ACADEMY OF ROMANIAN SCIENTISTS

ANNALS

SERIES ON MATHEMATICS AND ITS APPLICATIONS

VOLUME 3 **2011** **NUMBER 2**

ISSN 2066 – 6594

TOPICS:

- **ORDINARY AND PARTIAL DIFFERENTIAL EQUATIONS**
- **OPTIMIZATION, OPTIMAL CONTROL AND DESIGN**
- **NUMERICAL ANALYSIS AND SCIENTIFIC COMPUTING**
- **ALGEBRAIC, TOPOLOGICAL AND DIFFERENTIAL STRUCTURES**
- **PROBABILITY AND STATISTICS**
- **ALGEBRAIC AND DIFFERENTIAL GEOMETRY**
- **MATHEMATICAL MODELLING IN MECHANICS ENGINEERING SCIENCES**
- **MATHEMATICAL ECONOMY AND GAME THEORY**
- **MATHEMATICAL PHYSICS AND APPLICATIONS**

EDITURA

ACADEMIEI OAMENILOR DE ȘTIINȚĂ DIN ROMÂNIA

Annals of the Academy of Romanian Scientists

Series on Mathematics and its Applications

Annals of the Academy of Romanian Scientists
Series on Mathematics and its Applications
ISSN 2066 - 6594 Volume 3, Number 2 / 2011

CONTENTS

Annals of the Academy of Romanian Scientists
Series on Mathematics and its Applications
ISSN 2066 - 6594 Volume 3, Number 2 / 2011

THE CLASSICAL MAXIMUM PRINCIPLE. SOME OF ITS EXTENSIONS AND APPLICATIONS.*

Cristian - Paul Danet †

Abstract

The intention of this paper is to survey some extensions (the P function method) and applications of the classical maximum principle for elliptic operators.

MSC: 35B50, 35G15, 35J40.

keywords: maximum principle, P function method, higher order elliptic equations, plate theory.

1 Introduction

The intention of this paper is to survey some extensions (the P function method) and applications of the classical maximum principle for elliptic operators.

The maximum principle is one of the most useful and best known tools employed in the study of partial differential equations. The maximum principle enables us to obtain information about the uniqueness, approximation,

*Accepted for publication in revised form on January 17, 2011.

†cristiandanet@yahoo.com Department of Applied Mathematics, University of Craiova, Al. I. Cuza St., 13, 200585 Craiova, Romania; Paper written with financial support of Ministry of Education and Research Romania under Grant no. ID - PCE 1192 -09.

boundedness and symmetry of the solution, the bounds for the first eigenvalue, for quantities of physical interest (maximum stress, the torsional stiffness, electrostatic capacity, charge density etc), the necessary conditions of solvability for some boundary value problems, etc.

The first chapter specializes the maximum principle for partial differential equations to the one variable case. We present the one dimensional classical maximum principle and a new extension.
In chapter two, we present the classical maximum principle of Hopf for elliptic operators and some possible extensions (the P function method (in honour of L. Payne, see [43]) and give a number of applications.

The maximum principle occurs in so many places and in such varied forms that is impossible to treat all topics. We treat here only the classical maximum principle and one of its extensions, namely the P function method for the elliptic case.

2 The one dimensional case

The one dimensional maximum principle represents a generalization of the following simple result: *Let the smooth function u satisfy the inequality $u'' \geq 0$ in $\Omega = (\alpha, \beta)$. Then the maximum of u in Ω occurs on $\partial\Omega = \{\alpha, \beta\}$ (on the boundary of Ω)*, i.e.,

$$\max_{\overline{\Omega}} u = \max\{u(\alpha), u(\beta)\}.$$

Theorem 1. *(one dimensional weak maximum principle) Let $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$ be a nonconstant function satisfying $\mathrm{L}u \equiv u'' + b(x)u' \geq 0$ in Ω, with b bounded in closed subintervals of Ω. Then,*

$$\max_{\overline{\Omega}} u = \max_{\partial\Omega} u.$$

Drawing the graph of a function u satisfying $u'' \geq 0 \quad (u'' \neq 0)$ reveals us the interesting fact that at a point on $\partial\Omega$ (where u attains its maximum), the slope of u is nonzero. More precisely, $du/dn > 0$ at such a point. Here d/dn denotes the outward derivative on $\partial\Omega$, i.e.,

$$\frac{du}{dn}(\alpha) = -u'(\alpha), \ \frac{du}{dn}(\beta) = u'(\beta).$$

The next theorem is an extension of this result:

Theorem 2. *(one dimensional strong maximum principle) Let* $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$ *be a nonconstant function satisfying* $\mathrm{L}u \equiv u''+b(x)u'+c(x)u \geq 0$ *in* Ω, *with* b *and* c *bounded in closed subintervals of* Ω *and* $c \leq 0$ *in* Ω. *Then a nonnegative maximum can occur only on* $\partial\Omega$, *and* $du/dn > 0$ *there. If* $c \equiv 0$ *in* Ω *then,* u *takes its maximum on* $\partial\Omega$ *and* $du/dn > 0$ *there.*

The following simple counterexample shows that we have to impose some restrictions to c: The function $u(x) = e^{-x}\sin x$ satisfies

$$\mathrm{L}u \equiv u'' + 2u' + 3u \geq 0 \text{ in } \Omega = (0, \pi).$$

We see that the nonnegative u vanishes on $\partial\Omega$ and hence there can be no maximum principle. A result can still be proven if $c \geq 0$. The result is a version of Theorem 5 on page 9 in [65].

Theorem 3. *(one dimensional generalized maximum principle) Let* $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$ *be a nonconstant function satisfying* $\mathrm{L}u \equiv u''+c(x)u \geq 0$ *in* Ω. *Suppose that*

$$\sup_{\Omega} c < \frac{\pi^2}{(\operatorname{diam}\Omega)^2}. \tag{1}$$

Then, the function u/w_ε *cannot attain a nonnegative maximum in* Ω *unless it is a constant.* $\operatorname{diam}\Omega$ *represents the diameter of* Ω *and*

$$w_\varepsilon = \cos\frac{\pi(2x - \operatorname{diam}\Omega)}{2(\operatorname{diam}\Omega + \varepsilon)}\cosh(\varepsilon x),$$

where $\varepsilon > 0$ *is small.*

The proof follows from Theorem 5, page 9 in [65] and Lemma 2.1. [11]. Although our result is stated only for a particular operator L ($b \equiv 0$), is it more precise than the result stated for general operators $\mathrm{L}u \equiv u'' + b(x)u' + c(x)u$ (see Theorem 5, page 9 in [65]). The authors do not indicate when a maximum principle is valid. They state that a maximum principle is valid for " any sufficiently short interval Ω ".

The proofs of these theorems as well as their applications (uniqueness of the solution of the boundary value problem, approximation in boundary value problems, the classical Sturm- Liouville theory, existence for nonlinear equations via monotone methods) can be found in the excellent book of Protter and Weinberger [65].

Certain solutions of equations of higher order exhibit a maximum principle:

Theorem 4. *Let $1 \leq k \leq n-1$, $n \geq 2$ and $u \in C^n(\overline{\Omega})$ be a nonconstant function satisfying* $\mathrm{L}u \equiv u^n \geq 0$ *in* Ω. *Suppose that*

$$(-1)^{n-k} u^{(i)}(\alpha) \geq 0,\ i = 1, \ldots, k-1\ (\textit{if such } i \textit{ exist}),$$

$$(-1)^{n-k+j} u^{(j)}(\beta) \geq 0,\ j = 1, \ldots, n-k-1\ (\textit{if such } j \textit{ exist}).$$

Then, in the case $n-k$ even u attains its minimum value and in case $n-k$ odd u attains its maximum either at α or β.

The nontrivial proof is given in [76]. For n=4, k=2, Theorem 4 generalizes the maximum principle in [6]: *Let u satisfy the inequality $u^{(4)} \leq 0$ in Ω. If $u'(\alpha) \leq 0$, $u'(\beta) \geq 0$, then u attains its maximum at α or β.*
A maximum principle for general fourth order operators appears in [41].

3 The n dimensional case

In this section, we treat the n dimensional variants of results presented in section 1, some possible extensions for nonlinear equations and for equations of higher order as well as their applications.

We consider the linear operator (summation convention is assumed, i.e., summation from 1 to n is understood on repeated indices)

$$\mathrm{L}u = a^{ij}(x)u_{ij} + b^i(x)u_i + c(x)u,\ a^{ij}(x) = a^{ji}(x),$$

where $x = (x_1, ..., x_n) \in \Omega$, Ω is a bounded domain (unless otherwise stated) of $\mathbb{R}^n$, $n \geq 1$ and $u_i = \frac{\partial u}{\partial x_i}, u_{ij} = \frac{\partial^2 u}{\partial x_i \partial x_j}$.
The operator L is called elliptic at a point $x \in \Omega$ if the matrix $[a^{ij}(x)]$ is positive, i.e., if $\lambda(x)$ and $\Lambda(x)$ denote respectively the minimum and maximum eigenvalues of $[a^{ij}(x)]$, then

$$0 < \lambda(x)|\xi|^2 \leq a^{ij}(x)\xi_i\xi_j \leq \Lambda(x)|\xi|^2,$$

for all $\xi = (\xi_1, ..., \xi_n) \in \mathbb{R}^n - \{0\}$. If $\lambda \geq 0$, then L is called elliptic in Ω. If Λ/λ is bounded in Ω, we shall call L uniformly elliptic in Ω.

Theorem 5. *(weak maximum principle) ([25]). Let* L *be elliptic in* Ω. *Suppose that* $|b^i|/\lambda < +\infty$ *in* Ω, $i = 1, ..., n$. *If* $\mathrm{L}u \geq 0$ *in* Ω, $c = 0$ *in* Ω *and* $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$, *then the maximum of* u *in* $\overline{\Omega}$ *is achieved on* $\partial\Omega$, *that is:*

$$\max_{\overline{\Omega}} u = \max_{\partial\Omega} u. \tag{2}$$

Remarks: 1). Theorem 5 holds under the weaker hypothesis: the matrix $[a^{ij}]$ is nonnegative and the ratio $|b^k|/a^{kk}$ is locally bounded for some $k \in \{1, ..., n\}$.

2). The maximum principle for subharmonic functions goes back to Gauss (1838) ([17]). The first proof of a maximum principle for operators more general than the Laplace operator was proved in two dimensions by Paraf in 1892 ([42]).

Theorem 6. *(the strong maximum principle of E. Hopf) ([30]). Let* L *be uniformly elliptic,* $c = 0$ *and* $\mathrm{L}u \geq 0$ *in* Ω *(not necessarily bounded), where* $u \in C^2(\Omega)$. *Then, if* u *attains its maximum in the interior of* Ω, *then* u *is constant. If* $c \leq 0$ *and* c/λ *is bounded then* u *cannot attain a nonnegative maximum in the interior of* Ω, *unless* u *is constant.*

The proof is a consequence of the following useful result known as Hopf's lemma [30]:

Lemma 1. *Suppose that* L *is uniformly elliptic in* Ω, $c = 0$ *in* Ω *and* $\mathrm{L}u \geq 0$ *in* Ω. *Let* $x_0 \in \partial\Omega$ *be such that*
i) u *is continuous at* x_0,
ii) $u(x_0) > u(x)$ *for all* $x \in \Omega$,
iii) $\partial\Omega$ *satisfies an interior sphere condition at* x_0 *(i.e., there exists a ball* $B \subset \Omega$ *with* $x_0 \in \partial B$*).*
Then the outer normal derivative of u *at* x_0, *if it exists, satisfies the strict inequality*

$$\frac{\partial u}{\partial n}(x_0) > 0. \tag{3}$$

If $c \leq 0$ *and* c/λ *is bounded in* Ω, *then the same conclusion holds provided* $u(x_0) \geq 0$, *and if* $u(x_0) = 0$ *then, the same conclusion holds irrespective of the sign of* c.

We now restrict ourselves to the case $b^i \equiv 0$ and prove Danet, [11]:

Theorem 7. *(generalized maximum principle) Let* $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$ *satisfy the inequality* $\mathrm{L}u \equiv \Delta u + c(x)u \geq 0$*, where* $c \geq 0$ *in* Ω*. Suppose that*

$$\sup_{\Omega} c < C_1 = \frac{4n+4}{(\operatorname{diam}\Omega)^2}. \tag{4}$$

Then, the function u/w_1 *cannot attain a nonnegative maximum in* Ω*, unless it is a constant.*

Similarly, if Ω *lies in a slab of width* d *and*

$$\sup_{\Omega} c < C_2 = \frac{\pi^2}{d^2}, \tag{5}$$

we obtain a similar result for u/w_2*. Here*

$$w_1(x) = 1 - (\sup_{\Omega} c/2n)(x_1^2 + \cdots + x_n^2)$$

and

$$w_2 = \cos\frac{\pi(2x_i - d)}{2(d+\varepsilon)} \prod_{j=1}^{n} \cosh(\varepsilon x_j),$$

for some $i \in \{1, \ldots, n\}$*, where* $\varepsilon > 0$ *is small.*

Comments

1. A broad class of domains satisfy $\Omega \subset B_{\mathrm{diam}\Omega/2}$. For these domains C_1 may be replaced by $C_3 = 8n/(\mathrm{diam}\Omega)^2$.

2. We may improve the constant C_3 (i.e., choose a larger constant) if $\Omega = \{x \in \mathbb{R}^n \mid 0 < R <\mid x \mid< R+\varepsilon\}$, where $\varepsilon > 0$ is sufficiently small. A maximum principle holds if

$$\sup_{\Omega} c < C_4 = \frac{2(n-1)}{(\varepsilon+\delta)\mathrm{diam}\Omega}. \tag{6}$$

For sufficiently small ε we have $C_4 > C_3$.

3. A similar result was given in [65], Theorem 10, p.73. for general operators. The authors proved that if

$$\sup_{\Omega} \gamma < \frac{4}{d^2 e^2}, \tag{7}$$

then a similar maximum principle is valid. Here $\mathrm{L}u \equiv \Delta u + c(x)u$, $c \geq 0$ in Ω and Ω is supposed to lie in a strip of width d. Of course, Theorem 7 (valid only for the case $b^i \equiv 0, c > 0$) is sharper that their result, but does not hold for general operators.

4. We have to impose some restrictions to c. Otherwise, as the following example shows, the maximum principle (Theorem 7) is false. The function $u(x,y) = \sin x \sin y$ satisfies $u = 0$ on $\partial\Omega$ and is solution of the equation $\Delta u + 2u = 0$ in $\Omega = (0,\pi) \times (0,\pi)$. Of course, (4) does not hold.

The maximum principles that we have presented above are valid only for the class $C^2(\Omega) \cap C^0(\overline{\Omega})$, i.e., the results are valid for classical solutions. We may consider operators L of the divergence form

$$\mathrm{L}u \equiv (a^{ij}(x)_j u + b^i(x)u)_i + c^i(x)u_i + d(x)u,$$

whose coefficients $a^{ij},\ b^i,\ c^i,\ d,\ i,j = 1,2,...,n$ are assumed to be measurable functions on a domain $\Omega \subset \mathbb{R}^n$.

The divergence form has the advantage that the operator L may be defined for a significant broader class of functions than the class $C^2(\Omega)$.

Assume that u is weakly differentiable and that $a^{ij}D_j u + b^i u$ and that $c_i D_i u + du,\ i = 1,2,...,n$ are locally integrable. Then u satisfies in a weak sense $\mathrm{L}u = 0\,(\geq 0,\ \leq 0)$ in Ω if :

$$\mathcal{L}(u,\varphi) = \int_\Omega \left[(a^{ij}u_j + b^i u)\varphi_i - (c^i u_i + du)\varphi\right] dx = 0\,(\leq 0,\ \geq 0),$$

for all non-negative $\varphi \geq \in C_0^1(\Omega)$.
We shall assume that L is strictly elliptic in Ω and that L has bounded coefficients, i.e. there exists some constants Λ and $\nu \geq 0$ such that:

$$\sum_{i,j} |a^{ij}|^2 \leq \Lambda,\ \lambda^{-2}\sum_i \left(|b^i|^2 + |c^i|^2\right) + \lambda^{-1}|d| \leq \nu^2. \tag{8}$$

is valid in Ω.

We state now the weak maximum principle for weak solutions.

Theorem 8. *([4], [25]) Let $u \in W^{1,2}(\Omega) \cap C^0(\overline{\Omega})$ satisfy $\mathrm{L}u \geq 0$ in Ω. If*

$$\int_\Omega (d\varphi - b^i\varphi_i)dx \leq 0,\ \forall\ \varphi \geq 0,\ \varphi \in C_0^1(\Omega) \tag{9}$$

then,

$$\sup_{\Omega} u \leq \sup_{\partial\Omega} u^+.$$

Here $u^+ = \max\{u, 0\}$.

Extensions and application of this result are presented in the book of Gilbarg and Trudinger [25].

We now deal with a possible extension of the maximum principle, namely the P function method. The method consists in determining a function $\mathrm{P} = \mathrm{P}(x, u, \nabla u, \dots)$, satisfying a maximum principle, i.e.,

$$\max_{\overline{\Omega}} \mathrm{P} = \max_{\partial\Omega} \mathrm{P},$$

where u is a solution of the studied equation (boundary value problem). This powerful method has many applications of interest and represents the core of the paper.

I. **The second order case**

1. The St.-Venant problem. (the torsion problem)

First, we examine one of the simplest cases, the problem of the torsional rigidity of a beam

$$\begin{cases} \Delta u = -2 & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{10}$$

Theorem 9. *The function* $\mathrm{P}_1 = |\nabla u|^2 + 4u$ *takes its maximum value either at a critical point of* u *or at some point on the boundary, unless* P_1 *is a constant. If* Ω *is convex and smooth (*$\partial\Omega \in C^{2+\varepsilon}$*), then* P_1 *cannot take its maximum value on* $\partial\Omega$*. Moreover, if* Ω *degenerates to an infinite strip, then* $\mathrm{P}_1 \equiv$ *const. Similarly the function* $\mathrm{P}_2 = |\nabla u|^2$ *attains its maximum on* $\partial\Omega$*.*

The proof is due to L.E.Payne, [43] and follows from the differential inequality

$$\Delta \mathrm{P}_1 + \frac{1}{|\nabla u|^2}\{4\nabla \mathrm{P}_1 \cdot \nabla u + \frac{1}{2}|\nabla \mathrm{P}_1|^2\} \geq 0 \text{ in } \Omega,$$

and the maximum principle.

Theorem 10. *([83])*

The function $\mathrm{P}_3 = |\nabla u|^2 + (4/n)u$ *takes its maximum value at some point on the boundary, unless* P_3 *is a constant. Moreover,* P_3 *is identically constant in* Ω *if and only if* Ω *is a* n *dimensional ball.*

Remarks. 1. The simplest P function is $\mathrm{P} = u$ (the classical maximum principle).
2. There are no general methods to determine P functions. Sometimes we can check the one dimensional case in order to get an idea of what types of P functions we have to look in the n dimensional case. For example considering the one dimensional equation

$$u'' + 2 = 0 \text{ in } \Omega = (0, \alpha)$$

and multiplying it by u' and then integrating it we get that

$$\mathrm{P} = (u')^2 + 4u \equiv \text{const. in } \Omega.$$

This function is the one dimensional version of P_1.

Applications

a). Upper bound for the stress function u, if Ω is convex.

Let M be the unique critical point of u and Q a point on $\partial\Omega$, nearest to P. Let r measure the distance from M along the ray connecting M and Q. Hence

$$-\frac{du}{dr} \leq |\nabla u| \text{ in } \Omega. \tag{11}$$

From Theorem 9 we have $|\nabla u|^2 \leq 4(u_M - u(x))$ in Ω, where $u_M = \sup_\Omega u$. Using (11) we get

$$\int_0^{u_M} \frac{du}{2\sqrt{u_M - u}} \leq \int_Q^M dr = |MQ|.$$

Hence

$$\sqrt{u_M} \leq |MQ| \leq \rho,$$

where ρ is the radius of the largest ball contained in Ω.

Note that the following bound was also obtained using similar methods ([15])

$$u_M \leq \frac{\alpha}{\beta}\left[\frac{1}{\cos(\rho\sqrt{\beta})} - 1\right],$$

where $\alpha \geq 1+\sqrt{2}$ and $0<\beta<\pi^2/4\rho^2$.

A lower bound for u_M was given in [54] (in the case Ω convex and bidimensional). Further isoperimetric inequalities as well as bounds in terms of the stress function for the curvature of the level curves u=const are presented in [54].

b). Upper bound for the maximum stress.

An important quantity is the maximum stress $\sigma = \max_{\partial\Omega} |\nabla u|$. Since P_2 and P_3 attain their maximum value on the boundary of Ω and using standard calculations (see [78]) we get,

$$|\nabla u|^2 \leq \sigma \leq \frac{2}{nK(P)} \leq \frac{2}{nK_{min}}, \tag{12}$$

where $K_{min} = \min_{\partial\Omega} K$, K represents the average curvature of $\partial\Omega$ (the curvature if $n=2$) and P is a point on the boundary where P_2 assumes its maximum.

c). Upper bound for the average curvature of $\partial\Omega$.

Integrating (12) over Ω we obtain

$$K(P) \leq \frac{|\partial\Omega|}{n|\Omega|}, \tag{13}$$

where P is defined above, $|\partial\Omega|$ stands for the $n-1$ dimensional measure and $|\Omega|$ stands for the n dimensional measure.

Equation (13) tells us that at a point of maximum stress, the boundary must be sufficiently flat.

d). Upper bound for the torsional rigidity.

The torsional rigidity of Ω is $T = 2\int_\Omega u dx = \int_\Omega |\nabla u|^2 dx$. We have the following bound:

$$T \leq \frac{4}{3}|\Omega| u_M \leq \frac{4}{3}|\Omega|\rho^2.$$

e). An overdetermined St. - Venant problem.

We consider the problem (10) overdedermined by the boundary condition

$$K|\nabla u|^3 = \text{const.} > 0 \text{ on } \partial\Omega, \tag{14}$$

where Ω is a simply connected domain in $\mathbb{R}^2$ and K the curvature of $\partial\Omega$.

Makar-Limanov ([34]) introduced the function

$$\mathrm{P}_4 = u_{ij}u_iu_j - |\nabla u|^2 \Delta u + u((\Delta u)^2 - u_{ij}u_{ij}),$$

(u is a solution of St. - Venant problem (10)) and showed that it satisfies a maximum principle. A consequence is the convexity of level lines $\{u = \text{const.}\}$. Moreover, we have $\mathrm{P} = K|\nabla u|^3 \geq 0$ on $\partial\Omega$. and P_4 is constant in Ω if and only if Ω is an ellipse. The next theorem (Henrot and Philippin, [32]) tells us that ellipses are the only domains for which condition (14) holds.

Theorem 11. *The over determined problem (10), (14) is solvable only if Ω is an ellipse.*

The proof follows from the implication:

$$\mathrm{P}_4 = \text{const. on}\, \partial\Omega \Rightarrow \mathrm{P}_4 = \text{const. in}\, \overline{\Omega}.$$

Standard methods of investigation for overdetermined problems may not work (Serrin's moving plane method). In this case, we can take advantage of the P function method.

2. The membrane problem.

We are concerned now with eigenvalues of elastically supported membrane problem:

$$\begin{cases} \Delta u + \lambda u = 0 & \text{in}\, \Omega \subset \mathbb{R}^2 \\ \partial u/\partial n + \alpha u = 0 & \text{on}\, \partial\Omega, \end{cases} \tag{15}$$

where $\partial/\partial n$ is the outward normal derivative operator, α is a positive constant and is Ω simply connected, smooth and convex .

If α is large, Payne and Schaefer [53] derived a lower bound for the first eigenvalue λ_1

$$\lambda_1 > \rho^{-2}(\tan^{-1}(\alpha/\sqrt{\Lambda_1}))^2, \tag{16}$$

using that the P function $\mathrm{P}_5 = |\nabla u_1|^2 + \lambda_1 u_1^2$ takes its maximum either on $\partial\Omega$ or at an interior point at which $\nabla u = 0$. Here u_1 represents the first eigenfunction and ρ the radius of the largest inscribed disc. We see that the bound (16) involves Λ_1, the first eigenvalue for the problem,

$$\begin{cases} \Delta v + \Lambda v = 0 & \text{in}\, \Omega \\ v = 0 & \text{on}\, \partial\Omega. \end{cases} \tag{17}$$

If necessary, we can use upper bounds for Λ_1. A known bound for convex regions was given by Hersch [28]

$$\Lambda_1 \geq \frac{\pi^2}{4\rho^2}. \tag{18}$$

On the other hand if α is small we have

$$\lambda_1 > \rho^{-2}(\tan^{-1}(\alpha A/L)^{1/2})^2, \tag{19}$$

where L is the perimeter of Ω and A its area.

Bounds for eigenvalue of (15) have been previously obtained by Sperb [77], [78], [79], Payne and Weinberger [55].

Bounds for the first positive eigenvalue in the free membrane problem

$$\begin{cases} \Delta u + \mu u = 0 & \text{in } \Omega \subset \mathbb{R}^2 \\ \frac{\partial u}{\partial n} = 0 & \text{on } \partial\Omega, \end{cases}$$

are discussed in the book of Sperb [78].

3). A classical problem of electrostatics.

We consider the exterior Dirichlet problem

$$\begin{cases} \Delta u = 0 & \text{in } \Omega^* \equiv \mathbb{R}^3 - \overline{\Omega} \\ u = 1 & \text{on } \partial\Omega \\ u = O(\frac{1}{r}) \text{ as } r \to \infty. \end{cases} \tag{20}$$

u is the electrostatic potential of the conductor and r measures the distance from some origin inside Ω.

The following useful result was proven by Payne and Philippin [49].

Theorem 12. *Let H and h be harmonic functions in Ω, where $H \in C^1(\Omega)$, $h \in C^0(\Omega)$ and let $f(h)$ be a positive C^2 function. Assume that f satisfies*

$$[f^{n-2/2(n-1)}]'' \leq 0, \text{ if } n \geq 3,$$

$$[\log f]'' \leq 0, \text{ if } n = 2.$$

Then the function

$$\mathrm{P}_6 = \frac{\nabla H \cdot \nabla H}{f(h)},$$

assumes its maximum on $\partial\Omega$.

Theorem 12 tells that the function

$$\mathrm{P}_6 = \frac{\nabla u \cdot \nabla u}{u^4}, x \in \Omega^*$$

satisfies

$$\mathrm{P}_6 \leq \max_{\partial\Omega} \mathrm{P}_6, \tag{21}$$

with equality if Ω is a sphere. Moreover

$$C^{-2} \leq \max_{\partial\Omega} \mathrm{P}_6, \tag{22}$$

with equality if Ω is a sphere where, C is the capacity $C = \int_{\Omega^*} |\nabla u|^2 dx$.

At the point $P_0 \in \partial\Omega$, where P_6 assumes its maximum it follows from Hopf's lemma (lemma 1) that either Ω is a sphere or

$$\frac{\partial \mathrm{P}_6}{\partial n} = 2\frac{\partial u}{\partial n}\frac{\partial^2 u}{\partial n^2} - 4\Big(\frac{\partial u}{\partial n}\Big)^2 > 0.$$

For a smooth hypersurface S in $\mathbb{R}^n$ we have on S the relation (see [78], p.62)

$$\Delta u = \Delta_s u + (n-1)K\frac{\partial u}{\partial n} + \frac{\partial^2 u}{\partial n^2}, \tag{23}$$

where $\Delta_s u$ is the Laplacian in the induced metric of S and K the mean curvature (the curvature if $n = 2$).

From (23) we obtain on $\partial\Omega$

$$\frac{\partial^2 u}{\partial n^2} = 2K\frac{\partial u}{\partial n}.$$

Now it follows that

$$\frac{\partial u(P_0)}{\partial n} < K(P_0).$$

Since P_6 and $\frac{\partial u}{\partial n}$ take their maximum at the same point on the boundary it follows that either Ω is a sphere or

$$\max_{\partial\Omega} \frac{\partial u}{\partial n} < K(P_0) < \max_{\partial\Omega} K \equiv K_0. \tag{24}$$

A bound for the capacity C follows now from (22) and (24)

$$C \geq K_0^{-1}, \tag{25}$$

where the equality sign holds if Ω is a sphere.

An upper bound for the capacity is also given in [49]:

$$C \leq \frac{3|\Omega| K_0^2}{4\pi}, \tag{26}$$

where the equality sign holds if Ω is a sphere ($|\Omega| = \mathrm{vol}(\Omega)$).

Bounds for the derivatives of Green's function are also a consequence of Theorem 12. See for details [49].

4). Estimates for capillary free surfaces without gravity.

In the paper [36], Ma studied (using the P function method) the influence of boundary geometry and constant contact angle θ_0, $0 \leq \theta_0 < \pi/2$ (against the wall of the tube) on the size and shape for the capillary free surface without gravity.

Let Ω be a bounded, smooth and convex domain in $\mathbb{R}^2$ and let $K = \cos\theta_0 |\partial\Omega|/2|\Omega|$ be a given constant.

Consider the problem:

$$\begin{cases} \left(\frac{u_i}{\sqrt{1+|\nabla u|^2}}\right)_i = 2K & \text{in } \Omega \\ \partial u/\partial n = \cos\theta_0 \sqrt{1+|\nabla u|^2} & \text{on } \partial\Omega, \end{cases} \tag{27}$$

where $\partial u/\partial n$ denoted the directional derivative of u along the outer unit normal.

The graph of solution u of (27) describes a capillary free surface (having the nonparametric form $x_3 = u(x_1, x_2), (x_1, x_2) \in \Omega$) without gravity over the cross section Ω. We have the following result (Xi- Nan Ma)

Theorem 13. *If $u \in C^3(\Omega)$ is a solution of (27), then*

$$u(A) - u(C) \leq \frac{1-\sin\theta_0}{K}, \ k(A) \leq \frac{K}{\cos\theta_0} \tag{28}$$

$$u(B) - u(C) \geq \frac{1-\sin\theta_0}{K}, \ k(B) \geq \frac{K}{\cos\theta_0}. \tag{29}$$

If the equality sign holds in (28) and (29), then Ω is a disk of radius $\cos\theta_0/K$.

Here $A \in \partial\Omega$ is a point that corresponds to the minimum boundary value of u, $B \in \partial\Omega$ is a point that corresponds to the maximum boundary value of u and $C \in \Omega$ is the unique critical point of u.

If $S = \int_\Omega \sqrt{1 = |\nabla u|^2} dx$ is the area of the free capillary surface and $V = \int_\Omega u dx$ the volume of the liquid in the vertical tube, then we have the bound:

Theorem 14.

$$(\sin\theta_0 + 3Ku(A))|\Omega| - 3KV \leq S \leq (\sin\theta_0 + 3Ku(B))|\Omega| - 3KV.$$

Here and in the above mentioned result $A \in \partial\Omega$ is a point where u assumes its minimum on $\partial\Omega$, $B \in \partial\Omega$ is a point where u assumes its maximum on $\partial\Omega$, C is the unique critical point of u and K is the curvature of $\partial\Omega$.

The proofs follow from

Theorem 15. *If $u \in C^3(\Omega)$ is a solution of (27), then the function*

$$\mathrm{P}_7 = 2 - 2Ku - 2(1 + |\nabla u|^2)^{-\frac{1}{2}}$$

attains its minimum on the boundary of Ω.

Similar problems are treated in the paper of Payne and Philippin [46], e.g. equation of a surface of constant mean curvature, equation of the fluid in a capillary tube, equation of thin extensible film under the influence of gravity and surface tension. The authors obtain various bounds in terms of boundary data and geometry of Ω.

5). Equations of Monge - Ampère type.

We consider a class of Monge - Ampère equations

$$\det D^2 u = f(x, u, \nabla u) \tag{30}$$

with a prescribed contact angle boundary value on a bounded convex domain in two dimensions.

$$\partial u/\partial n = \cos\theta(x, u)\sqrt{1 + |\nabla u|^2} \quad \text{on } \partial\Omega,$$

where $D^2 u$ is the hessian matrix and $\theta(x, u) \in (0, \pi/2)$ is the wetting angle.

The existence of solutions for such boundary value problems is still open. Even the particular case is untreated in the literature.

$$\begin{cases} \det D^2u = c & \text{in } \Omega \subset \mathbb{R}^2 \\ \partial u/\partial n = \cos\theta_0\sqrt{1+|\nabla u|^2} & \text{on } \partial\Omega, \end{cases} \tag{31}$$

where Ω is convex, $c > 0$ is a constant and $\theta_0 \in (0, \pi/2)$.

Ma Xi-nan [35] gave a necessary condition of solvability for the problem (31).

Theorem 16. *Let $u \in C^2(\overline{\Omega}) \cap C^3(\Omega)$ be a strictly convex solution of problem (31). Under the above stated hypotheses on Ω, c, θ_0 we must have the relation*

$$K_0 \leq \max\{\sqrt{c}\cdot\cos\theta_0, \sqrt{c}\cdot\tan\theta_0\},$$

where

$$K_0 = \min_{\partial\Omega} K > 0$$

and K is curvature of $\partial\Omega$.

The proof is achieved by using the P function $\mathrm{P}_8 = |\nabla u|^2 - 2\sqrt{c}u$ (which satisfies a maximum principle) and introducing a curvilinear coordinate system.

Bounds for solutions and gradient of general Monge - Ampère equations (30) are presented in the work of Philippin and Safoui [58].

II. The higher order case

Miranda [39] was the first that showed that for the biharmonic equation $\Delta^2 u = 0$, where $u \in C^4(\Omega) \cap C^2(\overline{\Omega})$ is a function defined on a bounded plane domain the function $\mathrm{P}_9 = |\nabla u|^2 - u\Delta u$ takes its maximum value on the boundary of the domain, i.e.,

$$\max_{\overline{\Omega}} \mathrm{P}_9 = \max_{\partial\Omega} \mathrm{P}_9.$$

Since then many authors have extended the Miranda's result. For example, maximum principles for fourth order equations containing nonlinearities in u or Δu can be found in works of Payne [44], Schaefer [67], [70],[71]. Similar results are proved by H. Zhang and W. Zhang [84], Mareno [37], [38] (studied some equations from plate theory), Danet [8], Tseng and Lin [80], [10], [11] etc. (see the references). We will list only a few as an indication of the types of results that can be obtained.

1). Equations of fourth order arising in plate theory.

a). Von Kármán equations.

Assume that Ω is a bounded domain in the plane. We consider the von Kármán equations:

$$\begin{cases} \Delta^2\phi = -\frac{1}{2}[w,w] & \text{in } \Omega \\ \Delta^2 w = [w,\phi] + f(x,y) & \text{in } \Omega. \end{cases} \tag{32}$$

The equations (32) govern the equilibrium configuration of a thin elastic plate under stress. $f(x,y)$ represents nonconstant perpendicular loading terms. The function w denotes the deflection of the thin plate and ϕ represents the stress function. The operator $[\cdot,\cdot]$ is defined as follows:

$$[w,\phi] = w_{xx}\phi_{yy} - 2w_{xy}\phi_{xy} + w_{yy}\phi_{xx}.$$

Mareno [38] proved (the first that proved a maximum principle for such equations) that the P function

$$\mathrm{P}_{10} = |\nabla^2\phi|^2 + |\nabla^2 w|^2 - \phi_i\Delta\phi_i - w_i\Delta w_i + h(x,y)[|\nabla w|^2 + |\nabla\phi|^2] + f^2(x,y)$$

satisfies a maximum principle and as a consequence obtained the following bound:

$$\frac{2}{|\Omega|}\int_\Omega \Big(|\nabla^2\phi(x,y)|^2 + |\nabla^2 w(x,y)|^2\Big)dxdy \;\le\; |\nabla^2\phi(x_0,y_0)|^2 + |\nabla^2 w(x_0,y_0)|^2 + f^2(x_0,y_0),$$

for some point (x_0,y_0) on $\partial\Omega$, if $\phi = w = \partial\phi/\partial n = \partial w/\partial n = 0$ on $\partial\Omega$. Here $|\nabla^2 w| = w_{ij}w_{ij}$, and $h(x,y)$ is a smooth function.

b). An equation arising in plate theory.

We deal with the following equation

$$\Delta^2 u + k_1 u + k_2 u^3 = 0 \quad \text{in } \Omega \subset \mathbb{R}^{\mathrm{n}},\ \mathrm{n} \ge 2, \tag{33}$$

where $k_1, k_2 > 0$ are constants.

The equation (33) arises in the plate theory and in the bending of cylindrical shells [67].

The next maximum principle ([10]) will be used to obtain solution and gradient bounds for the equation (33)

Theorem 17. *Let u be a classical solution of (33). Then the function*

$$\mathrm{P}_{11} = (\Delta u)^2 + \frac{k_2}{2}u^4 + k_1 u^2$$

attains its maximum value on $\partial\Omega$.

If u satisfies (33) then, we have the following bounds

a).

$$\max_{\overline{\Omega}} |u| \leq \sqrt{\frac{1}{k_1}}\left(\max_{\partial\Omega} |\Delta u| + \sqrt{\frac{k_2}{2}} \max_{\partial\Omega} u^2 + \sqrt{k_1} \max_{\partial\Omega} |u| \right), \tag{34}$$

where $n \geq 2$.

b).

$$\max_{\overline{\Omega}} |\nabla u|^2 \leq \max_{\partial\Omega} |\nabla u|^2 + \frac{3+k_1}{2} \max_{\partial\Omega} u^2 + \frac{k_2}{2k_1} \max_{\partial\Omega} u^4 + \frac{2k_1+1}{2k_1} \max_{\partial\Omega} (\Delta u)^2, \tag{35}$$

where $n = 2$.

The hypothesis that is assumed over and over again in plate theory is convexity. Under this assumption, Schaefer [67] proved the uniqueness for the solution of

$$\begin{cases} \Delta^2 u + k_1 u + k_2 u^3 = 0 & \text{in } \Omega \\ u = \Delta u = 0 & \text{on } \partial\Omega, \end{cases} \tag{36}$$

where $\Omega \subset \mathbb{R}^2$ is a convex domain.

An application the maximum principle (Theorem 17) shows that the convexity assumption is redundant. Moreover, our result holds for $n \geq 2$.

The result reads as follows:

Theorem 18. *Let u be a classical solution of (36), where $\Omega \subset \mathbb{R}^n$ is an arbitrary domain. Then $u \equiv 0$ in Ω.*

Maximum principles for fourth and six order equations are presented in the author's paper [10] and [11].

2). The $m\,(> 4)$ order case.

We conclude this paper with a result for the general case due to the author [11].

Theorem 19. *Let u be a classical solution of equation*

$$\Delta^m u + a_0 u = 0$$

in Ω, where $\Omega \subset \mathrm{I\!R}^n, m$ even, $n \geq 2$.

Suppose that $a_0 > 0$, $\Delta a_0 \leq 0$ in Ω.

We define the function P_{10}

$$\mathrm{P}_{12} = \Big((\Delta^{m-1}u)^2 + (\Delta^{m-2}u)^2 + \cdots + u^2\Big)/a_0.$$

a). If

$$\max\{1 + \sup_{\Omega} a_0, 2\} + \sup_{\Omega} \frac{\Delta a_0}{a_0} \leq 0, \tag{37}$$

then, the function P_{12} attains its maximum value on $\partial\Omega$.

b). If

$$\max\{1 + \sup_{\Omega} a_0, 2\} + \sup_{\Omega} \frac{\Delta a_0}{a_0} < \frac{4}{d^2 e^2} \tag{38}$$

and if there exists $i \in \{1, \ldots, n\}$ such that $\frac{\partial}{\partial x_i}\big(\frac{1}{a_0}\big) \geq 0$ in Ω, then, the function P_{12}/w_3 attains its maximum value on $\partial\Omega$, where $w_3 = 1 - \beta e^{\alpha x_i}, \beta = \sup_{\Omega} c/\alpha^2$ and $\alpha > 0$ is a constant.

The proof follows from the generalized maximum principle, Theorem 7 and works also for the case m odd.

As an immediate consequence of the above mentioned maximum principle we obtain the uniqueness of the classical solution $(C^{2m}(\Omega) \cap C^{2m-2}(\overline{\Omega}), m \geq 3)$ of the boundary value problem

$$\begin{cases} \Delta^m u + (-1)^m a_0(x) u = f & \text{in } \Omega \\ u = g_1,\ \Delta u = g_2, \ldots,\ \Delta^{m-1} u = g_m & \text{on } \partial\Omega. \end{cases} \tag{39}$$

Moreover the following classical maximum principle holds for solutions of (39), if $g_2 = \cdots = g_m = 0$ on $\partial\Omega$ and $f = 0$ in Ω.

$$\max_{\overline{\Omega}} |u| \leq C \max_{\partial\Omega} |u|, \tag{40}$$

where $C > 1$ is a constant.

Note that the problem was solved for a more general problem, but under the restriction Ω is of class C^2 (see [73]).

Final remark. Below we collected many papers concerning the P function method for the interested reader (not all are quoted in this paper).

Acknowledgement. This research was supported by the Ministry of Education and Research Romania under Grant no. ID - PCE 1192 - 09.

References

[1] N. Anderson, A. M. Arthurs, Upper and lower bounds for the torsional stiffness of a prismatic bar, *Proc. Roy. Soc. London Ser. A* **328**:295-299, 1972.

[2] M. Berger, Geometry I, *Springer - Verlag*, 1987.

[3] A. Bennett, Symmetry in an overdetermined fourth order elliptic boundary value problem, *Siam J. Math. Anal.* **17**:1354-1358, 1986.

[4] M. Chicco, Principio di massimo forte per sottosoluzioni di equazioni ellitiche di tipo variazionale, *Boll. Un. Mat. Ital.* **22**:368-372, 1967.

[5] S. N. Chow, D. R. Dunninger, A maximum principle for n-metaharmonic functions, *Proc. Amer. Math. Soc.* **43**:79–83, 1974.

[6] S. N. Chow, D. R. Dunninger and A. Lasota, A maximum principle for fourth order ordinary differential equations, functions, *J. Differential Equations* **14**:101-105, 1973.

[7] C. Cosner, P. W. Schaefer, On the development of functionals which satisfy a maximum principle, *Appl. Anal.* **26**:45-60, 1987.

[8] C. - P. Danet, Uniqueness results for a class of higher - order boundary value problems, *Glasgow Math. J.* **48**: (2006), 547–552.

[9] C. - P. Danet, On the elliptic inequality L$u \leq 0$, *Math. Inequal. & Appl.* **11**:559-562, 2008.

[10] C. - P. Danet, Uniqueness in some higher order elliptic boundary value problems in n dimensional domains, El.J. Qualitative Th. Diff. Eq., no.54 (2011), p.1-12; http://www/math.u-szeged.hu/ejqt.de/.

[11] C. - P. Danet, On a m metaharmonic boundary value problem, submitted.

[12] D. R. Dunninger, Maximum principles for solutions of some fourth - order elliptic equations, *J. Math. Anal. Appl.* **37**: 655–658, 1972.

[13] D. R. Dunninger, Maximum principles for fourth order ordinary differential inequalities, *J. Math. Anal. Appl.* **82**:399-405, 1981.

[14] R. J. Duffin, On a question of Hadamard concerning super-biharmonic functions, *J. Math. Phys.* **27**:253-258, 1949.

[15] C. Enache and G. A. Philippin, Some maximum principles for a class of elliptic boundary value problems, *Math. Ineq. & Appl.*, **9**:695-706, 2006.

[16] L. E. Fraenkel, An introduction to Maximum Principles and Symmetry in Elliptic Problems, *Cambridge University Press*, 2000.

[17] C. F. Gauss, Allgemeine Theorie des Erdmagnetismus. Beobachtungen des magnetischen Vereins im Jahre 1838, *Leipzig*, 1839.

[18] S. Goyal and V. B. Goyal, Liouville - type and uniqueness results for a class of sixth-order elliptic equations, *J. Math. Anal. Appl.*, **139**:586–599, 1989.

[19] V. B. Goyal, Liouville - type results for fourth order elliptic equations, *Proc. Roy. Soc. Edinburgh A*, **103**:209-213, 1986.

[20] S. Goyal, V. B. Goyal, Liouville - type and uniqueness results for a class of sixth - order elliptic equations, *J. Math. Anal. Appl.*, **139**:586-599, 1989.

[21] S. Goyal, V. B. Goyal, Liouville - type and uniqueness results for a class of elliptic equations, *J. Math. Anal. Appl.* **151**:405-416, 1990.

[22] V. B. Goyal, P. W. Schaefer, Liouville theorems for elliptic systems and nonlinear equations of fourth order, *Proc. Roy. Soc. Edinburgh A*, **91**:235-242, 1982.

[23] V. B. Goyal, P. W. Schaefer, Comparison principles for some fourth order elliptic problems, in *Lecture Notes in Mathematics*, **964**, 272-279, Springer, 1982.

[24] V. B. Goyal, K. P. Singh, Maximum principles for a certain class of semi - linear elliptic partial differential equations, *J. Math. Anal. Appl.* **69**:1-7, 1979.

[25] D. Gilbarg, N. S. Trudinger, Elliptic Partial Differential Equations of Second Order, *Classics in Mathematics, Springer*, 2001.

[26] J. Hadamard, Mémoire sur le problème d'analyse relatif à l'équilibre des plaques élastiques encastrées, in:*Oeuvres de Jacques Hadamard, Tome II*, 515-641, Centre National de la Recherche Scientifique: Paris, 1968, Reprint of: *Mémoires présentés par divers savants à l'Académie des Sciences* **33**:1-128, 1908.

[27] Z. Hailiang, Maximum principles for a class of semilinear elliptic boundary - value problems, *Bull. Austral. Math. Soc.* **52**:169-176, 1995.

[28] J. Hersch, Sur la fréquence fondamentale d'une membrane vibrante:évaluations par défault et principe de maximum, *Z. angew. Math. Phys.* **11**:387-413, 1960.

[29] G. N. Hille, C. Zhou, Liouville theorems for even order elliptic systems, *Indiana Univ. Math. J.* **43**:383-410, 1994.

[30] E. Hopf, Elementare Bemerkungen über die Lösungen partieller Differential gleichungen zweiter Ordnung vom elliptischen Typus, *Sitz. Ber. Preuss. Akad. Wissensch. Berlin, Math.-Phys. Kl* **19**:147-152, 1927.

[31] E. Hopf, A remark on linear elliptic differential equations of second order, Proc. *Amer. Math. Soc.* **3**:791-793, 1952.

[32] A. Henrot and G. A. Philippin, Some overdetermined boundary value problems with elliptical free boundaries, *SIAM J. Math. Anal.* **28**:309-320, 1998.

[33] G. N. Hille, M. H. Protter, Maximum principles for a class of first - order elliptical systems, *J. Diff. Eq.* **24**:136-151, 1977.

[34] L. G. Makar-Limanov, Solutions of Dirichlet problem for the equation $\Delta u = -1$ in a convex region, *Math. Notes Acad. Sci. URSS*, **9**:52-53, 1971.

[35] X. Ma, A necessary condition of solvability for the capillarity boundary of Monge - Ampère equations in two dimensions, *Proc. Amer. Mat. Soc.* **127**:763-769, 1999.

[36] X. Ma, Sharp size estimates for capillary free surfaces without gravity, *Pacific J. Math.*, **192**:121-134, 2000.

[37] A. Mareno, Maximum principles for a fourth order equation from plate theory, *J. Math. Anal. Appl.* **343**:932-937, 2008.

[38] A. Mareno, Integral Bounds for von Kármán's equations, *Z. Angew. Math. Mech.* **90**:509-513, 2010.

[39] C. Miranda, Formule di maggiorazione e teorema di esistenza per le funzioni biarmoniche di due variabili, *Giorn. Mat. Battaglini* **78**:97-118, 1948.

[40] C. Miranda, Teorema del massimo modulo e teorema di esistenza e di unicita per il problema di Dirichlet relative alle equazioni ellitiche in due variabili, *Ann. Mat. Pura Appl.* **46**:265-311, 1958.

[41] A. Mohammed, G. Porru, Maximum principles for ordinary differential inequalities of fourth and sixth order, *J. Math. Anal. Appl.* **146**:408-419, 1990.

[42] M. A. Paraf, Sur le problème de Dirichlet et son extension au cas de l'équation linéare générale du second ordre, *Ann. Fac. Sci. Toulouse* **6**, Fasc. 47-Fasc. 54, 1982

[43] L. E. Payne, Bounds for the maximal stressin the Saint Venant torsion problem, *Indian J. Mech. Math.* special issue,51–59, 1968.

[44] L. E. Payne, Some remarks on maximum principles, *J. Analyse Math.* **30**:421-433, 1976.

[45] L. E. Payne, G. A. Philippin, Some applications of the maximum principle in the problem of torsional creep, *Siam. J. Appl. Math.* **33**:446-455, 1977.

[46] L. E. Payne, G. A. Philippin, Some maximum principles for nonlinear elliptic equations in divergence form with applications to capillary surfaces and to surfaces of constant mean curvature, *Nonlinear Analysis* **3**:193-211, 1979.

[47] L. E. Payne, G. A. Philippin, On maximum principles for a class of nonlinear second order elliptic equations, *J. Diff. Eq.* **37**: 39-48, 1980.

[48] L. E. Payne, G. A. Philippin, On gradient maximum principles for quasilinear elliptic equations, *Nonlinear Analysis* **23**:387-398, 1994.

[49] L. E. Payne, G. A. Philippin, On some maximum principles involving harmonic functions and their derivatives, *SIAM J. Math. Anal.* **10**:96-104, 1979.

[50] L. E. Payne, P. W. Schaefer On overdetermined boundary value problems for the biharmonic operator, *J. Math. Anal. Appl.* **187**:598-616, 1994.

[51] L. E. Payne, R. P. Sperb, I. Stakgold, On Hopf type maximum principles for convex domains, *Nonlinear Analysis* **1**:547-559, 1977.

[52] L. E. Payne, I. Stakgold, Nonlinear problems in nuclear reactor analysis, *Springer Lecture Notes in Mathematics*, **322**, 1973.

[53] L.E. Payne, P.W. Schaefer, Eigenvalue and eigenfunction inequalities for the elastically supported membrane, *Z. angew. Math. Phys.* **52**: 888-895, 2001.

[54] L.E. Payne and G. A. Philippin, Isoperimetric inequalities in the torsion and clamped membrane problems for convex plane regions, *SIAM J. Math. Anal.* **14**:1154-1162, 1983.

[55] L.E. Payne and H.F. Weinberger, Lower bounds for vibration frequencies of elastically supported membranes annd plates, *J. Soc. Ind. Appl. Math.* **5**:171-182, 1957.

[56] G. Philippin, Some remarks on the elastically suported membrane, *Z.A.M.P.* **29**:306-314, 1978.

[57] G. A. Philippin, L. E. Payne, On the conformal capacity problem, in Geometry of Solutions to Partial Differential Equations (Edited by G. Talenti), *Symposia Mathematica*, 119-136, Academic Press, 1989.

[58] G. A. Philippin, A. Safoui, Some maximum principles and symmetry results for a class of boundary value problems involving the Monge - Ampère equation, *Math. Mod. Meth. Appl. Sci.* **11**:1073-1080, 2001.

[59] G. A. Philippin, A. Safoui, Some minimum principles for a class of elliptic boundary value problems, *Appl. Anal.*, **83**: 231-241, 2004.

[60] G. A. Philippin, A. Safoui, Some applications of the maximum principle to a variety of fully nonlinear elliptic PDE's, *Z. angew. Math. Phys.* **54**:739-755, 2003.

[61] M. H. Protter, H. F. Weinberger, Maximum Principles in Differential Equations, *Prentice Hall Inc.*, 1967.

[62] P. Pucci, J. Serrin, A note on the strong maximum principle for elliptic differential inequalities, *J. Math. Pures Appl.* **79**:57-71, 2000.

[63] G.Porru, A. Safoui and S. Vernier-Piro, Best possible maximum principles for fully nonlinear elliptic partial differential equations, *J. Anal. and Appl.* **25**:421-434, 2006.

[64] P. Pucci, J. Serrin, H. Zou, A strong maximum principle and a compact support principle for singular elliptic inequalities, *J. Math. Pures Appl.* **78**:769-789, 1999.

[65] M. H. Protter and H. F. Weinberger, Maximum principles in differential equations, *Prentice - Hall, Inc., Englewood Cliffs, N. J.*, 1967.

[66] P. Quittner and P. Souplet, Superlinear parabolic problems, *Birkhäuser*, 2007.

[67] P. W. Schaefer, On a maximum principle for a class of fourth - order semilinear elliptic equations, *Proc. Roy. Soc. Edinburgh Sect. A* **77**:319-323, 1977.

[68] P. W. Schaefer, Uniqueness in some higher order elliptic boundary value problems, *Z. Angew. Math. Mech.* **29**:693-697, 1978.

[69] P. W. Schaefer, Some maximum principles in semilinear elliptic equations, *Proc. Amer. Math. Soc.* **98**:97-102, 1986.

[70] P. W. Schaefer, Pointwise estimates in a class of fourth - order nonlinear elliptic equations, *Z. A. M. P.* **38**:477-479, 1987.

[71] P. W. Schaefer, Solution, gradient, and laplacian bounds in some nonlinear fourth order elliptic equations, *Siam J. Math. Anal.* **18**:430-434, 1987.

[72] P. W. Schaefer (Editor), Maximum principles and eigenvalue problems in partial differential equations, *Longman Scientific and Technical*, 1988.

[73] P. W. Schaefer, W. Walter, On pointwise estimates for metaharmonic functions, *J. Math. Anal. Appl.* **69**:171-179, 1979.

[74] J. Serrin, The problem of Dirichlet for quasilinear elliptic differential equations with many independent variables, *Philos. Trans. Roy. Soc. London Ser. A* **264**:413-496, 1969.

[75] J. B. Serrin, A symmetry problem in potential theory, *Arch. Rat. Mech. Anal.* **43**:304-318, 1971.

[76] Q. Sheng and R. P. Argawal, Generalized maximum principle for higher order ordinary differential equations, *J. Math. Anal.*, **174**:476-479, 1993.

[77] R. P. Sperb, Untere un obere Schranken für den tiefsten Eigenwert der elastisch gestützen Membran, *Z. angew. Math. Phys.* **23**:231-244, 1972.

[78] R. P. Sperb, Maximum Principles and Their Applications, *Academic Press*, 1981.

[79] R. P. Sperb, An isoperimetric inequality for the first eigenvalue of the Laplacian under Robin boundary conditions, *General Inequalities* **6**, ISNM 103, 361-367, *Birkhäuser*, Basel 1992.

[80] S. Tseng, C-S. Lin, On a subharmonic functional of some even order elliptic problems, *J. Math. Anal. Appl.* **207**:127- 157, 1997.

[81] J. Urbas, Nonlinear oblique boundary value problem for Hessian equations in two dimensions, *Ann. Inst. Henri Poincaré*, **12**:507-575, 1995.

[82] J. R. L. Webb, Maximum principles for functionals associated with the solution of semilinear elliptic boundary value problems, *Z.A.M.P.* **40**:330-338, 1989.

[83] H. F. Weinberger, Remark on a preceding paper of Serrin, *Arch. Rat. Mech. Anal.* **43**:319-320, 1971.

[84] H. Zhang, W. Zhang, Maximum principles and bounds in a class of fourth - order uniformly elliptic equations, *J. Phys. A: Math. Gen.* **35**:9245-9250, 2002.

Annals of the Academy of Romanian Scientists
Series on Mathematics and its Applications
ISSN 2066 - 6594 Volume 3, Number 2 / 2011

MULTIGRID METHODS WITH CONSTRAINT LEVEL DECOMPOSITION FOR VARIATIONAL INEQUALITIES*

Lori Badea†

Abstract

In this paper we introduce four multigrid algorithms for the constrained minimization of non-quadratic functionals. These algorithms are combinations of additive or multiplicative iterations on levels with additive or multiplicative ones over the levels. The convex set is decomposed as a sum of convex level subsets, and consequently, the algorithms have an optimal computing complexity. The methods are described as multigrid V-cycles, but the results hold for other iteration types, the W-cycle iterations, for instance. We estimate the global convergence rates of the proposed algorithms as functions of the number of levels, and compare them with the convergence rates of other existing multigrid methods. Even if the general convergence theory holds for convex sets which can be decomposed as a sum of convex level subsets, our algorithms are applied to the one-obstacle problems because, for these problems, we are able to construct optimal decompositions. But, in this case, the convergence rates of the methods introduced in this paper are better than those of the methods we know in the literature.

AMS subject classification: 65N55, 65N30, 65J15

Keywords: domain decomposition methods, variational inequalities, multigrid and multilevel methods

---

*Accepted for publication in revised form on March 3, 2011.

†Institute of Mathematics of the Romanian Academy, P.O. Box 1-764, RO-014700 Bucharest, Romania (e-mail: lori.badea@imar.ro)

1 Introduction

The multigrid or multilevel methods for the constrained minimization of functionals have been studied almost exclusively for the complementarity problems. Such a method has been proposed by Mandel in [22], [23] and [11]. Related methods have been introduced by Brandt and Cryer in [8] and Hackbush and Mittelmann in [14]. The method has been studied later by Kornhuber in [16] and extended to variational inequalities of the second kind in [17] and [18]. A variant of this method using truncated nodal basis functions has been introduced by Hoppe and Kornhuber in [15] and analyzed by Kornhuber and Yserentant in [20]. Also, versions of this method have been applied to Signorini's problem in elasticity by Kornhuber and Krause in [19] and Wohlmuth and Krause in [27]. Evidently, the above list of citations is not exhaustive and, for further information, we recommend the review article [13] written by Gräser and Kornhuber. For the two-level method, global convergence rates have been established by Badea, Tai and Wang in [7], and for its additive variant by Badea in [3]. A global convergence rate has been also estimated by Tai in [24] for a subset decomposition method.

In [2], a projected multilevel method has been introduced for the constrained minimization of non quadratic functionals. The convex set may be a little more general than of one- or two-obstacle type. The drawback of this method is its sub-optimal computing complexity because the convex set, which is defined on the finest mesh, is used in the smoothing steps on the coarse levels. Multigrid methods with optimal computing complexity have been introduced in [4] (see also, [5]) for the two-obstacle problems. In these algorithms, the convex level sets are recursively constructed for each smoothing step of the iterations. In the present paper, we introduce four multilevel algorithms in which the convex set is decomposed as a sum of convex level subsets. These algorithms, like those introduced in [4], have an optimal computing complexity, and are combinations of additive or multiplicative iterations on levels with additive or multiplicative ones over the levels. Even if the general convergence theory holds for convex sets which can be decomposed as a sum of convex level subsets, these algorithms are applied for the constrained minimization problems of the one-obstacle type. To our knowledge, optimal decompositions as sums of convex level sets for more general convex sets (two-obstacle convex sets, for instance) is an open problem. The methods are described as multigrid V-cycles, but the results hold for W-cycle iterations, for instance.

Regarding the convergence study of the classical multigrid method, an estimate of the asymptotic convergence rate of $1-1/(1+CJ^3)$, J being the number of levels, has been proved by Kornhuber in [16] for the complementarity problems in the bidimensional space. For these problems, the same estimate, but for the global convergence rate, is obtained for the methods in [4] which are of the multiplicative type over the levels. The methods in that paper which are of the additive type over the levels have a global convergence rate of $1-1/(1+CJ^4)$. The global convergence rates of the methods introduced in this paper are better than those of the methods in [4]. We found, for the complementarity problems in $\mathbf{R}^2$, that the convergence rate of the methods which are of the multiplicative type over the levels is of $1-1/(1+CJ^2)$, and of $1-1/(1+CJ^3)$ for the methods of additive type over the levels.

The paper is organized as follows. In Section 2, we state four algorithms in a general framework of reflexive Banach spaces, and prove their convergence under some assumptions. In Section 3, we show that these algorithms can be viewed as multilevel methods for the constrained minimization of non quadratic functionals, if we associate finite element spaces to the level meshes and consider decompositions of the domain at each level. We prove that the assumptions made in the previous section hold for convex sets of one-obstacle type. If the decompositions of the domain are made using the supports of the nodal basis functions we get, in Section 4, the multigrid methods. This particular choice of the domain decompositions allows us to obtain better estimates for the convergence rate of the methods.

2 Abstract convergence results

We consider a reflexive Banach space V and $V_1,\ldots,V_J$, are some closed subspaces of V, where $V_J=V$. Let $K\subset V$ be a nonempty closed convex set, and we assume that there exist some closed convex sets $K_j\subset V_j$, $j=1,\ldots,J$ such that

$$K=K_1+\ldots+K_J \tag{2.1}$$

The algorithms we introduce will be combinations of additive or multiplicative algorithms over levels with additive or multiplicative algorithms on each level. To this end, we assume that at each level $1\le j\le J$ we have I_j closed subspaces of V_j, V_{ji}, $i=1,\ldots,I_j$, and we shall write $I=\max\limits_{j\in J} I_j$. Also, for a fixed $\sigma > 1$, we assume that there exists a constant C_1 such that

$$||\sum_{j=1}^{J}\sum_{i=1}^{I_j} w_{ji}|| \le C_1(\sum_{j=1}^{J}\sum_{i=1}^{I_j}||w_{ji}||^\sigma)^{\frac{1}{\sigma}} \tag{2.2}$$

for any $w_{ji} \in V_{ji}$, $j = J, \dots, 1$, $i = 1, \dots, I_j$. Evidently, we can take, for instance,

$$C_1 = (IJ)^{\frac{\sigma-1}{\sigma}} \tag{2.3}$$

but sharper estimations can be available in certain cases. In the case when we use multiplicative algorithms on the levels $1 \le j \le J$, we make the following

ASSUMPTION 2.1. *We assume that there exist two positive constants C_2 and C_3, and that any $w \in K$ can be written as $w = \sum_{j=1}^{J} w_j$, with $w_j \in K_j$, $j = 1, \dots, J$, such that*

- for any $v \in K$,

- and any $w_{ji} \in V_{ji}$ satisfying $w_j + \sum_{k=1}^{i} w_{jk} \in K_j$, $j = 1, \dots, J$, $i = 1, \dots, I_j$,

there exist $v_{ji} \in V_{ji}$, $j = 1, \dots, J$, $i = 1, \dots, I_j$, which satisfy

$$w_j + \sum_{k=1}^{i-1} w_{jk} + v_{ji} \in K_j \text{ for } j = 1, \dots, J,\ i = 1, \dots, I_j,$$

$$v - w = \sum_{j=1}^{J}\sum_{i=1}^{I_j} v_{ji} \text{ and } \sum_{j=1}^{J}\sum_{i=1}^{I_j}||v_{ji}||^\sigma \le C_2^\sigma||v-w||^\sigma + C_3^\sigma\sum_{j=1}^{J}\sum_{i=1}^{I_j}||w_{ji}||^\sigma.$$

If we use additive algorithms on the levels $1 \le j \le J$, we assume

ASSUMPTION 2.2. *We assume that there exists a constant $C_2 > 0$, and that any $w \in K$ can be written as $w = \sum_{j=1}^{J} w_j$, with $w_j \in K_j$, $j = 1, \dots, J$, such that for any $v \in K$,*

there exist $v_{ji} \in V_{ji}$, $j = 1, \dots, J$, $i = 1, \dots, I_j$, which satisfy

$$w_j + v_{ji} \in K_j \text{ for } j = 1, \dots, J,\ i = 1, \dots, I_j,$$

$$v - w = \sum_{j=1}^{J}\sum_{i=1}^{m} v_{ji} \text{ and } \sum_{j=1}^{J}\sum_{i=1}^{I_j}||v_{ji}||^\sigma \le C_2^\sigma||v-w||^\sigma.$$

REMARK 2.1. *In the proofs, for the writing uniformity, we shall consider in Assumption 2.2 a constant* $C_3 = 0$ *and inequality* $\sum_{j=1}^{J}\sum_{i=1}^{I_j}||v_{ji}||^\sigma \leq C_2^\sigma||v-w||^\sigma$ *will be written like in Assumption 2.1,* $\sum_{j=1}^{J}\sum_{i=1}^{I_j}||v_{ji}||^\sigma \leq C_2^\sigma||v-w||^\sigma + C_3^\sigma\sum_{j=1}^{J}\sum_{i=1}^{I_j}||w_{ji}||^\sigma$*, for any* $w_{ji}\in V_{ji}$.

Now, we consider a Gâteaux differentiable functional $F : V \to \mathbf{R}$, which is assumed to be coercive on K, in the sense that $\frac{F(v)}{||v||} \to \infty$, as $||v|| \to \infty$, $v \in K$, if K is not bounded. Also, we assume that there exist two real numbers $p,\ q > 1$ such that $\frac{p}{p-q+1} \leq \sigma \leq p$ and that, for any real number $M > 0$ there exist $\alpha_M,\ \beta_M > 0$ for which

$$\begin{array}{l} \alpha_M||v-u||^p \leq < F'(v)-F'(u), v-u > \text{ and} \\ ||F'(v)-F'(u)||_{V'} \leq \beta_M||v-u||^{q-1} \end{array} \tag{2.4}$$

for any $u, v \in V$ with $||u||, ||v|| \leq M$. Above, we have denoted by F' the Gâteaux derivative of F, and we have marked that the constants α_M and β_M may depend on M. It is evident that if (2.4) holds, then for any $u, v \in V$, $||u||, ||v|| \leq M$, we have

$$\alpha_M||v-u||^p \leq < F'(v)-F'(u), v-u > \leq \beta_M||v-u||^q.$$

Following the way in [12], we can prove that for any $u, v \in V$, $||u||, ||v|| \leq M$, we have

$$\begin{array}{l} < F'(u), v-u > + \frac{\alpha_M}{p}||v-u||^p \leq F(v)-F(u) \leq \\ < F'(u), v-u > + \frac{\beta_M}{q}||v-u||^q. \end{array} \tag{2.5}$$

Also, using the same techniques, we can prove that if F satisfies (2.4), then $1 < q \leq 2 \leq p$. We point out that since F is Gâteaux differentiable and satisfies (2.4), then F is a convex functional (see Proposition 5.5 in [10], pag. 25).

In certain cases, the second equation in (2.4) can be refined, and we assume that there exist some constants $0 < \beta_{jk} \leq 1$, $\beta_{jk} = \beta_{kj}$, $j,\ k = J, \ldots, 1$, such that

$$\langle F'(v+v_{ji}) - F'(v), v_{kl}\rangle \leq \beta_M\beta_{jk}||v_{ji}||^{q-1}||v_{kl}|| \tag{2.6}$$

for any $v \in V$, $v_{ji} \in V_{ji}$, $v_{kl} \in V_{kl}$ with $||v||$, $||v+v_{ji}||$, $||v_{kl}|| \leq M$, $i = 1, \ldots, I_j$ and $l = 1, \ldots, I_k$. Evidently, in view of (2.4), the above inequality holds for

$$\beta_{jk} = 1,\ j,\ k = J, \ldots, 1 \tag{2.7}$$

We consider the variational inequality

$$u \in K : \; < F'(u), v-u > \geq 0, \text{ for any } v \in K, \tag{2.8}$$

and since the functional F is convex and differentiable, it is equivalent with the minimization problem

$$u \in K : F(u) \leq F(v), \text{ for any } v \in K. \tag{2.9}$$

We can use, for instance, Theorem 8.5 in [21], pag. 251, to prove that problem (2.9) has a unique solution if F has the above properties. In view of (2.5), for a given $M > 0$ such that the solution $u \in K$ of (2.9) satisfies $||u|| \leq M$, we have

$$\frac{\alpha_M}{p}||v-u||^p \leq F(v) - F(u) \text{ for any } v \in K, \; ||v|| \leq M. \tag{2.10}$$

To solve problem (2.8), we propose four algorithms which are either of additive or multiplicative type from a level to another one, in combination with additive or multiplicative iterations on the levels. We first define the algorithm which is of the multiplicative type over the levels as well as on each level.

ALGORITHM 2.1. *We start the algorithm with a $u^0 \in K$ and decompose it as in Assumption 2.1 with $w = u^0$, $u^0 = u_1^0 + \ldots + u_J^0$, $u_j^0 \in K_j$, $j = 1, \ldots, J$. At iteration $n+1$, $n \geq 0$, assuming that we have $u^n \in K$, we decompose it as in Assumption 2.1 with $w = u^n$, $u^n = u_1^n + \ldots + u_J^n$, $u_j^n \in K_j$, $j = 1, \ldots, J$. Then, for $j \in J, \ldots, 1$,*

- we successively calculate, the corrections $w_j^{n+1} \in V_j$, $u_j^n + w_j^{n+1} \in K_j$, by the multiplicative algorithm: we first write $w_j^n = 0$, and for $i = 1, \ldots, I_j$, successively calculate $w_{ji}^{n+1} \in V_{ji}$, $u_j^n + w_j^{n+\frac{i-1}{I_j}} + w_{ji}^{n+1} \in K_j$, the solution of the inequality

$$\langle F'\left(u^n + \sum_{k=j+1}^{J} w_k^{n+1} + w_j^{n+\frac{i-1}{I_j}} + w_{ji}^{n+1}\right), v_{ji} - w_{ji}^{n+1}\rangle \geq 0 \tag{2.11}$$

for any $v_{ji} \in V_{ji}$, $u_j^n + w_j^{n+\frac{i-1}{I_j}} + v_{ji} \in K_j$, and write $w_j^{n+\frac{i}{I_j}} = w_j^{n+\frac{i-1}{I_j}} + w_{ji}^{n+1}$,

- then, we write, $u^{n+\frac{J-j+1}{J}} = u^{n+\frac{J-j}{J}} + w_j^{n+1}$.

The algorithm which is of multiplicative type over the levels and of the additive type on levels is written as,

ALGORITHM 2.2. *We start the algorithm with an $u^0 \in K$ and decompose it as in Assumption 2.2 with $w = u^0$, $u^0 = u_1^0 + \ldots + u_J^0$, $u_j^0 \in K_j$, $j = 1, \ldots, J$. At iteration $n+1$, $n \geq 0$, assuming that we have $u^n \in K$, we decompose it as in Assumption 2.2 with $w = u^n$, $u^n = u_1^n + \ldots + u_J^n$, $u_j^n \in K_j$, $j = 1, \ldots, J$. Then, for $j = J, \ldots, 1$,*

- we successively calculate, the corrections $w_j^{n+1} \in V_j$, $u_j^n + w_j^{n+1} \in K_j$, by the additive algorithm: we simultaneously calculate $w_{ji}^{n+1} \in V_{ji}$, $u_j^n + w_{ji}^{n+1} \in K_j$, the solution of the inequality

$$\langle F'\left(u^n + \sum_{k=j+1}^{J} w_k^{n+1} + w_{ji}^{n+1}\right), v_{ji} - w_{ji}^{n+1}\rangle \geq 0 \qquad (2.12)$$

for any $v_{ji} \in V_{ji}$, $u_j^n + v_{ji} \in K_j$, and write $w_j^{n+1} = \frac{r}{I}\sum_{i=1}^{I_j} w_{ji}^{n+1}$, with a fixed $0 < r \leq 1$.

- then, we write, $u^{n+\frac{J-j+1}{J}} = u^{n+\frac{J-j}{J}} + w_j^{n+1}$.

Now, the additive algorithm over levels and which is of multiplicative type on each level reads,

ALGORITHM 2.3. *We start the algorithm with an $u^0 \in K$ and decompose it as in Assumption 2.1 with $w = u^0$, $u^0 = u_1^0 + \ldots + u_J^0$, $u_j^0 \in K_j$, $j = 1, \ldots, J$. At iteration $n+1$, $n \geq 0$, assuming that we have $u^n \in K$, we decompose it as in Assumption 2.1 with $w = u^n$, $u^n = u_1^n + \ldots + u_J^n$, $u_j^n \in K_j$, $j = 1, \ldots, J$. Then we simultaneously calculate, for $j = 1, \ldots, J$, the corrections $w_j^{n+1} \in V_j$, $u_j^n + w_j^{n+1} \in K_j$, by the multiplicative algorithm:*

– we first write $w_j^n = 0$, and for $i = 1, \ldots, I_j$, successively calculate $w_{ji}^{n+1} \in V_{ji}$, $u_j^n + w_j^{n+\frac{i-1}{I_j}} + w_{ji}^{n+1} \in K_j$, the solution of the inequality

$$\langle F'\left(u^n + w_j^{n+\frac{i-1}{I_j}} + w_{ji}^{n+1}\right), v_{ji} - w_{ji}^{n+1}\rangle \geq 0 \qquad (2.13)$$

for any $v_{ji} \in V_{ji}$, $u_j^n + w_j^{n+\frac{i-1}{I_j}} + v_{ji} \in K_j$, and write $w_j^{n+\frac{i}{I_j}} = w_j^{n+\frac{i-1}{I_j}} + w_{ji}^{n+1}$, Then, we write $u^{n+1} = u^n + \frac{s}{J}\sum_{j=1}^{J} w_j^{n+1}$, with a fixed $0 < s \leq 1$.

Finally, the algorithm which is of additive type over the levels as well as on each level is written as,

ALGORITHM 2.4. *We start the algorithm with an* $u^0 \in K$ *and decompose it as in Assumption 2.2 with* $w = u^0$, $u^0 = u_1^0 + \ldots + u_J^0$, $u_j^0 \in K_j$, $j = 1, \ldots, J$. *At iteration* $n+1$, $n \geq 0$, *assuming that we have* $u^n \in K$, *we decompose it as in Assumption 2.2 with* $w = u^n$, $u^n = u_1^n + \ldots + u_J^n$, $u_j^n \in K_j$, $j = 1, \ldots, J$. *Then we simultaneously calculate, for* $j = 1, \ldots, J$, *the corrections* $w_j^{n+1} \in V_j$, $u_j^n + w_j^{n+1} \in K_j$, *by the additive algorithms:*

– we simultaneously calculate $w_{ji}^{n+1} \in V_{ji}$, $u_j^n + w_{ji}^{n+1} \in K_j$, *the solution of the inequality*

$$\langle F'\left(u^n + w_{ji}^{n+1}\right), v_{ji} - w_{ji}^{n+1}\rangle \geq 0 \tag{2.14}$$

for any $v_{ji} \in V_{ji}$, $u_j^n + v_{ji} \in K_j$, *and write* $w_j^{n+1} = \frac{r}{I}\sum_{i=1}^{I_j} w_{ji}^{n+1}$, *with a fixed* $0 < r \leq 1$.
Then, we write $u^{n+1} = u^n + \frac{s}{J}\sum_{j=1}^{J} w_j^{n+1}$, *with a fixed* $0 < s \leq 1$.

Like inequality (2.8), inequalities (2.11)–(2.14) are equivalent with minimization problems (see [6]).

The convergence result is given by

Theorem 2.1. *We consider that* V *is a reflexive Banach,* V_j, $j = 1, \ldots, J$, *are closed subspaces of* V, *and* V_{ji}, $i = 1, \ldots, I_j$, *are some closed subspaces of* V_j, $j = 1, \ldots, J$. *Let* K *be a non empty closed convex subset of* V *decomposed as in (2.1) where* K_j *are closed convex subsets of* V_j, $j = 1, \ldots, J$, *and* F *be a Gâteaux differentiable functional on* V *which is supposed to be coercive if* K *is not bounded, and satisfies (2.4). Also, we assume that Assumption 2.1 holds for Algorithms 2.1 and 2.3, and Assumption 2.2 holds for Algorithms 2.2 and 2.4. On these conditions, if* u *is the solution of problem (2.8) and* u^n, $n \geq 0$, *are its approximations obtained from the above described algorithms, then there exists* $M > 0$ *such that* $||u||$, $||u^n|| \leq M$, $n \geq 0$, *and the following error estimations hold:*

(i) if $p = q = 2$ *we have*

$$F(u^n) - F(u) \leq \left(\frac{\tilde{C}_1}{\tilde{C}_1 + 1}\right)^n \left[F(u^0) - F(u)\right], \tag{2.15}$$

$$||u^n - u||^2 \leq \frac{2}{\alpha_M}\left(\frac{\tilde{C}_1}{\tilde{C}_1 + 1}\right)^n \left[F(u^0) - F(u)\right], \tag{2.16}$$

where $\tilde{C}_1$ is given in (2.30), and

(ii) if $p > q$ we have

$$F(u^n) - F(u) \le \frac{F(u^0) - F(u)}{\left[1 + n\tilde{C}_2 \left(F(u^0) - F(u)\right)^{\frac{p-q}{q-1}}\right]^{\frac{q-1}{p-q}}}, \tag{2.17}$$

$$||u - u^n||^p \le \frac{p}{\alpha_M} \frac{F(u^0) - F(u)}{\left[1 + n\tilde{C}_2 \left(F(u^0) - F(u)\right)^{\frac{p-q}{q-1}}\right]^{\frac{q-1}{p-q}}}, \tag{2.18}$$

where $\tilde{C}_2$ is given in (2.34).

Proof. Step 1. We first prove the boundedness of the approximations u^n of u as well as of the corrections w_{ji}^{n+1} obtained from the above algorithms. If K is not bounded, using the coercivity and convexity of F, we get that there exists a $M > 0$, such that $||u||$, $||u^n||$, $||w_{ji}^{n+1}|| \le M$, $n \ge 0$, $j = J, \dots, 1$, $i = 1, \dots, I_j$, for the four Algorithms 2.1–2.4. The proof is similar with that given in [1], [3] or [4], and can be found in [6].

Step 2. Now, we study the boundedness of $\sum_{j=1}^J \sum_{i=1}^{I_j} ||w_{ji}^{n+1}||^p$. For Algorithm 2.1, in view of (2.5) and (2.11), we have

$$\frac{\alpha_M}{p}||w_{ji}^{n+1}||^p \le F\left(u^n + \sum_{k=j+1}^J w_k^{n+1} + w_j^{n+\frac{i-1}{I_j}}\right) - F\left(u^n + \sum_{k=j+1}^J w_k^{n+1} + w_j^{n+\frac{i}{I_j}}\right)$$

ie.,

$$\frac{\alpha_M}{p}\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^p \le F\left(u^n + \sum_{k=j+1}^J w_k^{n+1}\right) - F\left(u^n + \sum_{k=j}^J w_k^{n+1}\right) \tag{2.19}$$

Also, for Algorithm 2.2, from (2.12), we get

$$\frac{\alpha_M}{p}||w_{ji}^{n+1}||^p \le F\left(u^n + \sum_{k=j+1}^J w_k^{n+1}\right) - F\left(u^n + \sum_{k=j+1}^J w_k^{n+1} + w_{ji}^{n+1}\right)$$

But,

$$F\left(u^n+\sum_{k=j+1}^{J}w_k^{n+1}+w_j^{n+1}\right)=F\left(u^n+\sum_{k=j+1}^{J}w_k^{n+1}+\frac{r}{I}\sum_{i=1}^{I_j}w_{ji}^{n+1}\right)\leq$$
$$(1-\frac{rI_j}{I})F\left(u^n+\sum_{k=j+1}^{J}w_k^{n+1}\right)+\frac{r}{I}\sum_{i=1}^{I_j}F\left(u^n+\sum_{k=j+1}^{J}w_k^{n+1}+w_{ji}^{n+1}\right)$$

From the above two equations, we get

$$\begin{aligned}&\frac{r}{I}\frac{\alpha_M}{p}\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^p\leq F\left(u^n+\sum_{k=j+1}^{J}w_k^{n+1}\right)-\\&F\left(u^n+\sum_{k=j}^{J}w_k^{n+1}\right)\end{aligned}\tag{2.20}$$

By a similar proof, for Algorithm 2.3, using (2.13), we get

$$\frac{\alpha_M}{p}\sum_{i=1}^{I_J}||w_{ji}^{n+1}||^p\leq F\left(u^n\right)-F\left(u^n+w_j^{n+1}\right)\tag{2.21}$$

and, in view of (2.14), for Algorithm 2.4, we have,

$$\frac{r}{I}\frac{\alpha_M}{p}\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^p\leq F\left(u^n\right)-F\left(u^n+w_j^{n+1}\right)\tag{2.22}$$

Now, let us write

$$t=\begin{cases}1 & \text{for Algorithm 2.1}\\ \frac{r}{I} & \text{for Algorithm 2.2}\\ \frac{s}{J} & \text{for Algorithm 2.3}\\ \frac{s}{J}\frac{r}{I} & \text{for Algorithm 2.4}\end{cases}\tag{2.23}$$

For Algorithms 2.1 and 2.2, in view of (2.19) and (2.20), we can write

$$t\frac{\alpha_M}{p}\sum_{j=1}^{J}\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^p\leq F\left(u^n\right)-F\left(u^{n+1}\right)\tag{2.24}$$

With t in (2.23), the same equation holds for Algorithms 2.3 and 2.4. Indeed,

$$F\left(u^{n+1}\right) = F\left(u^n + \frac{s}{J}\sum_{j=1}^{J} w_j^{n+1}\right) \le (1-s)F(u^n) + \frac{s}{J}\sum_{j=1}^{J} F\left(u^n + w_j^{n+1}\right)$$

and (2.24) follows from (2.21) and (2.22).

Step 3. We now estimate $F(u^{n+1}) - F(u)$. For a given $j \in J$, we write $\overline{w}_j^{n+1} = \sum_{i=1}^{I_j} w_{ji}^{n+1}$. Evidently, for Algorithm 2.1, we have

$$F(u^{n+1}) = F\left(u^n + \sum_{j=1}^{J} \overline{w}_j^{n+1}\right)$$

For Algorithm 2.2, we get,

$$F(u^{n+1}) = F\left(u^n + \frac{r}{I}\sum_{j=1}^{J} \overline{w}_j^{n+1}\right) \le (1-\frac{r}{I})F\left(u^n\right) + \frac{r}{I}F\left(u^n + \sum_{j=1}^{J} \overline{w}_j^{n+1}\right)$$

It is clear that for Algorithm 2.3, we have

$$F(u^{n+1}) \le (1-\frac{s}{J})F(u^n) + \frac{s}{J}F\left(u^n + \sum_{j=1}^{J} \overline{w}_j^{n+1}\right)$$

Finally, for Algorithm 2.4, we get,

$$F(u^{n+1}) = F\left(u^n + \frac{s}{J}\frac{r}{I}\sum_{j=1}^{J} \overline{w}_j^{n+1}\right) \le$$
$$(1-\frac{s}{J}\frac{r}{I})F(u^n) + \frac{s}{J}\frac{r}{I}F\left(u^n + \sum_{j=1}^{J} \overline{w}_j^{n+1}\right)$$

From the above four equations we conclude that

$$F(u^{n+1}) \le (1-t)F(u^n) + tF\left(u^n + \sum_{j=1}^{J} \overline{w}_j^{n+1}\right) \tag{2.25}$$

where t is given in (2.23). With $v = u$ and $w = u^n$, we consider a decomposition $\sum_{j=1}^{J}\sum_{i=1}^{I_j} v_{ji}^n$ of $u - u^n$ as in Assumption 2.1, in the case of Algorithms

2.1 and 2.3, or as in Assumption 2.2, in the case of Algorithms 2.2 and 2.4. In Assumption 2.1, we take $w_{ji} = w_{ji}^{n+1}$, $j = J, \dots, 1$, $i = 1, \dots, I_j$, which are obtained from Algorithms 2.1 or 2.3. In view of (2.5), we have

$$
\begin{aligned}
&F\left(u^n + \sum_{j=1}^{J} \overline{w}_j^{n+1}\right) - F(u) + \frac{\alpha_M}{p}||u^n + \sum_{j=1}^{J} \overline{w}_j^{n+1} - u||^p \leq \\
&\langle F'\left(u^n + \sum_{j=1}^{J} \overline{w}_j^{n+1}\right), u^n + \sum_{j=1}^{J} \overline{w}_j^{n+1} - u\rangle = \\
&-\sum_{k=1}^{J}\sum_{i=1}^{I_k}\langle F'\left(u^n + \sum_{j=1}^{J} \overline{w}_j^{n+1}\right), v_{ki}^n - w_{ki}^{n+1}\rangle
\end{aligned}
\tag{2.26}
$$

For Algorithm 2.1, in view of (2.11) and (2.6), we have,

$$
\begin{aligned}
&-\langle F'\left(u^n + \sum_{j=1}^{J} \overline{w}_j^{n+1}\right), v_{ki}^n - w_{ki}^{n+1}\rangle \leq \\
&\langle F'\left(u^n + \sum_{l=k+1}^{J} \overline{w}_l^{n+1} + w_k^{n+\frac{i-1}{I_k}} + w_{ki}^{n+1}\right) - \\
&F'\left(u^n + \sum_{j=1}^{J} \overline{w}_j^{n+1}\right), v_{ki}^n - w_{ki}^{n+1}\rangle \leq \\
&\beta_M \sum_{j=1}^{J} \beta_{kj} \sum_{l=1}^{I_j} ||w_{jl}^{n+1}||^{q-1}||v_{ki}^n - w_{ki}^{n+1}||
\end{aligned}
$$

Above, we have added and subtracted the missing terms between $F'\left(u^n + \sum_{l=k+1}^{J} \overline{w}_l^{n+1} + w_k^{n+\frac{i-1}{I_j}} + w_{ki}^{n+1}\right)$ and $F'\left(u^n + \sum_{j=1}^{J} \overline{w}_j^{n+1}\right)$. Also, for Algorithm 2.2, from (2.12), we have

$$
\begin{aligned}
&-\langle F'\left(u^n+\sum_{j=1}^{J}\overline{w}_j^{n+1}\right), v_{ki}^n-w_{ki}^{n+1}\rangle \le \\
&\langle F'\left(u^n+\frac{r}{I}\sum_{j=k+1}^{J}\overline{w}_j^{n+1}+w_{ki}^{n+1}\right)-F'\left(u^n+\sum_{j=1}^{J}\overline{w}_j^{n+1}\right), v_{ki}^n-w_{ki}^{n+1}\rangle \le \\
&2\beta_M\sum_{j=1}^{J}\beta_{kj}\sum_{l=1}^{I_j}||w_{jl}^{n+1}||^{q-1}||v_{ki}^n-w_{ki}^{n+1}||
\end{aligned}
$$

Here, we have added and subtracted the missing terms between $F'(u^n)$ and $F'\left(u^n+\frac{r}{I}\sum_{j=k+1}^{J}\overline{w}_j^{n+1}+w_{ki}^{n+1}\right)$, between $F'(u^n)$ and $F'\left(u^n+\sum_{j=1}^{J}\overline{w}_j^{n+1}\right)$, and used the fact that $\frac{r}{I}\le 1$. Similarly, we get the above inequality from (2.13) for Algorithm 2.3, and from (2.14) for Algorithm 2.4. Consequently, in view of (2.26), we can write for all the four algorithms,

$$
\begin{aligned}
&F(u^n+\sum_{j=1}^{J}\overline{w}_j^{n+1})-F(u)+\frac{\alpha_M}{p}||u^n+\sum_{j=1}^{J}\overline{w}_j^{n+1}-u||^p \le \\
&2\beta_M\sum_{j=1}^{J}\sum_{k=1}^{J}\beta_{kj}\sum_{l=1}^{I_j}||w_{jl}^{n+1}||^{q-1}\sum_{i=1}^{I_k}||v_{ki}^n-w_{ki}^{n+1}|| \le \\
&2\beta_M I^{\frac{\sigma-1}{\sigma}+\frac{p-q+1}{p}}\sum_{j=1}^{J}\left(\sum_{k=1}^{J}\beta_{kj}(\sum_{i=1}^{I_k}||v_{ki}^n-w_{ki}^{n+1}||^\sigma)^{\frac{1}{\sigma}}\right)\left(\sum_{l=1}^{I_j}||w_{jl}^{n+1}||^p\right)^{\frac{q-1}{p}} \le \\
&2\beta_M I^{\frac{\sigma-1}{\sigma}+\frac{p-q+1}{p}}\left[\sum_{j=1}^{J}\left(\sum_{k=1}^{J}\beta_{kj}(\sum_{i=1}^{I_k}||v_{ki}^n-w_{ki}^{n+1}||^\sigma)^{\frac{1}{\sigma}}\right)^\sigma\right]^{\frac{1}{\sigma}}\cdot \\
&\left(\sum_{j=1}^{J}(\sum_{l=1}^{I_j}||w_{jl}^{n+1}||^p)^{\frac{q-1}{p}\frac{\sigma}{\sigma-1}}\right)^{\frac{\sigma-1}{\sigma}} \le 2\beta_M I^{\frac{\sigma-1}{\sigma}+\frac{p-q+1}{p}}J^{\frac{\sigma-1}{\sigma}-\frac{q-1}{p}}(\max_{k=1,\cdots,J}\sum_{j=1}^{J}\beta_{kj})\cdot \\
&(\sum_{j=1}^{J}\sum_{i=1}^{I_j}||v_{ji}^n-w_{ji}^{n+1}||^\sigma)^{\frac{1}{\sigma}}(\sum_{j=1}^{J}\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^p)^{\frac{q-1}{p}}
\end{aligned}
$$

Above, we have used the inequality (see Corollary 4.1 in [25])

$$||Ax||_{l^\sigma} \le (\max_i \sum_j |A_{ij}|)||x||_{l^\sigma} \tag{2.27}$$

where $A = (A_{ij})_{ij}$ is a symmetric matrix. In view of (2.2), Assumptions 2.1 and 2.2 and Remark 2.1, we have

$$\begin{aligned}
&(\sum_{j=1}^{J}\sum_{i=1}^{I_j}||v_{ji}^n - w_{ji}^{n+1}||^\sigma)^{\frac{1}{\sigma}} \le (\sum_{j=1}^{J}\sum_{i=1}^{I_j}||v_{ji}||^\sigma)^{\frac{1}{\sigma}} + (\sum_{j=1}^{J}\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^\sigma)^{\frac{1}{\sigma}} \le \\
&(C_2^\sigma||u-u^n||^\sigma + C_3^\sigma\sum_{j=1}^{J}\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^\sigma)^{\frac{1}{\sigma}} + (\sum_{j=1}^{J}\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^\sigma)^{\frac{1}{\sigma}} \le \\
&C_2||u-u^n|| + (1+C_3)(\sum_{j=1}^{J}\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^\sigma)^{\frac{1}{\sigma}} \le \\
&C_2||u-u^n-\sum_{j=1}^{J}\overline{w}_j^{n+1}|| + (1+C_1C_2+C_3)(IJ)^{\frac{p-\sigma}{p\sigma}}(\sum_{j=1}^{J}\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^p)^{\frac{1}{p}}
\end{aligned}$$

Therefore, we get

$$\begin{aligned}
&F(u^n+\sum_{j=1}^{J}\overline{w}_j^{n+1}) - F(u) + \frac{\alpha_M}{p}||u^n+\sum_{j=1}^{J}\overline{w}_j^{n+1}-u||^p \le \\
&2\beta_M I^{\frac{\sigma-1}{\sigma}+\frac{p-q+1}{p}} J^{\frac{\sigma-1}{\sigma}-\frac{q-1}{p}}(\max_{k=1,\cdots,J}\sum_{j=1}^{J}\beta_{kj})\cdot \\
&\left[C_2||u-u^n-\sum_{j=1}^{J}\overline{w}_j^{n+1}||(\sum_{j=1}^{J}\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^p)^{\frac{q-1}{p}} + \right. \\
&\left.(1+C_1C_2+C_3)(IJ)^{\frac{p-\sigma}{p\sigma}}(\sum_{j=1}^{J}\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^p)^{\frac{q}{p}}\right]
\end{aligned}$$

But, for any $\varepsilon > 0$, $p > 1$ and $x, y \ge 0$, we have $xy \le \varepsilon x^p + \frac{1}{\varepsilon^{\frac{1}{p-1}}} y^{\frac{p}{p-1}}$.

Consequently, we have

$$F(u^n+\sum_{j=1}^J \overline{w}_j^{n+1})-F(u)+\frac{\alpha_M}{p}||u^n+\sum_{j=1}^J \overline{w}_j^{n+1}-u||^p \le$$
$$2\beta_M I^{\frac{\sigma-1}{\sigma}+\frac{p-q+1}{p}} J^{\frac{\sigma-1}{\sigma}-\frac{q-1}{p}}(\max_{k=1,\cdots,J}\sum_{j=1}^J \beta_{kj})\cdot$$
$$\Bigg[C_2\varepsilon||u-u^n-\sum_{j=1}^J \overline{w}_j^{n+1}||^p+C_2\frac{1}{\varepsilon^{\frac{1}{p-1}}}(\sum_{j=1}^J\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^p)^{\frac{q-1}{p-1}}+$$
$$(1+C_1C_2+C_3)(IJ)^{\frac{p-\sigma}{p\sigma}}(\sum_{j=1}^J\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^p)^{\frac{q}{p}}\Bigg]$$

for any $\varepsilon>0$. With

$$\varepsilon=\frac{\alpha_M}{p}\frac{1}{2C_2\beta_M I^{\frac{\sigma-1}{\sigma}+\frac{p-q+1}{p}} J^{\frac{\sigma-1}{\sigma}-\frac{q-1}{p}}(\max_{k=1,\cdots,J}\sum_{j=1}^J \beta_{kj})} \tag{2.28}$$

the above equation becomes,

$$F(u^n+\sum_{j=1}^J \overline{w}_j^{n+1})-F(u)\le\frac{\frac{\alpha_M}{p}}{C_2\varepsilon}\cdot$$
$$\left[\frac{C_2}{\varepsilon^{\frac{1}{p-1}}}(\sum_{j=1}^J\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^p)^{\frac{q-1}{p-1}}+(1+C_1C_2+C_3)(IJ)^{\frac{p-\sigma}{p\sigma}}(\sum_{j=1}^J\sum_{i=1}^{I_j}||w_{ji}^{n+1}||^p)^{\frac{q}{p}}\right]$$

From this equation and (2.24)

$$F(u^n+\sum_{j=1}^J \overline{w}_j^{n+1})-F(u)\le\frac{\frac{\alpha_M}{p}}{C_2\varepsilon}\cdot\Bigg[\frac{C_2}{\varepsilon^{\frac{1}{p-1}}(t\frac{\alpha_M}{p})^{\frac{q-1}{p-1}}}(F(u^n)-F(u^{n+1}))^{\frac{q-1}{p-1}}+$$
$$\frac{(1+C_1C_2+C_3)(IJ)^{\frac{p-\sigma}{p\sigma}}}{(t\frac{\alpha_M}{p})^{\frac{q}{p}}}(F(u^n)-F(u^{n+1}))^{\frac{q}{p}}\Bigg]$$

with t in (2.23) and ε in (2.28). In view of the above equation and (2.25), we have

$$F(u^{n+1}) - F(u) \leq \tfrac{1-t}{t}(F(u^n) - F(u^{n+1})) + \frac{\frac{\alpha_M}{p}}{C_2\varepsilon}\cdot \\ \left[\frac{C_2}{\varepsilon^{\frac{1}{p-1}}(t\frac{\alpha_M}{p})^{\frac{q-1}{p-1}}}(F(u^n) - F(u^{n+1}))^{\frac{q-1}{p-1}} + \right. \\ \left. \frac{(1 + C_1C_2 + C_3)(IJ)^{\frac{p-\sigma}{p\sigma}}}{(t\frac{\alpha_M}{p})^{\frac{q}{p}}}(F(u^n) - F(u^{n+1}))^{\frac{q}{p}}\right] \tag{2.29}$$

Step 4. We prove error estimations (2.15)–(2.18). First, using (2.10), we see that error estimations in (2.16) and (2.18) can be obtained from (2.15) and (2.17), respectively. Now, if $p = q = 2$, then $\sigma = 2$, and from the above equation, we easily get equation (2.15), where

$$\tilde{C}_1 = \frac{1-t}{t} + \frac{1}{C_2t\varepsilon}\left[\frac{C_2}{\varepsilon} + 1 + C_1C_2 + C_3\right] \text{ with} \\ \varepsilon = \frac{\frac{\alpha_M}{2}}{2C_2\beta_M I(\max\limits_{k=1,\cdots,J}\sum\limits_{j=1}^{J}\beta_{kj})} \tag{2.30}$$

Finally, if $p > q$, from (2.29), we have

$$F(u^{n+1}) - F(u) \leq \tilde{C}_3(F(u^n) - F(u^{n+1}))^{\frac{q-1}{p-1}} \tag{2.31}$$

where

$$\tilde{C}_3 = \frac{1-t}{t}(F(u^0) - F(u))^{\frac{p-q}{p-1}} + \frac{\frac{\alpha_M}{p}}{C_2\varepsilon}\left[\frac{C_2}{\varepsilon^{\frac{1}{p-1}}(t\frac{\alpha_M}{p})^{\frac{q-1}{p-1}}} + \right. \\ \left. \frac{(1 + C_1C_2 + C_3)(IJ)^{\frac{p-\sigma}{p\sigma}}}{(t\frac{\alpha_M}{p})^{\frac{q}{p}}}(F(u^0) - F(u))^{\frac{p-q}{p(p-1)}}\right] \tag{2.32}$$

with ε in (2.28). From (2.31), we get

$$F(u^{n+1}) - F(u) + \frac{1}{\tilde{C}_3^{\frac{p-1}{q-1}}}(F(u^{n+1}) - F(u))^{\frac{p-1}{q-1}} \leq F(u^n) - F(u),$$

and we know (see Lemma 3.2 in [25]) that for any $r > 1$ and $c > 0$, if $x \in (0, x_0]$ and $y > 0$ satisfy $y + cy^r \leq x$, then $y \leq (\frac{c(r-1)}{crx_0^{r-1}+1} + x^{1-r})^{\frac{1}{1-r}}$.

Consequently, we have $F(u^{n+1}) - F(u) \le [\tilde{C}_2 + (F(u^n) - F(u))^{\frac{q-p}{q-1}}]^{\frac{q-1}{q-p}}$, from which,

$$F(u^{n+1}) - F(u) \le [(n+1)\tilde{C}_2 + (F(u^0) - F(u))^{\frac{q-p}{q-1}}]^{\frac{q-1}{q-p}}, \tag{2.33}$$

where

$$\tilde{C}_2 = \frac{p-q}{(p-1)(F(u^0) - F(u))^{\frac{p-q}{q-1}} + (q-1)\tilde{C}_3^{\frac{p-1}{q-1}}}. \tag{2.34}$$

Equation (2.33) is another form of equation (2.17). □

3 Multilevel Schwarz methods

We consider a family of regular meshes $\mathcal{T}_{h_j}$ of mesh sizes h_j, $j = 1, \dots, J$ over the domain $\Omega \subset \mathbf{R}^d$. We write $\Omega_j = \cup_{\tau \in \mathcal{T}_{h_j}} \tau$ and assume that $\mathcal{T}_{h_{j+1}}$ is a refinement of $\mathcal{T}_{h_j}$ on Ω_j, $j = 1, \dots, J-1$, and $\Omega_1 \subset \Omega_2 \subset \dots \subset \Omega_J = \Omega$. Also, we assume that, if a node of $\mathcal{T}_{h_j}$ lies on $\partial\Omega_j$, then it lies on $\partial\Omega_{j+1}$, too, that is, it lies on $\partial\Omega$. Besides, we suppose that $\mathrm{dist}(x_{j+1}, \Omega_j) \le Ch_j$, for any node x_{j+1} of $\mathcal{T}_{h_{j+1}}$, $j = 1, \dots, J-1$. In this section, C denotes a generic positive constant independent of the mesh sizes, the number of meshes, as well as of the overlapping parameters and the number of subdomains in the domain decompositions which will be considered later. Since the mesh $\mathcal{T}_{h_{j+1}}$ is a refinement of $\mathcal{T}_{h_j}$, we have $h_{j+1} \le h_j$, and assume that there exists a constant γ, independent of the number of meshes or their sizes, such that

$$1 < \gamma \le \frac{h_j}{h_{j+1}} \le C\gamma, \quad j = 1, \dots, J-1. \tag{3.1}$$

At each level $j = 1, \dots, J$, we consider an overlapping decomposition $\{\Omega_j^i\}_{1 \le i \le I_j}$ of Ω_j, and assume that the mesh partition $\mathcal{T}_{h_j}$ of Ω_j supplies a mesh partition for each Ω_j^i, $1 \le i \le I_j$. Also, we assume that the overlapping size for the domain decomposition at the level $1 \le j \le J$ is δ_j. In addition, we suppose that if ω_{j+1}^i is a connected component of Ω_{j+1}^i, $j = 1, \dots, J-1$, $i = 1, \dots, I_j$, then

$$\mathrm{diam}(\omega_{j+1}^i) \le Ch_j \tag{3.2}$$

Since $h_{j+1} \le \delta_{j+1}$, from (3.1), we also have

$$\frac{h_j}{\delta_{j+1}} \le C\gamma, \quad j = 1, \dots, J-1. \tag{3.3}$$

Finally, we assume that $I_1 = 1$.

At each level $j = 1, \dots, J$, we introduce the linear finite element spaces,

$$V_{h_j} = \{v \in C(\bar{\Omega}_j) : v|_\tau \in P_1(\tau),\ \tau \in \mathcal{T}_{h_j},\ v = 0 \text{ on } \partial\Omega_j\}, \tag{3.4}$$

and, for $i = 1, \dots, I_j$, we write

$$V^i_{h_j} = \{v \in V_{h_j} : v = 0 \text{ in } \Omega_j \backslash \Omega^i_j\}. \tag{3.5}$$

The functions in V_{h_j} $j = 1, \dots, J-1$, will be extended with zero outside Ω_j and the spaces will be considered as subspaces of $W^{1,\sigma}$, $1 \le \sigma \le \infty$. We denote by $||\cdot||_{0,\sigma}$ the norm in L^σ, and by $||\cdot||_{1,\sigma}$ and $|\cdot|_{1,\sigma}$ the norm and seminorm in $W^{1,\sigma}$, respectively.

We consider the obstacle problem

$$u \in K :\ < F'(u), v - u > \geq 0, \text{ for any } v \in K, \tag{3.6}$$

where

$$K = \{v \in V_{h_J} : \varphi \le v\}, \tag{3.7}$$

with $\varphi \in V_{h_J}$. We shall prove that Assumptions 2.1 and 2.2 hold for this type of convex set, and explicitly write the constants C_2 and C_3 as functions of the mesh and overlapping parameters. We can then conclude from Theorem 2.1 that if the functional F has the asked properties, then Algorithms 2.1–2.4 are globally convergent.

We first introduce the operators $I_{h_j} : V_{h_{j+1}} \to V_{h_j}$, $j = 1, \dots, J-1$, defined as follows. Let us denote by x_{ji} a node of $\mathcal{T}_{h_j}$, by ϕ_{ji} the linear nodal basis function associated with x_{ji} and $\mathcal{T}_{h_j}$, and by ω_{ji} the support of ϕ_{ji}. Given a $v \in V_{h_{j+1}}$, we write $I_{ji}v = \min_{x\in\omega_{ji}} v(x)$. Finally, we define $I_{h_j}v := \sum_{x_{ji} \text{node of } \mathcal{T}_{h_j}} (I_{ji}v)\phi_{ji}(x)$.

Remark 3.1. 1) In [24], similar operators, $I_{h_j} : V_{h_J} \to V_{h_j}$, are defined. For a $v \in V_{h_J}$, we write as above, $I_{ji}v = \min_{x\in\omega_{ji}} v(x)$ and $I_{h_j}v := \sum_{x_{ji} \text{node of } \mathcal{T}_{h_j}} (I_{ji}v)\phi_{ji}(x)$. These operators have the disadvantage that $I_{h_j}v$ can not be computed from $I_{h_{j+1}}v$. For this reason, in the case of the multigrid method, their definition is modified in [13] by taking $I_{ji}v = \min_{x\in \mathrm{Int}\omega_{ji}} v(x)$ in the place of $I_{ji}v = \min_{x\in\omega_{ji}} v(x)$.

2) Since the finite element spaces are linear, for a $v \in V_{h_j}$, we can take $I_{ji}v = \min\limits_{x\in\omega_{ji},\, x \text{ node of } \mathcal{T}_{h_{j+1}}} v(x)$ in the place of $I_{ji}v = \min_{x\in\omega_{ji}} v(x)$ in our above definition.

3) In [2] some more general operators $I_{h_j} : V_{h_{j+1}} \to V_{h_j}$ have been introduced. They coincide with those above defined ones for $v \geq 0$.

For a $v \in V_{h_J}$, we recursively define

$$v^J = v \text{ and } v^j = I_{h_j} v^{j+1},\ j = J-1, \ldots, 1 \tag{3.8}$$

Writing

$$C_{d,\sigma}(H,h) = \begin{cases} 1 & \text{if } d = \sigma = 1 \\ & \text{or } 1 \leq d < \sigma < \infty \\ (\ln \frac{H}{h} + 1)^{\frac{d-1}{d}} & \text{if } 1 < d = \sigma < \infty \\ (\frac{H}{h})^{\frac{d-\sigma}{\sigma}} & \text{if } 1 \leq \sigma < d < \infty, \end{cases} \tag{3.9}$$

we have the following result

Lemma 3.1. *Let $v^j, w^j \in V_{h_j}$, $j = J, \ldots, 1$ defined as in (3.8) for some $v, w \in V_{h_J}$, respectively. Then, for $j = J, \ldots, 1$, we have*

$$|v^j - w^j|_{1,\sigma} \leq CC_{d,\sigma}(h_j, h_J)|v - w|_{1,\sigma} \tag{3.10}$$

Proof. Equation (3.10) is evident for $j = J$. For a $j = J-1, \ldots, 1$, let $\omega_j(x_j)$ be the support of the nodal basis function in V_{h_j} corresponding to the node x_j of $\mathcal{T}_{h_j}$. Then there exist two nodes of $\mathcal{T}_{h_j}$, $x_j^1, x_j^2 \in \omega_j(x_j)$, such that

$$|v^j - w^j|^{\sigma}_{1,\sigma,\omega_j(x_j)} \leq Ch_j^{d-\sigma}|(v^j - w^j)(x_j^1) - (v^j - w^j)(x_j^2)|^{\sigma} \tag{3.11}$$

and let us assume that

$$|(v^j - w^j)(x_j^1) - (v^j - w^j)(x_j^1)| = (v^j - w^j)(x_j^1) - (v^j - w^j)(x_j^2) \tag{3.12}$$

Now, we have

$$\begin{aligned} &(v^j - w^j)(x_j^1) - (v^j - w^j)(x_j^2) = \\ &(I_{h_j} v^{j+1} - I_{h_j} w^{j+1})(x_j^1) + (I_{h_j} w^{j+1} - I_{h_j} v^{j+1})(x_j^2) \end{aligned}$$

and let

$$I_{h_j} w^{j+1}(x_j^1) = w^{j+1}(x_{j+1}^1) \text{ and } I_{h_j} v^{j+1}(x_j^2) = v^{j+1}(x_{j+1}^2)$$

where $x^1_{j+1} \in \omega_j(x^1_j)$ and $x^2_{j+1} \in \omega_j(x^2_j)$ are two nodes of $\mathcal{T}_{h_{j+1}}$, $\omega_j(x^1_j)$ and $\omega_j(x^2_j)$ being the supports of the nodal basis functions in V_{h_j} corresponding to the nodes x^1_j and x^2_j, respectively. Consequently, we get

$$\begin{aligned}&(v^j - w^j)(x^1_j) - (v^j - w^j)(x^2_j) \le \\ &(v^{j+1} - w^{j+1})(x^1_{j+1}) - (v^{j+1} - w^{j+1})(x^2_{j+1})\end{aligned}$$

Repeating the above reasoning, we get that, for $k = j, \ldots, J-1$, there exist $x^{k+1}_1 \in \omega_k(x^k_1)$ and $x^{k+1}_2 \in \omega_k(x^k_2)$ are two nodes of $\mathcal{T}_{h_{k+1}}$, $\omega_k(x^k_1)$ and $\omega_k(x^k_2)$ being the supports of the nodal basis functions in V_{h_k} corresponding to the nodes x^k_1 and x^k_2, respectively, such that

$$\begin{aligned}&(v^k - w^k)(x^k_1) - (v^k - w^k)(x^k_2) \le \\ &(v^{k+1} - w^{k+1})(x^{k+1}_1) - (v^{k+1} - w^{k+1})(x^{k+1}_2)\end{aligned} \tag{3.13}$$

From (3.11), (3.12) and (3.13), we get

$$|v^j - w^j|^\sigma_{1,\sigma,\omega_j(x_j)} \le Ch_j^{d-\sigma}[(v-w)(x^1_J) - (v-w)(x^2_J)]^\sigma \tag{3.14}$$

Since the radius of ω_k is less than h_k, and in view of (3.1), it follows that $\text{dist}(x_j, x^1_J)$, $\text{dist}(x_j, x^1_J) \le h_j + (h_j + \ldots + h_{J-1}) \le (1 + 1 + \frac{1}{\gamma} + \ldots + \frac{1}{\gamma^{J-1-j}})h_j \le \frac{2\gamma-1}{\gamma-1}h_j$. Therefore, if we write

$$\tilde{\omega}_j(x_j) = \bigcup_{\tau \in \mathcal{T}_{h_j},\ \text{dist}(x_j,\tau) \le \frac{\gamma}{\gamma-1}h_j} \tau,$$

then $x^1_J, x^2_J \in \tilde{\omega}_j(x_j)$. Subtracting and adding $(v-w)(x)$, $x \in \tilde{\omega}_j(x_j)$, in the right hand side of (3.14), integrating over $\tilde{\omega}_j(x_j)$, in view of Lemma 4.1 in [2], we have

$$\begin{aligned}&(\tfrac{2\gamma-1}{\gamma-1}h_j)^d|v^j - w^j|^\sigma_{1,\sigma,\omega_j(x_j)} \le Ch_j^{d-\sigma}\Big[||(v-w)(x^1_j) - (v-w)(x)||^\sigma_{0,\sigma,\tilde{\omega}_j(x_j)} + \\ &||(v-w)(x^2_j) - (v-w)(x)||^\sigma_{0,\sigma,\tilde{\omega}_j(x_j)}\Big] \le \\ &Ch_j^{d-\sigma}(2\tfrac{2\gamma-1}{\gamma-1}h_j)^\sigma C_{d,\sigma}(2\tfrac{2\gamma-1}{\gamma-1}h_j, h_J)^\sigma |v-w|^\sigma_{1,\sigma,\tilde{\omega}_j(x_j)},\end{aligned}$$

ie.,

$$|v^j - w^j|_{1,\sigma,\omega_j(x_j)} \le CC_{d,\sigma}(h_j, h_J)|v-w|_{1,\sigma,\tilde{\omega}_j(x_j)}$$

Finally, since the mesh $\mathcal{T}_{h_j}$ is regular and γ is independent of J and of the mesh parameters, then $\omega_j(x_j)$ and $\tilde{\omega}_j(x_j)$ contain a bounded number of simplexes of $\mathcal{T}_{h_j}$, which is also independent of J and of the mesh parameters. Consequently, we get (3.10). □

Another result we shall utilize is given by the following lemma.

Lemma 3.2. *For any* $v, w \in V_{h_{j+1}}$, $j = J-1, \dots, 1$, *we have*

$$||v - w - I_{h_j} v + I_{h_j} w||_{0,\sigma} \le C h_j C_{d,\sigma}(h_j, h_{j+1}) |v - w|_{1,\sigma} \tag{3.15}$$

Proof. As in the proof of the previous lemma, we denote by $\omega_j(x_j)$ the support of the nodal basis function ϕ_j in V_{h_j} corresponding to the node x_j of $\mathcal{T}_{h_j}$. For a $\tau \in \mathcal{T}_{h_j}$, we have

$$\begin{aligned} & ||v - w - I_{h_j} v + I_{h_j} w||_{0,\sigma,\tau} = \\ & || \sum_{x_j \text{ node of } \tau} [v - w - (I_{h_j} v - I_{h_j} w)(x_j)] \phi_j ||_{0,\sigma,\tau} \le \\ & \sum_{x_j \text{ node of } \tau} ||v - w - (I_{h_j} v - I_{h_j} w)(x_j)||_{0,\sigma,\tau} \end{aligned}$$

From the definition of I_{h_j}, there exist two nodes of $\mathcal{T}_{h_{j+1}}$, $x_{j+1}^1, x_{j+1}^2 \in \omega_j(x_j)$, such that $(I_{h_j} v)(x_j) = v(x_{j+1}^1)$ and $(I_{h_j} w)(x_j) = w(x_{j+1}^2)$. Therefore,

$$\begin{aligned} & ||v - w - I_{h_j} v + I_{h_j} w||_{0,\sigma,\tau} \le \\ & \sum_{x_j \text{ node of } \tau} ||v - w - v(x_{j+1}^1) + w(x_{j+1}^2)||_{0,\sigma,\omega_j(x_j)} \end{aligned}$$

Now, let $\omega_j(x_j)^+ = \{x \in \omega_j(x_j) : v - w - v(x_{j+1}^1) + w(x_{j+1}^2) \ge 0\}$ and $\omega_j(x_j)^- = \{x \in \omega_j(x_j) : v - w - v(x_{j+1}^1) + w(x_{j+1}^2) \le 0\}$. From the above equation, the definition of I_{h_j} and Lemma 4.1 in [2], we get

$$\begin{aligned} & ||v - w - I_{h_j} v + I_{h_j} w||_{0,\sigma,\tau} \le \\ & \sum_{x_j \text{ node of } \tau} \Big[||v - w - v(x_{j+1}^1) + w(x_{j+1}^1)||_{0,\sigma,\omega_j(x_j)^+}^{\sigma} + \\ & ||w - v - w(x_{j+1}^2) + v(x_{j+1}^2)||_{0,\sigma,\omega_j(x_j)^-}^{\sigma} \Big]^{1/\sigma} \le \\ & C h_j C_{d,\sigma}(h_j, h_{j+1}) \sum_{x_j \text{ node of } \tau} \Big[|v - w|_{1,\sigma,\omega_j(x_j)^+}^{\sigma} + |w - v|_{1,\sigma,\omega_j(x_j)^-}^{\sigma} \Big]^{1/\sigma} = \\ & C h_j C_{d,\sigma}(h_j, h_{j+1}) \sum_{x_j \text{ node of } \tau} |v - w|_{1,\sigma,\omega_j(x_j)} \end{aligned}$$

Since the mesh $\mathcal{T}_{h_j}$ is regular, $\omega_j(x_j)$ contains a bounded number of simplexes of $\mathcal{T}_{h_j}$, which is independent of J and of the mesh parameters. Consequently, inequality (3.15) can be obtained from the above equation. $\square$

Now, we consider a decomposition of $\varphi = \varphi_J + \ldots + \varphi_1$ with $\varphi_j \in V_{h_j}$, $j = J, \ldots, 1$, and define

$$K_j = \{v \in V_{h_j} : \varphi_j \leq v\}, \; j = J, \ldots, 1 \tag{3.16}$$

In this way, we get a decomposition of K as in (2.1). For a $v \in K$, with the notation in (3.8), we write

$$\begin{aligned} v_j &= \varphi_j + (v - \varphi)^j - (v - \varphi)^{j-1}, \; j = J, \ldots, 2 \\ v_1 &= \varphi_1 + (v - \varphi)^1 \end{aligned} \tag{3.17}$$

Evidently,

$$v_j \in K_j, \; j = J, \ldots, 1, \text{ and } v = v_J + \ldots + v_1 \tag{3.18}$$

We have the following

Lemma 3.3. *If $v_j, w_j \in K_j$, $j = J, \ldots, 1$, are defined as in (3.17) for some $v, w \in K$, respectively, then*

$$|v_j - w_j|_{1,\sigma} \leq C C_{d,\sigma}(h_{j-1}, h_J) |v - w|_{1,\sigma} \tag{3.19}$$

and

$$||v_j - w_j||_{0,\sigma} \leq C h_{j-1} C_{d,\sigma}(h_j, h_J) |v - w|_{1,\sigma} \tag{3.20}$$

where we take $h_0 = h_1$ for $j = 1$.

Proof. For $j = J, \ldots, 2$, in view of (3.10), we have

$$\begin{aligned} &|v_j - w_j|_{1,\sigma} = |(v - \varphi)^j - (v - \varphi)^{j-1} - (w - \varphi)^j + (w - \varphi)^{j-1}|_{1,\sigma} \leq \\ &C[C_{d,\sigma}(h_j, h_J) + C_{d,\sigma}(h_{j-1}, h_J)] |v - w|_{1,\sigma} \end{aligned}$$

ie., (3.19) holds for $j = J, \ldots, 2$. Also, by a similar proof, we get that (3.19) for $j = 1$. Now, using (3.15) and (3.10), for $j = J, \ldots, 2$, we get

$$\begin{aligned} &||v_j - w_j||_{0,\sigma} = \\ &||(v - \varphi)^j - I_{h_{j-1}}(v - \varphi)^j - (w - \varphi)^j + I_{h_{j-1}}(w - \varphi)^j||_{0,\sigma} \leq \\ &C h_{j-1} C_{d,\sigma}(h_{j-1}, h_j) |(v - \varphi)^j - (w - \varphi)^j|_{1,\sigma} \leq \\ &C h_{j-1} C_{d,\sigma}(h_{j-1}, h_j) C_{d,\sigma}(h_j, h_J) |v - w|_{1,\sigma} \end{aligned}$$

and therefore, (3.20) holds for $j = J, \ldots, 2$. For $j = 1$, from the classical Friedrichs-Poincaré inequality and (3.10), we have

$$\begin{aligned} &||v_1 - w_1||_{0,\sigma} = ||(v - \varphi)^1 - (w - \varphi)^1||_{0,\sigma} \leq \\ &C h_1 |(v - \varphi)^1 - (w - \varphi)^1|_{1,\sigma} \leq C h_1 C_{d,\sigma}(h_1, h_J) |v - w|_{1,\sigma}, \end{aligned}$$

ie., we obtained (3.20) for $j = 1$. $\square$

To prove that Assumption 2.1 holds, we associate to the decomposition $\{\Omega_j^i\}_{1\le i\le I_j}$ of Ω_j, some functions $\theta_j^i \in C(\bar{\Omega}_j)$, $\theta_j^i|_\tau \in P_1(\tau)$ for any $\tau \in \mathcal{T}_{h_j}$, $i = 1, \cdots, I_j$, such that

$$\begin{aligned} &0 \le \theta_j^i \le 1 \text{ on } \Omega_j, \\ &\theta_j^i = 0 \text{ on } \cup_{l=i+1}^{I_j} \Omega_j^l \backslash \Omega_j^i,\ \theta_j^i = 1 \text{ on } \Omega_j^i \backslash \cup_{l=i+1}^{I_j} \Omega_j^l \end{aligned} \tag{3.21}$$

Also, for Assumption 2.2, we associate a unity partition to each domain decomposition $\{\Omega_j^i\}_{1\le i\le I_j}$, $j = J, \ldots, 1$,

$$0 \le \theta_j^i \le 1 \text{ and } \sum_{i=1}^{I_j} \theta_j^i = 1 \text{ on } \Omega_j \tag{3.22}$$

with $\theta_j^i \in C(\bar{\Omega}_j)$, $\theta_j^i|_\tau \in P_1(\tau)$ for any $\tau \in \mathcal{T}_{h_j}$, $i = 1, \cdots, I_j$. Such functions θ_j^i with the above properties exist (see [2] or [26] p. 59, for instance). Moreover, since the overlapping size of the domain decomposition on a level $j = J, \ldots, 1$ is δ_j, the above functions θ_j^i can be chosen to satisfy

$$|\partial_{x_k} \theta_j^i| \le C/\delta_j, \text{ a.e. in } \Omega_j, \text{ for any } k = 1, \ldots, d \tag{3.23}$$

Finally, we recall some interpolation properties. For a $v \in V_{h_j}$ and a continuous functions θ which is of polynomial form on the elements of $\tau \in \mathcal{T}_{h_j}$, we have (see [9] and [28]),

$$||\theta v - L_{h_j}(\theta v)||_{0,\sigma} \le C h_j |\theta v|_{1,\sigma} \text{ and } |L_{h_j}(\theta v)|_{1,\sigma} \le C|\theta v|_{1,\sigma}$$

where L_{h_j} is the P_1-Lagrangian interpolation operator which uses the function values at the nodes of the mesh $\mathcal{T}_{h_j}$. Therefore, we have

$$||L_{h_j}(\theta v)||_{1,\sigma} \le C||\theta v||_{1,\sigma} \tag{3.24}$$

Now, we can prove

Proposition 3.1. *Assumption 2.1 holds with the constants C_2 and C_3 are given in (3.29) for the convex sets K_j, $j = J, \ldots, 1$, defined in (3.16).*

Proof. Let us consider $v,\ w \in K$ and let $v_j, w_j \in K_j$, $j = J, \ldots, 1$, be their decompositions defined as in (3.17), respectively. Also, let $w_{ji} \in V_{h_j}^i$

such that $w_j + w_{j1} + \ldots + w_{ji} \in K_j$, $j = J, \ldots, 1$, $i = 1, \ldots, I_j$. Now, for $j = J, \ldots, 2$, we define

$$\begin{aligned} v_{j1} &= L_{h_j}(\theta_j^1(v_j - w_j) + (1-\theta_j^1)w_{j1}) \\ v_{ji} &= L_{h_j}(\theta_j^i((v_j - w_j) - \sum_{l=1}^{i-1} v_{jl}) + (1-\theta_j^i)w_{ji}), \quad i = 2, \ldots, I_j \end{aligned}$$

with θ_j^i in (3.21). Like in Proposition 3.1 in [2], where we take $v = v_j$ and $w = w_j$, we can prove that

$$\begin{aligned} & v_{ji} \in V_{h_j}^i, \quad w_j + w_{j1} + \ldots + w_{ji-1} + v_{ji} \in K_j,\ i = 1, \ldots, I_j \\ & v_j - w_j = \sum_{i=1}^{I_j} v_{ji} \end{aligned} \tag{3.25}$$

We point out that here, the condition $w_j + w_{j1} + \ldots + w_{ji-1} + v_{ji} \in K_j$ can be proved by verifying that it is satisfied only at the nodes of $\mathcal{T}_{h_j}$. At the level $j = 1$, we do not have a domain decomposition, $I_1 = 1$, and we take

$$v_{11} = v_1 - w_1.$$

From this equation, (3.18) and (3.25), we get that the first two conditions of Assumption 2.1 are satisfied.

We estimate now the constants C_2 and C_3. Using Lemma 3.3 and the same techniques as in [2] or [4] (see [6] for details), we can write

$$||v_{ji}||_{1,\sigma}^{\sigma} \le CI^{\sigma} \left\{ C_{d,\sigma}(h_{j-1}, h_J)^{\sigma} |u - w|_{1,\sigma}^{\sigma} + \sum_{k=1}^{I_j} |w_{jk}|_{1,\sigma}^{\sigma} \right\} \tag{3.26}$$

for any $j = J, \ldots, 2$ and $i = 1, \ldots, I_j$. At the level $j = 1$, from Lemma 3.3, we have

$$||v_{11}||_{1,\sigma}^{\sigma} \le CC_{d,\sigma}(h_1, h_J)^{\sigma} |v - w|_{1,\sigma}^{\sigma} \tag{3.27}$$

From (3.26) and (3.27), we get

$$\begin{aligned} \sum_{j=1}^{J} \sum_{i=1}^{I_j} ||v_{ji}||_{1,\sigma}^{\sigma} \le CI^{\sigma+1} \Bigg\{ & \sum_{j=2}^{J} \sum_{i=1}^{I_j} |w_{ji}|_{1,\sigma}^{\sigma} + \\ & \left[\sum_{j=2}^{J} C_{d,\sigma}(h_{j-1}, h_J)^{\sigma} \right] |u - w|_{1,\sigma}^{\sigma} \Bigg\} \end{aligned} \tag{3.28}$$

Consequently, from (3.28), we get that the constants C_2 and C_3 can be written as

$$C_2 = CI^{\frac{\sigma+1}{\sigma}} \left[\sum_{j=2}^{J} C_{d,\sigma}(h_{j-1}, h_J)^\sigma \right]^{\frac{1}{\sigma}} \text{ and } C_3 = CI^{\frac{\sigma+1}{\sigma}} \tag{3.29}$$

□

Concerning Assumption 2.2 we have

Proposition 3.2. *Assumption 2.2 holds with the constants C_2 and C_3 are given in (3.32) for the convex sets K_j, $j = J, \ldots, 1$, defined in (3.16).*

Proof. Let us consider $v,\ w \in K$, and let $v_j, w_j \in K_j$, $j = J, \ldots, 1$, be their decompositions defined as in (3.17), respectively. Now, we define

$$\begin{aligned} & v_{ji} = L_{h_j}(\theta_j^i(v_j - w_j)),\ i = 1, \ldots, I_j,\ \text{for } j = J, \ldots, 2, \\ & \text{and } v_{11} = v_1 - w_1 \end{aligned} \tag{3.30}$$

with θ_j^i in (3.22). In view of (3.18) and (3.30), we get that the first two conditions of Assumption 2.2 hold.

We estimate now the constants C_2 and C_3. For $j = J, \ldots, 2$, from (3.23) and (3.24), we get

$$||v_{ji}||_{1,\sigma}^\sigma \le C(|v_j - w_j|_{1,\sigma}^\sigma + (1 + \frac{1}{\delta_j})^\sigma ||v_j - w_j||_{0,\sigma}^\sigma)$$

Using this equation, the proof is similar with that of the previous proposition. For $j = J, \ldots, 2$, in view of (3.19) and (3.20), we have

$$||v_{ji}||_{1,\sigma}^\sigma \le CC_{d,\sigma}(h_{j-1}, h_J)^\sigma |v - w|_{1,\sigma}^\sigma$$

and we use (3.27) for the estimation of $||v_{11}||_{1,\sigma}$. From these equations, we get

$$\sum_{j=1}^{J} \sum_{i=1}^{I_j} ||v_{ji}||_{1,\sigma}^\sigma \le CI \left[\sum_{j=2}^{J} C_{d,\sigma}(h_{j-1}, h_J)^\sigma \right] |v - w|_{1,\sigma}^\sigma \tag{3.31}$$

Consequently, the constants C_2 and C_3, can be written as

$$C_2 = CI^{\frac{1}{\sigma}} [\sum_{j=2}^{J} C_{d,\sigma}(h_{j-1}, h_J)^\sigma]^{\frac{1}{\sigma}} \text{ and } C_3 = 0 \tag{3.32}$$

□

The constants C_1 and β_{jk}, $j,\ k = J, \dots, 1$, can be taken as in (2.3) and (2.7), but better choices are available in the case of the multigrid methods in the next section. As we see form the above estimations, the convergence rates given in Theorem 2.1 depend on the functional F, the maximum number of the subdomains on each level, I, and the number J of levels. The number of subdomains on levels can be associated with the number of colors needed to mark the subdomains such that the subdomains with the same color do not intersect with each other. Since this number of colors depends in general on the dimension of the Euclidean space where the domain lies, we can conclude that our convergence rate essentially depends on the number J of levels.

We first estimate the constants C_1–C_3 as functions of J. To this end, in the remainder of this section, C will be a generic constant which does not depend on J. Writing $S_{d,\sigma}(J) = \left[\sum_{j=2}^{J} C_{d,\sigma}(h_{j-1}, h_J)^{\sigma}\right]^{\frac{1}{\sigma}}$ from (3.1) and (3.9), we can consider

$$S_{d,\sigma}(J) = \begin{cases} (J-1)^{\frac{1}{\sigma}} & \text{if } d = \sigma = 1 \\ & \text{or } 1 \le d < \sigma < \infty \\ CJ & \text{if } 1 < d = \sigma < \infty \\ C^J & \text{if } 1 \le \sigma < d < \infty \end{cases} \tag{3.33}$$

in our estimations. In this general framework, we take C_1, and β_{jk}, $j,\ k = J, \dots, 1$, as in (2.3) and (2.7),

$$C_1 = CJ^{\frac{\sigma-1}{\sigma}} \text{ and } \max_{k=1,\cdots,J} \sum_{j=1}^{J} \beta_{kj} = J \tag{3.34}$$

Also, from (3.29) and (3.32), we get

$$C_2 = CS_{d,\sigma}(J) \text{ and } C_3 = \begin{cases} C & \text{for Algorithms 2.1 and 2.3} \\ 0 & \text{for Algorithms 2.2 and 2.4} \end{cases} \tag{3.35}$$

As a consequence of Theorem 2.1 and Propositions 3.1 and 3.2 we have

Corollary 3.1. *Let us consider the finite element spaces V_{h_j} defined in (3.4) which are associated with the levels $j = 1, \dots, J$, and their subspaces $V_{h_j}^i$, $i = 1, \dots, I_j$, given in (3.5), which are associated with the level domain decompositions. Also, let K be the closed convex subset of $V = V_J$ given in (3.7), which is decomposed as a sum of the level closed convex sets $K_j \subset V_{h_j}$,*

$j = J, \dots, 1$, defined in (3.16). If F is a Gâteaux differentiable functional on V which is supposed to be coercive and to satisfy (2.4), then the approximation sequences u^n, $n \geq 0$ obtained from Algorithms 2.1–2.4 converge to the solution u of the one-obstacle problem (3.6) and the error estimations in Theorem 2.1 hold. The constants $\tilde{C}_1$ and $\tilde{C}_2$ in these error estimations depend on the number of levels J through the constants C_1–C_3 given in (3.34) and (3.35).

REMARK 3.2. 1) The results of this section have referred to problems in $W^{1,\sigma}$ with Dirichlet boundary conditions, and the functions corresponding to the coarse levels have been extended with zero outside the domains Ω_j, $j = J-1, \dots, 1$. Let us assume that the problem has mixed boundary conditions: $\partial\Omega_J = \Gamma_d \cup \Gamma_n$, with Dirichlet conditions on Γ_d and Neumann conditions on Γ_n. In this case, if a node of $\mathcal{T}_{h_j}$, $j = J-1, \dots, 1$, lies in $\mathrm{Int}(\Gamma_n)$, we have to assume that all the sides of the elements $\tau \in \mathcal{T}_{h_j}$ having that node are included in Γ_n.

2) Similar convergence results with those ones presented in this section can be obtained for problems in $(W^{1,s})^d$.

4 Multigrid methods

In the above multilevel methods a mesh is the refinement of that one on the previous level, but the domain decompositions are almost independent from one level to another. We obtain similar multigrid methods by decomposing the level domains by the supports of the nodal basis functions. Consequently, the subspaces $V_{h_j}^i$, $i = 1, \dots, I_j$, are one-dimensional spaces generated by the nodal basis functions associated with the nodes of $\mathcal{T}_{h_j}$, $j = J, \dots, 1$. In this case Algorithms 2.1–2.4 are V-cycle multigrid iterations in which the smoothing steps are performed by a combination of multiplicative methods with additive ones. Evidently, similar results can be given for the W-cycle multigrid iterations.

In this section, we show that the estimations given in (2.3) and (2.7) for the constants C_1 and β_{jk}, $j,\ k = J, \dots, 1$ can be improved in the case of the multigrid methods. Finally, we summarize the previous results by writing the convergence rates of the four algorithms as functions of the number J of the levels, for the varied values of the constants p, q, σ and d.

Concerning the constants β_{jk}, $j, k = J, \dots, 1$, in (2.6), we can prove (see [4] or [5], for instance) that, in the case of the multigrid methods, there exist such constants such that

$$\max_{k=1,\dots,J}\sum_{j=1}^{J}\beta_{kj}=C$$

where C is a constant independent of the meshes and their number. Also, the constant C_1 in (2.2) is estimated in Lemma 4.1 in [4] or [5],

$$C_1=(n!)^{\frac{1}{\sigma}}C^{\frac{n-1}{n}}\left(I\frac{\gamma^{\frac{d}{n}}}{\gamma^{\frac{d}{n}}-1}\right)^{\frac{n-1}{\sigma}}$$

where $n\in\mathbf{N}$, $n-1<\sigma\le n$, and C is a constant independent of the meshes and their number.

Now, we shall write the convergence rate of the multigrid Algorithms 2.1–2.4 in function of the number J of levels. To this end, we write the error estimations in Theorem 2.1 of the four algorithms using the above estimations of C_1 and $\max_{k=J,\dots,1}\sum_{j=1}^{J}\beta_{kj}$, and C_2 and C_3 given in (3.35). In order to be more conclusive, we limit ourselves to a typical example where

$$F(v)=\frac{1}{\sigma}||v||_{1,\sigma}^{\sigma}-L(v),\quad v\in W^{1,\sigma}(\Omega) \tag{4.1}$$

where L is a linear and continuous functional on $W^{1,\sigma}(\Omega)$, $\sigma>1$. In this case (see [1], for instance),

$$p=2,\ q=\sigma \text{ if } \sigma<2;\quad p=2,\ q=2 \text{ if } \sigma=2;\quad p=\sigma,\ q=2 \text{ if } \sigma>2$$

Evidently, we can use the same procedure for other problems, too.

For $\sigma=2$ and $p=q=2$, in view of (2.30), (2.23) and (3.35), we get

$$\tilde{C}_1(J)=\begin{cases} CS_{d,2}(J)^2 & \text{for Algorithms 2.1 and 2.2}\\ CJS_{d,2}(J)^2 & \text{for Algorithms 2.3 and 2.4}\end{cases} \tag{4.2}$$

and, from Theorem 2.1, we have

$$||u^n-u||_{1,2}^2\le\tilde{C}_0\left(1-\frac{1}{1+\tilde{C}_1(J)}\right)^n \tag{4.3}$$

where $\tilde{C}_0$ is a constant independent of J.

For $1 < q = \sigma < 2$ and $p = 2$, in view of (2.32), (2.23) and (3.35), we get

$$\tilde{C}_3(J) = \begin{cases} CJ^{\frac{(\sigma-1)(2-\sigma)}{\sigma}} S_{d,\sigma}(J)^2 & \text{for Algorithms 2.1 and 2.2} \\ CJ^{\frac{2(\sigma-1)}{\sigma}} S_{d,\sigma}(J)^2 & \text{for Algorithms 2.3 and 2.4} \end{cases} \tag{4.4}$$

From Theorem 2.1, we get that

$$||u^n - u||_{1,\sigma}^2 \le \tilde{C}_0 \frac{1}{\left(1 + n\tilde{C}_2(J)\right)^{\frac{\sigma-1}{2-\sigma}}} \tag{4.5}$$

where, in view of (2.34), we can take

$$\tilde{C}_2(J) = \frac{1}{1 + \tilde{C}_3(J)^{\frac{1}{\sigma-1}}} \tag{4.6}$$

For $p = \sigma > 2$ and $q = 2$, we get

$$\tilde{C}_3(J) = \begin{cases} CJ^{\frac{\sigma-2}{\sigma-1}} S_{d,\sigma}(J)^{\frac{\sigma}{\sigma-1}} & \text{for Algorithms 2.1 and 2.2} \\ CJS_{d,\sigma}(J)^{\frac{\sigma}{\sigma-1}} & \text{for Algorithms 2.3 and 2.4} \end{cases} \tag{4.7}$$

Finally, in this case, we have

$$||u^n - u||_{1,\sigma}^\sigma \le \tilde{C}_0 \frac{1}{\left(1 + n\tilde{C}_2(J)\right)^{\frac{1}{\sigma-2}}} \tag{4.8}$$

where

$$\tilde{C}_2(J) = \frac{1}{1 + \tilde{C}_3(J)^{\sigma-1}} \tag{4.9}$$

We make now some remarks on the above error estimations of the four algorithms. First, we point out that the above convergence results give global rate estimations. As we have expected, the multiplicative (over the levels) Algorithms 2.1 and 2.2 converge better than their additive variants, Algorithms 2.3 and 2.4. For the complementarity problems, we can compare the convergence rates of the four multigrid algorithms with the similar ones in the literature. In this case, $p = q = \sigma = d = 2$ in the above example, from (4.3) and (4.2), we get that the convergence rate of Algorithms 2.1 and 2.2 is of $1 - \frac{1}{1+CJ^2}$, and that of Algorithms 2.3 and 2.4 is of $1 - \frac{1}{1+CJ^3}$. These convergence rates are better, with a factor J, than those of the similar algorithms introduced in [4], which are of $1 - \frac{1}{1+CJ^3}$ and $1 - \frac{1}{1+CJ^4}$, respectively.

For the truncated monotone multigrid method, an asymptotic convergence rate of $1-\frac{1}{1+CJ^4}$, and under some conditions, of $1-\frac{1}{1+CJ^3}$, is found in [16] and [13]. An estimate of $1-\frac{1}{1+CJ^3}$ is also obtained in [16] for the asymptotic convergence rate of the standard monotone multigrid methods. In [13], it is mentioned that this asymptotic rate may be of $1-\frac{1}{1+CJ^2}$, or even of $1-\frac{1}{1+CJ}$, under some conditions.

Acknowledgment. This work has been supported by the CNCSIS under the contract ID-PCE nr. 566/2009.

References

[1] L. Badea, Convergence rate of a multiplicative Schwarz method for strongly nonlinear variational inequalities, in *Analysis and Optimization of Differential Systems*, V.Barbu et al. (eds.), Kluwer Academic Publishers, 2003, p. 31-42 (also available at http://www.imar.ro/ lbadea/).

[2] L. Badea, Convergence rate of a Schwarz multilevel method for the constrained minimization of non-quadratic functionals, *SIAM J. Numer. Anal.*, **44**, 2, 2006, p. 449-477.

[3] L. Badea, Additive Schwarz method for the constrained minimization of functionals in reflexive Banach spaces, in U. Langer et al. (eds.), *Domain decomposition methods in science and engineering XVII*, LNSE 60, Springer, 2008, p. 427-434.

[4] L. Badea, Multigrid methods for variational inequalities, *SIAM J. on Num. Anal*, submitted, 2010.

[5] L. Badea, Multigrid methods for variational inequalities, Preprint series of the Institute of Mathematics of the Romanian Academy, no. 1, 2010, available at http://www.imar.ro/ lbadea/.

[6] L. Badea, Multigrid methods with constraint level decomposition for variational inequalities, Preprint series of the Institute of Mathematics of the Romanian Academy, no. 3, 2010, available at http://www.imar.ro/ lbadea/.

[7] L. Badea, X.-C. Tai and J. Wang, Convergence rate analysis of a multiplicative Schwarz method for variational inequalities, *SIAM J. on Num. Anal*, **41**, 3, 2003, p. 1052-1073.

[8] A. Brandt and C. Cryer, Multigrid algorithms for the solution of linear complementary problems arising from free boundary problems, *SIAM J. Sci. Stat. Comput.*, **4**, 1983, p. 655-684.

[9] M. Dryja and O. B. Widlund, Domain decomposition algorithms with small overlap, *SIAM J. Sci. Comput.*, **15**, 3, 1994, p. 604-620.

[10] I. Ekeland and R. Temam, *Analyse convexe et problèmes variationnels*, Dunod, Paris, 1974.

[11] E. Gelman and J. Mandel, On multilevel iterative method for optimization problems, *Math. Program.*, **48**, 1990, p. 1-17.

[12] R. Glowinski, J. L. Lions and R. Trémolières, *Analyse numérique des inéquations variationnelles*, Dunod, Paris, 1976.

[13] C. Gräser and R. Kornhuber, Multigrid methods for obstacle problems, *J. Comput. Math.*, **27**, 1, 2009, p. 1-44.

[14] W. Hackbusch and H. Mittelman, On multigrid methods for variational inequalities, *Numer. Math.*, **42**, 1983, p. 65-76.

[15] R. Hoppe and R. Kornhuber, Adaptive multilevel methods for obstacle problems, *SIAM J. Numer. Anal.*, **31**, 2, 1994, p. 301-323.

[16] R. Kornhuber, Monotone multigrid methods for elliptic variational inequalities I. Numer. Math. **69**, (1994) pp. 167-184.

[17] R. Kornhuber, Monotone multigrid methods for elliptic variational inequalities II. Numer. Math. **72**, (1996) pp. 481-499

[18] Kornhuber, R., *Adaptive monotone multigrid methods for nonlinear variational problems.* Teubner-Verlag, Stuttgart (1997)

[19] R. Kornhuber and R. Krause, Adaptive multigrid methods for Signorini's problem in linear elasticity, *Comp. Visual. Sci.*, **4**, 2001, p. 9-20.

[20] R. Kornhuber and H. Yserentant, Multilevel methods for elliptic problems on domains not resolved by the coarse grid, *Contemporary Mathematics*, **180**, 1994, p. 49-60.

[21] J. L. Lions, *Quelques méthodes de résolution des problèmes aux limites non linéaires*, Dunod, Paris, 1969.

[22] Mandel, J., A multilevel iterative method for symmetric, positive definite linear complementarity problems, *Appl. Math. Opt.*, **11**, (1984), pp. 77-95.

[23] Mandel, J., Etude algébrique d'une méthode multigrille pour quelques problèmes de frontière libre, *C. R. Acad. Sci.*, **298**, Ser. I, (1984), pp. 469-472.

[24] X. C. Tai Rate of convergence for some constraint decomposition methods for nonlinear variational inequalities, *Numer. Math.*, **93**, 4, 2003, p. 755-786.

[25] X.-C. Tai and J. Xu, Global and uniform convergence of subspace correction methods for some convex optimization problems, *Math. Comp.*, 71, 2002, p. 105–124.

[26] A. Toselli and O. Widlund, *Domains Decomposition Methods - Algorithms and Theory*. Springer-Verlag, Berlin (2005).

[27] B. Wohlmuth and R. Krause, Monotone multigrid methods on nonmatching grids for nonlinear multibody contact problems, *SIAM Sci. Comput.*, **25**, 2003, p. 324-347.

[28] J. Xu, *Theory of multilevel methods*, PhD thesis, Cornell University, 1989

Annals of the Academy of Romanian Scientists
Series on Mathematics and its Applications
ISSN 2066 - 6594 Volume 3, Number 2 / 2011

TWO OPTIMAL CONTROL PROBLEMS IN CANCER CHEMOTHERAPY WITH DRUG RESISTANCE*

Werner Krabs† and Lothar von Wolfersdorf‡

Abstract

We investigate two well-known basic optimal control problems for chemotherapeutic cancer treatment modified by introducing a time-dependent "resistance factor". This factor should be responsible for the effect of the drug resistance of tumor cells on the dynamical growth for the tumor. Both optimal control problems have common pointwise but different integral constraints on the control. We show that in both models the usually practised bang-bang control is optimal if the resistance is sufficiently strong. Further, we discuss different optimal strategies in both models for general resistance.

MSC 2000: 49 J 15; 49 K 15; 92 C 50.

Keywords: Cancer chemotherapy, optimal control, drug resistance.

1 Introduction

Optimal control problems based on mathematical models for cancer chemotherapy have a long history and obtained a renewed interest in the last years

*Accepted for publication on March 13, 2011.

†Department of Mathematics, TU Darmstadt, Schloßgartenstr. 7, 64289 Darmstadt, Germany, e-mail: krabs@mathematik.tu-darmstadt-de

‡Lothar von Wolfersdorf passed away on November 30, 2010.

(cf. [2-6, 10-26]). There are further recent papers on mathematical models for immunotherapy and mixed immunotherapy and chemotherapy starting with papers by A. Kuznetsov and coworkers in the nineties (cf. [19,20], for instance) but which are not in our focus here. Instead the present paper follows the two pioneering papers by J.M. Murray in 1990 [16,17] (see also [24]) whose basic problems are modified in the following.

One difficulty in applying the considered optimal control problems in cancer chemotherapy is the occurance of optimal solutions which are seldom or not used in medical practice. A desired optimal solution by the physician is the bang-bang control consisting of a starting interval with maximal dose of drug followed by an interval of zero-therapy till the end of treatment (considering one cycle of the chemotherapeutic treatment). In particular, the therapy should theoretically end with an interval of zero-therapy to have the required minimum of the tumor cells population also somewhat later than at the practical end of treatment. To obtain optimal solutions of this type often a suitable choice of the objective functions is proposed (cf. [14, 17, 19, and 24]).

The aim of the present paper is to circumvent this difficulty taking into account the resistance of the tumor cells against drug (and further using the size of the tumor cells population at the final time as the natural objective function). Acquired and intrinsic resistance of the tumor cells against drug is an important but very complex phenomen in tumor therapy (cf. [8, 12]) and related deterministic models [4, 12, 14] and stochastic ones [2, 3] in dealing with it are developed recently. In our highly simplified model we only consider a summarizing effect of resistance by introducing a time-dependent "resistance factor" in front of the loss function of the tumor cells in the deterministic differential equation for the tumor growth. In particular, we do not distinguish between drug sensitive and resistant tumor cells like in [4, 12, 14].

Further, we deal with two basic problems where in each problem we have two restrictions, namely the usual pointwise inequality for the control function (which is in the dose of the drug administered per unit of time) and an integral inequality for the loss function of the normal cells in the first problem and for the drug dose itself in the second problem. To keep the mathematical analysis simple other restrictions like the pointwise limit for the size of the population of the normal cells like in [15, 17, and 24] are not taken into consideration. There is only one dynamic equation for the growth and suppression of the tumor cells and no one for the normal cells (but which could be easily supplemented).

Both optimal control problems show the desired effect that for (properly defined) “strong resistance” the above-named bang-bang control is the unique optimal control (cp. with the results in [3, 26], for instance). With respect to general resistance we have another picture. In the first problem in case of “weak resistance” the optimal control is the non desired “opposite” bang-bang control with starting interval of zero-therapy and final interval of maximal drug dose. On the other hand, in the second problem for general resistance an optimal control similar to the desired bang-bang control starting and ending with a subinterval of zero-therapy is to be expected as an example with Gompertzian growth show (cp. with other forms of optimal solutions in [14, 17], for instance). So, especially with respect to weak resistance the second problem seems preferable to the first problem.

The plan of the paper is as follows. After performing the mathematical modelling in Section 2 we investigate the first optimal control problem in Section 3 and the second optimal control problem with the example for Gompertzian growth in Section 4.

2 Mathematical Models

We denote the time-dependent number of cancer cells in the tumor by a function $T = T(t)$, $t \in I\!R_+$, which we assume to be differentiable with derivative $\dot{T}(t)$. The temporal development of the tumor cells population $T(t)$ in a given interval $[0, t_f]$ is governed by the differential equation

$$\dot{T}(t) = [f(T(t)) - \varphi(t)L(M(t))]T(t),\ t \in [0, t_f], \tag{2.1}$$

with the initial condition

$$T(0) = T_0 > 0. \tag{2.2}$$

The function $f = f(t), T \geq 0$ describes the dynamics of the tumor population $F(T) = f(T)T$ if there is no administration of drugs. We assume $f \in C^1(I\!R_+)$ with $f(T) > 0$, $f'(T) < 0$ for all relevant $T \geq 0$. This is fulfilled for many of the commonly used dynamics as Gompertz, logistic (Verhulst-Pearl) and other growth laws in an interval $[0, \theta]$ with maximal tumor population θ (cf. [5, 10, 16, 22, 24]).

By $L = L(M)$, $M \geq 0$ we denote the destruction rate of the drug level M. We assume that this loss function $L \in C^2(I\!R_+)$ satisfies $L(0) = 0$ and $L'(M) > 0$ for all relevant $M \geq 0$. This is fulfilled, for instance, for linear and fractional linear ("saturated") function L (cf. [10, 16, 24]). The drug level function $M = M(t)$ obeys the linear differential equation

$$\dot{M}(t) = -\delta M(t) + V(t),\ t \in [0, t_f], \tag{2.3}$$

and the initial condition

$$M(0) = 0 \tag{2.4}$$

with a positive drug decay rate δ where $V(t)$ denotes the drug dose that is administered per unit of time at time $t \in [0, t_f]$. In the following we assume $\delta \geq 0$, thus including the mathematical limit case $\delta = 0$ of no drug decay.

The drug dosis $V = V(t)$ per unit of time is considered as the control function in the model. We assume it to be a bounded measurable function, i.e. $V \in L^\infty(0, t_f)$, and to satisfy the pointwise condition

$$0 \leq V(t) \leq A \ \text{ for a.a. } \ t \in [0, t_f] \tag{2.5}$$

where $A > 0$ is a prescribed constant, the maximum drug dosis per unit of time. Further, below we require additionally an integral condition which we regard as responsible for the compatibility of the treatment.

The new feature in this model is the introduction of the function $\varphi = \varphi(t)$, $t \in [0, t_f]$, which is assumed to be in $C^1[0, t_f]$ satisfying $\varphi(t) > 0$ in $[0, t_f)$ and normed by $\varphi(0) = 1$. The factor φ in Eq. (2.1) should - in a most simple way - describe the total effect of inner influences (like drug resistance) and other ones (like accompanying therapies) on the destruction rate of the tumor cells by the drug during the treatment. Especially, the influence of the drug resistance of the tumor cells will be expressed by a function $\varphi \in C^1[0, t_f]$ with $\varphi(0) = 1$, $\dot{\varphi}(t) \leq 0$ in $[0, t_f]$ and $\varphi(t) > 0$ in $[0, t_f]$. (We call such a function a "resistance factor" in the following).

The integral condition in *problem 1* now reads

$$\int_0^{t_f} \varphi_0(t) L_0(M(t))dt \leq B \tag{2.6}$$

with a prescribed constant $B > 0$ where $L_0(M)$ is the destruction rate of the normal cells for which we assume the same properties as for the loss function $L(M)$ of the tumor cells above and $\varphi_0 \in C^1[0, t_f]$ with $\varphi_0(t) > 0$ in $[0, t_f)$, $\varphi_0(0) = 1$ is a weight function possessing the analogous meaning for the normal cells as φ for the tumor cells. In *problem 2* we simply require that

$$\int_0^{t_f} V(t)dt \leq B \tag{2.7}$$

with a given constant $B > 0$.

The aim of chemotherapeutic treatment is to make the tumor cells population $T(t_f)$ at the end of the treatment as small as possible. In view of Eq. (2.1) this can be written in the usual form of the minimum condition

$$\int_0^{t_f} [f(T(t)) - \varphi(t)L(M(t))]T(t)dt \to \min. \tag{2.8}$$

We further remark that for a given $V \in L^\infty(0, t_f)$ the solution of (2.3), (2.4) has the form

$$M(t) = M[V](t) = \int_0^t e^{\delta(s-t)} V(s)ds, t \in [0, t_f]$$

which implies $M \in C[0, t_f]$. By our assumptions on f, φ, L we then have $T \in C^1[0, t_f]$ for the corresponding solution $T = T[V]$ of Eq. (2.1).

Our *optimal control problems* are now defined by the minimum condition (2.8) for the state equations (2.1) - (2.4) with the constraints (2.5), (2.6) (*problem 1*) or (2.5), (2.7) (*problem 2*). Here the integral constraints (2.6) and (2.7) can be taken, respectively, in the form

$$Q(t_f) \le B \quad \text{or} \quad U(t_f) \le B \tag{2.9}$$

where the additional state functions $Q = Q(t)$ and $U = U(t)$ are given by the integrals

$$Q(t) = \int_0^t \varphi_0(s) L_0(M(s)ds \quad \text{and} \quad U(t) = \int_0^t V(s)ds,$$

respectively, or equivalently by the additional state equations

$$\dot{Q}(t) = \varphi_0(t) L_0(M(t)), \; t \in [0, t_f] \tag{2.10}$$

with $Q(0) = 0$ and

$$\dot{U}(t) = V(t), t \in [0, t_f] \tag{2.11}$$

with $U(0) = 0$, respectively.

These optimal control problems always have solutions as follows by adapting the *existence proof* by J.M. Murray in [16] on the basis of Theorem 5.4.4 in [1] (taking into account that for the admissible control $V(t) = 0$ in $[0, t_f]$ the state equation (2.1) has a continuous solution $T(t)$ in $[0, t_f]$ with finite $T(t_f)$ and because of the finite interval $[0, t_f]$ also the parameter $\delta = 0$ in Eq. (2.3) is possible).

Finally, we simplify the mathematical analysis for our problems slightly by applying the usual substitution $y = \ell n T$ for $T > 0$. Then the differential equation (2.1) is transformed into

$$\dot{y}(t) = f(e^{y(t)}) - \varphi(t) L(M(t)), t \in [0, t_f] \tag{2.12}$$

and the initial condition (2.2) reads

$$y(0) = y_0 = \ell n T_0. \tag{2.13}$$

The minimum condition (2.8) takes the form

$$\int_0^{t_f} [f(e^{y(t)}) - \varphi(t)L(M(t))]dt \to \text{ Min.} \tag{2.14}$$

The optimal control problems to be solved are then given by the minimum condition (2.14) for the state equations (2.12), (2.13), (2.3), (2.4), and (2.10) or (2.11), respectively, under the constraints (2.5), (2.9).

3 Solutions of the first problem

We determine optimal solutions of problem (2.12 - 2.14), (2.5), (2.9), (2.10) as usual with the aid of the maximum principle [9]. The Hamiltonian of the problem is given by

$$\begin{aligned} &H(t, y, M, Q, V, p_1, p_2, p_3, \lambda_0) \\ &= (f(e^y) - \varphi(t)L(M))(p_1 - \lambda_0) \\ &+(V - \delta M)p_2 + \varphi_0(t)L_0(M)p_3 \end{aligned} \tag{3.1}$$

with the parameter λ_0 and the adjoint state functions p_k, $k = 1, 2, 3$. If $(\hat{y}, \hat{M}, \hat{Q}, \hat{V})$ is an optimal quadruple there exist a number $\lambda_0 \geq 0$ and three functions $p_k \in C^1[0, t_f]$, $k = 1, 2, 3$ with $(\lambda_0, p_1, p_2, p_3) \neq (0, 0, 0, 0)$ satisfying the differential equations

$$\dot{p}_1(t) = -f'(e^{\hat{y}(t)})e^{\hat{y}(t)}(p_1(t) - \lambda_0) \tag{3.2}$$

$$\dot{p}_2(t) = \varphi(t)L'(\hat{M}(t))(p_1(t) - \lambda_0) + \delta p_2(t) - \varphi_0(t)L_0'(\hat{M}(t))p_3(t) \tag{3.3}$$

$$\dot{p}_3(t) = 0 \tag{3.4}$$

in $[0, t_f]$ and the transversality conditions in t_f

$$p_1(t_f) = 0, p_2(t_f) = 0, \text{ and } p_3(t_f) \leq 0, p_3(t_f)(\hat{Q}(t_f) - B) = 0 \tag{3.5}$$

such that for a.a. $t \in [0, t_f]$ the maximum condition

$$\hat{V}(t)p_2(t) = \max_{0 \le V \le A} [V p_2(t)] \tag{3.6}$$

is valid. From (3.4) and (3.5) it follows that p_3 is a nonpositive constant which vanishes if $\hat{Q}(t_f) < B$.

We define $\tilde{p}_1(t) = p_1(t) - \lambda_0$ and

$$g(t) = -f'(e^{\hat{y}(t)})e^{\hat{y}(t)} > 0, \ t \in [0, t_f]. \tag{3.7}$$

Then from (3.2) we have $\dot{\tilde{p}}_1(t) = g(t)\tilde{p}_1(t)$ which gives

$$\tilde{p}_1(t) = \tilde{p}_1(0) \exp\Big(\int_0^t g(s)ds\Big), \ t \in [0, t_f]. \tag{3.8}$$

From $p_1(t_f) = 0$ we obtain

$$\tilde{p}_1(t_f) = \tilde{p}_1(0) \exp\Big(\int_0^{t_f} g(t)dt\Big) = -\lambda_0 \le 0 \tag{3.9}$$

which shows that $\tilde{p}_1(t) \le 0$ for all $t \in [0, t_f]$.

We further put

$$h(t) = \varphi(t)L'(\hat{M}(t))\tilde{p}_1(t) - \varphi_0(t)L_0'(\hat{M}(t))p_3. \tag{3.10}$$

From (3.3) we get

$$\dot{p}_2(t) = \delta p_2(t) + h(t), \ t \in [0, t_f] \tag{3.11}$$

which yields $p_2(t) = e^{\delta t}H(t)$ with

$$H(t) = p_2(0) + \int_0^t e^{-\delta s}h(s)ds, \ t \in [0, t_f]. \tag{3.12}$$

In view of $p_2(t_f) = 0$ we have

$$p_2(0) = -\int_0^{t_f} e^{-\delta t} h(t)dt \tag{3.13}$$

which implies

$$p_2(t) = -e^{\delta t} \int_t^{t_f} e^{-\delta s} h(s)ds, \ t \in [0, t_f]. \tag{3.14}$$

Now we distinguish the two cases

$$B \geq Q_A(t_f) \equiv \int_0^{t_f} \varphi_0(t) L_0(M_A(t))dt \tag{3.15}$$

where

$$M_A(t) = \frac{A}{\delta}[1 - e^{-\delta t}] \text{ if } \delta > o, At \text{ if } \delta = 0$$

is the solution of (2.3), (2.4) for $V(t) = A$ a.e. in $[0, t_f]$, and

$$B < Q_A(t_f) \equiv \int_0^{t_f} \varphi_0(t) L_o(M_A(t))dt. \tag{3.16}$$

If (3.15) is fulfilled we have the optimal solution $\hat{V}(t) = A$ a.e. in $[0, t_f]$. So we can assume (3.16) in the following. In this case the equality

$$\hat{Q}(t_f) \equiv \int_0^{t_f} \varphi_0(t) L_0(\hat{M}(t))dt = B \tag{3.17}$$

for the optimal solutions must hold. We prove this by contradiction. If $\hat{Q}(t_f) < B$ we have $p_3 = 0$. In the anormal case $\lambda_0 = 0$ by (3.9), (3.8) and (3.10), (3.14) this implies $p_1(t) = \tilde{p}_1(t) = 0$ and $p_2(t) = 0$ in $[0, t_f]$ which contradicts the condition $(\lambda_0, p_1, p_2, p_3) \neq (0, 0, 0, 0)$. In the normal case $\lambda_0 > 0$ by (3.8) we would have $\tilde{p}_1(t) < 0$ and hence by (3.10) also $h(t) < 0$ in $[0, t_f]$ which by (3.14) yields $p_2(t) > 0$ in $[0, t_f]$. The maximum condition (3.6) then would give $\hat{V}(t) = A$ for a. a. $t \in [0, t_f]$. This implies $Q_A(t_f) = \hat{Q}(t_f) < B$, a contradiction to (3.16). Therefore, (3.17) and $p_3 < 0$ hold true.

We further show because of the condition (3.17) the abnormal case $\lambda_0 = 0$ cannot occur. Namely, from $\lambda_0 = 0$ as before we obtain $\tilde{p}_1(t) = p_1(t) = 0$ in $[0, t_f]$ implying

$$h(t) = -\varphi_0(t) L_0'(\hat{M}(t)) p_3 > 0, \; t \in [0, t_f]$$

from (3.10). By (3.14) it follows that $p_2(t) < 0$ in $[0, t_f]$. Then (3.6) yields $\hat{V}(t) = 0$ for a.a. $t \in [0, t_f]$ which by (2.3), (2.4) leads to $\hat{M}(t) = 0$ in $[0, t_f]$ and by $L(0) = 0$ to $\hat{Q}(t_f) = 0$, a contradiction to (3.17). Summing up, in the case (3.16) equality (3.17) is valid and we have $\lambda_0 > 0$, $\tilde{p}_1(t) < 0$ in $[0, t_f]$ and $p_3 < 0$.

We introduce the functions

$$q(t) = \frac{L_0'(\hat{M}(t))}{L'(\hat{M}(t))}, \Delta(t) = \varphi(t)\tilde{p}_1(t) - \varphi_0(t) q(t) p_3 \tag{3.18}$$

so that by (3.10) we have $h(t) = L'(\hat{M}(t))\Delta(t)$ with $sign h(t) = sign \Delta(t)$. Now we make the *assumption* that Δ is a strictly increasing function in $[0, t_f]$ which is fulfilled if

$$\frac{d}{dt}\Delta(t) \equiv \frac{d}{dt}[\varphi(t)\tilde{p}_1(t) - \varphi_0(t) q(t) p_3] > 0 \text{ in } (0, t_f). \tag{3.19}$$

We distuingish the three cases

(i) $\Delta(0) \geq 0$

(ii) $\Delta(0) < 0, \Delta(t_f) \leq 0$

(iii) $\Delta(0) < 0, \Delta(t_f) > 0$.

In case (i) we have $\Delta(t) > \Delta(0) \geq 0$ in $[0, t_f]$ implying $h(t) > 0$ in $(0, t_f]$ and $p_2(t) < 0$ in $[0, t_f)$ by (3.14). The condition (3.6) then yields the solution $\hat{V}(t) = 0$ a.e. in $[0, t_f]$ which is not possible.

In case (ii) we have $\Delta(t) < \Delta(t_f) \leq 0$ in $[0, t_f]$ which gives $h(t) < 0$ in $[0, t_f]$ and $p_2(t) > 0$ in $[0, t_f]$ by (3.14) again. In view of (3.6) then $\hat{V}(t) = A$ a.e. in $[0, t_f]$ which is also not allowed in the case of (3.16).

It remains the case (iii). By the strict monotonicity of Δ there exists exactly one $t_1 \in (0, t_f)$ with $\Delta(t_1) = 0$, $\Delta(t) < 0$ in $[0, t_1)$, and $\Delta(t) > 0$ in $(t_1, t_f]$. This implies the analogous inequalities for h. Therefore the function H in (3.12) is strictly decreasing from $H(0) = p_2(0)$ to $H(t_1) < p_2(0)$ and then strictly increasing from $H(t_1)$ to $H(t_f) = 0$. If now $p_2(0) \leq 0$ were true we would get $H(t) < 0$ and hence $p_2(t) < 0$ in $(0, t_f)$. This would imply $\hat{V}(t) = 0$ a.e. in $[0, t_f]$ again. Therefore, it must be $p_2(0) > 0$. Then there exists exactly one $t_0 \in (0, t_1)$ with $H(t_0) = 0$, $H(t) > 0$ in $[0, t_0)$ and $H(t) < 0$ in $(t_0, t_f]$ which implies $p_2(t) > 0$ in $[0, t_0)$ and $p_2(t) < 0$ in $(t_0, t_f]$. The maximum condition (3.6) yields the unique optimal solution

$$\hat{V}(t) = \begin{cases} A & \text{for a.a.} \quad t \in [0, t_0) \\ 0 & \text{for a.a.} \quad t \in (t_0, t_f] \end{cases} \tag{3.20}$$

where $t_0 \in (0, t_f)$ can be defined as the (unique) solution of the equation

$$\int_0^{t_f} \varphi_0(t) L_0(\hat{M}(t)) dt = B \tag{3.21}$$

with

$$\hat{M}(t) = \begin{cases} \frac{A}{\delta}[1 - \bar{e}^{\delta t}] & \text{if} \quad \delta > 0, At \quad \text{if} \quad \delta = 0 \quad \text{for} \quad t \in [0, t_0] \\ \frac{A}{\delta}[e^{\delta t_0} - 1]\bar{e}^{\delta t} & \text{if} \quad \delta > 0, At_0 \quad \text{if} \quad \delta = 0 \quad \text{for} \quad t \in (t_0, t_f] \end{cases}$$

following from (3.17) and (2.3), (2.4) with (3.20).

We summarize the result in

THEOREM 3.1

(i) Let (3.15)be fulfilled. Then problem 1 has the unique optimal solution $\hat{V}(t) = A$ *a.e. in* $[0, t_f]$.

ii) Let (3.16) be fulfilled and the function Δ *in (3.18) strictly increasing. Then problem 1 has the unique optimal solution (3.20) with (3.21).*

REMARKS. The monotonicity assumption on Δ in Theorem 3.1 is an implicit condition on φ (and φ_0) since the functions $\tilde{p}_1$by (3.8) and q by (3.18) in general depend on the optimal solution $\hat{V}$ of the problem with the function

φ in (2.1) (and φ_0 in (2.6)). But this dependence can be well derived from (3.20) with (3.21) and Eqs. (2.1), (2.3). Moreover, in the important particular case $L = cL_0$ with a constant $c > 0$ and $L_0 \in C^1(I\!R_+)$ (cf. [16, 17]) we have $q(t) = c$ and for $\varphi_0(t) = 1$ in $[0, t_f]$ the sufficient condition (3.19) reduces to the simple condition

$$\dot{\varphi}(t) + g(t)\varphi(t) < 0 \text{ in } (0, t_f) \tag{3.22}$$

with the positive function $g = -f'(\hat{T})\hat{T}$ by (3.7). Further, in the special case of Gompertzian growth $f(T) = \lambda \ell n \frac{\theta}{T}(\lambda, \theta > 0)$ we have $g = \lambda$, a constant which is independent of the optimal solution $\hat{V}$.

Condition (3.22) is for instance satisfied, if

$$\varphi(t) = \exp(-(\int_0^t g(s)ds + \gamma t)),\ t \in [0, t_f],$$

for some $\gamma > 0$.

In general, it remains the dependence of Δ on the (negative) parameter p_3 or equivalently on the (positive) quotient $p_3/\tilde{p}_1(0)$ which are not directly expressed by the optimal solution $\hat{V}$. To avoid this dependence we derive a further sufficient criterion for the optimal solution (3.20) in the sequel.

LEMMA 3.2
Under the conditions (3.16) and

$$\frac{d}{dt}[\rho(t)p_2(t)] < 0 \text{ in } (0, t_f) \tag{3.23}$$

with a nonnegative function $\rho \in C^1(0, t_f)$ the optimal solution of problem 1 is uniquely determined and has the form (3.20).

Proof. Because of (3.23) the optimal solution cannot contain singular parts in subintervals where $\dot{p}_2(t) = p_2(t) = 0$ and parts of the form

$$\hat{V}(t) = \begin{cases} 0 & \text{a.e. in } [t_1, \tau) \\ A & \text{a.e. in } (\tau, t_2] \end{cases}$$

with $0 \le t_1 < \tau < t_2 \le t_f$ where $p_2(t) \le 0$ in (t_1, τ), $p_2(\tau) = 0$, $p_2(t) \ge 0$ in (τ, t_2) and $\dot{p}_2(\tau) \ge 0$. Further, the solutions $\hat{V}(t) = 0$ a.e. in $[0, t_f]$ and

$\hat{V}(t) = A$ a.e. in $[0, t_f]$ are not possible in view of (3.16) with (3.17). This proves the lemma.

In view of (3.11) the condition (3.23) can be written in the form

$$[\dot{\rho}(t) + \delta g(t)]p_2(t) + \rho(t)h(t) < 0 \text{ in } (0, t_f) \tag{3.24}$$

with h defined in (3.10). Taking

$$\rho(t) = \exp(\int_0^t \mu(s)ds), \ \mu \in C(0, t_f)$$

and $r(t) = \mu(t) + \delta \in C(0, t_f)$ condition (3.24) simply writes

$$r(t)p_2(t) + h(t) < 0 \text{ in } (0, t_f).$$

By (3.10) and (3.14) this means

$$\Delta_1(t) + p_3\Delta_2(t) < 0 \text{ in } (0, t_f) \tag{3.25}$$

where

$$\begin{aligned}
\Delta_1(t) &= \varphi(t)L'(\hat{M}(t))\tilde{p}_1(t) - r(t)\int_t^{t_f} e^{\delta(t-s)}\varphi(s)L'(\hat{M}(s))\tilde{p}_1(s)ds, \\
\Delta_2(t) &= r(t)\int_t^{t_f} e^{\delta(t-s)}\varphi_0(s)L_0'(\hat{M}(s))ds - \varphi_0(t)L_0'(\hat{M}(t)).
\end{aligned}$$

We now choose $r \in C(0, t_f)$ such that $\Delta_2(t) = 0$ in $(0, t_f)$, i.e.

$$r(t) = \frac{e^{-\delta t}\varphi_0(t)L_0'(\hat{M}(t))}{\int_t^{t_f} e^{-\delta s}\varphi_0(s)L_0'(\hat{M}(s))ds}.$$

Then (3.25) simplifies to the condition $\Delta_1(t) < 0$ in $(0, t_f)$ or defining further the quotient

$$q_1(t) = \frac{\tilde{p}_1(t)}{\tilde{p}_1(0)} = \exp(-\int_0^t f'(\hat{T}(s))\hat{T}(s)ds) > 0 \tag{3.26}$$

by (3.7), (3.8) to the *integral inequality*

$$\begin{aligned} & q_1(t)\varphi(t) \int_t^{t_f} e^{-\delta s}\varphi_0(s)L_0'(\hat{M}(s))ds \\ & > q(t)\varphi_0(t) \int_t^{t_f} e^{-\delta s}\varphi(s)L'(\hat{M}(s))q_1(s)ds \end{aligned} \tag{3.27}$$

in $(0, t_f)$ where q is defined in (3.18). A sufficient condition for (3.27) is the *differential condition*

$$\begin{aligned} & \tfrac{d}{dt}[q_1(t)\varphi(t)] \int_t^{t_f} e^{-\delta s}\varphi_0(s)L_0'(\hat{M}(s))ds \\ & < \tfrac{d}{dt}[q(t)\varphi_0(t)] \int_t^{t_f} e^{-\delta s}\varphi(s)L'(\hat{M}(s))q_1(s)ds \text{ in } 0, (t_f). \end{aligned} \tag{3.28}$$

Summing up we obtain

THEOREM 3.2

i) *Let (3.16) and (3.27) with (3.18), (3.26) be fulfilled. Then problem 1 has the unique optimal solution (3.20) with (3.21).*

ii) *The integral condition (3.27) is satisfied if the differential condition (3.28) is valid.*

REMARKS. The conditions (3.27) and (3.28) do not contain the unknown parameters p_3 and $\tilde{p}_1(0)$. In the particular case $L = cL_0$ with $\varphi_0(t) = 1$ from (3.28) we get the condition (3.22) again.

Finally, we briefly deal with the cases where in Theorem 3.1 the function Δ is strictly decreasing and the inequalities (3.19) and (3.27) in Theorems 3.1 and 3.2, respectively, are fulfilled with the opposite signs. In particular, this is the case if $L = cL_0$ and $\varphi(t) = \varphi_0(t) = 1$ on $[0, t_f]$. Then the above analysis shows that the unique optimal solution of problem 1 is

$$\hat{V}(t) = \begin{cases} 0 & \text{for} \quad a.a. t \in [0, t_*) \\ A & \text{for} \quad a.a. t \in (t_*, t_f] \end{cases} \tag{3.29}$$

where $t \in (0, t_f)$ is the (unique) solution of the equation

$$\int_{t_*}^{t_f} \varphi_0(t) L_0(\hat{M}(t)) dt = B$$

with $\hat{M}(t) = 0$ for $t \in [0, t_*]$ and

$$\hat{M}(t) = \frac{A}{\delta}[d^{\delta(t_* - t} - 1] \text{ if } \delta > 0,\ A(t - t_*) \text{ if } \delta = 0$$

for $t \in [t_*, t_f]$ following from (3.17) and (2.3), (2.4) with (3.29) again.

In case of the conditions (3.19) or (3.27), (3.28) for a resistance factor φ (with some associated φ_0) we say that we have *strong resistance* of the tumor cells against the drug, and in case of these conditions with the opposite sign *weak resistance*.

4 Solutions of the second problem

Problem (2.12 - 2.14), (2.5), (2.9), (2.11) possesses the Hamiltonian

$$\begin{aligned} &H(t, y, M, U, V, p_1, p_2, p_3, \lambda_0) \\ &= (f(e^y) - \varphi(t)L(M))(p_1 - \lambda_0) + (V - \delta M)p_2 + Vp_3 \end{aligned} \tag{4.1}$$

with the parameter λ_0 and the adjoint state functions $p_k, k = 1, 2, 3$. If $(\hat{y}, \hat{M}, \hat{U}, \hat{V})$ is an optimal quadruple, by the maximum principle [9], there exist a number $\lambda_0 \geq 0$ and three functions $p_k \in C^1[0, t_f]$, $k = 1, 2, 3$ with $(\lambda_0, p_1, p_2, p_3) \neq (0, 0, 0, 0)$ satisfying the differential equations

$$\dot{p}_1(t) = -f'(e^{\hat{y}(t)})e^{\hat{y}(t)}(p_1(t) - \lambda_0) \tag{4.2}$$

$$\dot{p}_2(t) = \varphi(t)L'(\hat{M}(t))(p_1(t) - \lambda_0) + \delta p_2(t) \tag{4.3}$$

$$\dot{p}_3(t) = 0 \tag{4.4}$$

in $[0, t_f]$ and the transversality conditions in t_f

$$p_1(t_f) = 0, p_2(t_f) = 0, \text{ and } p_3(t_f) \leq 0, p_3(t_f)(\hat{U}(t_f) - B) = 0 \tag{4.5}$$

such that for a.a. $t \in [0, t_f]$ the maximum condition

$$\hat{V}(t)(p_3(t) + p_3) = \max_{0 \leq V \leq A} [V(p_2(t) + p_3)] \tag{4.6}$$

holds. By (4.4), (4.5) p_3 is a nonpositive constant which vanishes if $\hat{U}(t_f) < B$.

We remark that in the limit case $\delta = 0$ in view of (2.3), (2.4) and (2.11) the quantities U and M coincide. Hence U, p_3 could be omitted and formally $p_2(t) + p_3$ replaced by a new $p_2(t)$.

We define the functions $\tilde{p}_1$ and g as in problem 1 with the relations (3.7) - (3.9). Further we have the relations (3.11) - (3.14) for p_2 if we replace the function h in (3.10) by

$$h_0(t) = \varphi(t) L'(\hat{M}(t), \tilde{p}_1(t), \; t \in [0, t_f]. \tag{4.7}$$

Discussing the optimal solutions of problem 2 we distinguish the two cases $B \geq t_f A$ and $B < t_f A$. For $B \geq T_f A$ the obvious solution is $\hat{V}(t) = A$ for a.a. $t \in [0, t_f]$. For $B < t_f A$ we have the equality

$$\hat{U}(t_f) = \int_0^{t_f} \hat{V}(t) dt = B \tag{4.8}$$

and the inequalities $p_3 < 0$, $\lambda_0 > 0$, and $\tilde{p}_1(t) < 0$ in $[0, t_f]$ which can be shown as above in problem 1. By (4.7) this implies $h_0(t) < 0$ in $[0, t_f]$ which by (3.11) and (3.14) gives

$$\dot{p}_2(t) - \delta p_2(t) < 0, p_2(t) > 0 \text{ in } [0, t_f). \tag{4.9}$$

If additionally $\varphi(t_f) > 0$ then also $h_0(t_f) < 0$ and consequently $\dot{p}_2(t_f) < 0$.

From (4.9) we obtain a first result about the form of the optimal solutions in the case $B < t_f A$.

LEMMA 4.1

For $B < t_f A$ the optimal solutions of problem 2 do not contain a solution part of the form

$$\hat{V}(t) = A \text{a.e. for } t \in [\tau, t_f], \tau \in [0, t_f) \tag{4.10}$$

and if $\varphi(t_f) > 0$ they do not contain singular parts in intervals of the form $[\tau, t_f]$ with $\tau \in [0, t_f)$.

Proof: The assertion (4.10) for $\tau = 0$ follows from (4.14). For $\tau > 0$ we have $p_2(t) + p_3 \geq 0$ in $(\tau, t_f]$ and $p_2(\tau) + p_3 = 0$ implying $\dot{p}_2(\tau) \geq 0$, but since $p_2(\tau) = -[p_2(t_f) + p_3] \leq 0$ by (4.9) it must be $\dot{p}_2(\tau) < \delta p_2(\tau)$ and $\dot{p}_2 < 0$.

The proof for the singular parts is a consequence of the condition $\dot{p}_2(t) = 0$ in $[\tau, t_f]$ which leads to a contradiction to $\dot{p}_2(t_f) < 0$ from (4.9).

Lemma 4.1 shows that the optimal solutions of problem 2 end with an interval of zero-therapy if $\varphi(t_f) > 0$.

We now give a *sufficient condition* for the optimal solutions being of the (in practice desired) bang-bang control type.

LEMMA 4.2

Under the conditions $B < t_f A$ and

$$\dot{p}_2(t) < 0 \text{ in } (0, t_f) \tag{4.11}$$

the optimal solution of problem 2 is uniquely determined and has the form

$$\hat{V}(t) = \begin{cases} A & \text{for a.a.} \quad t \in [0, t_0) \\ 0 & \text{for a.a.} \quad t \in (t_0, t_f] \end{cases} \tag{4.12}$$

where $t_0 = B/A \in (0, t_f)$.

Proof. Since $\dot{p}_2(t) \neq 0$ in $(0, t_f)$ the optimal solution does not contain singular parts. Further, it does not contain parts of the forms

$$\hat{V}(t) = 0 \text{ for a.a. } t \in [0, \tau], \tau \in (0, t_f]$$

and

$$\hat{V}(t) = \begin{cases} 0 & \text{for a.a.} \quad t \in (t_1, \tau) \\ A & \text{for a.a.} \quad t \in (\tau, t_2) \end{cases} \quad (0 \leq t_1 < \tau < t_2 \leq t_f)$$

The first one is impossible for $\tau = t_f$ because of (4.8) and for $\tau < t_f$ since we could have $p_2(t) + p_3 \leq 0$ in $(0, \tau)$ and $p_2(\tau) + p_3 \geq 0$ implying $\dot{p}_2(\tau) \geq 0$. For the second one we obtain $p_2(t) + p_3 \leq 0$ in (t_1, τ) and $p_2(t) + p_3 \geq 0$ in (τ, t_2) yielding $\dot{p}_2(\tau) \geq 0$ again. This proves the form (4.12) of the optimal solution $\hat{V}$ with unique value t_0 following from (4.8).

The proof can also be given directly by using the fact that (4.11) implies $p_2(t) > 0$ for all $t \in [0, t_f)$ and discussing the two cases $p_2(0) + p_2 \leq 0$ and $p_2(0) + p_3 > 0$.

By equations (3.11) and (4.7) the condition (4.11) is equivalent to

$$\delta p_2(t) < \phi(t) \text{ in } (0, t_f)$$

where

$$\phi(t) = -h_0(t) = -\varphi(t) L'(\hat{M}(t)\tilde{p}_1(t) > 0 \text{ in } [0, t_f) \tag{4.13}$$

with $\phi(t_f) \geq 0$ and by (3.14) equivalent to the *integral inequality*

$$\psi(t) \equiv \phi(t) - \delta \int_t^{t_f} e^{\delta(t-s)} \phi(s) ds > 0 \text{ in } (0, t_f). \tag{4.14}$$

If (4.14) holds the optimal solution is given by (4.12). In particular, this is fulfilled for all positive functions φ in the limit case $\delta = 0$ suggesting that (4.14) is not a too strong condition on φ for sufficiently small $\delta > 0$.

This can be underlined in the simple case of Gompertz growth (cf. [7, 10, 17, 23 - 25]) where

$$f(T) = \lambda \ell n \frac{\theta}{T}, T > 0, (\lambda > 0, \theta > 0)$$

and a linear loss function

$$L(M) = kM, M \geq 0, (k > 0).$$

In this case we find that

$$g(t) = -f'(e^{\hat{y}(t)}) e^{\hat{y}(t)} = \lambda, h_0(t) = k\varphi(t)\tilde{p}_1(0)e^{\lambda t}, t \in [0, t_f],$$

and (4.14) turns out to be equivalent with

$$\varphi(t)e^{(\lambda-\delta)t} - \delta \int\limits_t^{t_f} \varphi(s)e^{(\lambda-\delta)s}ds > 0 \text{ for all } t \in (0, t_f). \tag{4.15}$$

If we put

$$\varphi(t) = e^{-(\lambda-\delta)t}, t \in [0, t_f],$$

and assume that $\lambda > \delta$, then it follows that $\varphi \in C^1[0, t_f]$,

$$\varphi(0) = 1, \dot{\varphi}(t) < 0 \text{ and } \varphi(t) > 0 \text{ for all } t \in [0, t_f].$$

Further (4.15) turns out to be equivalent to

$$(1 - \delta(t_f - t) > 0 \text{ for all } t \in (0, t_f)) \Longleftrightarrow \delta t_1 < 0.$$

This shows that (4.15) can be satisfied for sufficiently small $\delta > 0$ and a suitable choice of φ.

The inequality (4.14) is fulfilled if we have

$$\dot{\phi}(t) < 0 \text{ in } (0, t_f), \tag{4.15}$$

since integration by parts of the integral in (4.14) yields

$$\psi(t) = e^{\delta t}[e^{-\delta t_f}\phi(t_f) - \int\limits_t^{t_f} e^{-\delta s}\dot{\phi}(s)ds] > 0 \text{ in } (0, t_f)$$

due to $\phi(t_f) \geq 0$ and (4.15). Differentiating (4.13) and using $\dot{\tilde{p}}_1 = g\tilde{p}_1$ we further have

$$\dot{\phi}(t) = -\tilde{p}_1(t)L'(\hat{M}(t))[\dot{\varphi}(t) + \{g(t) + m(t)\}\varphi(t)]$$

where

$$m(t) = \frac{1}{L'(\hat{M}(u)}\frac{d}{dt}[L'(\hat{M}(t))] = \frac{L''(\hat{M}(u)\dot{\hat{M}}(t)}{L'(\hat{M}(t))}. \tag{4.16}$$

Therefore, in view of $\tilde{p}_1(t) < 0$ in $[0, t_f]$ and $L'(M) > 0$, condition (4.15) is equivalent to the *differential inequality*

$$\dot{\varphi}(t) + [g(t) + m(t)]\varphi(t) < 0 \text{ in } (0, t_f) \tag{4.17}$$

where $m = m(t)$ is given by (4.16) and $g = g(\hat{T})$ by (3.17), i.e.

$$g(t) = -f'(\hat{T}(t))\hat{T}(t) < 0, t \in [0, t_f]. \tag{4.18}$$

Condition (4.17) has the same form as condition (3.22) and is like this in general an implicit condition on φ.

Summing up, by Lemma 4.2 and (4.14) - (4.18) we obtain

THEOREM 4.3

(i) *Under the conditions $B < t_f A$ and (4.14) the optimal solution of problem 2 is uniquely determined and has the form (4.12).*

ii) *Asumption (4.14) is satisfied if the condition (4.17) with (4.18) and (4.16) holds true.*

REMARKS. For a linear loss function L we have $m(t) = 0$ in $[0, t_f]$ and the condition (4.17) reduces to (3.22). As for problem 1 we say in case of (4.14) for a resistance factor φ that there is a *strong resistance* of the tumor cells against the drug.

We conclude the paper working out the simple case of Gompertz growth (cf. [7, 10, 17, 23 - 25])

$$f(T) = \lambda \ell n \frac{\theta}{T} (\lambda > 0, \theta > 0) \tag{4.19}$$

with a linear loss function $L(M) = kM$ $(k > 0)$ as an example for what can happen for general resistance.

In this case we have for $y = \ell n T$ the explicit expression

$$\begin{aligned} y(t) &= \ell n T_0 \cdot e^{-\lambda t} + \lambda \ell n \theta [1 - e^{-\lambda t}] \\ &- k \int_0^t e^{-\lambda(t-s)} \varphi(s) M(s) ds \end{aligned}$$

and the minimum condition for $y(t_f)$ leads to the maximum condition

$$\int_0^{t_f} e^{\lambda s}\varphi(s)M(s)ds \longrightarrow \max$$

which can be written in the form

$$\int_0^{t_f} p(t)V(t)dt \longrightarrow \max \tag{4.20}$$

where

$$p(t) = e^{\delta t}\int_0^{t_f} e^{(\lambda-\delta)s}\varphi(s)ds. \tag{4.21}$$

The maximum problem (4.20) where $V \in L^\infty(0,t_f)$ satisfies the restrictions (2.5) and (2.7) is a linear problem of the form of the Neyman-Pearson lemma and can be solved in explicit form. Let be $B < t_f A$. For (4.19) the condition (4.17) is equivalent to the inequality

$$\frac{d}{dt}[e^{\lambda t}\varphi(t)] < 0 \text{ in } (0,t_f).$$

If this is fulfilled the problem has the solution (4.12). We consider further the opposite case that

$$\frac{d}{dt}[e^{\lambda t}\varphi(t)] > 0 \text{ in } (0,t_f) \tag{4.22}$$

incorporating the limit case of non-resistance that $\varphi(t) \equiv 1$ on $[0,t_f]$. For $\delta > 0$ the function (4.21) has the derivative

$$\dot{p}(t) = e^{\delta t}[F(t) - C], t \in [0,t_f]$$

where

$$C = e^{(\lambda-\delta)t_f}\varphi(t_f), F(t) = \int_t^{t_f} e^{-\delta s}\frac{d}{ds}[e^{\lambda s}\varphi(s)]ds.$$

Under the assumption (4.22) the function F is strictly decreasing in $[0, t_f]$ from the value

$$F(0) = \int_0^{t_f} e^{-\delta s} \frac{d}{ds}[e^{\lambda s}\varphi(s)]ds > 0$$

to $F(t_f) = 0$. Hence we have two cases (i) $F(0) \leq C$ where $\dot{p}(t) < 0$ in $(0, t_f)$ so that $p(t)$ is strictly decreasing in $[0, t_f]$ and (ii) $F(0) > C$ where there exists a unique $t_0 \in (0, t_f)$ such that $p(t)$ is strictly increasing in $[0, t_0]$ and strictly decreasing in $[t_0, t_f]$ till $p(t_f) = 0$. In case (i) the optimal solution is given by (4.12). In case (ii) the optimal solution has the form

$$\hat{V}(t) = \begin{cases} 0 & \text{a.e. in} \quad [0, t_1) \quad \text{and} \quad (t_2, t_f] \\ A & \text{a.e. in} \quad (t_1, t_2) \end{cases} \tag{4.23}$$

where t_1, t_2 with $0 < t_1 < t_0 < t_2 < t_f$ are uniquely determined by the equations

$$A(t_2 - t_1) = B, p(t_1) = p(t_2).$$

In the particular case $\varphi(t) = 1$ on $[0, t_f]$ we have

$$p(t) = \begin{cases} (e^{(\lambda-\delta)t_f} - e^{(\lambda-\delta)t})e^{\delta t} \text{ if } \lambda > \delta \\ (t_f - t)e^{\delta t} \text{ if } \lambda = \delta \\ (e^{(\lambda-\delta)t} - e^{(\lambda-\delta)t_f})e^{\delta t} \text{ if } \lambda < \delta \end{cases}$$

and

$$C = e^{(\lambda-\delta)t_f}, F(0) = \begin{cases} \frac{\lambda}{\lambda-\delta}[e^{(\lambda-\delta)t_f} - 1] \text{ if } \lambda \neq \delta \\ \lambda t_f \text{ if } \lambda = \delta. \end{cases}$$

Therefore, case (i) occurs if $\lambda t_f \leq 1$ for $\lambda = \delta$, $\delta e^{(\lambda-\delta)t_f} \leq \lambda$ for $\lambda > \delta$, and $\lambda e^{(\delta-\lambda)t_f} \leq \delta$ for $\lambda < \delta$, and case (ii) under the opposite inequalities.

We remark hat the optimal solution (4.23) in case (ii) starts and ends with an interval of zero-therapy.

References

[1] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.

[2] A.J. Coldman, J.H. Goldic, A model for the resistance of tumor cells to cancel chemotherapeutic agents, Math. Biosci. 65 (1983) 291-307.

[3] A. J. Coldman, J.M. Murray, Optimal control for a stochastic model of cancer chemotherapy, Math. Biosci. 168 (2000) 187-200.

[4] M.I.S. Costa, J.M. Boldrini, and R.C. Bassanezi, Drug kinetics and drug resistance in optimal chemotherapy, Math. Biosci. 125 (1995) 191-209.

[5] M. Eisen, Mathematical Models in Cell Biology and Cancer Chemotherapy, Lecture Notes in Biomathematics, Vol. 30, Springer-Verlag, Berlin-Heidelberg-New York, 1979.

[6] K.R. Fister, J.C. Panetta, Optimal control applied to competing chemotherapeutic cell-kill strategies, SIAM J. Appl. Math. 63 (2003) 1954-1971.

[7] C.L. Frenzen, J.D. Murray, A cell kinetics justification for Gompertz' equation, SIAM J. Appl. Math. 46 (1986) 614-629.

[8] J.H. Goldie, Drug resistance in cancer: A perspective, Cancer and Metastasis Reviews 20 (2001) 63-68.

[9] A.D. Joffe, V.M. Tichomirov, Theorie der Extremalaufgaben, Deutscher Verlag der Wissenschaften, Berlin, 1979.

[10] W. Krabs, S.W. Pickl, Modelling, Analysis and Optimization of Biosystems, Springer-Verlag, Berlin–Heidelberg–New York, 2007.

[11] U. Ledzewicz, T. Brown and H. Schüttler, Comparison of optimal controls for a model in cancer chemotherapy with L_1- and L_2-type objectives, Optimization Methods and Software 19 (2004) 339-350.

[12] U. Ledzewicz, H. Schüttler, Drug resistance in cancer chemotherapy as an optimal control problem, Discrete and Continuous Dynamical Systems, Ser. B 6 (2006) 129-150.

[13] R.B. Martin, M.E. Fisher, R.F. Minchin, and K.L. Teo, A mathematical model of cancer chemotherapy with an optimal selection of parameters, Math. Biosci. 99(1990)205-230.

[14] R.B. Martin, M.E. Fisher, R.F. Minctin, and K.L. Teo, Low-intensity combination chemotherapy maximizes host survival time for tumors containing drug-resistant cells, Math. Biosci. 110 (1992) 221-252.

[15] A.S. Matveev, A.V. Savkin, Application of optimal control theory to analysis of cancer chemotherapy regimes, Systems and Control Letters 46 (2002) 311-321.

[16] J.M. Murray, Optimal control for a cancer chemotherapy problem with general growth and loss functions, Math. Biosci. 98 (1990) 273-287.

[17] J.M. Murray, Some optimal control problems in cancer chemotherapy with a toxicity limit, Math. Biosci. 100 (1990) 49-67.

[18] J.M. Murray, Optimal drug regimes in cancer chemotherapy for single drugs that block progression through the cell cycle, Math. Biosci. 123 (1994) 183-193.

[19] L.G. de Pillis, W. Gu, K.R. Fister, T. Head, K. Maples, A. Murugan, T. Neal, and K. Yoshida, Chemotherapy for tumors: An analysis of the dynamics and a study of quadratic and linear optimal controls, Math. Biosci. 209 (2007) 292-315.

[20] L.G. de Pillis, A. Radunskaya, A mathematical tumor model with immune resistance and drug therapy: An optimal control approach, J. Theor. Medicine 3 (2001), 79-100.

[21] L.G. de Pillis, A. Radunskaya, The dynamics of an optimally controlled tumor model: A case study, Math. Comput. Modelling 37 (2003) 1221-1244.

[22] G.W. Swan, Optimal control analysis of a cancer chemotherapy problem, IMA J. Math. Appl. Med. Biol. 4 (1987) 171-184.

[23] G.W. Swan, Applications of Optimal Control Theory in Biomedicine, Marcel Dekker Inc., New York – Basel 1984.

[24] G.W. Swan, Role of Optimal Control Theory in Cancer Chemotherapy, Math. Biosci. 101 (1990) 237-284.

[25] G.W. Swan, T.L. Vincent, Optimal control analysis in the chemotherapy of IgG multiple myeloma, Bull. Math. Biology 39 (1977) 317-337.

[26] A. Swierniak, U. Ledzewicz, and H. Schättler, Optimal control for a class of compartmental models in cancer chemotherapy, Int. J. Appl. Math. Comput. Sci. 13 (2003) 357-368.

Annals of the Academy of Romanian Scientists
Series on Mathematics and its Applications
ISSN 2066 - 6594 Volume 3, Number 2 / 2011

AROUND AN INEQUALITY, OR TWO, OF KY FAN*

Charles Horvath†

Abstract

In 1957 Ky Fan gave in [5] a necessary and sufficient condition, known as Fan's Consistency Condition, for a finite system of convex inequalities to have a solution. This result has been somewhat overshadowed by the famous Fan's Inequality which is equivalent to Brouwer's Fixed Point Theorem. Another result which bears Fan's name, but which is not due to him, is Fan's Lopsided Inequality which Aubin and Ekeland prove in [1] using Fan's Inequality.
We first prove a fairly general, but elementary result, Theorem 2.1.1, from which we derive both Fan's Theorem for finite systems of convex inequalities and Fan's Lopsided Inequality whose proof, therefore, does not require Brouwer's Fixed Theorem. We show that Theorem 2.1.1 is equivalent to Fan's Theorem for finite systems of convex inequalities; consequently, the Lopsided Inequality is a consequence of Fan's Theorem for finite systems of convex inequalities.
A number of well known and important results are proved along the way. The paths leading from Fan's 1957 theorem to those results are, we hope, simple enough to demonstrate that it deserves to be as well known as its younger and powerful cousin, Fan's Inequality.

MSC: 46A22, 46N10, 49J35, 49J40.

Keywords: convex inequalities, Fan's consistency condition. fixed points, Mazur-Orlicz.

*Accepted for publication in revised form on May 10, 2011
†Université de Perpignan, département de mathématiques, horvath@univ-perp.fr.

1 Introduction

Apart from Theorem 2.1.1, very little that is not already very well known will be found in this note. From Theorem 2.1.1 one proves Fan's Theorem on systems of inequalities for convex functions, Theorem 2.2.3. This result of Ky Fan is over half a century hold and it has been somewhat left aside after the appearance of Ky Fan's Inequality which, in its many different forms, has become a standard tool from mathematical economics to partial differential equations. But, as long as one does not deal with results that are at least as strong Brouwer's Fixed Point Theorem, Fan's result on finite systems of convex inequalities can be very versatile. We give rather simple proofs, starting from Theorem 2.2.3, or an equivalent formulation, of such results as the Kakutani Fixed Point Theorem for commutative families of continous affine maps, from which one can derive Day's Theorem on the existence of invariant means on compact topological semigroups and the Mazur-Orlicz Theorem. Proposition 2.2.6, whose proof from Fan's theorem is short and direct, leads to simple proofs of Stamppachia's and Lax-Milgram's Theorems (details are left to the reader).
In the last section, we give a proof of Fan's Lopsided Inequality, Theorem 3.0.6 using Theorem 2.1.1, and therefore Fan's Theorem.

2 A lopsided minsup inequality

2.1 The main result

Theorem 2.1.1 *Let X be a compact topological space and $f : X \times X \to \mathbb{R}$ such that:*

(1) *for all $y \in X$ $x \mapsto f(x, y)$ is lower semicontinuous;*

(2) *for all nonempty finite subset $S \subset X$ and for all $(x_1, x_2) \in X \times X$ there exists $x_3 \in X$ such that*

$$\forall y \in S \quad f(x_3, y) \le \frac{1}{2} f(x_1, y) + \frac{1}{2} f(x_2, y).$$

(3) *for all $(x, y) \in X \times X$*

$$f(x, y) + f(y, x) \le 0.$$

Then

$$\min_{x\in X}\sup_{y\in X} f(x,y) \le 0.$$

Proof. Let us begin with two remarks:

(A) The set $\mathbb{D}_n$ of dyadic elements of the standard n-dimentional simplex Δ_n is dense in Δ_n, where by $\mathbb{D}_n$ we mean the set of elements $(d_0,\cdots,d_n)\in\Delta_n$ such that each d_i is of the form $k_i/2^{m_i}$ where k_i and m_i are positive integers.

(B) Hypothesis (2) can be generalized as follows: for all nonempty finite subset $S\subset X$, for all $(x_0,\cdots,x_n)\in X^{n+1}$ and for all $(d_0,\cdots,d_n)\in\mathbb{D}_n$ there exists $x_{n+1}\in X$ such that, for all $y\in S$,

$$f(x_{n+1},y) \le \sum_{i=0}^{n} d_i f(x_i,y). \tag{2.1}$$

To prove (2.1) one can proceed by induction starting with $n=1$: we have to see that if d is dyadic number then there exists $x_3\in X$ such that, for all $y\in S$, $f(x_3,y)\le df(x_1,y)+(1-d)f(x_2,y)$.
Let $\mathcal{D}$ be the set of dyadic numbers in the interval $[0,1]$ and let $\mathcal{D}_i$, $i\in\mathbb{N}$ be those dyadic numbers which can be written as $\frac{k}{2^i}$ with k being an integer not greater than 2^i; we have $\mathcal{D}=\cup_{i\in\mathbb{N}}\mathcal{D}_i$. Since $\mathcal{D}_0=\{0,1\}$ there is nothing to prove if $d\in\mathcal{D}_0$. Also, since $\mathcal{D}_1=\left\{0,\frac{1}{2},1\right\}$ the existence of x_3 either holds trivially or by hypothesis (2). If $d\in\mathcal{D}_{k+1}$ but $d\notin\mathcal{D}_k$ we can write $d=\frac{1}{2}d_1+\frac{1}{2}d_2$ with d_1 and d_2 in $\mathcal{D}_k$. By the induction hypothesis we can find $x_{3,1}$ and $x_{3,2}$ such that, for all $y\in S$, $f(x_{3,i},y)\le d_if(x_1,y)+(1-d_i)f(x_2,y)$; by hypothesis (2) there exists x_3 such that, for all $y\in S$, $f(x_3,y)\le\frac{1}{2}f(x_{3,1},y)+\frac{1}{2}f(x_{3,2},y)$. This concludes the proof of (2.1) for $n=1$.

For $n=m+1$ we can assume that $d_{m+1}\ne 1$ and we set for $i\le m$, $d'_i=\dfrac{d_i}{\sum_{j=0}^m d_j}$ and we find $x'_{m+1}\in X$ such that; for all $y\in S$, $f(x'_{m+1},y)\le\sum_{i=0}^m d'_if(x_i,y)$. Since a sum of dyadic numbers is a dyadic number we can find $x_{m+2}\in X$ such that, for all $y\in S$,
$f(x_{m+2},y)\le\left(\sum_{j=0}^m d_j\right)f(x'_{m+1},y)+d_{m+1}f(x_{m+1},y)$.

Let us now proceed with the proof. Let $S = \{y_0, \cdots, y_n\}$ be an arbitrary nonempty finite subset of X and define on the standard n-dimensional simplex Δ_n a bilinear form as follows: $B_S(u,v) = \sum_{i,j} u_i f(y_i,y_j)v_j$. From condition (3) we have,

$$\forall u \in \Delta_n \quad B_S(u,u) \leq 0. \tag{2.2}$$

From Von Neumann's Minimax Theorem for bilinear forms, there exists a saddle point $(u_S, v_S) \in \Delta_n \times \Delta_n$ for B_S. From

$$\forall\, (u,v) \in \Delta_n \times \Delta_n \quad B_S(u_S,v) \leq B_S(u,v_S)$$

and

$$B_S(v_S,v_S) \leq 0$$

we obtain

$$\forall v \in \Delta_n \quad B_S(u_S,v) \leq 0 \tag{2.3}$$

Let $\varepsilon > 0$ be an arbitrary positive number. Since the elements of $\mathbb{D}_n$ are dense in Δ_n one can find $\bar{u}_S \in \mathbb{D}_n$ such that

$$\forall v \in \Delta_n \quad B_S(\bar{u}_S,v) \leq \varepsilon \tag{2.4}$$

which can also be written as

$$\forall j \in \{0,\cdots,n\} \quad \sum_{i=0}^{n} \bar{u}_{S,i} f(y_i,y_j) \leq \varepsilon. \tag{2.5}$$

From (B) with $x_i = y_i$ there exists $x_{S,\varepsilon} \in X$ such that,

$$\forall y \in S \quad f(x_{S,\varepsilon},y) \leq \varepsilon. \tag{2.6}$$

To complete the proof let, for all $\varepsilon > 0$ and all $y \in S$,

$$[f(-,y) \leq \varepsilon] = \{x \in X : f(x,y) \leq \varepsilon\}.$$

We have shown that the family of sets $\big\{[f(-,y) \leq \varepsilon] : y \in X\big\}$ has the finite intersection property; by hypothesis (1) all the sets in question are closed. By compactness of X the set $\cap_{y\in X}[f(-,y) \leq \varepsilon]$ is not empty, and also compact. For $0 \leq \varepsilon' \leq \varepsilon$ we obviously have $\cap_{y\in X}[f(-,y) \leq \varepsilon'] \subset \cap_{y\in X}[f(-,y) \leq \varepsilon]$ and consequently $\cap_{\varepsilon>0} \cap_{y\in X} [f(-,y) \leq \varepsilon] \neq \emptyset$. This concludes the proof. $\square$

An equivalent, but more general, formulation of Theorem 2.1.1 can be given without hypothesis (3).
Let $\lambda = \sup_{(x,y)\in X\times X} \dfrac{f(x,y)+f(y,x)}{2}$ and, to obtain a non trivial result, assume that $\lambda \neq +\infty$. Now let $g(x,y) = f(x,y) - \lambda$ and notice that (1) and (2) hold for g if they hold for f and that (3) holds for g.

Theorem 2.1.2 *Let X be a compact topological space and $f : X \times X \to \mathbb{R}$ such that:*

(1) *for all $y \in X$ $x \mapsto f(x,y)$ is lower semicontinuous;*

(2) *for all nonempty finite subset $S \subset X$ and for all $(x_1,x_2) \in X \times X$ there exists $x_3 \in X$ such that*

$$\forall y \in S \quad f(x_3,y) \leq \frac{1}{2}f(x_1,y) + \frac{1}{2}f(x_2,y).$$

Then

$$\exists x_0 \in X \ \textit{such that}\ \forall y \in X \quad f(x_0,y) \ \leq \sup_{(x,y)\in X\times X} \frac{f(x,y)+f(y,x)}{2}.$$

Let us say that a function $f : X \times Y \to \mathbb{R}$ defined on the product of two arbitrary sets X and Y is **finitely midconvex in its first variable** if condition (2) of Theorem 2.1.1 holds; one can similarly define what it means to be **finitely midconcave in its second variable**.

A given function $f : X \times Y \to \mathbb{R}$ is finitely midconvex in its first variable exactly if the family $\{S(x_1, x_2 : y) : y \in Y\}$ has the finite intersection property, where
$$S(x_1,x_2;y) = \{x \in X : f(x,y) \leq \tfrac{1}{2}f(x_1,y) + \tfrac{1}{2}f(x_2,y)\}.$$
Furthermore, if X is a compact topological space and if $f : X \times Y \to \mathbb{R}$ is lower semicontinuous in its first variable then, for all $y \in Y$ and for all $x_1, x_2 \in X$, the set $S(x_1, x_2 : y)$ is compact.
In conclusion, if X is a compact topological space and if $f : X \times Y \to \mathbb{R}$ is finitely midconvex and lower semicontinuous in its first variable then $\cap_{y\in Y} S(x_1,x_2;y) \neq \emptyset$, that is, there exists $x_3 \in X$ such that, for all $y \in Y$, $f(x_3,y) \leq \frac{1}{2}f(x_1,y) + \frac{1}{2}f(x_2,y)$.

Assume now that f is both lower semicontinuous and finitely midconvex in its first variable.

Take an arbitrary real number $t \in [0,1]$ and a sequence $(d_n)_{n\in\mathbb{N}}$ of dyadic numbers in $[0,1]$ which converges to t; for all n there exists $x_{3,n} \in X$ such that, for all $y \in Y$, $f(x_{3,n}, y) \le (1-d_n)f(x_1,y)+d_n f(x_2,y)$ and therefore, by compactness of X and lower semicontinuity of $f(-,y)$, there exists $x_{3,t} \in X$ such that, for all $y \in Y$, $f(x_{3,t},y) \le (1-t)f(x_1,y) + tf(x_2,y)$. In other words, f is **convexlike in its first variable** that is; for all $x_1, x_2 \in X$ and for all $t \in [0,1]$, there exists $x_3 \in X$ such that, for all $y \in Y$, $f(x_3,y) \le (1-t)f(x_1,y) + tf(x_2,y)$. In conclusion, *assuming compactness of X and lower semicontinuity in the first variable, being finitely midconvex in the first variable or being convexlike in the first variable are equivalent conditions and these are in turn equivalent to*

$$\forall n \in \mathbb{N} \quad \forall (x_0, \cdots, x_n) \in X^{n+1} \quad \forall u \in \Delta_n \quad \exists \hat{x} \in X$$
$$\text{such that}$$
$$\forall y \in X \quad f(\hat{x}, y) \le \sum_{i=0}^{n} u_i f(x_i, y). \tag{2.7}$$

One could similarly define what it means for f to be concave like in its second variable and reach a similar conclusion with respect to functions which are finitely midconcave in the second variable.

2.2 Some results that can be derived from the Main Theorem

Proposition 2.2.1 *Let X and Y be two compact topological spaces and $f, g : X \times Y \to \mathbb{R}$ two functions such that:*

(1) *f is lower semicontinuous and finitely midconvex in its first variable;*

(2) *g is upper semicontinuous and finitely midconcave in its second variable;*

(3) $\forall (x,y) \in X \times Y$ $f(x,y) \le g(x,y)$.

Then

$$\exists (x_0, y_0) \in X \times Y \quad \textit{such that} \quad \forall (x,y) \in X \times Y \quad f(x_0, y) \le g(x, y_0).$$

Proof. Apply Theorem 2.1.1 to the compact topological space $Z = X \times Y$ and the function $F\big((x_1,y_1),(x_2,y_2)\big) = f(x_1,y_2) - g(x_2,y_1)$. □

Taking $f = g$ in Proposition 2.2.1 one obtains Proposition 2.2.2 below. On the one hand, Proposition 2.2.2 clearly implies Von Neumann's Minimax

Theorem, on the other hand, Theorem 2.1.1 was derived from Von Neumann's Minimax Theorem . Proposition 2.2.2 and Von Neumann's Minimax Theorem are therefore equivalent.

Proposition 2.2.2 *Let X and Y be two compact topological spaces and let $f : X \times Y \to \mathbb{R}$ be a function which is lower semicontinuous and finitely midconvex in its first variable and upper semicontinuous and finitely midconcave in its second variable. Then,*

$$\exists (x_0, y_0) \in X \times Y \quad \textit{such that} \quad \forall (x, y) \in X \times Y \quad f(x_0, y) \leq f(x_0, y_0) \leq f(x, y_0).$$

Propositions 2.2.1 and 2.2.2, in a somewhat more general version involving 4 functions, are due to Granas and Liu [8].

In [5] Fan proved the following existence theorem for finite systems of inequalities:

Theorem 2.2.3 (Fan's Theorem) *Let $f_i : X \to \mathbb{R}$, $i \in \{0, \cdots, n\}$, be a finite family of lower semicontinuous functions defined on a compact convex subset of a linear topological vector space.*
Assume that the following condition holds:

$$\forall u \in \Delta_n \quad \exists x \in X \quad \textit{such that} \quad \sum_{i=0}^{n} u_i f_i(x) \leq 0 \tag{2.8}$$

then

$$\exists x_0 \in X \quad \textit{such that} \quad \forall i \in \{0, \cdots, n\} \quad f_i(x_0) \leq 0.$$

The proof of Fan's Theorem can be found on page 41 of [7]. Fan's Theorem follows from Theorem 2.1.1. We prove a somewhat more general result, which is implicitely contained in Fan's paper. First, let us say that a family of functions $f_i : X \to \mathbb{R}$, $i \in I$ is a **finitely midconvex family** if the function $F : X \times I \to \mathbb{R}$ defined by $F(x, i) = f_i(x)$ is finitely midconvex in its first variable. In case I is a finite set the adjective "finitely" is dropped. Let us say that **Fan's consistency condition** holds for the family $\mathcal{F} = \{f_i : i \in I\}$ if for all finite subsets $\{f_0, \cdots, f_n\}$ of $\mathcal{F}$ condition (2.8) of Fan's Theorem holds.

Theorem 2.2.4 *Let $\mathcal{F}$ be a finitely midconvex family of lower semicontinuous functions defined on a compact topological space. If Fan's consistency condition holds then*

$$\exists x_0 \in X \quad \textit{such that} \quad \sup_{f \in \mathcal{F}} f(x_0) \leq 0.$$

Proof. For all $f \in \mathcal{F}$ let $[f \leq 0] = \{x \in X : f(x) \leq 0\}$. Since X is compact and the elements of $\mathcal{F}$ are lower semicontinuous we have to show that the family $\big\{[f \leq 0] : f \in \mathcal{F}\big\}$ has the finite intersection property. Given a finite subfamily $\{f_0, \cdots, f_n\}$ of $\mathcal{F}$ the function $\varphi : X \times \Delta_n \to \mathbb{R}$ defined by $\varphi(x, u) = \sum_{i=0}^n u_i f_i(x)$ is finitely midconvex and lower semicontinuous in its first variable and finitely midconcave and upper semicontinuous in its second variable.
By Proposition 2.2.2 there exists $(x_0, u_0) \in X \times \Delta_n$ such that, for all (x, u) in $X \times \Delta_n$, $\varphi(x_0, u) \leq \varphi(x_0, u_0) \leq \varphi(x, u_0)$.
From Fan's consistency condition, there exists $x^\star \in X$ such that $\varphi(x^\star, u_0) \leq 0$ and therefore, $\varphi(x_0, u_0) \leq 0$; we have shown that $\sup_{u \in \Delta_n} \varphi(x_0, u) \leq 0$, that is $x_0 \in \cap_{i=0}^n [f_i \leq 0]$. □

To close this circle of ideas let us see that Theorem 2.1.1 can be deduced from Theorem 2.2.4.

Given $f : X \times X \to \mathbb{R}$ as in Theorem 2.1.1 take X itself as the set of indices and let $f_y(x) = f(x, y)$. If Fan's Consistency Condition holds for the family $\{f_y : y \in X\}$ we are done.
If Fan's Consistency Condition does not hold then there exists a finite subset $\{y_0, \cdots, y_n\}$ of X and there exists $u \in \Delta_n$ such that,

$$\forall x \in X \quad \sum_{i=0}^n u_i f(x, y_i) > 0. \tag{2.9}$$

From (2.9) and $f(x, y_i) + f(y_i, x) \leq 0$ we have

$$\forall x \in X \quad \sum_{i=0}^n u_i f(y_i, x) < 0 \tag{2.10}$$

and from (2.7), there exists $\hat{y} \in X$ such that

$$\forall x \in X \quad f(\hat{y}, x) \leq \sum_{i=0}^n u_i f(y_i, x) \tag{2.11}$$

and therefore, from (2.10),

$$\forall x \in X \quad f(\hat{y}, x) < 0. \tag{2.12}$$

But (2.12) clearly implies that Fan's Consistency Condition holds (and it also implies the conclusion of Theorem 2.1.1). We have reached a contradiction and therefore Fan's Consistency Condition holds. □

Fan's Theorem is a non linear version of Fourier's Theorem on systems of linear inequalities, a classical result of linear programming, from which one can derive Von Neumann's Minimax Theorem for bilinear forms, or the well known Farkas Lemma; all these results are equivalent, in the sense that they can all be derived from any given one of them. Fourier's Theorem can be proved in a completely elementary way, as in [11]. Here is a short proof from Fan's Theorem.

Theorem 2.2.5 (Fourier) *Let A be an $m \times n$ matrix and $B \in \mathbb{R}^m$ then, either the system of linear inequalities $AX \geq B$ has a solution or there exists $Y \in \mathbb{R}^m_+$ such that $A^tY = 0$ and $Y^tB > 0$.*

To see that Theorem 2.2.5 follows from Fan's Theorem, let $f_i(X) = b_i - \sum_{j=1}^n a_{ij}x_j$, $i \in \{1, \cdots, n\}$. If there is no solution in $\mathbb{R}^n$ to the system of inequalities $f_i(X) \leq 0$ then, for all integer $k > 0$, there exists $Y_k \in \Delta_{n-1}$ such that, for all $X \in \mathbb{R}^n$ of norm not exceeding k, $\sum_{i=1}^n y_{n,k} f_i(X) > 0$, that is, $Y_k^tB - Y_k^tAX > 0$. We can assume that the sequence $(Y_k)_{k\in\mathbb{N}}$ converges to some $Y^\star \in \Delta_{n-1}$. For any given $X \in \mathbb{R}^n$, we will have, for $k > \|X\|$, $Y_k^tB - Y_k^tAX > 0$; and consequently $Y^{\star t}B - Y^{\star t}AX \geq 0$. We must have $Y^{\star t}A = 0$, otherwise we can choose X such that $Y^{\star t}AX > 0$, and therefore $Y^{\star t}A(rX) > 0$ for all $r > 0$. This proves Fourier's Theorem.

Von Neumann's Minimax Theorem for bilinear forms is easily derived from Fourier's Theorem. There are many elementary proofs of the Von Neumann's Minimax Theorem for bilinear forms.

Proposition 2.2.6 (Weak Fan Inequality) *Let $f : C \times C \to \mathbb{R}$ be a function defined on a compact convex subset of some linear space. Assume that the following conditions hold:*

(1) $\forall y \in C \quad x \mapsto f(x, y)$ *is lower semicontinuous and convex on C;*

(2) $\forall x \in C \quad y \mapsto f(x,y)$ *is concave on* C*;*

(3) $\forall x \in C \quad f(x,x) \leq 0$.

Then, there exists $x_0 \in C$ *such that, for all* $y \in C$*,* $f(x_0,y) \leq 0$*.*

Proof. Let us see that Fan's Consistency Condition holds for the family $\mathcal{F} = \{f(-,y) : y \in C\}$. Otherwise, there exists $y_0, \cdots, y_n \in C$ and $(u_0, \cdots, u_n) \in \Delta_n$ such that, for all $x \in C$,

$$\sum_{i=0}^{n} u_i f(x, y_i) > 0. \tag{2.13}$$

Let $\hat{x} = \sum_{i=0}^{n} u_i y_i$ and, in (2.13), take $x = \hat{x}$. From the second hypothesis we then have $f(\hat{x}, \hat{x}) > 0$ which is contradiction with hypothesis (3). □

Along with Theorem 4.2 of [7] page 65 (which can be seen as an elementary proof of the weak compactness of closed convex subsets of a Hilbert space.), Proposition 2.2.6 can be used to easily prove such results as the Stampacchia or the Lax-Milgram theorems. In [7] these results are derived from a weak form of the KKM Lemma.

2.3 Fan's Theorem and Fixed Points

Lemma 2.3.1 (Markov's Theorem) *A linear map* $X \mapsto PX$ *from* Δ_n *to itself has a fixed point.*

Proof. For $(X,Y) \in \Delta_n \times \Delta_n$ let $f_Y(X) = X^t\big(P^t - I\big)Y$; if Fan's Consistency Condition holds for the family $\mathcal{F} = \{f_Y : Y \in \Delta_n\}$ then there exists $X_0 \in \Delta_n$ such that, for all $Y \in \Delta_n$, $X_0^t\big(P^t - I\big)Y \leq 0$ which is equivalent to $X_0^t\big(P^t - I\big) \leq 0$ or, $PX_0 - X_0 \leq 0$. Since both PX_0 and X_0 belong to Δ_n equality must hold.

For a contradiction, assume that Fan's Consistency Condition does not hold. Then, there exists $Y_0, \cdots, Y_k \in \Delta_n$ and $u \in \Delta_k$ such that, for all $X \in \Delta_n$, $\sum_{i=0}^{k} u_k X^t\big(P^t - I\big)Y_k > 0$. With $\hat{Y} = \sum_{i=0}^{k} u_k Y_k$ we obtain

$$\forall X \in \Delta_n \quad X^t\big(P^t - I\big)\hat{Y} > 0 \tag{2.14}$$

which is equivalent to

$$\forall i \in \{1, \cdots, n\} \quad \sum_{j=1}^{n} p_{j,i}\hat{y}_j > \hat{y}_i. \tag{2.15}$$

Let $\left\|\hat{Y}\right\| = \max\{\hat{y_1}, \cdots, \hat{y_n}\}$ and choose i_0 such that $\hat{y_{i_0}} = \left\|\hat{Y}\right\|$. Since the entries of P are non negative and since each column sums up to 1 we obtain from (2.15) $\left\|\hat{Y}\right\| > \left\|\hat{Y}\right\|$. □

A similar elementery proof of Markov's theorem based on Farkas Theorem can be found in [6]. Farkas Theorem is easily derived from Fourier's Theorem.

Theorem 2.3.2 (Kakutani) *Let C be a convex compact subspace of a locally convex topological vector space and let $\mathcal{F}$ be a commutative family of continuous affine maps from C into itself. Then the members of $\mathcal{F}$ have a common fixed point.*

Proof. (a) Let T be an arbitrary, but fixed, element of $\mathcal{F}$. We show that T has a fixed point. Let W be an arbitrary convex neighborhood of the origin. By compactness there is a finite subset $\{x_0, \ldots, x_n\}$ of C such that $C \subset \bigcup_{i=0}^{n} (x_i + W)$. For each index $i \in \{0, \cdots, n\}$ choose an index $\varphi(i) \in \{0, \cdots, n\}$ such that

$$T(x_i) \in x_{\varphi(i)} + W. \tag{2.16}$$

Let T_W be the unique affine map from Δ_n to itself such that $T_W(e_i) = e_{\varphi(i)}$ where $e_0, \ldots, e_n$ are the vertices of Δ_n and let $p_w = \sum_{i=0}^{n} \mu_i e_i$ be a fixed point of T_W. From $p_w = T_W(p_w)$ we have $\sum_{i=0}^{n} \mu_i e_i = \sum_{i=0}^{n} \mu_i e_{\varphi(i)}$. Let $U_W : \Delta_n \to C$ be the unic affine function such that $U(e_i) = x_i$; from $U_W(p_w) = U_W\big(T_W(p_w)\big)$ it follows that

$$\sum_{i=0}^{n} \mu_i x_i = \sum_{i=0}^{n} \mu_i x_{\varphi(i)}. \tag{2.17}$$

Since W is convex we have from (2.16)

$$\sum_{i=0}^{n} \mu_i \left(T(x_i) - x_{\varphi(i)}\right) \in W \tag{2.18}$$

and since T is affine

$$\sum_{i=0}^{n} \mu_i \left(T(x_i) - x_{\varphi(i)}\right) = T\Big(\sum_{i=0}^{n} \mu_i x_i\Big) - \sum_{i=0}^{n} \mu_i x_{\varphi(i)} \tag{2.19}$$

$$= T\Big(\sum_{i=0}^{n} \mu_i x_i\Big) - \sum_{i=0}^{n} \mu_i x_i \tag{2.20}$$

We have shown that for any neighborhood W of the origin there is a point $x \in C$ such that $T(x) - x \in W$. By the compactness of C and the continuity of T we can infer that T has a fixed point.

For each element T of $\mathcal{F}$ let $\mathrm{Fix}(T)$ be the set of fixed points of T. Each of these sets is closed in C and therefore compact, they are also convex since each map T is affine and we have shown that they are not empty.

The proof can now be completed as in [7]. The commutativity of $\mathcal{F}$ implies that for all finite subsets $\{T_1, \dots, T_n\}$ and all T_0 of $\mathcal{F}$ the inclusion $T_0\left(\bigcap_{i=1}^n \mathrm{Fix}\,(T_i)\right) \subset \left(\bigcap_{i=1}^n \mathrm{Fix}\,(T_i)\right)$ holds. Now a straightforward induction shows that the family $\{\mathrm{Fix}\,(T) : T \in \mathcal{F}\}$ has the finite intersection property and therefore by compactness the set $\bigcap \{\mathrm{Fix}\,(T) : T \in \mathcal{F}\}$ is not empty. $\square$

A simple proof of Theorem 2.3.2 making explicit use of Fan's Theorem can be found on page 43 of [7].

2.4 Invariant means and the Mazur-Orlicz Theorem

Day's theorem on the existence of invariant means on compact topological semigroups is usually proved via the Hahn-Banach theorem, [2], [10]. It is obtained here as a direct consequence of Kakutani's Theorem, and therefore, indirectly, as a consequence of Fan's Theorem. From Day's Theorem we derive, following [3] with a slight adaptation, the Mazur-Orlicz Theorem. There is a very short step from the Mazur-Orlicz theorem to the Hahn-Banach theorem.

We give the theorem of Mazur-Orlicz a somewhat geometrical formulation which is readily seen to be equivalent to the standard formulation.

Let G be an abelian semigroup and let $B(G)$ be the space of all bounded real valued functions on G. An invariant mean on G is a real valued linear function m on $B(G)$ such that

$$m(1) = 1,$$

$$m(f) \geq 0 \quad \text{if} \quad f \geq 0$$

and

$$m(f_g) = m(f)$$

for all $f \in B(G)$ and all $g \in G$, where $f_g(x) = f(gx)$.

Theorem 2.4.1 (Day) *If G is an abelian semigroup then there is an invariant mean on G.*

Proof. With the norm $\|f\| = \sup_{x \in G} \mid f(x) \mid$ the space of bounded functions on G is a Banach space. Let E be the Banach space of bounded linear functionals on $B(G)$ and let C be the subset of the unit ball of E consisting of positive functionals taking the value 1 on the constant function 1 of $B(G)$. If $g \in G$ and $f \in B(G)$ then $f \mapsto f(g)$ defines an element of C. Consequently C is not empty and it is obviously a closed and convex subset of the unit ball of E. For the weak topology C is therefore compact.

Now for $g \in G$, $L \in C$ and $f \in B(G)$ let $T_g(L)(f) = L(f_g)$. Then $\{T_g : g \in G\}$ is a commutative family of continuous affine maps on C. By Kakutani's theorem there is an element m of C such that for each $g \in G$ one has $T_g(m) = m$. This m is an invariant mean on G. $\square$

Theorem 2.4.2 *Let $p : G \to \mathbb{R}$ be a subadditive map defined on an abelian semigroup G (respectively, an abelian group G) and let $C \subseteq G \times \mathbb{R}$ be an additive subset (i.e. if $(x, r), (x', r') \in C$ then $(x+x', r+r') \in C$). Then, there exists an additive function (respectively, a group homomorphism) $f : G \to \mathbb{R}$ such that*

$$(i)\ \forall x \in G\ \ f(x) \leq p(x)$$

and

$$(ii)\ \forall (x, r) \in C\quad r \leq f(x)$$

if and only if

$$\forall (x, r) \in C \quad r \leq p(x).$$

The necessity of the condition is obvious. Let us show that this condition is sufficient.

For all $x \in G$ let $P(x) = \inf \{p(x + y) - r : (y, r) \in C\}$. From the subadditivity of p it follows that,

$$\forall x, x' \in G \quad P(x + x') - P(x') \leq p(x). \tag{2.21}$$

Since C is an additive subset of $G \times \mathbb{R}$ we have, from the definition of P,

$$\forall x \in G \quad \forall (y, r), (y', r') \in C \quad P(x) + r \leq p(x + y + y') - r'. \tag{2.22}$$

Taking the infimum over $(y', r') \in C$ gives

$$\forall x \in G \quad \forall (y, r) \in C \quad r \leq P(x + y) - P(x). \tag{2.23}$$

Let m be an invariant mean on G. For $x \in G$ let $f(x) = m\left(P_x - P\right)$ where, for all $y \in E$, $P_x(y) = P(x+y)$.

From (2.21) and (2.23) we have

$$\forall x \in G \quad f(x) \leq p(x) \quad \text{and} \quad \forall (y,r) \in C \quad r \leq f(x) \tag{2.24}$$

For all $x, x' \in E$ one has

$$\begin{aligned} f(x) &= m\left(P_x - P\right) \\ &= m\left(\left(P_x - P\right)_{x'}\right) \\ &= m(P_{(x+x')} - P) - m\left(P_{x'} - P\right) = f(x+x') - f(x'). \end{aligned}$$

We have shown that f is additive and consequently, a group homorphism if G is a group. □

In Theorem 2.4.2 one does not have to assume that C is an additive subset of $G \times \mathbb{R}$ since, for an arbitrary $S \subset G \times \mathbb{R}$, Theorem 2.4.2 holds with S instead of C if and only if it holds with C being the additive subset of $G \times \mathbb{R}$ spanned by S.

Theorem 2.4.3 (Mazur-Orlicz) *Let $p : E \to \mathbb{R}$ be a subadditive and positively homogeneous map defined on real vector space E and let $C \subseteq E \times \mathbb{R}$ a convex cone. Then there is a linear function $f : E \to \mathbb{R}$ such that*

$$(i)\ \forall x \in E\ \ f(x) \leq p(x)$$

and

$$(ii)\ \forall (x,r) \in C \quad r \leq f(x)$$

if and only if

$$\forall (x,r) \in C \quad r \leq p(x).$$

Proof A convex cone in $E \times \mathbb{R}$ is an additive subset. The function $f : E \to \mathbb{R}$ defined in the proof Theorem 2.4.2 is a group homomorphism. Lemma 2.4.4 below shows that f is linear. □

Lemma 2.4.4 *Let $p : E \to \mathbb{R}$ be a subadditive and positively homogeneous map defined on real vector space E. If $f : E \to \mathbb{R}$ is an additive map such that*

$$\forall x \in E \quad f(x) \leq p(x)$$

then f is linear.

Proof. Since f is a group homomorphism, we have, for all $(x, r) \in E \times \mathbb{Q}$, $f(rx) = rf(x)$.

Take $(x, t) \in E \times \mathbb{R}$ and assume that $f(x) \geq 0$; which implies $p(x) \geq 0$. Then

$$\begin{aligned} tf(x) - f(tx) &= \inf \{rf(x) : r > t, r \in \mathbb{Q}\} - f(tx) \\ &= \inf \{rf(x) - f(tx) : r > t, r \in \mathbb{Q}\} \\ &= \inf \{f((r-t)x) : r > t, r \in \mathbb{Q}\} \\ &\leq \inf \{p((r-t)x) : r > t, r \in \mathbb{Q}\} \\ &= \inf \{(r-t)p(x) : r > t, r \in \mathbb{Q}\} = 0. \end{aligned}$$

We have shown that

$$\forall (x, t) \in E \times \mathbb{R} \text{ such that } f(x) \geq 0 \text{ one has } tf(x) \leq f(tx) \tag{2.25}$$

which also shows that if $t \geq 0$ and $f(x) \geq 0$ then $f(tx) \geq 0$. Therefore, in (2.25), for $t > 0$, we can replace x by tx and t by $1/t$ to obtain, $f(tx) \leq tf(x)$. Since $f(0) = 0$ we have

$$\forall (x, t) \in E \times \mathbb{R}_+ \text{ such that } f(x) \geq 0 \text{ one has } tf(x) = f(tx). \tag{2.26}$$

Finally, $f(-x) = -f(x)$ implies $f(tx) = tf(x)$ for all $(x, t) \in E \times \mathbb{R}$. □

Another proof of the Mazur-Orlicz Theorem using Kakutani's Fixed Point Theorem for commuting families of affine maps can be found in both [7] on pages 70 to 73, and also in [12]. Those proofs make explicit use of Tychonov's Theorem on the compactness of an arbitrary product of compact spaces.

Proposition 2.2.2 appears in [12] as Theorem 2.1 under the additional hypothesis that X is a compact convex subset of some topological vector space

3 Ky Fan's lopsided inequality

Lemma 3.0.5 *Let $\varphi : C \times C \to \mathbb{R}$ be a function defined on a convex subset C of some topological vector space and assume that the following conditions hold:*

(1) $\forall x \in C$ *the partial map* $\varphi(x, -)$ *is concave;*

(2) $\forall (x, y) \in C \times C$ *the map* $t \mapsto \varphi\big((1-t)y + tx, y\big)$ *is lower semicontinuous on* $[0, 1]$*;*

(3) $\forall x \in C \quad \varphi(x, x) \leq 0.$

Then, for all $x_0 \in C$ *such that* $\inf_{y \in C} \varphi(y, x_0) \geq 0$ *we also have* $\sup_{y \in C} \varphi(x_0, y) \leq 0.$

Proof. Assume that $\inf_{y \in C} \varphi(y, x_0) \geq 0$. Take an arbitrary element $y \in C$ and let $\eta(t) = (1-t)x_0 + ty$ for $t \in [0, 1]$. From $0 \leq \varphi(\eta(t), x_0)$ and (3) we obtain, for all $0 \leq t < 1$,

$$0 \leq \varphi(\eta(t), x_0) - \frac{1}{1-t}\varphi(\eta(t), \eta(t)). \tag{3.1}$$

Since $\varphi(\eta(t), -)$ is concave we obtain from (3.1)

$$\forall t \in [0, 1[\quad 0 \leq -\frac{t}{1-t}\varphi(\eta(t), y) \tag{3.2}$$

or, equivalently,

$$\forall t \in [0, 1[\quad \varphi(\eta(t), y) \leq 0 \tag{3.3}$$

and finally, since $t \mapsto \varphi(\eta(t), y)$ is lower semicontinuous on $[0, 1]$, $\varphi(x_0, y) \leq 0$. $\square$

Theorem 3.0.6 (Fan's Lopsided Inequality) *Let C be a convex subset of a topological vector space and let $\varphi : C \times C \to \mathbb{R}$ be a function such that:*

(1) $\forall x \in C \quad \varphi(x, -)$ *is upper semicontinuous and concave on C;*

(2) $\forall (x, y) \in C^2 \quad t \mapsto \varphi\big((1-t)y + tx, y\big)$ *is lower semicontinuous on* $[0, 1]$,

(3) $\forall (x, y) \in C^2 \quad 0 \leq \varphi(x, y) + \varphi(y, x)$*;*

(4) $\forall x \in C \quad \varphi(x, x) \leq 0$*;*

(5) $\exists y_0 \in C$ *such that* $\{x \in C : \varphi(x, y_0) \leq 0\}$ *is contained in a compact and convex subset* K_0 *of* C.

Then

$$\exists x_0 \in C \text{ such that } \sup_{y \in C} \varphi(x_0, y) \leq 0.$$

Proof. Let $\mathcal{F}$ be the family of compact and convex subsets of C containing K_0. To each $K \in \mathcal{F}$ associate the function $f_K : K \times K \to \mathbb{R}$ defined by $f_K(x,y) = -\varphi(y,x)$. By Theorem 2.1.1 there exists $x_K \in K$ such that $\inf_{y \in K} \varphi(y, x_K) \geqslant 0$. By Lemma 3.0.5 we also have $\sup_{y \in K} \varphi(x_K, y) \leqslant 0$.

Let $A_K = \{x \in K : \sup_{y \in K} \varphi(x,y) \leqslant 0\}$ and $F_K = \overline{A_K}$; from (4) we have $y_0 \in K_0$ and from (5) $A_K \subset K_0$. F_K is therefore a nonempty compact subset of C. Notice also that if $K \subset K'$ then $A_{K'} \subset A_K$ since, if $x \in K'$ and $\sup_{y \in K'} \varphi(x,y) \leqslant 0$ then $x \in K_0 \subset K$. Consequently, if $K \subset K'$ then $F_{K'} \subset F_K$.

Let us see that the family $\{F_K : K \in \mathcal{F}\}$ has the finite intersection property. Let Δ_m be the standard m-dimensional simplex and, given $K_0, \cdots, K_m$ in $\mathcal{F}$, let

$$K = \left\{ \sum_{i=0}^{m} t_i x_i : (t_0, \cdots, t_m) \in \Delta_m \text{ and } (x_0, \cdots, x_m) \in \prod_{i=0}^{m} K_i \right\}.$$

Since K is compact and convex and, for all $i \in \{0, \cdots, m\}$, $K_i \subset K$ we have $K \in \mathcal{F}$ and $F_K \subset \cap_{i=0}^{m} F_{K_i}$.

We have shown that $\bigcap_{K \in \mathcal{F}} F_K \neq \emptyset$. Let $x^\star$ be an arbitrary point of $\bigcap_{K \in \mathcal{F}} F_K$.

For all $y \in C$ let $K(y) = \{(1-t)y + tx : x \in K_0\}$ and fix an arbitrary $\bar{y}$ in C.

From $K(\bar{y}) \in \mathcal{F}$ and $x^\star \in F_{K(\bar{y})}$ we have, for all neighborhood U of $x^\star$ in C,

$$U \cap \{x \in K(\bar{y}) : \sup_{y \in K(\bar{y})} \varphi(x,y) \leqslant 0\} \neq \emptyset.$$

Hypothesis (3) implies that $\inf_{y \in K(\bar{y})} \varphi(y, x_U) \geqslant 0$ for all $x_U \in U$ such that $\sup_{y \in K(\bar{y})} \varphi(x_U, y) \leqslant 0$. Since U is an arbitrary neigborhood of $x^\star$ we have shown that $x^\star$ belongs to the closure of $\{x \in K(\bar{y}) : \inf_{y \in K(\bar{y})} \varphi(y,x) \geqslant 0\}$.

By hypothesis, for all $y \in K(\bar{y})$, $\varphi(y, -)$ is upper semicontinuous on $K(\bar{y})$ and therefore $\{x \in K(\bar{y}) : \inf_{y \in K(\bar{y})} \varphi(y, x) \geqslant 0\}$ is closed in C.

We have shown that $\inf_{y \in K(\bar{y})} \varphi(y, x^\star) \geqslant 0$; another application of Lemma 3.0.5 yields $\sup_{y \in K(\bar{y})} \varphi(x^\star, y) \leqslant 0$ and in particular $\varphi(x^\star, \bar{y}) \leqslant 0$. Since $\bar{y}$ was an arbitrary element of C this concludes the proof. □

Ky Fan's Inequality, fixed point theorems and variational inequalities are all closely related. Let us give, without proofs, two classical results that can be derived without much difficulty from Theorem 3.0.6; the first is the Minty-Browder Theorem on the surjectivity of monotone operators, the second is the Hilbert space version of the fixed point theorem for nonexpansive maps of Browder-Goehde-Kirk.

Theorem 3.0.7 *Let E be a reflexive Banach space, $g : E \to \mathbb{R}$ a lower semicontinuous function and $A : E \to E^\star$ such that:*

(1) *A is weakly continuous on the finite dimensional subspaces of E;*

(2) $\forall (x, y) \in E \times E$ $\langle Ax - Ay, x - y\rangle \geq 0$*;*

(3) $\exists y_0 \in E$ *such that* $\lim\limits_{\|x\| \to \infty} \dfrac{\langle Ax, x - y_0\rangle + g(x)}{\|x\|} = \infty$

Then, for all $y^\star \in E^\star$ there exists $x_0 \in E$ such that

$$\forall y \in E \quad \langle A(x_0) - y^\star, x_0 - y\rangle + g(x_0) \leq g(y).$$

Proof. Apply Theorem 3.0.6 to $\varphi(x, y) = \langle A(x) - y^\star, x - y\rangle + g(x) - g(y)$. □

Theorem 3.0.8 *Let $f : C \to C$ be a function defined on a closed bounded convex subset C of a Hilbert space H. If f is nonexpansive, that is,*
$\forall (x, y) \in C \times C \quad \|f(x) - f(y)\| \leq \|x - y\|$*, then f has a fixed point.*

Proof. Apply Theorem 3.0.6 to $\varphi(x, y) = \langle x - f(x), x - y\rangle$. □

Acknowledgement. A la mémoire de Ky Fan avec respect et gratitude.

References

[1] J.P. Aubin and I. Ekeland, *Applied Nonlinear Analysis*, John Wiley & Sons, 1984.

[2] S. K. Berberian, *Lectures in Functional Analysis and Operator Theory*, G.T.M. Springer Verlag, 1974.

[3] W. Chojnacki, *Sur un Théorème de Day, un Théorème de Mazur-Orlicz et une Généralisation de Quelques Théorèmes de Silverman* , Colloquium Mathematicum, vol. L, Fasc. 2 (1986), 257–262.

[4] W. Chojnacki, *Erratum à l'article "Sur un Théorème de Day, un Théorème de Mazur-Orlicz et une Généralisation de Quelques Théorèmes de Silverman"* , Colloquium Mathematicum, vol. LXII, Fasc. 1 (1992).

[5] K. Fan *Existence theorems and extreme solutions for inequalities concerning convex functions or linear transformations*, Math. Z., Vol. 68 (1957), 205-217.

[6] J. Franklin, *Mathematical Methods of Economics*, Amer. Math. Monthly, Vol. 90, nb. 4 (1983), 229–244.

[7] A. Granas and J. Dugundji, *Fixed Point Theory*, Springer 2003.

[8] A. Granas and F. C. Liu, *Some Minimax Theorems without Convexity*, in Nonlinear and Convex Analysis, proceedings in honour of Ky Fan, Edited by Bor-Luh Lin and Stephen Simons, Lectures notes in pure and applied mathematics, Vol. 107, Marcel Dekker, 1987.

[9] G. Maltese, *A Simple Proof of the Fundamental Theorem of Finite Markov's Chains*, Amer. Math. Monthly, vol. 93, nb. 8 (1986), 629–630.

[10] F. P. Greenleaf, *Invariant Means on Topological Groups*, Van Nostrand Mathematical studies 16, 1969.

[11] A. S. Solodovnikov, *Systems of Linear Inequalities*, Little Mathematical Library, Mir Publishers, Moscow, 1979.

[12] W. Takahashi, *Fan's existence theorem for inequalities concerning convex functions and its applications*, Minimax Theory and its Applications, B. Ricceri and S. Simons (eds) (1998), 241-260.

Annals of the Academy of Romanian Scientists
Series on Mathematics and its Applications
ISSN 2066 - 6594 Volume 3, Number 2 / 2011

A NOTE ON SOME BOUNDARY VALUE PROBLEMS FOR HIGHER ORDER DIFFERENTIAL INCLUSIONS*

Aurelian Cernea†

Abstract

We establish Filippov existence theorems for solutions of certain boundary value problems associated to some higher order differential inclusions.

MSC: 34A60

keywords: differential inclusion, boundary value problem, measurable selection.

1 Introduction

This paper is concerned with differential inclusions of the form

$$\mathcal{D}x \in F(t,x), \tag{1.1}$$

where $\mathcal{D}$ is a differential operator and $F(.,.) : [0,1] \times \mathbf{R} \to \mathcal{P}(\mathbf{R})$ is a set-valued map.

In the last years we observe a remarkable amount of interest in the study of existence of solutions of several boundary value problems associated to

*Accepted for publication in revised form on April 13, 2011.
†acernea@fmi.unibuc.ro, Faculty of Mathematics and Informatics, University of Bucharest, Academiei 14, 010014 Bucharest, Romania

problem (1.1). Most of these existence results are obtained using fixed point techniques and are based on an integral form of the right inverse to the operator $\mathcal{D}$. This means that for every f the unique solution y of the equation $\mathcal{D}y = f$ can be written in the form $y = \mathcal{R}f$, when the operator $\mathcal{R}$ has nonnegative Green's function.

For a first order differential inclusion defined by a lipschitzian set-valued map with nonconvex values, Filippov's theorem ([7]) consists in proving the existence of a solution starting from a given almost solution. Moreover, the result provides an estimate between the starting almost solution and the solution of the differential inclusion.

The aim of this note is to show that Filippov's ideas can be suitably adapted in order to obtain the existence of solutions for the following problems

$$x^{(n)} - \lambda x \in F(t,x), \quad a.e.\ (I) \tag{1.2}$$

with boundary conditions of the form

$$x^{(i)}(0) - x^{(i)}(T) = \mu_i, \quad i = 0,1,...,n-1, \tag{1.3}$$

and

$$(p(t)x'(t))' \in F(t,x(t)) \quad a.e.\ (I), \tag{1.4}$$

with boundary conditions of the form

$$\alpha x(0) - \beta \lim_{t\to 0+} p(t)x'(t) = 0, \quad \gamma x(T) + \delta \lim_{t\to T-} p(t)x'(t) = 0, \tag{1.5}$$

where $I = [0,T]$, $\lambda \in \mathbf{R}$, $\mu_i \in \mathbf{R}$, $i = 0,1,...,n-1$, $F : I \times \mathbf{R} \to \mathcal{P}(\mathbf{R})$ is a set-valued map, $p(.) : I \to (0,\infty)$ is a continuous function and $\alpha, \beta, \gamma, \delta$ are nonnegative reals with $\alpha\delta + \beta\gamma + \gamma\alpha \int_0^T \frac{dt}{p(t)} \neq 0$.

Existence results obtained using fixed point techniques for problem (1.2)-(1.3) may be found in [2,3] and for problem (1.4)-(1.5) may be found in [4,5,9,10]. The results in the present paper are improvements of previous existence theorems from our papers [3] respectively, [4].

The paper is organized as follows: in Section 2 we recall some preliminary facts that we need in the sequel and in Section 3 we prove our main results.

2 Preliminaries

Let (X,d) be a metric space. We recall that the Pompeiu-Hausdorff distance of the closed subsets $A,B\subset X$ is defined by

$$d_H(A,B)=\max\{d^*(A,B),d^*(B,A)\},\quad d^*(A,B)=\sup\{d(a,B);a\in A\},$$

where $d(x,B)=\inf_{y\in B}d(x,y)$.

In what follows $C(I,\mathbf{R})$ is the Banach space of all continuous functions from I to $\mathbf{R}$ with the norm $||x(.)||_C=\sup_{t\in I}|x(t)|$, $AC^i(I,\mathbf{R})$ is the space of i-times differentiable functions $x:I\to\mathbf{R}$ whose i-th derivative $x^{(i)}(.)$ is absolutely continuous, $AC^1_p(I,\mathbf{R})$ is the space of continuous functions $x:I\to\mathbf{R}$ such that $p(.)x'(.)$ is absolutely continuous and $L^1(I,\mathbf{R})$ is the Banach space of integrable functions $u(.):I\to\mathbf{R}$ endowed with the norm $||u(.)||_1=\int_0^T|u(t)|dt$.

A function $x(.)\in AC^{n-1}(I,\mathbf{R})$ is called a solution of problem (1.2)-(1.3) if there exists a function $v(.)\in L^1(I,\mathbf{R})$ with $v(t)\in F(t,x(t))$, *a.e.* (I) such that $x^{(n)}(t)-\lambda x(t)=v(t)$, *a.e.* (I) and $x(.)$ satisfies conditions (1.3).

We consider the Green function $G(.,.):I\times I\to\mathbf{R}$ associated to the periodic boundary problem

$$x^{(n)}-\lambda x=0,\quad x^{(i)}(0)-x^{(i)}(T)=0,\quad i=0,1,...,n-1.$$

For the properties of $G(.,.)$ we refer to [2].

The next result is well known.

Lemma 2.1. ([2]) *If $v(.):[0,T]\to\mathbf{R}$ is an integrable function then the problem*

$$\begin{gathered}x^{(n)}(t)-\lambda x(t)=v(t)\quad a.e.\ (I)\\ x^{(i)}(0)-x^{(i)}(T)=\mu_i,\quad i=0,1,...,n-1.\end{gathered}$$

has a unique solution $x(.)\in AC^{n-1}(I,\mathbf{R})$ given by

$$x(t)=P_\mu(t)+\int_0^T G(t,s)v(s)ds,$$

where

$$P_\mu(t)=\sum_{i=0}^{n-1}\frac{\partial^i}{\partial t^i}G(t,0)\mu_{n-1-i}.\tag{2.1}$$

A function $x(.) \in AC^1_p(I,\mathbf{R})$ is called a solution of problem (1.4)-(1.5) if there exists a function $v(.) \in L^1(I,\mathbf{R})$ with $v(t) \in F(t,x(t))$, *a.e.* (I) such that $(p(t)x'(t))' = v(t)$, *a.e.* (I) and conditions (1.5) are satisfied.

Lemma 2.2. ([9]) *If $v(.) : [0,T] \to \mathbf{R}$ is an integrable function then the problem*

$$(p(t)x'(t))' = v(t) \quad a.e.\ (I),$$
$$\alpha x(0) - \beta \lim_{t\to 0+} p(t)x'(t) = 0, \quad \gamma x(T) + \delta \lim_{t\to T-} p(t)x'(t) = 0$$

has a unique solution $x(.) \in AC^1_p(I,\mathbf{R})$ given by

$$x(t) = \int_0^T G_1(t,s)v(s)ds,$$

where

$$G_1(t,s) := \frac{1}{\rho}\begin{cases} (\beta+\alpha\int_0^s \frac{du}{p(u)})(\delta+\gamma\int_t^T \frac{du}{p(u)}) & if \quad 0\le s<t\le T \\ (\beta+\alpha\int_0^t \frac{du}{p(u)})(\delta+\gamma\int_s^T \frac{du}{p(u)}) & if \quad 0\le t<s\le T \end{cases}$$

and $\rho := \alpha\delta + \beta\gamma + \gamma\alpha\int_0^T \frac{dt}{p(t)} \neq 0$.

Finally, we recall a selection result which is a version of the celebrated Kuratowski and Ryll-Nardzewski selection theorem ([8]).

Lemma 2.3. ([1]) *Consider X a separable Banach space, B is the closed unit ball in X, $H : I \to \mathcal{P}(X)$ is a set-valued map with nonempty closed values and $g : I \to X, L : I \to \mathbf{R}_+$ are measurable functions. If*

$$H(t) \cap (g(t) + L(t)B) \neq \emptyset \quad a.e.(I),$$

then the set-valued map $t \to H(t)\cap(g(t)+L(t)B)$ has a measurable selection.

In the sequel we assume the following conditions on F.

Hypothesis 2.4. (i) *$F(.,.) : I\times\mathbf{R} \to \mathcal{P}(\mathbf{R})$ has nonempty closed values and for every $x \in \mathbf{R}$ $F(.,x)$ is measurable.*

(ii) *There exists $L(.) \in L^1(I,\mathbf{R})$ such that for almost all $t \in I, F(t,\cdot)$ is L(t)-Lipschitz in the sense that*

$$d_H(F(t,x),F(t,y)) \le L(t)|x-y| \quad \forall\ x,y \in \mathbf{R}.$$

3 The main results

We are now ready to prove the main result of this paper.

Denote $L_0 := \int_0^T L(s)ds$ and $M_0 := \sup_{t,s\in I}|G(t,s)|$.

Theorem 3.1. *Assume that Hypothesis* 2.4 *is satisfied and* $M_0L_0 < 1$. *Let* $y(.) \in AC^{n-1}(I,\mathbf{R})$ *be such that there exists* $q(.) \in L^1(I,\mathbf{R})$ *with* $d(y^{(n)}(t) - \lambda y(t), F(t,y(t))) \leq q(t)$, *a.e.* (I). *Denote* $\tilde{\mu}_i = y^{(i)}(0) - y^{(i)}(T)$, $i = 0,1,...,n-1$.

Then there exists $x(.) : I \to \mathbf{R}$ *a solution of* (1.2)-(1.3) *satisfying for all* $t \in I$

$$|x(t) - y(t)| \leq \frac{1}{1 - M_0L_0}\sup_{t\in I}|P_\mu(t) - P_{\tilde{\mu}}(t)| + \frac{M_0}{1-M_0L_0}\int_0^T q(t)dt,$$

where $P_\mu(t)$ *is defined in* (2.1).

Proof. The set-valued map $t \to F(t,y(t))$ is measurable with closed values and

$$F(t,y(t)) \cap \{y^{(n)}(t) - \lambda y(t) + q(t)[-1,1]\} \neq \emptyset \quad a.e.\ (I).$$

From Lemma 2.3 it follows that there exists a measurable selection $f_1(t) \in F(t,y(t))$ *a.e.* (I) such that

$$|f_1(t) - y^{(n)}(t) + \lambda y(t)| \leq q(t) \quad a.e.\ (I) \tag{3.2}$$

Define $x_1(t) = P_\mu(t) + \int_0^T G(t,s)f_1(s)ds$ and one has

$$|x_1(t) - y(t)| \leq \sup_{t\in I}|P_\mu(t) - P_{\tilde{\mu}}(t)| + M_0||q||_1.$$

We claim that it is enough to construct the sequences $x_n(.) \in C(I,\mathbf{R})$, $f_n(.) \in L^1(I,\mathbf{R})$, $n \geq 1$ with the following properties

$$x_n(t) = P_\mu(t) + \int_0^T G(t,s)f_n(s)ds, \quad t \in I, \tag{3.3}$$

$$f_n(t) \in F(t,x_{n-1}(t)) \quad a.e.\,(I),\ n \geq 1, \tag{3.4}$$

$$|f_{n+1}(t) - f_n(t)| \leq L(t)|x_n(t) - x_{n-1}(t)| \quad a.e.\,(I), n \geq 1. \tag{3.5}$$

If this construction is realized then from (3.2)-(3.5) we have for almost all $t \in I$

$$|x_{n+1}(t) - x_n(t)| \leq \int_0^T |G(t,t_1)|.|f_{n+1}(t_1) - f_n(t_1)|dt_1 \leq$$

$$M_0 \int_0^T L(t_1)|x_n(t_1) - x_{n-1}(t_1)|dt_1 \leq M_0 \int_0^T L(t_1) \int_0^T |G(t_1,t_2)|.$$

$$|f_n(t_2) - f_{n-1}(t_2)|dt_2 \leq M_0^2 \int_0^T L(t_1) \int_0^T L(t_2)|x_{n-1}(t_2) - x_{n-2}(t_2)|dt_2 dt_1$$

$$\leq M_0^n \int_0^T L(t_1) \int_0^T L(t_2)... \int_0^T L(t_n)|x_1(t_n) - y(t_n)|dt_n...dt_1 \leq$$
$$\leq (M_0L_0)^n (\sup_{t\in I} |P_\mu(t) - P_{\tilde{\mu}}(t)| + M_0||q||_1).$$

Therefore $\{x_n(.)\}$ is a Cauchy sequence in the Banach space $C(I,\mathbf{R})$, hence converging uniformly to some $x(.) \in C(I,\mathbf{R})$. Therefore, by (3.5), for almost all $t \in I$, the sequence $\{f_n(t)\}$ is Cauchy in $\mathbf{R}$. Let $f(.)$ be the pointwise limit of $f_n(.)$.

Moreover, one has

$$|x_n(t) - y(t)| \leq |x_1(t) - y(t)| + \textstyle\sum_{i=1}^{n-1} |x_{i+1}(t) - x_i(t)| \leq \sup_{t\in I} |P_\mu(t) - P_{\tilde{\mu}}(t)| + M_0||q||_1 + \sum_{i=1}^{n-1} (\sup_{t\in I} |P_\mu(t) - P_{\tilde{\mu}}(t)| + M_0||q||_1)(M_0L_0)^i \leq \frac{\sup_{t\in I} |P_\mu(t)-P_{\tilde{\mu}}(t)|+M_0||q||_1}{1-M_0L_0}. \tag{3.6}$$

On the other hand, from (3.2), (3.5) and (3.6) we obtain for almost all $t \in I$

$$|f_n(t) - y^{(n)}(t) + \lambda y(t)| \leq \textstyle\sum_{i=1}^{n-1} |f_{i+1}(t) - f_i(t)| + |f_1(t) - y^{(n)}(t) + \lambda y(t)| \leq L(t) \frac{\sup_{t\in I} |P_\mu(t)-P_{\tilde{\mu}}(t)|+M_0||q||_1}{1-M_0L_0} + q(t).$$

Hence the sequence $f_n(.)$ is integrably bounded and therefore $f(.) \in L^1(I,\mathbf{R})$.

Using Lebesgue's dominated convergence theorem and taking the limit in (3.3), (3.4) we deduce that $x(.)$ is a solution of (1.1). Finally, passing to the limit in (3.6) we obtained the desired estimate on $x(.)$.

It remains to construct the sequences $x_n(.), f_n(.)$ with the properties in (3.3)-(3.5). The construction will be done by induction.

Since the first step is already realized, assume that for some $N \geq 1$ we already constructed $x_n(.) \in C(I, \mathbf{R})$ and $f_n(.) \in L^1(I, \mathbf{R}), n = 1, 2, ...N$ satisfying (3.3),(3.5) for $n = 1, 2, ...N$ and (3.4) for $n = 1, 2, ...N - 1$. The set-valued map $t \to F(t, x_N(t))$ is measurable. Moreover, the map $t \to L(t)|x_N(t) - x_{N-1}(t)|$ is measurable. By the lipschitzianity of $F(t, .)$ we have that for almost all $t \in I$

$$F(t, x_N(t)) \cap \{f_N(t) + L(t)|x_N(t) - x_{N-1}(t)|[-1, 1]\} \neq \emptyset.$$

From Lemma 2.3 we obtain that there exists a measurable selection $f_{N+1}(.)$ of $F(., x_N(.))$ such that

$$|f_{N+1}(t) - f_N(t)| \leq L(t)|x_N(t) - x_{N-1}(t)| \quad a.e.\ (I).$$

We define $x_{N+1}(.)$ as in (3.3) with $n = N + 1$. Thus $f_{N+1}(.)$ satisfies (3.4) and (3.5) and the proof is complete.

Remark 3.2. In [3], using Covitz-Nadler set-valued contraction principle ([6]) one obtains another Filippov type existence result for problem (1.2)-(1.3). More exactly, according to Theorem 3.1 in [3], for any $\varepsilon > 0$ there exists $x_\varepsilon(.)$ a solution of problem (1.2)-(1.3) satisfying for all $t \in I$

$$|x_\varepsilon(t) - y(t)| \leq \frac{1}{1 - M_0 L_0} \sup_{t \in I} |P_\mu(t) - P_{\tilde{\mu}}(t)| + \frac{M_0}{1 - M_0 L_0} \int_0^T q(t)dt + \varepsilon \quad (3.7)$$

Obviously, the estimate in (3.1) is better than the one in (3.7). Moreover, in [3] it is required that the set-valued map $F(., .)$ satisfy an additional hypothesis, namely $d(0, F(t, 0)) \leq L(t)$ a.e. (I).

We are concerned now with the boundary value problem (1.4)-(1.5).

Set $M_1 := \max_{t,s \in I} |G_1(t, s)|$.

Theorem 3.3. *Assume that Hypothesis* 2.4 *is satisfied and* $M_1 L_0 < 1$. *Let* $y(.) \in AC_p^1(I, \mathbf{R})$ *be such that there exists* $q(.) \in L^1(I, \mathbf{R})$ *with* $d((p(t)y(t))', F(t, y(t))) \leq q(t)$, *a.e.* (I), $\alpha y(0) - \beta \lim_{t \to 0+} p(t)y'(t) = 0$, $\gamma y(T) + \delta \lim_{t \to T-} p(t)y'(t) = 0$.

Then there exists $x(.) : I \to \mathbf{R}$ *a solution of* (1.1)-(1.2) *satisfying for all* $t \in I$

$$|x(t) - y(t)| \leq \frac{M_1}{1 - M_1 L_0} \int_0^T q(t)dt. \quad (3.8)$$

The proof of Theorem 3.3 is similar to the one of Theorem 3.1.

Remark 3.4. In [4], using Covitz-Nadler set-valued contraction principle one obtains another Filippov type existence result for problem (1.4)-(1.5). More precisely, according to Theorem 3.1 in [4], for any $\varepsilon > 0$ there exists $x_\varepsilon(.)$ a solution of problem (1.4)-(1.5) satisfying for all $t \in I$

$$|x_\varepsilon(t) - y(t)| \leq \frac{M_1}{1 - M_1 L_0} \int_0^T q(t)dt + \varepsilon. \tag{3.9}$$

Obviously, the estimate in (3.8) is better than the one in (3.9).

References

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

[2] M. Benchohra, J. R. Graef, J. Henderson, S. K. Ntouyas. Nonresonance impulsive higher order functional nonconvex-valued differential inclusions. *Electron. J. Qual. Theory Differ. Equ.* N. 13, 2002.

[3] A. Cernea. On the existence of solutions for a higher order differential inclusion without convexity. *Electron. J. Qual. Theory Differ. Equ.* N. 8, 2007.

[4] A. Cernea. On a boundary value problem for a Sturm-Liouville type differential inclusion. *J. Sys. Sci. Complexity.* 23:390-394, 2010.

[5] Y.K. Chang, W.T. Li. Existence results for second order impulsive functional differential inclusions. *J. Math. Anal. Appl.* 301:477-490, 2005.

[6] H. Covitz, S.B. Nadler jr.. Multivalued contraction mapping in generalized metric spaces. *Israel J. Math.* 8:5-11, 1970.

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

[8] K. Kuratowski, C. Ryll-Nardzewski. A general theorem on selectors. *Bull. Acad. Pol. Sci.* 13:397-403, 1965.

[9] Y. Liu, J. Wu, Z. Li. Impulsive boundary value problems for Sturm-Liouville type differential inclusions. *J. Sys. Sci. Complexity.* 20:370-380, 2007.

[10] Y. Sun, B. L. Xu, L. S. Liu. Positive solutions of singular boundary value problems for Sturm-Liouville equations. *J. Sys. Sci. Math. Sci.* 25:69-77, 2005.

INFORMATION FOR AUTHORS
SUBMITTING MANUSCRIPTS IN ELECTRONIC FORM

The **ANNALS OF THE ACADEMY OF ROMANIAN SCIENTISTS,** series on **MATHEMATICS AND ITS APPLICATIONS** covers the following fields:

- Ordinary and partial differential equations
- Optimization, optimal control and design
- Numerical analysis and scientific computing
- Algebraic, topological and differential structures
- Probability and statistics
- Algebraic and differential geometry
- Mathematical modelling in mechanics and engineering sciences
- Mathematical economy and game theory
- Mathematical physics and applications

We encourage the publication of contributed papers, short notes, survey articles and reviews, with a novel and correct content, in any area of mathematics and its applications. Two issues are published every year, both in electronic and printed form. Special issues devoted to International Conferences or other important mathematical events will be also considered.

The manuscript (in English) should be sent via email as a PDF file, either to the series editors or to any member of the Editorial Board together with a cover letter stating the intention to publish it in Mathematics and its Applications. The submission of a manuscript entails the assurance that it has not been published or it is not under consideration for publication elsewhere and that all the coauthors have approved its submission to Mathematics and its Applications. A template that precises the style and the format of the manuscripts may be found at

www.mathematics-and-its-applications.com

where the online version is published.

After refereeing is complete, an official letter of acceptance or refusal is sent by email to the corresponding author together with the observations of the reviewers. For publication, the LATEX and PDF files of the revised paper are required, which may be edited by our Secretariate to follow the general presentation rules of the journal. If the manuscript includes figures, they should be added as separate files in the same email message. All **line drawings** and **photos** will be in **grey scale** in the printed version and may be coloured in the electronic version.Very long papers may be divided in several parts and published in subsequent issues of the journal.

A PDF file with the proofs will be sent to the corresponding author with indications for corrections.

For communication, the name, phone number and e-mail address of the author to whom proofs should be sent must be indicated. The corresponding author is responsible for informing all co-authors that the manuscript has been submitted for this series.