alpha = 1, ..., m$. Then a function x will change in $\bar{x}(\tau) = x(w(\tau))$. Consider the Jacobian matrix of the transformation, $J = \left(\dfrac{\partial w^\alpha}{\partial \tau^\beta} \right)$ and assume that $det(J)$ is not zero at all points of the domain Ω. In the new variables, the domain Ω becomes Ω_τ, the volume element is transformed as

$$dt^1 ... dt^m = det(J)\, d\tau^1 ... d\tau^m$$

and the partial derivatives in variables t^α become in the new variables

$$\frac{\partial x^i}{\partial t} = \left(\frac{\partial \bar{x}^i}{\partial \tau^\beta} \frac{\partial \tau^\beta}{\partial t^\alpha} \right) = \frac{\partial \bar{x}^i}{\partial \tau} J^{-1}\,.$$

Let us consider a non-autonomous multitime control problem given by a controlled functional

$$I(u) = \int_{\Omega} L(t, x(t), u(t)) dt^1 ... dt^m$$

and a non-autonomous PDE system

$$\frac{\partial x^i}{\partial t^\alpha}(t) = X^i_\alpha(t, x(t), u(t)) \,.$$

We may transform this non-autonomous problem by a change of the multitime in two ways.

8.1. Change the problem

The multitime controlled functional I becomes

$$I_\tau = \int_{\Omega_\tau} L(w(\tau), \bar{x}(\tau), \bar{u}(\tau))\, det(J)\, d\tau^1 ... d\tau^m$$

and the constraints change in

$$\frac{\partial \bar{x}^i}{\partial \tau^\alpha} = \frac{\partial x^i}{\partial t^\beta} \frac{\partial w^\beta}{\partial \tau^\alpha} = X^i_\beta(w(\tau), \bar{x}(\tau), \bar{u}(\tau)) \frac{\partial w^\beta}{\partial \tau^\alpha} \,,$$

or

$$\frac{\partial \bar{x}^i}{\partial \tau} = X^i(w(\tau), \bar{x}(\tau), \bar{u}(\tau))\, J \,.$$

Then we obtain the Lagrange functional, with adjoint vectors $q_i = (q_i^\alpha)$,

$$\mathcal{I} = \int_{\Omega_\tau} \left[L(w(\tau), \bar{x}(\tau), \bar{u}(\tau)) + Tr\, q_i(\tau) \left(X^i(w(\tau), \bar{x}(\tau), \bar{u}(\tau))\, J - \frac{\partial \bar{x}^i}{\partial \tau} \right) \right]$$
$$\times \det(J)\, d\tau^1 ... d\tau^m \,.$$

The new Hamiltonian is

$$\mathcal{H}_1 = \big(L(w(\tau), \bar{x}(\tau), \bar{u}(\tau)) + Tr\, q_i(\tau) X^i(w(\tau), \bar{x}(\tau), \bar{u}(\tau))\, J \big) \det(J)$$
$$= \mathcal{H} \det(J)$$

and the corresponding variational equations are:

$$\frac{\partial}{\partial \tau^\alpha}(\det(J)\, q_i^\alpha) = -\det(J) \frac{\partial \mathcal{H}}{\partial \bar{x}^i},\ \frac{\partial \bar{x}^i}{\partial \tau^\alpha} = \frac{\partial \mathcal{H}}{\partial q_i^\alpha},\ \frac{\partial \mathcal{H}}{\partial u} = 0 \,.$$

8.2. Change the variables in the Lagrange functional

The Lagrange functional, with adjoint vectors $p_i = (p_i^\alpha)$, is

$$J = \int_{\Omega} \left[L(t, x(t), u(t)) + p_i^\alpha(t) \left(X^i_\alpha(t, x(t), u(t)) - \frac{\partial x^i}{\partial t^\alpha} \right) \right] dt^1 ... dt^m \,.$$

Changing the multitime by $t = w(\tau)$, the Lagrange functional becomes

$$\mathcal{J} = \int_{\Omega_\tau} \left[L(w(\tau), \bar{x}(\tau), \bar{u}(\tau)) + Tr\, \bar{p}_i(\tau) \left(X^i(w(\tau), \bar{x}(\tau), \bar{u}(\tau)) - \frac{\partial \bar{x}^i}{\partial \tau} J^{-1} \right) \right]$$
$$\times \det(J)\, d\tau^1 ... d\tau^m$$

$$= \int_{\Omega_\tau} \left[L(w(\tau), \bar{x}(\tau), \bar{u}(\tau)) + Tr\, J^{-1}\, \bar{p}_i(\tau) \left(X^i(w(\tau), \bar{x}(\tau), \bar{u}(\tau))\, J - \frac{\partial \bar{x}^i}{\partial \tau} \right) \right] \times \det(J)\, d\tau^1 ... d\tau^m \,.$$

The new Hamiltonian is

$$\mathcal{K}_1 = \big(L(w(\tau), \bar{x}(\tau), \bar{u}(\tau)) + Tr\, \bar{p}_i(\tau) X^i(w(\tau), \bar{x}(\tau), \bar{u}(\tau)) \big) \det(J) = \mathcal{K} \det(J) \,.$$

For the two ways commute, the costates p_i and q_i must be related in a change of variable following the rule

$$\bar{p}_i(\tau) = J\, q_i(\tau) \,.$$

8.3. Conversion to problems over a fixed interval

By the multitime transformation $s^\alpha = \frac{1}{T^\alpha}\, t^\alpha$, where $T^\alpha = t_f^\alpha$, for constants $t_f^\alpha > 0$, a free-end multitime problem is converted to problem over the fixed interval $\Omega_{01} = [0,1]^m$. The unknown end multitime T is represented by an additionally state variable $T = (T^\alpha)$, for which $\frac{\partial T^\alpha}{\partial s^\beta} = 0$ and $T(0) = t_f$ is assumed. The evolution PDEs will be

$$\frac{\partial \bar{x}}{\partial t^\alpha} = \delta_{\alpha\beta} T^\beta X_\alpha \,, \ \frac{\partial T^\alpha}{\partial s^\beta} = 0 \,,\ T(0) = t_f .$$

Using the Jacobian $\Delta = T^1 \cdots T^m$, it follows

$$J = \int_{\Omega_{0t_f}} L(x(t), x_\gamma(t)) dt^1 ... dt^m = \int_{\Omega_{01}} L(x(T^\alpha\, s^\alpha), u(T^\alpha\, s^\alpha))\, \Delta\, ds^1 ... ds^m$$

$$= \int_{\Omega_{01}} T^1 \cdots T^m\, L(\bar{x}(s), \bar{u}(s))\, ds^1 ... ds^m \,.$$

Denoting q_α^β the costates associated with the variables T^α we have the following new extended Lagrangian

$$\mathcal{L} = T^1 \cdots T^m \left(L(\bar{x}, \bar{u}) \ + p_i^\alpha X_\alpha^i T^\alpha - p_i^\alpha x_\alpha^i - q_\alpha^\beta T_\beta^\alpha \right)$$
$$= T^1 \cdots T^m \left(\mathcal{H} - p_i^\alpha x_\alpha^i - q_\alpha^\beta T_\beta^\alpha \right),$$

where $\mathcal{H} = L(\bar{x}, \bar{u}) \ + p_i^\alpha X_\alpha^i T^\alpha$ is the new Hamiltonian.
The variational Euler equations with respect to $\bar{x}, p, \bar{u}, T$ and q, respectively give us

$$\frac{\partial \mathcal{H}}{\partial \bar{x}^i} + \frac{\partial p_i^\alpha}{\partial s^\alpha} = 0 \,,\ \frac{\partial \mathcal{H}}{\partial p_i^\alpha} - \bar{x}_\alpha^i = 0 \,,\ \frac{\partial \mathcal{H}}{\partial \bar{u}} = 0 \,,$$

$$\frac{\partial}{\partial T^\alpha} (T^1 \cdots T^m\ \mathcal{H}) + \frac{\partial q_\alpha^\beta}{\partial s^\beta} = 0 \,,\ T_\beta^\alpha = 0 \,.$$

Let us consider that there exist functions Q_α such that

$$\frac{\partial q_\alpha^\beta}{\partial s^\beta} = \frac{\partial^m Q_\alpha}{\partial s^1 ... \partial s^m} \,.$$

Then we have, by an integral on Ω_{01},

$$\int_{\Omega_{01}} \frac{\partial}{\partial T^\alpha} (T^1 \cdots T^m\ \mathcal{H})\, ds^1 ... ds^m \ = - \int_{\Omega_{01}} \frac{\partial^m Q_\alpha}{\partial s^1 ... \partial s^m}\ ds^1 ... ds^m \ = Q_\alpha(1) = 0 \,.$$

8.4. Generating costates

In a multitime optimal control problem there exist generating costates p_i such that

$$p_i^\alpha = \frac{\partial^{m-1} p_i}{\partial t^1 ... \widehat{\partial t^\alpha} ... \partial t^m}$$

(analogously for q). So we have

$$\frac{1}{m}\frac{\partial \mathcal{H}}{\partial \bar{x}^i} + \frac{\partial^m \bar{p}_i}{\partial s^1 ... \partial s^m} = 0 \text{ and } \frac{1}{m}\frac{\partial \mathcal{H}}{\partial T^\alpha} + \frac{\partial^m q_\alpha}{\partial s^1 ... \partial s^m} = 0\,.$$

By an integral on Ω_{01} we obtain

$$T^1 ... \widehat{T^\alpha} ... T^m \int_{\Omega_{01}} H\, ds^1 ... ds^m \; = -m \int_{\Omega_{01}} \frac{\partial^m q_\alpha}{\partial s^1 ... \partial s^m} ds^1 ... ds^m \; = q_\alpha(1) = 0\,.$$

But $H^*(t) \equiv ct$ and hence $H^*(t) \equiv 0$.

Let us consider the duality relation $\frac{\partial^m p_i}{\partial t^1 ... \partial t^m} = \frac{\partial p_i^\alpha}{\partial t^\alpha}$ (divergence form, complete integrability condition).

1) If $p_i^\alpha(t)$ are given, then

$$p_i(t) = \int_{\partial\Omega_{0t}} p_i^\alpha\, \omega_\alpha = \int_{\Omega_{0t}} \frac{\partial p_i^\alpha}{\partial t^\alpha}(t)\, dt^1 ... dt^m\,, \quad \text{with } p_i|_{t^\beta = 0} = 0,$$

where $\omega_\alpha = (-1)^{\alpha-1} dt^1 ... \widehat{dt^\alpha} ... dt^m$.
Generally: If

$$\omega = dt^1 \wedge \cdots \wedge dt^m, \;\; \omega_\alpha = \frac{\partial}{\partial t^\alpha} \rfloor \omega,$$

then

$$\int_{\partial\Omega_{0t}} p_i^\alpha\, \omega_\alpha = \int_{\Omega_{0t}} d(p_i^\alpha\, \omega_\alpha)$$

$$= \int_{\Omega_{0t}} \frac{\partial p_i^\alpha}{\partial t^\beta}\, dt^\beta \wedge \omega_\alpha = \int_{\Omega_{0t}} \frac{\partial p_i^\alpha}{\partial t^\alpha}\, \omega = \int_{\Omega_{0t}} \frac{\partial^m p_i}{\partial t^1 ... \partial t^m}\, \omega$$

$$= p_i(t) - \Sigma_\alpha p_i(t)|_{t^\alpha = 0} + \Sigma_{\alpha \neq \beta} p_i(t)|_{t^\alpha = 0, t^\beta = 0} - ... + (-1)^m p_i(0).$$

2) If $p_i(t)$ is given, then we can take

$$p_i^\alpha(t) = \frac{1}{m}\frac{\partial^{m-1} p_i}{\partial t^1 ... \hat{\partial t}^\alpha ... \partial t^m}(t).$$

9. Conversion of free end multitime problems to problems over fixed interval

The control problems considered so far are free end multitime problems, as the end multitime t_f of the interval $\Omega_{0t_f} = [0, t_f]$ is unspecified. By the multitime transformation $s^\alpha = \frac{1}{t_f^\alpha}\, t^\alpha$ (no sum) for constants $t_f^\alpha > 0$, such problems are converted to problems over the fixed interval $\Omega_{01} = [0, 1]$. The transformed problems are called fixed end multitime problems. The unknown end multitime t_f is represented by an addition state variable $y = (y_\alpha)$, for which $\frac{\partial y_\alpha}{\partial s^\beta} = 0$ and $y(0) = t_f$ is assumed.

Definition 9.1. [Transformed multitime-optimal control problem with fixed end multitime] A multitime optimal control problem is considered. Let $z = (\bar{x}, y)$ be the extended state, $M \times \mathbb{R}^m_+$ the extended state space, and

$$d\bar{x}^i = y_\alpha X_\alpha ds^\alpha,\ dy = 0$$

the extended control system. The problem to find an initial condition $y(0) = t_f$ and an input map $\bar{u}(\cdot)$ such that a solution $z(\cdot)$ results which satisfies $z(0) = (x_0, t_f)$ and $z(1) = (x_f, t_f)$ and gives the minimal value of the cost function $I(z(\cdot), \bar{u}(\cdot)) = t_f$ is called transformed multitime-optimal control problem with fixed end multitime. Any solution $(z(\cdot), \bar{u}(\cdot))$ to this problem is called transformed multitime-optimal solution.

Let us consider a free end multitime functional

$$J = \int_{\Omega_{0t_f}} L(x(t), x_\gamma(t))\ dt^1...dt^m.$$

We introduce the changing of variables $t^\alpha = t^\alpha_f\, s^\alpha$, that moves Ω_{0t_f} to Ω_{01} and

$$\frac{\partial x}{\partial t^\gamma} = \frac{\partial x}{\partial s^\alpha}\frac{\partial s^\alpha}{\partial t^\gamma} = \frac{1}{t^\gamma_f}\frac{\partial x}{\partial s^\gamma}.$$

Using the Jacobian $\Delta = t^1_f ... t^m_f$, it follows

$$J = \int_{\Omega_{0t_f}} L(x(t), x_\gamma(t))\ dt^1...dt^m = \Delta \int_{\Omega_{01}} L(x(t^\alpha_f\, s^\alpha), \frac{1}{t^\gamma_f}\, x_\gamma(t^\alpha_f\, s^\alpha))\, ds^1...ds^m.$$

In this way, the free end multitime variational problem is changed into a fixed end multitime variational problem.
Let us consider a free end controlled multitime functional

$$I(u) = \int_{\Omega_{0t_f}} L(x(t), u(t))\ dt^1...dt^m.$$

We introduce the changing of variables $t^\alpha = t^\alpha_f\, s^\alpha$, that moves Ω_{0t_f} to Ω_{01}. Using the Jacobian $\Delta = t^1_f ... t^m_f$, it follows

$$I = \int_{\Omega_{0t_f}} L(x(t), u(t))\ dt^1...dt^m = \Delta \int_{\Omega_{01}} L(x(t^\alpha_f\, s^\alpha), u(t^\alpha_f\, s^\alpha))\, ds^1...ds^m.$$

In this way, the free end controlled multitime problem is changed into a fixed end multitime problem.

Remark 9.2. The evolution PDEs are

$$\frac{\partial \bar{x}}{\partial t^\alpha} = y_\alpha X_\alpha,\ \frac{\partial y_\alpha}{\partial s^\beta} = 0,\ y(0) = t_f.$$

10. Conclusions

We start with a single-time optimal control problem. The Hamiltonian is a function used to solve such a problem for a dynamical system. It was introduced by Lev Pontryagin for single-time optimal control problems as part of his maximum principle. The idea is that a necessary condition for solving an optimal control problem is that the control should be chosen so as to optimize the Hamiltonian. From Pontryagin's maximum principle, special conditions for the Hamiltonian can be derived. When the final time t_f is fixed and the Hamiltonian does not depend explicitly on time (is autonomous), we have $H(x^*(t), u^*(t), p^*(t)) \equiv \text{ct}$, or if the terminal time is free, then $H(x^*(t), u^*(t), p^*(t)) \equiv 0$. Further, if the terminal time tends to infinity, a transversality condition on the Hamiltonian applies and $\lim_{t\to\infty} H(t) = 0$.

The main question: do some of these properties from uni-temporal problems survive for multi-temporal problems? Our goal was to provide positive answers where possible, which we did in this paper.

In order to give positive answers, we had to go through the following steps of original research: any solution of the anti-trace problem is solution of multitime maximum principle, weak and strong multitime maximum principle, multitime-invariant dynamical systems and change rate of Hamiltonian, Hamiltonian tensor, change of variables in multitime optimal control, generated costates. All these combine ideas from differential geometry, multitemporal variational calculus and optimal multitemporal control, topics to which we have made an essential contribution in recent years [5], [7]-[3].

Acknowledgments. The authors are indebted to the reviewers who insisted on getting an improved version both scientifically and linguistically.

References

[1] Barbu, V., *Necessary conditions for multiple integral problem in the calculus of variations*, Math. Ann., **260**(1982), 175-189.

[2] Berkovitz, L., *Lower semicontinuity of integral functionals*, Trans. Amer. Math. Soc., **192**(1974), 51-57.

[3] Iliuta, M., Udriste, C., Tevy, I., *Multitime control strategies for skilled movements*, Balkan J. Geom. Appl., **18**(2013), no. 2, 31-46.

[4] Mauder, M., *Time-Optimal Control of the Bi-Steerable Robot, A Case Study in Optimal Control of Nonholonomic Systems*, PhD Thesis, Julius-Maximilians-University at Würzburg, 2012.

[5] Pirvan, M., Udriste, C., *Optimal control of electromagnetic energy*, Balkan J. Geom. Appl., **15**(2010), no. 1, 131-141.

[6] Rund, H., *Pontryagin functions for multiple integral control problems*, J. Optim. Theory Appl., **18**(1976), no. 4, 511-520.

[7] Udriste, C., *Multitime controllability, observability and bang-bang principle*, J. Optim. Theory Appl., **139**(2008), no. 1, 141-157.

[8] Udriste, C., *Simplified multitime maximum principle*, Balkan J. Geom. Appl., **14**(2009), no. 1, 102-119.

[9] Udriste, C., *Multitime maximum principle for curvilinear integral cost*, Balkan J. Geom. Appl., **16**(2011), no. 1, 128-149.

[10] Udriste, C., *Multitime optimal control for quantum systems*, Proceedings of Third International Conference on Lie-Admissible Treatments of Irreversible Processes (ICLATIP-3), Kathmandu University, Dhulikhel, Nepal, January 3-7, 2011, 203-218.

[11] Udriste, C., *Multitime maximum principle for curvilinear integral cost*, Balkan J. Geom. Appl., **16**(2011), no. 1, 128-149.

[12] Udriste, C., Bejenaru, A., *Multitime optimal control with area integral costs on boundary*, Balkan J. Geom. Appl., **16**(2011), no. 2, 138-154.

[13] Udriste, C., Tevy, I., *Multitime dynamic programming for curvilinear integral actions*, J. Optim. Theory Appl., **146**(2010), Online First.

[14] Udriste, C., Tevy, I., *Multitime dynamic programming for multiple integral actions*, J. Glob. Optim., **51**(2011), no. 2, 345-360.

[15] Udriste, C., Tevy, I., *Sturm-Liouville operator controlled by sectional curvature on Riemannian manifolds*, Balkan J. Geom. Appl., **17**(2012), no. 2, 129-140.

[16] Udriste, C., Tevy, I., *Minirobots moving at different partial speeds*, Mathematics, MDPI, **1036(8)**(2020), 1-17.

Constantin Udriste
University Politehnica of Bucharest,
Faculty of Applied Sciences,
Department of Mathematics-Informatics,
Splaiul Independentei 313,
Bucharest 060042, Romania
and
Academy of Romanian Scientists,
3 Ilfov RO-050044,
Bucharest, Romania
e-mail: udriste@mathem.pub.ro, anet.udri@yahoo.com

Ionel Tevy
University Politehnica of Bucharest,
Faculty of Applied Sciences,
Department of Mathematics-Informatics,
Splaiul Independentei 313,
Bucharest 060042, Romania
e-mail: tevy@mathem.pub.ro, vascatevy@yahoo.fr

Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 241–243

Book reviews

Mark Elin, Simeon Reich and David Shoikhet, Numerical range of holomorphic mappings and applications, Cham: Birkhäuser xiv + 229 p. 2019.
ISBN 978-3-030-05019-1/hbk; 978-3-030-05020-7/ebook.[1]

For a closed linear operator A defined on a dense subspace $\mathcal{D}_A$ of a Hilbert space $\mathcal{H}$ the numerical range is defined by

$$V(A) = \{\langle Ax, x\rangle : x \in \mathcal{D}_A,\ \|x\| = 1\}.$$

This definition was extended by Lumer (1961) to a complex Banach space X:

$$V(A) = \{\langle Ax, x^*\rangle : x \in \mathcal{D}_A,\ \|x\| = 1,\ x^* \in J(x)\},$$

where $J(x) = \{x^* \in X^* : \langle x, x^*\rangle = \|x\|^2 = \|x^*\|^2\}$ and X^* is the dual of X.

In both cases the numerical radius of the operator A is defined by

$$|V(A)| = \sup\{|\lambda| : \lambda \in V(A)\}.$$

The operator A is called Hermitian if $V(A) \subset \mathbb{R}$ and dissipative if

$$\{\operatorname{Re}\lambda : \lambda \in V(A)\} \subset (-\infty, 0],$$

or, equivalently, $\|(tI - A)x\| \geq t\|x\|$ for all $x \in \mathcal{D}_A$ and all $t > 0$.

Numerical ranges turned out to be an essential tool in the study of semigroups of linear operators on Banach spaces, mainly due to the famous Lumer-Phillips theorem (1961): if A is dissipative and for some $\lambda_0 > 0$ (hence, for all $\lambda > 0$) $(\lambda_0 I - A)\mathcal{D}_A = X$, then A is the infinitesimal generator of a C_0-semigroup of linear contractions on X.

The theory of linear semigroups of operators on a Banach space is presented in the first chapter of the book, including the Hille-Yosida and Lumer-Phillips theorems, the analytic extension of semigroups of linear operators, as well as an overview of ergodic theory with emphasis on some classical and recent results on Cesàro and Abel averages.

Harris (1971) defined the numerical range for holomorphic functions in the following way. For a convex domain $\mathcal{D}$ in a Banach space X and $x \in \partial\mathcal{D}$ put

$$Q(x) = \{\ell \in X^* : \ell(x) = 1, \operatorname{Re}\ell(y) \leq 1,\ \forall y \in \mathcal{D}\}.$$

[1]Our colleague, Prof. Gabriela Kohr, enthusiastically embarked on writing this review. Unfortunately, her untimely death, a sad loss for all of us, prevented her from completing this task. This review is dedicated to her fond memory (S.C.).

For a holomorphic function $h : \mathcal{D} \to X$ admitting a continuous extension to $\overline{\mathcal{D}}$ define the numerical range of h on $\mathcal{D}$ by

$$V_{\mathcal{D}}(h) = \{\ell(h(x)) : x \in \partial\mathcal{D},\ \ell \in Q(x)\}\,,$$

and let $|V_{\mathcal{D}}(h)| = \sup\{|\lambda| : \lambda \in V_{\mathcal{D}}(h)\}$ be the numerical radius of the function h on $\mathcal{D}$.

In the second chapter, *Numerical range*, the authors generalize Harris' theory of the numerical range of holomorphic mappings. The main properties of the so-called quasi-dissipative mappings and their growth estimates are studied, including a nonlinear analog of the Lumer-Phillips theorem in the study of nonlinear semigroups and their applications to evolution equations and to geometric properties of holomorphic mappings in finite- and infinite-dimensional Banach spaces.

Another area of applications is that of fixed points, treated in the third chapter, 3. *Fixed points of holomorphic mappings.* The classical result in this field is the Grand Fixed Point Theorem (as it is called by the authors) due to Denjoy and Wolff (1926), but some extensions of the Earle-Hamilton and Bohl-Poincaré-Krasnoselskii Theorems, including their connections with Schwarz-Pick systems of pseudometrics and pseudo-contractive mappings, are presented as well. As the authors mention, another goal is to prove existence and uniqueness of fixed points of holomorphic mappings (not necessarily bounded) acting on the open unit ball of a Banach space.

A good companion in reading this chapter is the book by

S. Reich and D. Shoikhet, *Nonlinear semigroups, fixed points, and geometry of domains in Banach spaces*, Imperial College Press, London, 2005,

where, in the fifth chapter, Denjoy-Wolff type results are presented at length.

Ch. 4, *Semigroups of holomorphic mappings*, is concerned with certain autonomous dynamical systems acting on the open unit ball of a complex Banach space. This study is motivated by the fact that if such a system is differentiable with respect to time, then its derivative is a holomorphically dissipative mapping.

In Ch. 5, *The ergodic theory for holomorphic mappings*, the ergodic properties of a holomorphic mapping around its fixed points are studied. Special attention is paid to the so-called power bounded, dissipative and pseudo-contractive mappings.

The last chapter of the book, Ch. 6. *Applications*, is devoted to applications of the numerical range to diverse geometric and analytic problems – radii of starlikeness and spirallikeness, semigroups of composition operators on H^p-spaces, etc.

Combining methods from various areas of mathematical analysis (understood in a wide sense – functional analysis, operator theory, operator equations, holomorphic vector functions) and presenting both classical results and new developments, many of the latter due to the authors, this fine book reflects the authors' encyclopedic knowledge of mathematics as well as their ability to present the results in an accessible and clear manner. The book sheds a new light on the numerical range of holomorphic mappings and its applications and invites people, especially young researchers, to push further research in these areas.

S. Cobzaş

Dorin Andrica and Ovidiu Bagdasar, Recurrent sequences. Key results, applications, and problems, Problem Books in Mathematics. Cham: Springer 2020, xiv+402 p.
ISBN 978-3-030-51501-0/hbk; 978-3-030-51502-7/ebook.

The recurrence is a central theme in many fields of mathematics, primarily in the study of dynamical systems, but also in the theory of algorithms, numerical analysis, etc. It has deep applications in biology, physics, computer science, signal processing, economics.

The present book is devoted to sequences defined by recurrence relations, both in the real and complex field of numbers. The first two chapters, 1. *Introduction to recurrence relations* and 2. *Basic recurrent sequences*, are concerned with some fundamental results on recurrence sequences (including existence and uniqueness) and the basic recurrent sequences and polynomials – Fibonacci, Lucas, Pell, or Lucas-Pell. Homographic recurrences defined by linear fractional transformations in the complex plane are also discussed in the second chapter. The reading of these chapters, as well as of Chpter 5, requires only college algebra, complex numbers, analysis, and basic combinatorics, while for Chapters 3, 4, and 6, some basic results in number theory, linear algebra, and complex analysis are needed.

Chapter 3, *Arithmetic and trigonometric properties of some classical recurrent sequences*, is concerned with further properties and formulae for some classical recurrent sequences, while Chapter 4, *Generating functions*, treats the important topic of generating functions, both ordinary and exponential, for classical recurrent sequences. This chapter also contains a new version of Cauchy's integral formula, obtained by the authors, with applications to exact integral formulae for the coefficients of some classical polynomials as well as for some classical sequences.

Chapter 5, *More on second-order linear recurrent sequences*, is mainly concerned with the important class of Horadam sequences, including graphical representations in the complex plane of the orbits of these sequences. Here many original results of the authors are included.

Chapter 6, *Higher order linear recurrent sequences*, presents the dynamics of complex linear recurrent sequences of higher order and investigates the periodicity, geometric structure, and enumeration of the periodic patterns.

Chapter 7, *Recurrences in Olympiad training,* contains 123 olympiad training problems involving recurrent sequences, solved in detail in Chapter 8. *Solutions to proposed problems.*

Written in a clear and alive style, the book contains many results concerning recurrent sequences, reflecting the current research in the field and including authors' contributions. The theoretical results are illustrated by numerous examples and diagrams and practical applications to algebra, number theory, geometry of the complex plane, discrete mathematics, or combinatorics, are given.

The book is of interest to researchers working in this area and in related domains, but college or university students and their instructors will also find a lot of useful material in it.

S. Cobzaş