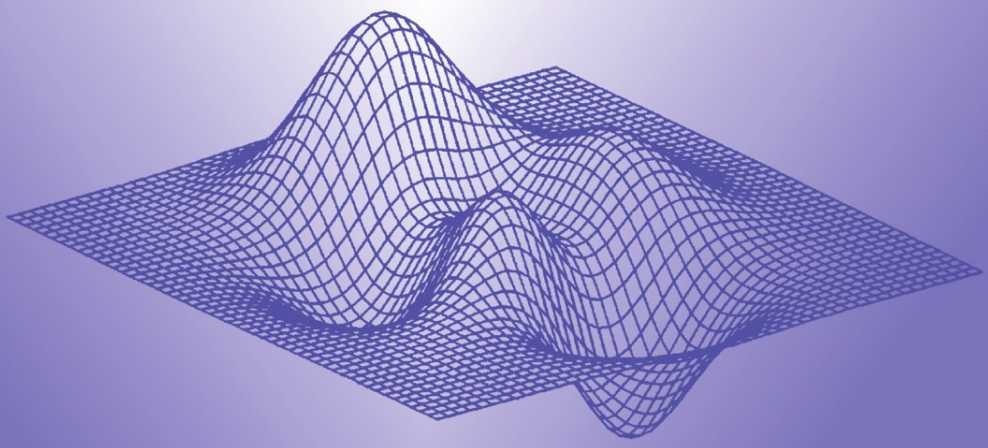
NONCONVEX OPTIMIZATION AND ITS APPLICATIONS

Variational Analysis and Applications

Edited by
Franco Giannessi
Antonino Maugeri



Springer

VARIATIONAL ANALYSIS AND APPLICATIONS

Nonconvex Optimization and Its Applications

VOLUME 79

VARIATIONAL ANALYSIS AND APPLICATIONS

Edited by

FRANCO GIANNESSI
University of Pisa, Italy

ANTONINO MAUGERI
University of Catania, Italy

Library of Congress Cataloging-in-Publication Data

A C.I.P. record for this book is available from the Library of Congress.

ISBN-10: 0-387-24209-0 ISBN-13: 978-0387-24209-5
e-ISBN-10: 0-387-24276-7 e-ISBN-13: 978-0387-24276-7

Printed on acid-free paper.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 11366270

springeronline.com

Contents

Preface xi

PART 1

The Work of G. Stampacchia in Variational Inequalities 3
J.-L. Lions

In Memory of Guido Stampacchia 31
M.G. Garroni

The Collaboration between Guido Stampacchia and Jacques-Louis Lions On Variational Inequalities 33
E. Magenes

In Memory of Guido Stampacchia 39
O.G. Mancino

Guido Stampacchia 47
S. Mazzone

Memories of Guido Stampacchia 79
L. Nirenberg

In Memory of Guido Stampacchia 81
C. Sbordone

Guido Stampacchia, My Father 83
G. Stampacchia

PART 2

Convergence and Stability of a Regularization Method for Maximal Monotone Inclusions and its Applications to Convex Optimization 89
Ya. I. Alber, D. Butnariu and G. Kassay

Partitionable Mixed Variational Inequalities 133
E. Allevi, A. Gnudi, I.V. Konnov and E.O. Mazurkevich

Irreducibility of the Transition Semigroup Associated with the Two Phase Stefan Problem 147
V. Barbu and G. Da Prato

On Some Boundary Value Problems for Flows with Shear Dependent Viscosity 161
H. Beirão da Veiga

Homogenization of Systems of Partial Differential Equations 173
A. Bensoussan

About the Duality Gap in Vector Optimization 195
G. Bigi and M. Pappalardo

Separation of Convex Cones and Extremal Problems 205
V.Boltyanski

Infinitely Many Solutions for the Dirichlet Problem via a Variational Principle of Ricceri 215
F.Cammaroto, A.Chinnì, and B. Di Bella

A Density Result on the Space VMO_{ω} 231
A.O. Caruso and M.S. Fanciullo

Linear Complementarity since 1978 239
R.W. Cottle

Variational Inequalities in Vector Optimization 259
G.P. Crespi, I. Ginchev and M. Rocca

Variational Inequalities for General Evolutionary Financial Equilibrium 279
P. Daniele

Variational Control Problems with Constraints via Exact Penalization 301
V.F. Demyanov, F. Giannessi and G.Sh. Tamasyan

Continuous Sets and Non-Attaining Functionals in Reflexive Banach Spaces 343
E. Ernst and M. Théra

Existence and Multiplicity Results for a Nonlinear Hammerstein Integral Equation 359
F. Faraci

Differentiability of Weak Solutions of Nonlinear Second Order Parabolic Systems with Quadratic Growth and Non Linearity $q \geq 2$ 373
L. Fattorusso

An Optimization Problem with Equilibrium Constraint in Urban Transport 393
P. Ferrari

Sharp Estimates for Green's Functions: Singular Cases 409
M.G. Garroni

First-Order Conditions for $C^{0,1}$ Constrained Vector Optimization 427
I. Ginchev, A. Guerraggio and M. Rocca

Global Regularity for Solutions to Dirichlet Problem for Elliptic Systems with Nonlinearity $q \geq 2$ and with Natural Growth 451
S. Giuffrè and G. Idone

Optimality Conditions for Generalized Complementarity Problems 465
S. Giuffrè, G. Idone and A. Maugeri

Variational Inequalities for Time Dependent Financial Equilibrium with Price Constraints 477
S. Giuffrè and S. Pia

Remarks about Diffusion Mediated Transport: Thinking about Motion in Small Systems 497
S. Hastings and D. Kinderlehrer

Augmented Lagrangian and Nonlinear Semidefinite Programs 513
X.X. Huang, X.Q. Yang and K.L. Teo

Optimality Alternative: a Non-Variational Approach to Necessary Conditions 531
A.D. Ioffe

A Variational Inequality Scheme for Determining an Economic Equilibrium of Classical or Estended Type 553
A. Jofre, R.T. Rockafellar and R.J.-B. Wets

On Time Dependent Vector Equilibrium Problems 579
A. Khan and F. Raciti

On Some Nonstandard Dynamic Programming Problems of Control Theory 589
A.B. Kurzhanski and P. Varaiya

Properties of Gap Function for Vector Variational Inequality 605
S.J. Li and G.Y. Chen

Zero Gravity Capillary Surfaces and Integral Estimates 633
G.M. Lieberman

Asymptotically Critical Points and Multiple Solutions in the Elastic Bounce Problem 651
A. Marino and C. Saccon

A Branch-and-Cut to the Point-to-Point Connection Problem on Multicast Networks 665
C.N. Meneses, C.A.S. Oliveira and P.M. Pardalos

Variational Inequality and Evolutionary Market Disequilibria: The Case of Quantity Formulation 681
M. Milasi and C. Vitanza

Numerical Approximation of Free Boundary Problem by Variational Inequalities. Application to Semiconductor Devices 697
M. Morandi Cecchi and R. Russo

Sensitivity Analysis for Variational Systems 723
B.S. Mordukhovich

Stable Critical Points for the Ginzburg Landau Functional on Some Plane Domains 745
M.K. Venkatesha Murthy

The Distance Function to the Boundary and Singular Set of Viscosity solutions of Hamilton-Jacobi Equation 765
L. Nirenberg

L^p -Regularity for Poincaré Problem and Applications 773
D.K. Palagachev

Minimal Fractions of Compact Convex Sets 791
D. Pallaschke and R. Urbański

On Generalized Variational Inequalities 813
B. Panicucci and M. Pappalardo

Bounded (Hausdorff) Convergence: Basic Facts and Applications 827
J.-P. Penot and C. Zălinescu

Control Processes with Distributed Parameters in Unbounded Sets. Approximate Controllability with Variable Initial Locus 855
G. Pulvirenti, G. Santagati and A. Villani

Well Posedness and Optimization Problems 889
L. Pusillo

Semismooth Newton Methods for Shape-Preserving Interpolation, Option Price and Semi-Infinite Programs 905
L. Qi

Hölder Regularity Results for Solutions of Parabolic Equations 921
M.A. Ragusa

Survey on the Fenchel Problem of Level Sets 935
T. Rapcsák

Integral Functionals on Sobolev Spaces Having Multiple Local Minima 953
B. Ricceri

Aspects of the Projector on Prox-Regular Sets 963
Stephen M. Robinson

Application of Optimal Control Theory to Dynamic Soaring of Seabirds 975
G. Sachs and P. Bussotti

On The Convergence of the Matrices Associated to the Adjugate Jacobians 995
C. Sbordone

Quasi-Variational Inequalities Applied to Retarded Equilibria in Time-Dependent Traffic Problems 1007
L. Scrimali

Higher Order Approximation Equations for the Primitive Equations of the Ocean 1025
E. Simonnet, T. Tachim Medjo and R. Temam

Hahn-Banach Theorems and Maximal Monotonicity 1049
S. Simons

Concrete Problems and the General Theory of Extremum 1085
V.M. Tikhomirov

Numerical Solution for Pseudomonotone Variational Inequality Problems by Extragradient Methods 1101
F. Tinti

Regularity and Existence Results for Degenerate Elliptic Operators 1129
C. Vitanza and P. Zamboni

Vector Variational Inequalities and Dynamic Traffic Equilibria 1141
X.Q. Yang and H. Yu

A New Proof of the Maximal Monotonicity of the Sum using the Fitzpatrick Function 1159
C. Zălinescu

Contributors 1173

Preface

This Volume contains the (refereed) papers presented at the 38th Conference of the School of Mathematics "G.Stampacchia" of the "E.Majorana" Centre for Scientific Culture of Erice (Sicily), held in Memory of G. Stampacchia and J.-L. Lions in the period June 20 - July 1, 2003.

The presence of 130 participants from 15 Countries has greatly contributed to the success of the meeting.

The School of Mathematics was dedicated to Stampacchia, not only for his great mathematical achievements, but also because He founded it.

The core of the Conference has been the various features of the Variational Analysis and their motivations and applications to concrete problems. Variational Analysis encompasses a large area of modern Mathematics, such as the classical Calculus of Variations, the theories of perturbation, approximation, subgradient, subderivates, set convergence and Variational Inequalities, and all these topics have been deeply and intensely dealt during the Conference. In particular, Variational Inequalities, which have been initiated by Stampacchia, inspired by Signorini Problem and the related work of G. Fichera, have offered a very great possibility of applications to several fundamental problems of Mathematical Physics, Engineering, Statistics and Economics.

The pioneer work of Stampacchia and Lions can be considered as the basic kernel around which Variational Analysis is going to be outlined and constructed.

The Conference has dealt with both finite and infinite dimensional analysis, showing that to carry on these two aspects disjointly is unsuitable for both.

The book is divided into two parts. The former contains the reproduction - under kind permission of J.Wiley - of a paper presented in 1978 at "E.Majorana" Centre by J.-L.Lions on the work of Stampacchia just after His death, and - in alphabetic order - reminiscences and comments on the mathematical achievements of Stampacchia. The latter contains - in alphabetic order - the other papers presented at the Conference.

We want to express our deep gratitude to all those who took part in the Conference. Special mention should once more be made of the "E. Majorana" Centre, which offered its facilities and stimulating environment for the meeting. We are all indebted to the "E.Majorana" Centre, the Municipality of Erice, the Italian National Group for Mathematical Analysis, Probability and Applications (GNAMPA), the University of Catania, the Faculty of Sciences and the Dept.of Mathematics and Computer Science of University of Catania, the University of Messina, the University of Pisa, the Dept.of Mathematics of University of Pisa, the University of Reggio Calabria (DIMET), for their financial support. We are grateful to Dr.J.Martindale of Kluwer Publ.Co. and to Professor P.M.Pardalos for having proposed to publish this book. We want also to thank L. Lucarelli Co. for the typing.

F.Giannessi A.Maugeri

PART 1

THE WORK OF G. STAMPACCHIA IN VARIATIONAL INEQUALITIES*

J.-L. Lions

1. INTRODUCTION

An introductory survey on variational inequalities should have been made here by G. Stampacchia.

All those of you who knew him, who had the pleasure to share with him long and stimulating discussions, who knew his warm personality, will share my emotion and my sorrow.

In what follows, I will try to present some of his main ideas and his main contributions *in the field of variational inequalities,* the main topic of the meeting mentioned in the preface and where he was looking forward to participating and lecturing.

Therefore, I will not speak of his previous contributions; a general report with a complete bibliography will be presented by E. Magenes in the *Bollettino dell'Unione Matematica Italiana.*[1]

In the field of partial differential equations and functional analysis, in 1958 he published a survey with E. Magenes (*Annali Scuola Normale Superiore Pisa,* 12 (1958), 247-357), which had a very deep influence on the teaching of partial differential equations (PDE), and he made very important

* Re-printed from "Variational Inequalities and Complementarity Problems. Theory and Applications", Edited by R. W. Cottle, F. Giannessi, J.-L. Lions, J. Wiley, 1980, pp. 1-24.

[1] Vol. 15-A, No. 3, 1978, pp. 715-756.

contributions to the study of second-order elliptic operators, in particular those without any smoothness hypothesis on the coefficients. It was while he was working on deep questions of regularity of solutions when coefficents are only assumed to be bounded and measurable, and on problems of potential theory, that he was led, at the beginning of the 60's, to variational inequalities.

2. VARIATIONAL INEQUALITIES

The following result is now classical [1]: let V be a Hilbert space on $\mathbb{R}$; let $a(u,v)$ be a continuous bilinear form on V, which is *not necessarily symmetric,* and which is V-elliptic, i.e. which satisfies

$$a(v,v) \geq \alpha \|v\|^2, \quad \alpha > 0, \quad \forall v \in V \tag{1}$$

($\|\ \|$ denotes the norm in V.) Let K be a closed convex subset of V, $K \neq \varnothing$, and let $v \to (f,v)$ be a continuous linear form on V; then, there exists a unique element $u \in K$ such that

$$a(u,v-u) \geq (f,v-u) \quad \forall v \in K \tag{2}$$

This (2) is what is called a *variational inequality* (in short VI).

Let us remark that:

(i) if $K = V$, (2) is equivalent to

$$a(u,v) = (f,v), \quad \forall v \in V \tag{3}$$

and the above result gives the Lax-Milgram lemma;

(ii) if a is symmetric $\left(\text{i.e. } a(u,v) = a(v,u) \forall u,v \in V\right)$ then (2) is equivalent to

$$\frac{1}{2}a(u,u) - (f,u) = \min_{v \in K}\left[\frac{1}{2}a(v,v) - (f,v)\right] \tag{4}$$

The idea of the original proof of Stampacchia is as follows:

(I) the result is immediate, according to (ii) above, if a is symmetric;

(II) if (2) is proven for $a(u,v)$, it will also be proven for $a(u,v) + p(u,v)$ where $p(u,v)$ is a not too large perturbation of $a(u,v)$;

(III) with this in mind, one introduces

$$\alpha(u,v)=\frac{1}{2}\left[a(u,v)+a(v,u)\right]$$
$$\beta(u,v)=\frac{1}{2}\left[a(u,v)-a(v,u)\right] \tag{5}$$

and, for $0\le\theta\le 1$,

$$a_\theta(u,v)=\alpha(u,v)+\theta\beta(u,v) \tag{6}$$

By virtue of (I), the result is true for $\theta=0$; using (II) one checks that the result is true for $a_\theta(u,v)$, $0\le\theta\le\theta_0$, where θ_0 is a constant depending only on *a,* and one proceeds in this way.

The main application that Stampacchia had in mind at the beginning of this theory was to *potential theory*; he was at that time giving a series of lectures in Leray's seminar [2, 3]. Let us give one example extracted from one of his works (see [2]).

3. APPLICATION TO POTENTIAL THEORY

Let Ω be a bounded open set of $\mathbb{R}^n$; we consider the classical Sobolev spaces

$$H^2(\Omega)=\left\{v\left|v,\frac{\partial v}{\partial x_1},\ldots,\frac{\partial v}{\partial x_n}\in L^2(\Omega)\right.\right\}$$

$$H_0^1(\Omega)=\left\{v\left|v\in H^1(\Omega),v=0\text{ on }\Gamma\right.\right\}$$

Let $a_{ij}(x)$ be a family of functions such that

$$a_{ij}\in L^\infty(\Omega), \tag{7}$$

$$\sum_{i,j=1}^{n}a_{ij}(x)\xi_i\xi_j\ge\alpha\sum_{i=1}^{n}\xi_i^2,\quad \alpha>0,\quad \xi_1\in\mathbb{R} \tag{8}$$

and let us define, $\forall u,v\in H^1(\Omega)$:

$$a(u,v)=\sum_{i,j}\int_{\Omega}a_{ij}(x)\frac{\partial u}{\partial x_j}\frac{\partial v}{\partial x_i}\mathrm{d}x \tag{9}$$

We do *not* assume symmetry ($a_{ij}\neq a_{ij}$ in general) and we do *not* assume regularity on the $a_{ij.}$.

Let us define (in a vague manner for the moment) the set K as follows: let E be a closed subset of Ω and let us set

$$K=\left\{v\middle|v\in H_0^1(\Omega),v\geq 1 \text{ on } E\right\} \tag{10}$$

The precise meaning of '$v\geq 1$ on E' is as follows: we say that $v\geq 1$ on E in the sense of $H_0^1(\Omega)$ if there exists a sequence of smooth functions u_m in $H_0^1(\Omega)$ such that:

(i) $u_m\to v$ in $H_0^1(\Omega)$

(ii) $u_m\geq 1$ on E

If K is not empty, then there exists a unique element $u\in K$ satisfying

$$a(a,v-u)\geq 0 \quad \forall v\in K \tag{11}$$

3.1 Interpretation of (11)

Stampacchia shows that

$$a(u,v)=\int_{\Omega}v\mathrm{d}\mu, \quad \forall v\in H_0^1(\Omega)\cap C^0(\Omega) \tag{12}$$

where

$\mathrm{d}\mu$ is a positive measure, with support in the boundary ∂E of E (13)

The fact that one has (12) with a positive *measure* is very simple: let φ be any smooth function with, say, compact support in Ω, and such that $\varphi\geq 0$;

then if u is the solution of (11), it is clear that $v = u + \varphi$ belongs to K, so that by using this choice of v in (11) we obtain

$$a(u, \varphi) \geq 0 \quad \text{for every } \varphi \geq 0 \tag{14}$$

The result follows using a theorem of Schwartz (every distribution which is greater than or equal to 0 is a (positive) *measure*).

The main point consists of showing that μ has its support in ∂E.

One shows first that

$$u = 1 \quad \text{on } E \tag{15}$$

(in the sense of $H_0^1(\Omega)$). In order to do that, Stampacchia used a simple technique, but a very powerful one, which is now one of the classical tools of the theory of partial differential equations. Let us define

$$w = \inf \{u, 1\} \tag{16}$$

One checks that $w \in K$ and that

$$a(w, w-u) = 0 \tag{17}$$

Indeed, either $u \geq 1$ and then $w = 1$, or $u \leq 1$ and then $w - u = 0$, so that

$$a_{ij}(x) \frac{\partial w}{\partial x_j}(x) \frac{\partial (w-u)}{\partial x_i}(x) = 0$$

in either case. We can take $v = w$ in (11), and from (11) and (17) we deduce that

$$a(w-u, w-u) \leq 0$$

Since[2]

$$a(w-u, w-u) \geq \alpha \|w-u\|^2$$

it follows that

$$w = u$$

[2] If we set $\|v\|^2 = \sum_{i=1}^{n} \int_\Omega \left(\frac{\partial v}{\partial x_i} \right)^2 dx$

hence (15) follows.

Let us take now φ as a smooth function in $H_0^1(\Omega)$, with support in $\complement E$; then $v = u \pm \varphi \in K$, and taking $v = u \pm \varphi$ in (11) gives

$$a(u,\varphi) = 0, \quad \forall \varphi \text{ with compact support in } \complement E \tag{18}$$

Then (13) follows from (15) and (18).

The measure μ is called the *capacitary measure* of E with respect to $a(u,v)$ and to Ω, and $\mu(1)$ in the corresponding *capacity* of E.

In reference [2](see also reference [3]), Stampacchia proceeds to study the properties of this capacity. He introduces, among other things, the notion of regular points with respect to A and shows that this notion is in fact *independent* of A (in the class of elliptic operators), so that it is equivalent with the Wiener condition (relative to $A = -\Delta$). (A is the second-order elliptic operator associated with a).

The techniques and the ideas of Stampacchia gave rise to several interesting contributions in potential theory[3].

4. A NUMBER OF VARIANTS AND EXTENSIONS

Let us return now to (8). A number of theoretical questions immediately present themselves.

A first natural question is connected with (10); if one considers, instead of $\frac{1}{2}\, a(v,v) - (f,v)$, a *general* convex function $J(v)$ defined in a *Banach* space, one is lead to a V I of the form (if J is differentiable).

$$(J'(u), v-u) \geq 0, \quad \forall v \in K$$

$$u \in K$$

It is then natural to replace the operator J' by a *monotonic operator.*

This led to the paper of Hartman and Stampacchia [4] where they study VI in reflexive Banach spaces, for non-linear partial differential operators of the types of those introduced (in increasing order of generality) by Minty[4],

[3] R. M. Hervé. *Ann. Inst. Fourier,* 14 (1964),493-508; M. Hervé, R. M. Hervé. *Ann. Inst. Fourier* 22 (1972), 131-145; A. Ancona. J. Mat. Pures et Appl., 54 (1975), 75-124.

[4] G. Minty. *Duke math. J.,* 29 (1962), 341-6.

Browder[5], and Leray and Lions[6]. The good abstract notion for *abstract* operators A leading to "well set" elliptic V I

$$\begin{aligned} &(A(u), v-u) \geq 0, \quad \forall v \in K \\ &u \in K \end{aligned} \tag{19}$$

was introduced by Brezis[7] with the notion of *pseudo-monotonic* operators.

As always in the work of Stampacchia, there is a *motivation* for the "abstract" part of the work[4]; we shall return to that.

Another question, motivated by the so-called "unilateral boundary conditions" arising in elasticity (or "Signorini's problem"; see Fichera[8]) is whether the coerciveness hypothesis (7) can be relaxed. This has been studied by Stampacchia and Lions [5, 6]. Let us mention here *one* result: suppose that $a(u, v)$ is given as in section 2 but that it satisfies, instead of (7), the much weaker condition:

$$a(v,v) \geq 0, \quad \forall v \in V \tag{20}$$

We *assume* that (8) allows *at least one* solution, and we denote by X the set of *all* solutions; one checks immediately that X is a *closed* convex *set*; let $b(u,v)$ be a continuous bilinear form on V such that

$$b(v,v) \geq \beta \|v\|^2, \quad \beta > 0, \quad \forall v \in V \tag{21}$$

Let $v \to (g,v)$ be a continuous linear form on V; for every $\varepsilon > 0$, there exists (according to the result (l) of section 2) a unique $u_\varepsilon \in K$ such that

$$a(u_\varepsilon, v-u_\varepsilon) + \varepsilon b(u_\varepsilon, v-u_\varepsilon) \geq (f + \varepsilon g, v-u_\varepsilon), \quad \forall v \in K \tag{22}$$

Then, as $\varepsilon \to 0$, $u_\varepsilon \to u_0$ in V, where u_0 is the solution of

$$\begin{aligned} &b(u_0, v-u_0) \geq (g, v-u_0), \quad \forall v \in X \\ &u_0 \in X \end{aligned} \tag{23}$$

[5] F. Browder. *Bull. Am. Math. Soc.*, 71 (1965), 780-5.
[6] J. Leray and J. L. Lions. *Bull. Soc. Math. Fr.*, 93 (1965), 97-107.
[7] H. Brezis. *Ann. Inst. Fourier*, 18 (1968),115-75.
[8] G. Fichera. *Mem. Accad. Naz. Lincei*, 8 (1964), 91-140.

This result is used in reference [6], among other things, to solve the unilateral problem.

Still another natural question is the *evolution analogue* of (2): find a function $t \to u(t)$, where t is the *time,* such that[9]

$$u(t) \in K \tag{24}$$

$$\left(\frac{\partial u(t)}{\partial t}, v - u(t)\right) + a(u(t), v - u(t)) \geq (f(t), v - u(t)) \qquad \forall v \in K \tag{25}$$

$$u(t)\big|_{t=0} = u(0) = u^0 \qquad \text{is given (in } K) \tag{26}$$

When $K = V$, (25) reduces to

$$\left(\frac{\partial u(t)}{\partial t}, v\right) + a(u(t), v) = (f(t), v), \quad \forall v \in V \tag{27}$$

It is the variational form of "abstract" parabolic equations.

This problem has been introduced in [6]; it was considerably extended and deepened in the work of Brezis[10]; many examples arising from mechanics have been studied[11]; this problem is also connected with *non-linear semi groups*[12] .

One difficulty which arises in connection with (25) is in the *definition* of what we mean by a *solution* of a VI and an important remark is now in order: let A be a non-linear operator from a reflexive Banach space V into its dual V', and let us assume that A is *monotonic*, i.e.

$$((A(u) - A(v), u - v) \geq 0, \quad \forall u, v \in V \tag{28}$$

Then if u is a solution of the V I

[9] We do not define in detail the function spaces where u can be taken.

[10] H. Brezis. *NATO Summer School,* Venice, June 1968; H. Brezis. *J. Math. Pures et Appl.*, 51 (1972), 1-168.

[11] G. Duvaut and J. L. Lions. "*Les inéquations en Mécanique et en Physique*". Dunod, Paris (1972).

[12] H. Brezis. "Opérateurs maximaux monotones et semi groupes de contractions dans les espaces de Hilbert". North-Holland, Amsterdam (1973).

$$\begin{aligned} &(A(u),v-u)\geq(f,v-u), \quad \forall v\in K \\ &u\in K \end{aligned} \tag{29}$$

one has

$$\begin{aligned} &(A(v),v-u)\geq(f,v-u), \quad \forall v\in K \\ &u\in K \end{aligned} \tag{30}$$

This is obvious, since

$$(A(v),v-u)=(A(u),v-u)+(A(v)-A(u),v-u)\geq(A(u),v-u)$$

(using (28)); but the *reciprocal* property is true, provided A is hemi-continuous (i.e. $\lambda\to(A(u+\lambda v),w)$ is continuous $\forall u,v,w\in V$). Indeed, if $\hat{v}$ is given in K, and if we choose in (30)

$$v=(1-\theta)u+\theta\hat{v}, \quad \theta\in]\,0,1]$$

we obtain, after dividing by θ:

$$(A((1-\theta)u+\theta\hat{v}),\hat{v}-u)\geq(f,\hat{v}-u) \tag{31}$$

By virtue of the hemi-continuity, we can let $\theta\to 0$ in (31) and we obtain (29) (with $\hat{v}$ instead of v).

This remark allows one to define *weak* solutions, or generalized solutions, of VI; it is used in the paper with Lewy [7] (we shall return to that) and it can also be used for (25) (to "replace" $\partial u/\partial t$ by $\partial v/\partial t$).

5. THE OBSTACLE PROBLEM

In section 4 we indicated very briefly *some* of the problems in variational inequalities which were under study in the years 1966-68; it was at about this time, may be a little earlier, that Stampacchia started working on a problem which is simple, beautiful and deep - and which led to important discoveries some of them being reported in this book.

This is the so-called "obstacle" problem. Let us consider $a(u,v)$ to be given by (9) and let us define

$$K=\left\{v\middle|v\in H_0^1(\Omega), v\geq\psi, \psi \text{ given in } \Omega\right\} \tag{32}$$

Of course, one has to specify in (32) the class where ψ is given, so that, in particular, K is not empty; the function ψ represents the *obstacle.* The corresponding VI (2) has a unique solution and the problem is as follows.

(1) How to interpret the V I?

(2) What are the regularity properties of the solution u?

In solving (l) and (2) a free-boundary problem will appear and the next question will be the following.

(3) What are the regularity properties of the free boundary?

Let us explain the basic idea of the work with Brezis [8] in a simple particular case. According to (30) the VI can be written[13]:

$$(Av, v-u)\geq(f, v-u), \quad \forall v\in K \tag{33}$$

where K is given by (32). Let us assume that

$$\begin{aligned} &\psi\in H^1(\Omega) \quad \psi\leq 0 \quad \text{on } \Gamma \\ &\text{and} \\ &A\psi\leq 0 \end{aligned} \tag{34}$$

Everything is based on a particular choice of v in (33).
For $\varepsilon>0$ we *define* u_ε as the solution in $H_0^1(\Omega)$ of

$$\begin{aligned} &\varepsilon Au_\varepsilon+u_\varepsilon=u \quad \text{in } \Omega \\ &u_\varepsilon=0 \quad \text{on } \Gamma \end{aligned} \tag{35}$$

Let us allow for the moment - *this is the crucial point* - that

$$u_\varepsilon\in K \quad (\text{i.e. } u_\varepsilon\geq\psi \quad \text{in } \Omega) \tag{36}$$

Then one can choose $v=u_\varepsilon$ in (33) and after dividing by ε it gives

$$(Au_\varepsilon, Au_\varepsilon)\leq(f, Au_\varepsilon)$$

Hence, it follows that

[13] $Av=-\sum\frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial v}{\partial x_j}\right)\in H^{-1}(\Omega)$ (dual space of $H_0^1(\Omega)$)

$$\|Au_\varepsilon\|_{L^2(\Omega)} \le \|f\|_{L^2(\Omega)} \tag{37}$$

It is a simple matter to check that $u_\varepsilon \to u$ in $H_0^1(\Omega)$ as $\varepsilon \to 0$, so from (37) one obtains that

$$Au \in L^2(\Omega) \tag{38}$$

Therefore, if we set $Au = \hat{f}\left(\hat{f} \neq f \text{ in general!}\right)$, one can think of u as being given by the solution of the Dirichlet's boundary value problem

$$Au = \hat{f} \quad \left(\hat{f} \in L^2(\Omega)\right), \quad u = 0 \text{ on } \Gamma$$

It follows that if the coefficients of A are smooth enough and if the boundary Γ of Ω is smooth enough, then

$$u \in H^2(\Omega) \tag{39}$$

that is

$$\frac{\partial^2 u}{\partial x_i \partial x_j} \in L^2(\Omega), \quad \forall i, j$$

Let us verify now that (36) holds. We write (35) as

$$\varepsilon A\left(u_\varepsilon - \psi\right) + \left(u_\varepsilon - \psi\right) + \varepsilon A\psi = u - \psi \tag{40}$$

and we take scalar products with $\left(u_\varepsilon - \psi\right)^-$ (where, in general, $v^- = \sup(-v, 0)\big)$. We obtain, since $\left(u_\varepsilon - \psi\right)^- \in H_0^1(\Omega)$ and since $a\left(v, v^-\right) = -a\left(v^-, v^-\right)$, $\left(v, v^-\right) = -\left(v^-, v^-\right)$:

$$\begin{aligned} &-\varepsilon a\left(\left(u_\varepsilon - \psi\right)^-, \left(u_\varepsilon - \psi\right)^-\right) - \left\|\left(u_\varepsilon - \psi\right)^-\right\|^2_{L^2(\Omega)} + \varepsilon\left(A\psi, \left(u_\varepsilon - \psi\right)^-\right) = \\ &= \left(u - \psi, \left(u_\varepsilon - \psi\right)^-\right) \end{aligned} \tag{41}$$

Since $A\psi \le 0$, then $\left(A\psi, \left(u_\varepsilon - \psi\right)^-\right) \le 0$, and (41) gives:

$$\left(u-\psi,(u_\varepsilon-\psi)^-\right)+\left\|(u_\varepsilon-\psi)^-\right\|^2_{L^2(\Omega)}+\varepsilon a\left((u_\varepsilon-\psi)^-,(u_\varepsilon-\psi)^-\right)\le 0$$

But $\left(u-\psi,(u_\varepsilon-\psi)^-\right)\ge 0$ and therefore $(u_\varepsilon-\psi)^-=0$, i.e. $u_\varepsilon\ge\psi$.

The above analysis can be extended, as we show below. Before doing so, let us apply (38) to the interpretation of the VI. One shows easily that u is characterized by

$$\begin{aligned} &Au-f\ge 0\\ &u-\psi\ge 0\\ &(Au-f)(u-\psi)=0 \qquad \text{in } \Omega \end{aligned} \tag{42}$$

and of course

$$u=0 \quad \text{on } \Gamma$$

Consequently, there are two sets in Ω:

$$\begin{aligned} &\text{the } \textit{coincidence} \text{ set, where } \quad u=\psi\\ &\\ &\text{the } \textit{equilibrium} \text{ set, where } \quad Au=f \end{aligned} \tag{43}$$

At least in the two-dimensional case, one can think of this problem as giving the displacement of a membrane subjected to forces f and required to stay above the obstacle ψ.

The membrane touches the obstacle on the coincidence set. The two regions are separated by a "surface" S, which is a free surface; S is not given, and on S one has two "boundary" conditions. If $\psi\in H^2(\Omega)$, one has

$$\begin{aligned} &u=\psi\\ &\text{and}\\ &\frac{\partial u}{\partial x_i}=\frac{\partial\psi}{\partial x_i} \quad \forall i \text{ on } S \end{aligned} \tag{44}$$

A natural and important question is now: under suitable hypotheses on f and on ψ, is it true that

$$Au\in L^p(\Omega)\,? \tag{45}$$

This is, of course, important for the regularity of u in spaces like

$$W^{2,p}(\Omega)=\left\{v\,\middle|\,v,\frac{\partial v}{\partial x_i},\ \frac{\partial^2 v}{\partial x_i \partial x_j}\in L^p(\Omega)\right\}$$

for p large. For the study of this problem, a more "abstract" presentation is in order, always following the work of Brezis and Stampacchia.

6. AN ABSTRACT REGULARITY THEOREM. APPLICATION TO THE OBSTACLE PROBLEM

We consider the VI

$$\begin{aligned}&(Av,v-u)\geq(f,v-u),\quad \forall v\in K\\ &u\in K\end{aligned}\tag{46}$$

where $K\subset V$, and V is a Hilbert space. The situation extends to cases where A is non-linear and where V is a reflexive Banach space. Let us consider a space X such that

$$V\subset X\subset V'\tag{47}$$

with continuous embedding, each space being dense in the following one.

Example 1

$V=H_0^1(\Omega)$, $V'=H^{-1}(\Omega)$ and $X=L^p(\Omega)$, for p large enough. The problem considered by Stampacchia and Brezis is: when can we conclude that

$$Au\in X\,?$$

One introduces a *duality mapping* J from $X\to X'(V\subset X'\subset V')$, i.e. a (non-linear) mapping from $X\to X'$ such that

$$(J(u),u)=\|J(u)\|_{X'}\|u\|_X$$

$\|J(u)\|_{X'}$ is a strictly increasing function of $\|u\|_X$ and goes to $+\infty$ as $\|u\|_X\to\infty$.

Example 2
If $X = L^p(\Omega), J(u) = |u|^{p-2}u$.

If X is a Hilbert space that we identify with its dual $\left(X = L^2(\Omega)\right.$ in the example), then J = identity. The crucial hypothesis is now:

$$\begin{aligned} &\text{One can find a duality mapping } J \text{ from } X \to X' \text{ such that} \\ &\forall \varepsilon > 0 \text{ and } \forall u \in K, \text{ there exists } u_\varepsilon \text{ such that} \\ &u_\varepsilon \in K, \quad Au_\varepsilon \in X \text{ and } u_\varepsilon + \varepsilon JAu_\varepsilon = u \end{aligned} \tag{48}$$

One can then take $v = u_\varepsilon$ in (46) and using the properties of J one obtains that

$$\|Au\|_X \leq \text{ constant}$$

Hence, it follows that

$$Au \in X \tag{49}$$

Application to the problem.
We take $J(u) = J_p(u) = |u|^{p-2}u$ and we consider the equation (48), i.e. since

$$J^{-1} = J_{p'} \text{ and } \frac{1}{p} + \frac{1}{p'} = 1$$

then

$$Au_\varepsilon + J_{p'}\left(\frac{u_\varepsilon - u}{\varepsilon}\right) = 0 \tag{50}$$

We want to show that $u_\varepsilon \geq \psi$. We use the same technique as in section 5, i.e. we multiply by $(u_\varepsilon - \psi)^-$. We obtain

$$-a\left((u_\varepsilon - \psi)^-, (u_\varepsilon - \psi)^-\right) + \left(A\psi, (u_\varepsilon - \psi)^-\right) + \left(J_{p'}\left(\frac{u_\varepsilon - u}{\varepsilon}\right), (u_\varepsilon - \psi)^-\right) = 0 \tag{51}$$

But

$$\left(J_{p'}\left(\frac{u_\varepsilon - u}{\varepsilon}\right),\ (u_\varepsilon - \psi)^-\right) = -\frac{1}{\varepsilon^{p'-1}}\int_\Omega |u_\varepsilon - u|^{p'-2}\left[(u_\varepsilon - \psi)^-\right]^2 dx$$

$$-\frac{1}{\varepsilon^{p'-1}}\int_\Omega |u_\varepsilon - u|^{p'-2}(u-\psi)(u_\varepsilon - \psi)^- dx$$

so that (51) gives (since $A\psi \leq 0$):

$$\int_\Omega |u_\varepsilon - u|^{p'-2}\left[(u_\varepsilon - \psi)^-\right]^2 dx \leq 0 \tag{52}$$

Therefore, either $u_\varepsilon(x) - u(x) = 0$, and hence $u_\varepsilon(x) \geq u(x) \geq \psi(x)$, or $(u_\varepsilon(x) - \psi(x))^- = 0$, i.e. $u_\varepsilon(x) \geq \psi(x)$.

It will follow that, under reasonable assumptions, the solution u of the obstacle problem satisfies

$$u \in W^{2,p}(\Omega) \tag{53}$$

Simple one-dimensional examples show that one cannot obtain L^p estimates for higher-order derivatives. One can study the regularity in *Schauder* spaces; we refer the reader to the book of Kinderlehrer and Stampacchia[14] and to other chapters of this book. See also the report at the International Congress of Mathematicians, Vancouver, 1974, made by Kinderlehrer.

Remark Due to the physical interpretation of the obstacle problem, it is quite natural to consider the problem of *minimal surfaces with obstacle.* This has been considered by Nitsche[15] and by Giusti, [16] Giaquinta and Pepe, [17]and for *surfaces with mean curvature fixed,* it has been considered by Mazzone. [18]

[14] An introduction to variational inequalities and their applications. Academic Press, New York, 1980.

[15] J.C. Nitsche. "Vorlesungen uber Minimalflachen". Grundlehr. Math. Wiss., vol. 199, Springer, Berlin (1975).

[16] E. Giusti. "Minimal surfaces with obstacles". *CIME course on Geometric Measure theory and Minimal surfaces,* Rome,1973,pp. 119-53.

[17] M. Giaquinta and L. Pepe. "Esistenza e regolarità per il problema dell'area minima con ostacoli in *n* variabili". *Ann. Scu. Norm. Sup., Pisa,* 25 (1971), 481-507.

[18] S. Mazzone. "Un problema di disequazioni variazionali per superficie di curvatura media assegnata". *Boll.Unione Mat. Ital.,* 7 (1973), 318-29.

7. AN INEQUALITY FOR THE OBSTACLE PROBLEM

In the above proof of the regularity for the obstacle problem, the hypothesis '$A\psi \leq 0$' is much too restrictive. This can be overcome in several ways.

(1) One can introduce more flexibility in the abstract hypothesis of section 4 (see [48]); following Brezis and Stampacchia, one introduces families of operators $B_\varepsilon : V \to X, C_\varepsilon : V \to X'$, which are *bounded* as $\varepsilon \to 0$ and such that the equation

$$u_\varepsilon + \varepsilon J\left(Au_\varepsilon + B_\varepsilon u_\varepsilon\right) = u + \varepsilon C_\varepsilon u_\varepsilon \tag{54}$$

has a solution $u_\varepsilon \in K$ such that $Au_\varepsilon \in X$; one then obtains the conclusion (same proof) that

$$Au \in X \tag{55}$$

and this allows one to obtain regularity results similar to the above results under the assumption:

$$A\psi \text{ is a measure on } \bar{\Omega}; \quad \sup\left\{A\psi, 0\right\} \in L^p\left(\Omega\right) \tag{56}$$

(2) One can use penalty arguments.

(3) One can use an inequality given in Lewy and Stampacchia [9, 10] that we now explain.

Let u be the solution of the obstacle problem. Then one has

$$f \leq Au \leq \max\left\{A\psi, f\right\} \tag{57}$$

One does not restrict the generality in taking

$$f = 0 \tag{58}$$

(Indeed if ω is defined by $A\omega = f, \omega = 0$ on Γ, then it suffices to work on $u - \omega$ instead of u). Then one has to show that

$$0 \leq Au \leq \max\left\{A\psi, 0\right\} \tag{59}$$

Actually, one can obtain a more precise result [9].Let us introduce $\theta(s) = 1$

for $s \le 0, \theta(s) = 0$ for $s > 0$. Then there exists a unique function u in $W^{2,p}(\Omega)$ such that

$$\begin{aligned} &Au = \max\{A\psi, 0\}\theta(u-\psi) \quad \text{in } \Omega \\ &u = 0 \quad \text{in } \Gamma \end{aligned} \tag{60}$$

and u is the solution of the obstacle problem. (Of course (59) follows from (60).)

Proof: The proof of (60) is in two essential steps:

Step1 One considers an approximation of (60). Let $\theta_n(s)$ be a sequence of Lipschitz continuous functions approximating θ :

$$\theta_n(s) = \begin{cases} 1 & \text{if } s \le 0 \\ 1 - ns & \text{if } 0 \le s \le 1/n \\ 0 & \text{if } s > 1/n \end{cases} \tag{61}$$

One considers the equation

$$Au_n = \max\{A\psi, 0\}\theta_n(u_n - \psi), \quad u_n = 0 \quad \text{on } \Gamma \tag{62}$$

One proves that this equation has a solution by a fixed-point argument: given $\omega \in H_0^1(\Omega)$ one defines $\hat{u}$ as the solution of

$$A\hat{u} = \max\{A\psi, 0\}\theta_n(\omega - \psi) \tag{63}$$

One verifies that $\omega \to \hat{u} = T(\omega)$ maps a suitable ball Σ of $H_0^1(\Omega)$ into itself and that T is continuous. One has also, if $\max\{A\psi, 0\} \in L^p(\Omega)$, and if the coefficients of A are smooth enough, that $\hat{u} \in W^{2,p}(\Omega)$ One then verifies that the mapping T is compact from $\Sigma \to \Sigma$, and hence it has a fixed point, which is a solution u_n of (62).

One has also obtained in this manner that

$$u_n \text{ remains in a bounded set of } W^{2,p}(\Omega) \tag{64}$$

Step 2 The second step in the proof consists of proving that

$$u_n \geq \psi \tag{65}$$

Assume that one can find x_0 where $u_n(x_0) < \psi(x_0)$. Then one can find an open set G around x_0 such that

$$\begin{aligned} &u_n < \psi \quad \text{on } G \\ &\text{and} \\ &u_n = \psi \quad \text{on } \partial G \end{aligned} \tag{66}$$

Since on G one has $u_n < \psi$, equation (62) reduces to

$$Au_n = \max\{A\psi, 0\} \quad \text{on G}$$

hence

$$\begin{aligned} &A(u_n - \psi) \geq 0 \quad \text{on } G \\ &u_n - \psi = 0\mathbb{Z} \quad \text{on } \partial G \end{aligned} \tag{67}$$

Therefore, by the maximum principle, $u_n - \psi \geq 0$ in G, in contradiction to $u_n(x_0) < \psi(x_0)$. Hence (65) follows.

One can then pass to the limit in n, using (64) and (65), and one shows that u is the solution of the VI of the obstacle problem and that u satisfies (60).

Application to the regularity.

The application is obvious, and actually it is already implicitly contained in (64).

Remark A systematic use of inequality (57) or of similar inequalities with other boundary conditions, or with parabolic operators, is made in the work of Mosco, Troianiello, Joly, Hanouzet and others.

8. ELASTO-PLASTIC PROBLEM

Another important VI arises in the theory of elasto-plastic materials; the physical problem corresponds to dimension 2.

One considers the same bilinear form $a(u,v)$ as in section 3 and the convex set

$$K=\left\{v \mid v \in H_0^1(\Omega), \mid \nabla v(x) \mid \leq 1 \text{ a.e. in } \Omega\right\} \tag{68}$$

(a.e. = almost everywhere).

The following result is due to Brezis and Stampacchia [8] : if $f \in L^p(\Omega)$, the solution of the VI corresponding to (68) satisfies

$$Au \in L^p(\Omega) \tag{69}$$

One uses the idea of section 6. One has to consider then the equation

$$u_\varepsilon + \varepsilon J_p A u_\varepsilon = u \tag{70}$$

where u is given in $K, u_\varepsilon = 0$ on Γ and to show that it has a solution u_ε in K such that $Au_\varepsilon \in L^p(\Omega)$. In fact, if $u_\varepsilon \in K$ then it is bounded and (70) implies that $Au_\varepsilon \in L^\infty(\Omega)$. We write (70) in the form

$$Au_\varepsilon = J_{p'}\left(\frac{u-u_\varepsilon}{\varepsilon}\right) \tag{71}$$

and one has to show that there is a solution such that $\mid \nabla u_\varepsilon(x) \mid \leq 1$ a.e.

More generally, let $\lambda \to \theta(\lambda)$ be a strictly increasing function such that $\theta(0)=0$ and let us consider the equation:

$$Au = \theta(f-u), \qquad u \in H_0^1(\Omega) \tag{72}$$

Brezis and Stampacchia [8] show that this problem has a solution and that if $\mid \nabla f(x) \mid \leq 1$ a.e.,it follows that $u \in K$.

The proof rests on a comparison lemma and on several technical ideas in order to obtain estimates on ∇u; the authors consider first the case when Ω. is convex and then the general case.

This problem of elasto-plasticity has been the object of a large number of interesting works. Two of the main questions are as follows.

(i) Let us consider $\delta(x)$,which is the distance of x to Γ, and let us define

$$K_1 = \{v \mid v \in H_0^1(\Omega),\ |v(x)| \leq \delta(x)\} \tag{73}$$

Let us consider

$$a(u,v) = \int_\Omega \nabla u \nabla v \, dx + \lambda \int_\Omega uv dx \tag{74}$$

Let u (and, respectively, $\tilde{u}$) be the solution of

$$\begin{aligned} &a(u, v-u) \geq (f, v-u) \quad \forall\, v \in K \\ &u \in K \end{aligned} \tag{75}$$

and, respectively,

$$\begin{aligned} &a(\tilde{u}, v-\tilde{u}) \geq (f, v-\tilde{u}), \quad \forall\, v \in K_1 \\ &\tilde{u} \in K_1 \end{aligned} \tag{76}$$

Then, under suitable hypotheses on f and λ it has been proven by Brezis and Sibony[19] that

$$\tilde{u} = u \tag{77}$$

Their proof uses, in an essential way, an idea of Hartman and Stampacchia [4].

When Ω is multiply connected, the formulation of the elasto-plastic problem has to be slightly changed with respect to the above - one has to consider functions which are *constant* on the boundaries of the "holes" of Ω (see Lanchon[20]).

(ii) Another important question is connected with the *regularity of the free boundary,* that is the regularity of the boundary between the elastic region (where $|\nabla u(x)| < 1$) and the plastic region (where $|\nabla u(x)| = 1$). Let us refer the reader to Caffarelli and Friedman[21] and to the bibliography therein.

[19] H.Brezis and M.Sibony. "Equivalence de deux I.V."*Arch.Ration.Mech.Anal.,* 41 (1971), 254-65.

[20] H.Lanchon. *J.Mécanique,* 13 (1974), 267-320.

[21] L.A. Caffarelli and A. Friedman. "The free boundary for elastic plastic torsion problems". To appear.

9. HODOGRAPH METHOD AND VI

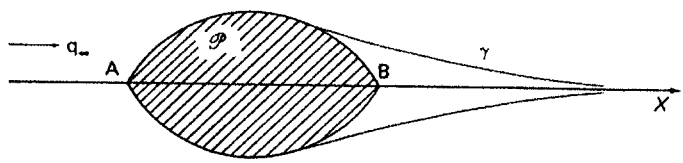
Towards the end of the 1960s, as the free-boundary problems solved by the technique of VI were becoming understood (with, of course, still many questions unanswered - in particular, of regularity - at that time), another type of question came into the picture: given a free-boundary problem arising from mathematical physics[22], when can it be formulated (and hopefully solved) by the technique of VI.

This type of question was raised, in particular, in problems of infiltration through porous media. It was observed in 1971 by Baiocchi[23] that by an appropriate transformation of the unknown function, it was possible to reduce, at least in some case, the problem of infiltration through porous media to a VI.

This idea gave rise to a large number of papers. Some of the most interesting among them are those of Brezis and Stampacchia [12, 13].

One considers in the plane x, y the flow of a perfect fluid (assumed to be steady, irrotational and incompressible) around a profile P, which is symmetric with respect to y. The flow is assumed to be uniform at infinity, i.e. if $\mathbf{q} = \{u, v\}$ denotes the velocity

$$\mathbf{q}(x,y) \to \{q_\infty, 0\} \quad \text{as } |x|+|y| \to \infty \tag{78}$$



(see the figure), one has

$$\text{div } \mathbf{q}=0, \qquad \text{rot } \mathbf{q}=0 \tag{79}$$

$$\mathbf{q} \text{ is tangential to } P \text{ along } \partial P \tag{80}$$

[22] We do not speak here of the free-boundary problems arising in the theory of optimal control.

[23] C. Baiocchi. "Sur un problème a frontière libre traduisant le filtrage de liquides à travers des milieux poreux". *C. R. Acad. Sci., Paris* 273 (1971), 1215-7.

There is a wake denoted by $\gamma;\gamma$ is a free boundary and along γ one has the two conditions

$$\begin{aligned} &\mathbf{q} \text{ is a tangent to } \gamma \\ &|\mathbf{q}| = q_\infty \end{aligned} \tag{81}$$

One introduces the stream function ψ by

$$u = \psi_{y,} \quad v = -\psi_x$$

Then

$$\Delta\psi = 0 \tag{82}$$

and by symmetry it suffices to consider the problem for $y > 0$.

Boundary conditions are (see the figure)

$$\psi = 0 \qquad \text{on M A B} \tag{83}$$

$$\psi = 0, \qquad |\nabla_\psi| = q_\infty \quad \text{on } \gamma \tag{84}$$

One considers the hodograph transformation

$$x, y \to u, v \to \theta, q$$

where

$$\tan\theta = v/u, \qquad q = |\mathbf{q}|$$

One considers ψ as a function of the independent variables

$$\theta \quad \text{and} \quad \sigma = -\log q \tag{85}$$

Then, one has

$$\Delta\psi = \frac{\partial^2\psi}{\partial\theta^2} + \frac{\partial^2\psi}{\partial\sigma^2} = 0 \tag{86}$$

In the hodograph transformation, the part of the boundary where one has only *one* condition (i.e. MAB) becomes a free boundary and the part of the boundary where one has *two* conditions (i.e. γ) becomes *known* (in fact, a part of the σ axis) and one ends up with a problem of the following nature.

Let Ω be an open set in $\mathbb{R}^n$; find $D \subset \Omega$ and a function ψ defined in D such that

$$-\Delta\psi = \varphi \quad \text{on } D \quad (\varphi \text{ given in } \Omega) \tag{87}$$

$$\psi \text{ satisfies a standard boundary condition on } \partial D \cap \partial\Omega \tag{88}$$

$$\begin{aligned} &\psi = 0 \text{ on } S = \partial D \cap \Omega \\ &\partial\psi/\partial\nu = \pi \cdot \nu \text{ on } S \end{aligned} \tag{89}$$

where π is a given vector field in Ω, ν is the normal to S extended to D; S is the free boundary (in the "hodograph" plane).

By a transformation of an unknown function of the type of that introduced by Baiocchi, one can reduce - at least in some case - this type of problem to a VI (see Brezis and Stampacchia [12,13,19]).

The analogous problem in a finite strip has been solved by a research student of Stampacchia[24].

10. FOURTH-ORDER VI

In a paper [17] with Brezis, Stampacchia studied the regularity of fourth-order VI; a very simple remark shows that - essentially - one cannot go farther than third-order derivative estimates. Indeed, if one defines

$$K_1 = \left\{ v \mid v \in H_0^1(\Omega) \cap H^2(\Omega), \alpha \le \Delta v \le \beta \right\}$$

where α and β are constants, $\alpha < 0 < \beta$, and if one considers the VI

$$\begin{aligned} &\left(\Delta u, \Delta(v-u)\right) \ge (f, v-u) \quad \forall v \in K_1 \\ &u \in K_1 \end{aligned} \tag{90}$$

then, if $f \in L^2(\Omega)$, one has $u \in W^{3,p}(\Omega)$ for every p finite but "nothing

[24] F. Tomarelli. *Graduation Thesis,* Scuola normale Superiore, Pisa (1978).

better". This is immediate: let us introduce F by

$$\Delta F = f, \quad F \in H_0^1(\Omega)$$

If we set $\Delta u = \hat{u}, \Delta v = \hat{v}$, then (90) becomes equivalent to

$$(\hat{u}, \hat{v} - \hat{u}) \geq (F, \hat{v} - \hat{u}) \quad \forall \hat{v}$$

$$\alpha \leq \hat{v} \leq \beta \quad \text{and} \quad \alpha \leq \hat{u} \leq \beta$$

Then, if $\lambda \to P(\lambda)$ denotes the projection $R \to [\alpha, \beta]$,we have

$$\hat{u} = P(F)$$

so that $u \in W^{1,\infty}(\Omega)$ and u has *exactly* the regularity properties given above. If one considers instead of K_1 the convex set K_2 given by

$$K_2 = \{v \mid v \in H_0^2(\Omega),\ \alpha \leq \Delta v \leq \beta\}$$

and if we denote by u the solution of the VI (90) where K_1 is replaced by K_2, then, again one has the same regularity result, namely, $u \in W^{3,p}(\Omega)$ for every p finite and essentially "nothing better". The proof consists of showing that there exists $z \in L^1(\Omega)$, such that $\Delta z = 0$ and such that $\hat{u}$ (with the same notations as above) can be represented by

$$\hat{u} = P(F + z)$$

For the "obstacle problem", i.e. the same problem with K_1 or K_2 replaced by

$$K_3 = \{v \mid v \geq \phi, v \in H_0^2(\Omega)\}$$

it has been shown by Frehse[25] that the corresponding solution belongs to $H^3_{\text{local}}(\Omega)$ (assuming that ϕ is smooth), a result which has been recently

[25] J. Frehse. *Hamburg Univ. Math. Sem., Abhard.*, 36 (1971),140-9.

improved by Caffarelli and Friedman[26] (these authors also study the regularity of the free boundary).

Further remarks concerning this problem with the convex set K_2 can be found in Torelli[27].

11. INFILTRATION IN POROUS MEDIA

We now briefly report on a posthumous work [20] with Brezis and Kinderlehrer on infiltration in porous media. This type of problem has been studied in particular by Baiocchi[28] and by Alt[29]. The method introduced in reference [18] is, roughly speaking, as follows. We are given an open set $\Omega \subset \mathbb{R}^2$ with boundary $\partial\Omega$ which consists of three parts: $\partial\Omega = S_1 \cup S_2 \cup S_3$; we want to find $p \in H^1(\Omega), p \geq 0$, and $g \in L^\infty(\Omega)$,such that

$$\begin{aligned} &g = 1 \qquad \text{if } p > 0 \\ &g \in [0,1] \qquad \text{if } p = 0 \end{aligned} \tag{91}$$

$$p = \text{given function on } S_2 \cup S_3 \tag{92}$$

$$\int_\Omega \left(\nabla p \nabla \zeta + g \frac{\partial \zeta}{\partial y} \right) \mathrm{d}x\, \mathrm{d}y \leq 0, \quad \forall \zeta \in Z \tag{93}$$

where Z is defined as follows :

$$Z = \left\{ \zeta \mid \zeta \in \mathrm{C}^1(\bar{\Omega}), \zeta \geq 0 \text{ on } S_2, \zeta = 0 \text{ on } S_3 \right\}$$

In order to prove that there exists $p \in W^{1,s}_{\mathrm{loc}}(\Omega)\ \forall$ finite s, and that there exists $g \in L^\infty(\Omega)$ such that (91), (92) and (93) are true, the authors in [18] introduce the following approximation procedure (of the penalty type). Let

[26] LA. Caffarelli and A. Friedman. "The obstacle problem for the biharmonic operator". Ann. Scuola Normale Superiore, Pisa, Classe Scienze, (IV), 6, 1979, pp. 151-184.

[27] A. Torelli. "Some regularity results for a family of variational inequalities". Ann. Scuola Normale Superiore, Pisa, Classe Scienze, (IV), 6 (1979), pp. 497-510.

[28] C. Baiocchi. *C. R. Acad. Sci., Paris,* 278 (1974), 1201-4: See also the book of Baiocchi and Capelo. Disequazioni Variazionali e Quasi variazionali. Applicazioni a problemi di frontiera libera.Published by Unione Matematica Italiana, Univ. of Bologna (1978).

[29] H. W. Alt. *Arch Ration. Mech. Anal.,* 64 (1977), 111-26.

$H_\varepsilon(\lambda)$ be defined by

$$H_\varepsilon(\lambda)=\begin{cases}1 & \text{if } \lambda \geq \varepsilon \\ \lambda\varepsilon & \text{if } 0 \leq \lambda \leq \varepsilon \\ 0 & \text{if } \lambda \leq 0\end{cases} \tag{94}$$

and let us consider the problem of finding $p_\varepsilon \in H^1(\Omega)$ such that

$$p_\varepsilon = \text{given on } S_2 \cup S_3 \qquad \text{(same values as in (92)} \tag{95}$$

$$\int_\Omega \left(\nabla p_\varepsilon \nabla \zeta + H_\varepsilon(p_\varepsilon) \frac{\partial \zeta}{\partial y} \right) dx\, dy = 0, \quad \forall\, \zeta \in H^1(\Omega)$$

$$\text{such that} \quad \zeta = 0 \quad \text{on } S_2 \cup S_3 \tag{96}$$

The authors prove (i) that (95) and (96) have a unique solution; and (ii) that p_ε converges, as $\varepsilon \to 0$, to a solution of the problem. (The uniqueness of p is an open question, except in particular cases, solved by Caffarelli and Rivière[30].

The existence in (95) and (96) follows easily from Schauder's fixed-point theorem.

For the uniqueness, if p_ε and $\hat{p}_\varepsilon$ are two solutions, one introduces

$$q = p_\varepsilon - \hat{p}_\varepsilon$$

and it is enough to prove that $q \leq 0$ (since then, by exchanging the roles of p_ε and $\hat{p}_\varepsilon, q = 0$)

One verifies that

$$\left| \int_\Omega (\text{grad } q)\ (\text{grad } \zeta)\ dx\, dy \right| \leq L \int_\Omega |q|\, |\zeta_y|\, dx\, dy \tag{97}$$

where L depends on ε (but ε is fixed for the time being).

Then one chooses, for any $\delta > 0$,

[30] L. Caffarelli and N. Rivière, "Existence and uniqueness for the problem of filtration through a porous media". Notices A.M.S., 24, A-576.

$$\zeta = \frac{1}{q}(q-\delta)^+$$

and after some computations, one shows that this implies (97).

12. CONCLUSION

In the above report, we have not spoken of a number of related works. Let us briefly cite some further related work not mentioned in the above report: on the regularity of solution of VI [20,21]; on the obstacle problem with the obstacles irregular [22] (work with A. Vignoli); or when the boundary conditions are of mixed type [23] (with V. Murthy). Stampacchia was also interested in the numerical aspects of the solution of VI as shown in [24,25].

For many years, he wanted to write a book, one which would be an introduction to VI and would also carry the reader close to the frontiers of research. Work on this book was undertaken in collaboration with D. Kinderlehrer in July 1976; this book was nearly completed at the time of his death [26].

Guido Stampacchia has left to us a beautiful example of a mathematician, working with very good taste on strongly motivated problems; he introduced elegant abstract methods but only when necessary and without artificial generality; he introduced, and masterfully used, some techniques which already belong to the classical tools of analysis.

REFERENCES

This is *not* a complete bibliography of the works of Guido Stampacchia, for which we refer the reader to the biography written by E. Magenes *(Boll. Unione Mat. Ital., 1978),* but rather a list of his majorcontributions in the domain of variational inequalities.

[1] "Formes bilinéaires coercitives sur les ensemble convexes". *C R. Acad. Sci., Paris,*258 (1964), 4413-6.

[2] "Equations elliptiques du second ordre à coefficients discontinus". *Sem. J. Leray, College de France* (1963/64).

[3] "Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus". *Ann. Inst. Fourier,* 15 (1965), 189-258.

[4] with Ph. Hartman. "On some non linear elliptic differential-functional equations".*Acta Math.,* 115 (1966), 271-310.

[5] with J. L. Lions. "Inéquations variationnelles non coercives". *C. R. Acad.Sci., Paris,*261 (1965),25-27.
[6] with J. L. Lions. "Variational Inequalities". *Commun. Pure & Appl. Math.,*20 (1967),493-519.
[7] with H. Lewy. "On existence and smoothness of solutions of some non-coercive variational inequalities". *Arch. Rat. Mech. Anal.,* 41 (1971), 241-53.
[8] with H. Brezis. "Sur la regularité de la solution d'inéquations elliptiques".*Bull.Soc.Math. Fr.,* 96 (1968),153-80.
[9] "Variational inequalities". In *Theory and Applications of Monotone operators*; *Proc. NATO Adv. Study Inst.* Venice, June 1968.
[10] with H. Lewy. "On the regularity of the solution of a variational inequality".*Commun. Pure & Appl. Math.,* 22,(1969), 153-88.
[11] with H. Brezis and L. Nirenberg. "A remark on Ky Fan's minimax principle". *Boll. Unione Mat. Ital.,* 6 (1972), 293-300.
[12] with H. Brezis. "Une nouvelle méthode pour l'étude d'écoulements stationnaires". *C R. Acad. Sci.,Paris,* 276 (1973),129-32.
[13] with H. Brezis. "The hodograph method in fluid dynamic in the light of variational inequalities". *Arch. Ration Mech. Anal.,* 61 (1976), 1-18.
[14] with H. Lewy. "On the smoothness of superharmonics which solve a minimum problem". *J. Anal. Math.,* 23 (1970), 227.36.
[15] "On the filtration of a fluid through a porous medium with variable cross section" .*Russ. Math. Surv.,* 29 (1974), 89-102.
[16] with D. Kinderlehrer. "A free boundary problem in the plane". *Ann. Inst. Fourier,*XXV (1975), 323-44.
[17] with H. Brezis. "Remark on some fourth order variational inequalities". *Ann. Scu. Norm. Sup., Pisa,* 4 (1977), 363-71.
[18] with H. Brezis and D. Kinderlehrer. "Sur une nouvelle formulation du problème de l'écoulement à travers une digue". *C R. Acad. Sci., Paris,* (1978).
[19] with H. Brezis. "Problèmes elliptiques avec frontière libre" . *Sem. Goulaouic Schwartz,*December 1972.
[20] with H. Brezis and D. Kinderlehrer. "On the regularity of solutions of VI." *Proc. Int.Con/. on Functional Analysis and Related Topics,* Tokyo, 1969, 285-9.
[21] "Regularity of solutions of some V.I.". In *Non linear Functional Analysis; Proc. Symp. Pure Math.,* 18, part I (1970), 271-81.
[22] "A remark on V.I. for a second order non linear differential operator with non Lipschitz obstacles". *Boll. Unione Mat. Ital.,* 5 (1972),123-31.
[23] "A V.I. with mixed boundary conditions". *Isr. J. Math.,* 13 (1972), 188-334.
[24] "On a problem of numerical analysis connected with the theory of V.I.". *Symp.Math.,* **IO** (1972),281-93.
[25] *"Programmazione convessa e disequazioni variazionali"*, Ist. di Calcolo delle Probabilità, Univ. Roma, (1973).
[26] with D. Kinderlehrer. *An Introduction to Variational Inequalities and Their Applications.* Academic Press, New York, 1980.

IN MEMORY OF GUIDO STAMPACCHIA

M.G. Garroni
Dept. of Mathematics, University of Rome "La Sapienza", Rome, Italy

Dear friends and colleagues, I am happy to have the occasion to take part in this international Congress in memory of Guido Stampacchia, that starts today.

Many of you, at least those closer to him, have come here not only to honour the memory of a great mathematician, but also in the name of the friendship and the affection for a friend and, in the case of the younger participants, for a real master. The personality of Stampacchia was both strong and simple, open and helpful. His human qualities are well known to all of you. I do not want to recall them with words that may sound conventional. All the participants here, who more, who less, have had the occasion to experiment, know and appreciate them, both within the common mathematical activity and outside the scientific work. I know that everybody shares my feelings at the only mention of them. Our being here together is already an implicit commemoration, more effective then any speech, and a testimony of admiration and of sincere unchanged affection.

What I just mentioned well explains the nature of this Congress: an International Conference and at the same time, how to say?, an almost

familial gathering. This is a gathering of the scholars that have worked with Guido Stampacchia, who have shared his directions of research, who have known him well and not only superficially, who have been his students. In this Congress there is only one conference dedicated to the aspects of Stampacchia's own mathematical work, the one given by Professor Magenes. This is just as Stampacchia would have liked. The rest of the congress consists of scientific communications mostly dealing with the continuation of his research. This is the best commemoration for a man who has devoted his entire life to research.

On my side, I only want to add a personal recollection. The place: Pisa. The time: the year 1966. I had just come out from a difficult period in Rome, due to a worsening of human relationships that had rendered my activity at the University unbearable. (I was not responsible for that situation, but that is not important.) I was tired and discouraged, and also exhausted from the birth of my younger daughter Adriana. Well, at that time Ennio De Giorgi, a close friend of mine from the years of our common University studies, introduced me to Professor Stampacchia. That was a decisive acquaintance: Stampacchia encouraged me not to leave the University, a decision that I was on the verge of taking, and to go back to scientific work.

I then started working again, this time under his direction. He was satisfied with my work, and gave me his full scientific and personal support. Until Stampacchia moved to the University of Rome, I went to Pisa at least once a month. On those occasions I had the chance to know people such as Nirenberg, Levy, Brézis, Kinderlehrer, and other important mathematicians that used to visit Guido Stampacchia. My scientific life (and not only scientific) changed completely.

In my opinion, this is not a small reason for my gratitude towards a master.

Thank you very much.

THE COLLABORATION BETWEEN GUIDO STAMPACCHIA AND JACQUES-LOUIS LIONS ON VARIATIONAL INEQUALITIES

E. Magenes
Dept. of Mathematics, University of Pavia, Italy

It is a motive of deep emotion for me to recall the collaboration between Stampacchia and Lions, since they were among my dearest friends; with each of whom I had the fortune of working together scientifically. The friendship with Stampacchia began in 1941, when we met as students at Scuola Normale Superiore in Pisa. As for Lions, the first time the three of us met was at Nice during the "I Congrès International des Mathématiciens d'expression latine", held during September 1957. Since then, we have met very many times; also in Erice in 1971 (June 27-July 7)

Concerning the collaboration between Stampacchia and Lions on Variational Inequalities, I consider here the classic paper [10] (announced essentially in [9]). But also after this paper their exchange of ideas and their discussions have been continuous, during their frequent meetings; a synthetic exposition is the lecture "The Work of G.Stampacchia in Variational Inequalities" delivered by Lions in Erice, during the International Conference held in June 19-30, 1978, published in [5] and then also in [7] (see, moreover, the paper by S. Mazzone in this Volume [11]).

Let V be a Hilbert space over $\mathbb{R}$; let a(u,v) be a continuous bilinear form on V; let K be a closed convex subset of V and finally let f be an element of the dual space V' (dual space of V). The problem studied in [10] is the following one:

Problem 1: find $u \in K$ which satisfies the "Variational Inequality":

$$a(u, v-u) \geq \langle f, v-u \rangle , \quad \forall v \in K , \tag{1}$$

where <.,.> denotes the duality pairing between V' and V.

It is well known that, if a(u,v) is also symmetric (i.e. a(u,v)=a(v,u), $\forall u, v \in V$), then (1) is equivalent to the following minimization problem:

$$\frac{1}{2}a(u,u) - < f,u > = \min_{v \in K} \{\frac{1}{2} a(v,v) - < f,v >\}$$

Moreover, if a(u,v) is again symmetric and also *coercive* on V, i.e.

$$a(v,v) \geq \alpha \| v \|^2 , \quad \forall v \in V \tag{2}$$

(α, a positive constant, $\| \ \|$ denotes the norm in V), then Problem 1 has a unique solution.

Stampacchia had long been interested in Calculus of Variations, since he was student of L. Tonelli in Pisa; and he was able to prove that Problem 1 has one and only one solution even if a(u,v) is coercive, but not necessarily symmetric, and to deduce some important applications to the theory of "capacitary potential" theory (see [12], [13]).

In the paper [10] Stampacchia and Lions try to study the existence of at least one solution of Problem 1; to this end, they replace the hypothesis (2) with the more general condition:

$$a(v,v) \geq 0 , \ \forall v \in V. \tag{3}$$

Such a condition had been inspired by the papers of G. Fichera about the "Signorini Problem" in elasticity theory ([6], [7]), where the related bilinear form a(u,v) is symmetric and satisfies (3), but not (2).

The main results of [10] can be summarized as follows (the hypothesis being still that a(u,v) is bilinear continuous on V, satisfies (3) and K is a closed and convex set of V):

1°) the set X of all solutions of Problem 1 is a (possibly empty) closed and convex set;

2°) approximation of the set X by regularizations: $\forall \varepsilon > 0$, let us consider the problem which consits in finding $u_\varepsilon \in K$, such that:

$$a(u_\varepsilon, v - u_\varepsilon) + \varepsilon\beta(u_\varepsilon, v - u_\varepsilon) \geq < f + \varepsilon g, v - u_\varepsilon >, \ \forall v \in K, \tag{4}$$

where $\beta(u,v)$ is a bilinear continuous coercive form on V and g a fixed element of V': problem (4) has one and only one solution $u_\varepsilon \in K$, since $a(u,v)+\varepsilon\beta(u,v)$ is coercive on V. Stampacchia and Lions showed that

$$u_\varepsilon \text{ stronghly converges in V to } u_0 \text{ as } \varepsilon \text{ tends to zero,} \tag{5}$$

where u_0 is the unique solution in X of the inequality

$$\beta(u_0, v-u_0) \geq \langle g, v-u_0 \rangle, \quad \forall v \in X; \tag{6}$$

3°) if moreover X is also bounded, then Problem 1 has at least one solution;

4°) semicoercive forms: let us assume that the norm $\|v\|$ of v is equivalent to $p_0(v)+p_1(v)$ where $p_0(v)$ is a norm in V, with respect to which V is a pre-Hilbert space and $p_1(v)$ is a seminorm on V, the space $Y=\{v \in V, p_1(v)=0\}$ has a finite dimension, and there exists a constant c_1, such that:

$$\inf_{y \in Y} p_0(v-y) \leq c_1 p_1(v).$$

Moreover, let a (u,v) be a continuous bilinear form on V which is semicoercive, i.e.

$$a(v,v) \geq c_2 (p_1(v))^2, \ \forall v \in V, \quad (c_2 \text{ positive constant});$$

let K be a closed and convex subset of V containing $\{0\}$; finally, let $f \in V'$ be such that $f = f_0 + f_1$, with $f_0 \in V', f_1 \in V'$ and satisfying, if $Y \cap K \neq \{0\}$, the following conditions:

$$\langle f_0, y \rangle < 0, \quad \forall y \in Y \cap K, \ y \neq 0$$
$$|\langle f_1, y \rangle| \leq c_3 p_1(v), \ \forall v \in V \qquad (c_3 \text{ positive constant}).$$

Then, there exists at least one solution of Problem 1.

The proofs of the results 1°), 2°), 3°), 4°) are given in the Sections 3,4,5 of [10]. In Section 6 of [10] some examples of applications to different problems for elliptic partial differential equations are given. In particular, as an application of 4°), Section 6 contains an example, which is related to the above mentioned "Signorini Problem" (for this, besides Fichera's papers [6] and [7], I suggest to refer, for more details, to the papers [1] by C. Baiocchi,

[2] by C. Baiocchi-F. Gastaldi-F. Tomarelli and [3] by C. Baiocchi-G. Buttazzo-F. Gastaldi-F. Tomarelli).

Finally, it seems to me important to point out also Section 7 of [10], in which for the first time the "Evolution Variational Inequalities" are introduced. The starting point is the following one: more in general (with respect to Problem 1), we can consider a bilinear form $a(u,\varphi)$ defined for $u \in V$ and $\varphi \in \Phi$, where Φ is a Hilbert space strictly contained in V (then a(u,u) has no meaning $\forall u \in V$). If now K is a closed subset of V and $u,\varphi \to a(u,\varphi)$ is a bilinear continuous form on $V \times \Phi$, the problem, analogous to Problem 1, is the following one: to find $u \in K$, such that:

$$a(u,\varphi - u) \geq \langle f,\varphi - u \rangle, \quad \forall \varphi \in \Phi \cap K. \tag{7}$$

The main example of this situation - studied in Section 7 of [10] - is a *Parabolic Evolution Inequality*, which we will describe here, taking for simplicity the heat operator:

$$\frac{\partial u}{\partial t} - \Delta u + cu \quad (c > 0).$$

Let Ω be an open, bounded, regular subset of $\mathbb{R}^n$, and V and Φ the spaces

$$V = \{v : v \in L_2(0,+\infty ; H^1(\Omega)), v'(=\frac{\partial v}{\partial t}) \in L_2(0,+\infty ; H^{-1}(\Omega)), v(0) = 0\} \tag{8}$$

and

$$\Phi = \{\varphi \in L_2(0,+\infty ; H^1(\Omega)), \varphi(0) = 0\}, \tag{9}$$

which are Hilbert spaces equipped with the obvious scalar products. Moreover let $\Sigma = \partial\Omega \times]0,+\infty[$ be the lateral boundary of $\Omega \times]0,+\infty[$; for $v \in V$ we can find the "trace" $v|_\Sigma$ (and we have:

$$v|_\Sigma \in L_2(0,+\infty; H^{1/2}(\partial\Omega)).$$

Let us take

$$K = \{v \in V : v|_\Sigma = 0\}, \tag{10}$$

which is a convex and closed subset of V. Then Stampacchia and Lions prove that, if $f \in L_2(\Omega)$, then there exists one and only one $u \in K$, solution of the Parabolic Variational Inequality:

$$\int_0^{+\infty} \int_\Omega \{\nabla u \nabla(\varphi - u) + cu(\varphi - u)\} dxdt - \int_0^{+\infty} \int_\Omega u\varphi dxdt$$
$$\geq \int_0^{+\infty} \int_\Omega f(\varphi - u) dxdt, \quad \forall \varphi \in \Phi \cap K. \tag{11}$$

where ∇v is the gradient of v with respect to the x variables.

Moreover, they give an interpretation of (12), proving that the solution u of (12) defines a "weak" solution of the parabolic equation:

$$\frac{\partial u}{\partial t} - \Delta u + cu = f \qquad \text{in } \Omega \times]0, +\infty[, \tag{12}$$

which satisfies the initial condition u(x,0)=0 and the "unilateral condition" on Σ:

$$u \geq 0, \frac{\partial u}{\partial \upsilon} \geq 0, u.\frac{\partial u}{\partial \upsilon} = 0 \quad (\upsilon \text{ exterior normal on } \partial\Omega).$$

This first result has been the starting point for further developments of the variational Evolution Inequalities due to H. Brezis in several important papers (see mainly [5]) and others.

REFERENCES

[1] Baiocchi C, "Disequazioni variazionali non coercive (non-coercive Variational Inequalities)". Atti Convegni Accademia Lincei, Vol.77,1986,pp. 153-157.

[2] Baiocchi C., Gastaldi G. and Tomarelli F., "Some existence results on non-coercive Variational Inequalities". Annali Scuola Normale Superiore, Vol.XIII,1986,pp. 617-659.

[3] Baiocchi C., Buttazzo G., Gastaldi F. and Tomarelli F., "General existence results for unilateral problems in Continuum Mechanics". Archivies of Rational Mechanics and Analysis, Vol. 100,No. 2,1988, pp.148-189.

[4] Brézis H., "Problemes unilateraux". Jour. Mathématique Pure et Appliquée, Vol.9,No. 51,1972,pp. 1-168.

[5] Cottle R. W., Giannessi F. and Lions J.-L, (Eds.), "Variational Inequalities and Complementarity Problems.Theory and Applications". John Wiley, Chichester, 1980.

[6] Fichera G., "Sul problema elastostatico di Signorini con ambigue condizioni al contorno (On the Signorini elastostatic problem with ambiguous boundary conditions)". Rendiconti Accademia Lincei, serie 8, Vol. 34,1963,pp. 138-142.

[7] Fichera G., "Problemi elastostatici con vincoli unilaterali: il problema di Signorini con ambigue condizioni al contorno (Elastostatic problems with unilateral constraints: the Signorini problem with ambiguous boundary conditions)". Memorie Accademia Lincei, serie 9, Vol. 7,1964,pp. 91-140.

[8] Lions J.-L., "The work of Stampacchia in Variational Inequalities". Bollettino Unione Matematica Italiana, serie 5, Vol. 15-A,1978,pp. 22-39.

[9] Lions J.-L.and Stampacchia G., "Inequation Variationnelles non coercives", Comptes Rendus Accademie Sciences, Paris, Vol.261,1965, pp.25-27.

[10] Lions J.-L.and Stampacchia G., "Variational Inequalities". Communications on Pure and Applied Mathematics,Vol. XX,1967,pp. 493-519.

[11] Mazzone S., " Guido Stampacchia". This volume.

[12] Stampacchia G., "Formes bilineaires non coercives sur les ensembles convexes". Comptes Rendus Accademie Sciences Paris,Vol. 258,1964,pp. 4413-4416.

[13] Stampacchia G., "Le problème de Dirichlet pour les equations elliptique". Annales Institut Fourier, Vol. XV,1965,pp. 189-258.

IN MEMORY OF GUIDO STAMPACCHIA

O.G. Mancino
Dept. of Applied Mathematics, University of Pisa, Pisa, Italy

Certainly, the remarkable human qualities, the successful academic career and the important scientific production of Guido Stampacchia are well known, especially to the national and international mathematical community. However, I don't think superfluous to recall briefly who he was and what he did, so that we all can reflect on the greatness of the Man we lost, on the honor we had in knowing him and how he enriched us both from a human and a scientific point of view.

It is therefore with deep love and sincere gratitude that I am going to remember Guido Stampacchia as a master and a scientist. Because someone else, more qualified than I am in the fields cultivated by Stampacchia, will talk about his scientific work and the influence it has in the present days, it will be enough for me to outline his academic career, sketch the main subjects of his research activity and recall some stages of my collaboration with him.

Stampacchia was born on March 26, 1922 in Napoli, the town he always loved. Besides in Pisa where, as a student at the Scuola Normale Superiore he had Leonida Tonelli as a master, Stampacchia completed his University studies in Napoli, where he took his degree in Mathematics in November 1944 and was a pupil of Renato Caccioppoli and Carlo Miranda. He was an assistant professor since January 1946 until December 1952, firstly at the Istituto Navale and then at the University.

In December 1952 he became Appointed Professor of Mathematical Analysis at the University of Genova and in November 1960 he was called by the University of Pisa, where he taught at the Faculties of Science and Engineering. In November 1968, he went to Rome, where he was professor at the Faculty of Science of the University "La Sapienza" and Director of the Istituto per le Applicazioni del Calcolo "M. Picone" of the National

Research Council. Finally, on November 1970, he returned to Pisa as Professor at the Scuola Normale Superiore.

Stampacchia was fellow of the Unione Matematica Italiana since 1948, member of the Società Italiana di Scienze, Lettere ed Arti of Napoli since 1954 and he won the Feltrinelli Award in 1966. Besides, he was corresponding fellow of the Accademia Nazionale dei Lincei since 1968, member of the Scientific Commission of the U.M.I. from 1964 to 1976 and President of the U.M.I. from 1967 to 1973.

Stampacchia was member of the managing board of the International Summer Center for Mathematics and of the editorial board of many scientific reviews, among which "Advances in Mathematics", "Applied Mathematics and Optimization", "Calcolo" and the "Annali della Scuola Normale Superiore". Since 1974 he was the Director of this last review, as well.

Stampacchia has left an indelible trace in the mathematical world with his 86 publications. Because of the originality, the depth and the importance of his scientific contributions, Stampacchia was invited to deliver lectures in many international Congresses, and courses or seminars in many Italian and foreign Universities.

His scientific work is mainly concerned with boundary-value problems in ordinary differential equations; the calculus of variations of multiple integrals and its connection with partial differential equations; variational inequalities and their application to problems of mathematical physics and numerical analysis. Enrico Magenes and Jacques Louis Lions have summarized, in a masterly way, the scientific work of Guido Stampacchia into two articles appeared in 1978 on Bull. U.M.I.. Therefore, I'll limit myself to personal memories.

***

I first met Stampacchia in Pisa during the spring of 1965. I was a researcher at the Centro Studi Calcolatrici Elettroniche. In that occasion, Stampacchia told me about his note "Formes bilinéaires coercitives sur les ensembles convexes", published on the C. R. Acad. Sc. Paris in 1964. His opinion was that it would be possible to infer from the paper a method for solving systems of equations, and he asked me to realize the method.

So I did, and my results appeared in two papers: "Sui sistemi lineari contenenti un parametro", published on Calcolo in 1966, and "Resolution by Iteration of Some Nonlinear Systems", appeared on the Journal of the A.C.M. in 1967. In the first paper, I consider a system of linear equations with a coefficient matrix of the form $A+tB$, where A and B are $n\times n$ matrices, A symmetric and positive definite, B skew. Then, I present an

iterative method for solving a sequence of such systems, generated by assigning discrete increasing real values to the parameter t starting with $t=0$.

In the second paper, I consider a system of nonlinear equations satisfying certain conditions. Then I prove that such a system admits a unique solution, present an iterative method for solving it, and find a formula which gives an excellent starting point.

***

In November 1968 Stampacchia went to Rome, but our collaboration continued. Our main results appear in the joint paper "Convex Programming and Variational Inequalities", published on J.O.T.A. in 1972. The following results, obtained in this paper, are worth of being considered.

Let $x=(x_1,x_2,...,x_n)$ be a generic point of n-dimensional Euclidean space $\mathbf{R}^n$, (x,y) the scalar product in $\mathbf{R}^n$, $\mathbf{K}\neq\varnothing$ a closed, convex subset of $\mathbf{R}^n$, $\partial\mathbf{K}$ the boundary of $\mathbf{K}$ and $F(x)$ a continuous mapping from $\mathbf{R}^n$ into $\mathbf{R}^n$. The mapping F is said *monotone* if:

$$(F(x')-F(x''),x'-x'')\geq 0 \tag{1}$$

for every pair of points $x',x''\in\mathbf{R}^n$, and *strictly monotone* if in relation (1) equality holds when and only when $x'=x''$.

Given x, the equation:

$$(F(x),y-x)=0 \tag{2}$$

represents a hyperplane if $F(x)\neq 0$. A hyperplane passing through $x\in\partial\mathbf{K}$ is called a *supporting plane* for $\mathbf{K}$ in x, if it leaves the whole convex set $\mathbf{K}$ on the same part.

Let $f(x)$ be a continuously differentiable convex function in $\mathbf{R}^n$. The gradient of $f(x)$ is a continuous, monotone mapping of $\mathbf{R}^n\to\mathbf{R}^n$. The problem of finding a point $x_K\in\mathbf{K}$ such that:

$$f(x_K)=\min_{x\in K} f(x) \tag{3}$$

is equivalent to that of finding a point $x_K\in\mathbf{K}$ for which:

$$(\operatorname{grad} f(x_K),x-x_K)\geq 0 \qquad \forall x\in\mathbf{K}, \tag{4}$$

which is a particular case of the more general problem of finding a point x such that:

$$x \in \mathbf{K}, \qquad (F(x), y - x) \geq 0 \qquad \forall y \in \mathbf{K}. \tag{5}$$

The relation (5) is called a *variational inequality* and any point satisfying it is called a *solution* of the variational inequality. We have the following results.

Theorem 1 *If $F(x)$ is strictly monotone in $\mathbf{K}$, then there exists at most one solution of the variational inequality (5).*

Theorem 2 *If a solution x of (5) is an interior point of $\mathbf{K}$, then (5) is equivalent to $F(x) = 0$.*

Supposing that $F(x)$ is strictly monotone in $\mathbf{R}^n$ and that for every non-empty, closed, convex subset of $\mathbf{R}^n$, the variational inequality (5) admits a solution, we have the following

Theorem 3 *Let $\mathbf{K}_1$ and $\mathbf{K}_2$ be two non-empty, closed and convex subsets of $\mathbf{R}^n$ such that $\mathbf{K}_1 \subset \mathbf{K}_2$, and let x_{K_1} and x_{K_2} be the two solutions of the corresponding variational inequalities, i.e.:*

$$\begin{aligned} x_{K_1} \in \mathbf{K}_1 \qquad & (F(x_{K_1}), y - x_{K_1}) \geq 0 \qquad && \forall y \in \mathbf{K}_1 \\ x_{K_2} \in \mathbf{K}_2 \qquad & (F(x_{K_2}), y - x_{K_2}) \geq 0 \qquad && \forall y \in \mathbf{K}_2. \end{aligned}$$

Then, if x_{K_1} is interior to $\mathbf{K}_1$, we have: $x_{K_2} = x_{K_1}$. If otherwise $x_{K_1} \in \partial \mathbf{K}_1$, the hyperplane $(F(x_{K_1}), y - x_{K_1}) = 0$ separates x_{K_2} from $\mathbf{K}_1$, i.e.:

$$\left(F\left(x_{K_1}\right), x_{K_2} - x_{K_1}\right) \leq 0.$$

More precisely, we find:

$$\left(F\left(x_{K_1}\right), x_{K_2} - x_{K_1}\right) < 0$$

unless $x_{K_2} = x_{K_1}$.

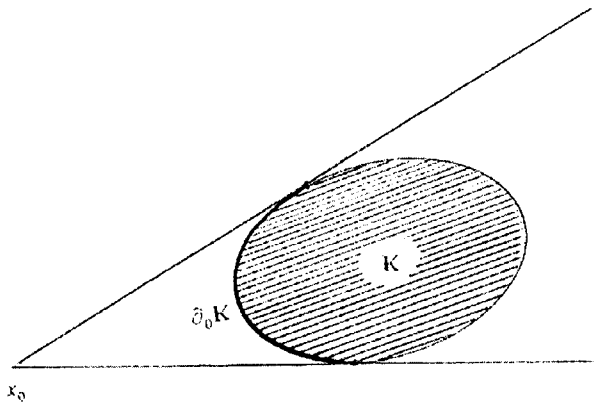
As a consequence of this theorem, if x_0 is the solution of the equation $F(x) = 0$, then we have the following

Theorem 4 *Let x_K be the solution of (5). Then:*

i) if $x_0 \in \mathbf{K}$, x_K coincides with x_0;

ii) if instead $x_0 \notin \mathbf{K}$, the hyperplane (2) with $x = x_K$ strictly separates x_0 from $\mathbf{K}$.

From a geometrical point of view, proposition ii) means that x_K belongs to the subset $\partial_0\mathbf{K}$ of $\partial\mathbf{K}$ consisting of all the points of $\partial\mathbf{K}$ at which there exists a supporting plane separating x_0 from $\mathbf{K}$ (see figure below).



Let us now consider the case in which $\mathbf{K}$ is the set of points in $\mathbf{R}^n$ for which the constraints:

$$h_i(x) \le 0 \qquad (i = 1,2,...,m) \tag{6}$$

hold, the $h_i(x)$'s being linear functions of the type $h_i(x) = a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n - a_{i,n+1}$, such that any set of $s \le n$ functions $h_i(x)$ constitutes a set of linearly independent functions. The boundary of $\mathbf{K}$ is composed by the points of $\mathbf{K}$ for which at least one of the functions $h_i(x)$ vanishes.

If $\mathbf{K}'$ denotes the subset of $\mathbf{K}$ obtained by imposing some equality constraints, for instance:

$$\mathbf{K}' = \{x \in \mathbf{R}^n | h_i(x) = 0, i = 1,2,...,k; h_j(x) \le 0, j = k+1, k+2,...,m\}$$

and $x_{K'}$ is a point such that:

$$x_{K'} \in \mathbf{K}', \quad (F(x_{K'}), x - x_{K'}) \ge 0 \quad \forall x \in \mathbf{K}',$$

then we have:

Theorem 5 *Point* $x_{K'}$ *is solution of the variational inequality relative to* **K** *if there exist non-negative constants* α_i *such that:*

$$F_r(x_{K'}) + \sum_{i=1}^{k} a_{i,r} a_i = 0 \qquad (r = 1,2,...,n), \tag{7}$$

where $F_r(x_{K'})$ *is the generic component of* $F(x_{K'})$.

Using the previous theorems, we reduce the resolution of the variational inequality (5) under the constraints (6) to the resolution of variational inequalities:

$$x_{E_s} \in E_s, \quad (F(x_{E_s}), x - x_{E_s}) \geq 0 \quad \forall x \in E_s \tag{8}$$

where E_s are linear manifolds defined by s constraints, among those given, in the form of equality, chosen by means of a procedure quite close to the method of Theil - Van de Panne for solving the quadratic programming problem.

Assuming, without loss of generality, that:

$$E_s = \{x \in \mathbf{R}^n : h_1(x) = 0, ..., h_s(x) = 0\}$$

and setting:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{s1} & a_{s2} & \cdots & a_{sn} \end{pmatrix},$$

$$d = \begin{pmatrix} a_{1,n+1} \\ a_{2,n+1} \\ \cdots \\ a_{s,n+1} \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \cdots \\ u_s \end{pmatrix}$$

we have the following:

Theorem 6 *The system:*

$$F(x)+A^T u=0, \quad Ax=d \tag{9}$$

is equivalent to the variational inequality (8).

Therefore, the resolution of (5) under (6) comes down to the repeated application of the Lagrange multipliers method, since we can write (9) as:

$$\begin{cases} F(x)+\sum_{i=1}^{s} u_i \nabla h_i(x)=0 \\ h_i(x)=0, \quad (i=1,2,...,s). \end{cases} \tag{10}$$

Remark Obviously, if $\hat{x}$ and $\hat{u}$ are two vectors such that:

$$F(\hat{x})+A^T\hat{u}=0, \quad A\hat{x}=d,$$

we have $\hat{x}=x_K$ if $\hat{u}_i \geq 0 (i=1,2,...,s)$.

***

These memories well show how Stampacchia was generous with his ideas.

As his assistant at the Faculty of Engineering and his friend since our first meeting, I knew Stampacchia very well. He was an excellent teacher; his lessons were uncommonly lucid and precise. The numerous students, who had him as a professor, remember him with high esteem and admiration. Nevertheless, during examinations, he was apparently severe.

Guido was a far-sighted man, sincerely democratic and friendly with his colleagues. He loved classic music and good cooking, but was a heavy smoker. He had a refined humour and liked to express witty remarks usually in his colourful Neapolitan language.

Guido and I often met and talked about Mathematics, our children, politics, and so on. I remember that the last topic we discussed together was "The role and the function of mathematicians in the technological society and the industry". We did not conclude our discussion because of the sudden death of Stampacchia on April 27, 1978, in Paris where he was as a visiting professor.

The last time I saw Guido was in the mortuary Chapel of the Misericordia in Via S. Frediano in Pisa, the day before the funeral ceremony. This took place at the Scuola Normale Superiore at the presence of close

relatives, friends and colleagues. Professor Carlo Miranda described the human and scientific figure of his great pupil by deep-felt words.

Guido now rests in Napoli, in the small British Cemetery, which he himself chose as his last home.

GUIDO STAMPACCHIA

Silvia Mazzone*
Dept. of Mathematics, University of Rome "La Sapienza", Rome, Italy

1. SCIENTIFIC EDUCATION AND FIRST RESEARCH ACTIVITY AT THE SCUOLA NORMALE SUPERIORE IN PISA AND AT THE UNIVERSITY OF NAPLES.

Guido Stampacchia was born on March 26th, 1922 in Naples, to the family of Emanuele Stampacchia and Giulia Campagnano. Giulia belonged to a Jewish family of Florentine origin,[1] that owned a factory which made hand embroidered household linen and linen garments for women. The Stampacchia family had origins in Lecce and practiced the Valdese Christian religion: the father, Emanuele, managed an iron tools factory which he was forced to sell off at the time of the war in Ethiopia, as a result of his refusal to join the fascist party. The young Guido got essentially a lay education, although as a child he attended the Valdese church together with his two sisters. He obtained his high school certification in classical subjects (maturità classica) at the age of eighteen from the Liceo-Ginnasio Gian Battista Vico in Naples, obtaining the excellent mark of 9/10 only in Mathematics and Physics. In spite of the classical studies he had followed, he intended to dedicate himself to

* Acknowledgment: I am very grateful to Sara Stampacchia and Enrico Magenes, who read a first draft of this writing and made a number of comments supplying useful information. My thanks are also due to M.K.Venkatesha Murthy for his help in the English translation of my Italian text and to Maria Giovanna Garroni, Louis Nirenberg and Paolo Podio Guidugli who read the final version.

[1] In some documents of the twenty years of fascism, the surname Campagnano was changed into Campagna, probably because of the then existing racist reasons.

mathematics and hence broadened his preparation in Mathematics and Physics by studying "the basic principles of the program prescribed for the scientific high school, seeking .... to find a logical process".[2]

In the autumn of 1940 he was admitted as an internal alumnus to the Scuola Normale Superiore of Pisa, for the undergraduate course in Pure Mathematics of the Classe di Scienze (Science Faculty), having secured the fifth position in the competitive entrance examination[3]; after that he completed brilliantly all the examinations in the curriculum of the first three years as required for the students of the Scuola Normale. In particular, he had Francesco Cecioni and Salvatore Cherubino among his teachers in the first three years of his university career, while in the third year he followed the courses of Leonida Tonelli on Analisi superiore (Advanced Analysis) and of Lamberto Cesari on Teoria delle funzioni (Theory of Functions). The latter, having graduated in 1933 under the supervision of Tonelli, was a professor in charge of a course from 1938, while Tonelli, the undisputed master of the Mathematical School in Pisa, taught the courses of Analisi infinitesimale and Analisi superiore, and maintained the chair of Analisi superiore during the three years – from November 1939 to October 1942 – when he had moved to the University of Rome. The advanced courses of Tonelli were concerned with trigonometric series, integral equations and calculus of variations, in a three-year cycle; when Stampacchia attended it, the course on Analisi superiore was dedicated to calculus of variations. During this period, the assistants of Tonelli at the University were Jaures Cecconi, who then became a professor at Genova, Landolino Giuliano, who later became a professor at the Naval Academy at Livorno, and Emilio Baiada who, as a professor, later taught at Palermo and Modena.

The courses at the Scuola Normale were organized by Tonelli and Giuliano. Tonelli used to organize two seminars, one for first level students and one for advanced students: respectively, Esercitazioni di Analisi e Geometria (Tutorial Sessions in Analysis and in Geometry) and Conferenze di teoria delle funzioni (Conferences in the Theory of Functions). He was the editor-in-chief of the Annali della Scuola Normale Superiore; later he was also the Director of the Scuola Normale, during the academic year 1943-44, contributing greatly to the survival of the Scuola in such a critical and difficult moment for Pisa and the whole country. Stampacchia followed his tutorial course of Analysis and

[2] G. Stampacchia, *Nota sugli studi e le tendenze personali*, sent along with the application for admission to the Scuola Normale of Pisa, conserved in the personal files of G. Stampacchia as a student, in the Archives of the Scuola.

[3] The topic of the written test in Analysis, for the competitive examination of the Scuola, was the theory of real numbers.

Geometry in 1941-42, took the examination in Theory of Functions the following year and, in 1943, passed the discussion on Ordinary Differential Equations, obtaining the mention of excellent preparation and excellent aptitude. Giuliano taught the courses of Complementary Mathematics I and II attended by Stampacchia, who took the relative examinations in his first and second year. During his stay at the Scuola Normale, Stampacchia had among his fellow students Giuseppe Colombo, Mario Dolcher, Jacopo Barsotti, who had joined the Scuola before him, Enrico Magenes, Roberto Conti, who joined a year later, and, finally, Aldo Andreotti, who was admitted in October 1942.

On the 24th of March 1943, Stampacchia, who had been drafted, informed the Administration of the Scuola Normale that he was expected to report at the Regia Aeronautica (the Royal Air Command) by the 28th of that month; anyway, he managed to take the examinations at the University of Pisa during the summer session of June 1943, securing marks of 30/30 *cum laude* from both Cesari and Tonelli. At the end of June he was sent to Rome to attend a course for Sergeant Cadet in technical specializations. He remained there until the 8th of September 1943, when he joined the Resistance Movement against the Germans, in the defense of Rome. After an adventurous trip, he regrouped with his parents and sisters in Isernia and returned to Naples only after the liberation of the city. He was assigned by the Liberation Army to administrative duties, and eventually discharged in June 1945.

In the meantime, taking advantage of a special Ministerial decree due to the war, he completed his fourth year of studies at the University of Naples. There he graduated, on the 28th of November 1944, obtaining his Laurea, on behalf of the University of Pisa, with the mark of 110/110 *cum laude*, discussing a thesis on ordinary differential equations, written under the guidance of Renato Caccioppoli. His thesis[4] was concerned with an adaptation of an approximation procedure for Volterra integral equations due to Tonelli[5] to boundary value problems for systems of ordinary differential equations.[6]

In the fall of 1944 he won a scholarship from the University of Naples for new graduates in mathematics, which allowed him to continue his studies under the direction of Renato Caccioppoli and Carlo Miranda; in addition, he worked as a voluntary assistant to the chair of Analisi Matematica (Mathematical Analysis) where he did tutorial work for the

[4] Stampacchia 1947, *Ma* [1], p. 418. Here and in the following *Ma*[·] is used to indicate the numbering of the scientific publications of G. Stampacchia as introduced by Magenes 1978b.

[5] See G. Sansone, *Equazioni differenziali nel campo reale*, v. 1, pp. 45-48.

[6] Stampacchia 1947, *Ma*[1], pp. 413-414.

course of Analisi Algebrica (Algebraic Analysis). Meanwhile he also prepared himself for the Esame di Licenza (Final Examination) in Analisi Superiore at the Scuola Normale, which he passed with 70/70 *cum laude* on November 19th, 1945. As the topic of his examination, he had chosen the semicontinuity of the double integral in the calculus of variations $\iint_D f(x,y,z,s)dxdy$, with the integrand depending on the mixed second-order derivatives of the unknown function. Thus, from his very beginning as a researcher, Stampacchia's interest for a topic very much studied by Tonelli and his school was evident, a topic on which he was to obtain very significant results within a span of few years.

In the academic year 1945-46, he shared with Jacopo Barsotti the first place for a position of Perfezionamento (Specialization) at the Scuola Normale in the Faculty of Sciences, but he declined, in order to accept a position of assistant at the Naval Institute at Naples, where Caccioppoli and Miranda were in charge of the courses in Analysis. A reason behind this decision, in addition to personal ones, was a certain dissatisfaction with his studies, which he so describes: "During the period I stayed in Pisa, I tried in several ways to develop a special interest in something, but I wandered among many different topics due to an absolute lack of a guide. Thereby, the reasons for the sacrifices I would have to make if I remained in Pisa became all the more meaningless."[7]

Stampacchia worked in Naples willingly and with satisfaction: having found in Caccioppoli a second teacher, he continued the studies in Differential Equations and Calculus of Variations he had begun with Tonelli. During the years 1945-46, 1946-47 and 1947-48 he fulfilled his teaching duties at the Naval Institute giving tutorial courses in Algebraic and Infinitesimal Analysis, as assistant in charge. At the same time, as a voluntary assistant, he helped with the courses on Analysis at the University, until he was given charge of the course entitled Istituzioni di Matematica (Principles of Mathematics) at the Faculty of Sciences, which he taught from November 1948 to December 1952. Moreover, in the academic year 1948-49, he held a CNR (National Research Council) scholarship to work on Calculus of Variations and methods of Functional Analysis at the Mathematics Institute of the University of Naples.

[7] G. Stampacchia to L. Russo, Naples 21-1-1946, conserved in the personal files of G. Stampacchia as a student, in the Archives of the Scuola Normale. Enrico Magenes remembers with emotion a visit he made with Stampacchia, on the first days of November 1945, to the house in Asciano Pisano where Leonida Tonelli had retired for health reasons. On this occasion, Tonelli advised Stampacchia to leave Pisa and benifit the possibility of working with Miranda and especially with Caccioppoli, whose charm had strongly attracted Stampacchia (Magenes 2000, p. 27). Leonida Tonelli died after a few months on 12 March 1946.

The scientific and economic independence he achieved in Naples permitted him to realize his desire, already evidenced when he gave up the specialization at the Scuola Normale, and marry Sara Naldini, his fellow student at the University. After the wedding was celebrated, in October 1948, the couple settled in Naples in the family house, where their children Mauro and Renata were born, in 1949 and 1951.

At the age of twenty seven, after having been declared eligible in a competition for the position of assistant in which he participated in November 1948, he was appointed as assistant with tenure to the chair of Mathematical Analysis at the University of Naples, on July 1st, 1949. Further, in the month of April 1951, he obtained the Libera Docenza (Habilitation) in Mathematical Analysis: for this examination, he presented 14 papers ranging from ordinary differential equations to the the theory of functions of real variables, from calculus of variations to partial differential equations.[8]

In the field of ordinary differential equations, which was the topic of his graduation thesis, Stampacchia studied, generalizing a problem posed by Nicoletti at the end of the eighteenth century, the problem of determining the solutions of a first order system of nonlinear differential equations, in the case when the boundary conditions were also given in a nonlinear form (*Ma*[1], [5]). In the Piccole Note of Unione Matematica Italiana (UMI), he gave a condition under which the solution of an equation of order n depending on a parameter, together with conditions to identify a polynomial of degree n, is reduced to an ordinary boundary value problem for an equation of order $n+1$ (*Ma*[7]). Further, in a note appeared in the Rendiconti Lincei, he gave a functional interpretation of the Peano phenomenon concerning the lack of uniqueness in the Cauchy problem (*Ma*[9]). Stampacchia returned to the study of ordinary differential equations during the fifties, when he solved the problem of determining an integral curve for a first order system lying on an n dimensional manifold (*Ma*[21]); and when, in a lecture delivered at Catania in May 1956, he gave a survey of the theory of boundary value problems for systems of ordinary differential equations (*Ma*[26]).

Among his first research works are the study of the Goursat problem at the large for a second order nonlinear hyperbolic partial differential equation in two variables (*Ma*[10]) and, due to his closeness with Caccioppoli, the paper (*Ma*[2]), in which he proves, for rectifiable surfaces, a conjecture about the uniqueness of the definition of area which fulfils lower semicontinuity condition.

[8] For details and an exaustive examination of these papers, see Magenes 1978a, pp. 717-722 (XIII-XVIII).

The work of Stampacchia in the Calculus of Variations, stimulated by the results obtained and the techniques developed by Tonelli, which he had rapidly mastered during his university career, continued with applications of the direct methods to double integrals depending on generic differential operators acting on the unknown function. In one of his earliest papers, semicontinuity was considered (*Ma*[3]); this work was completed by determining the conditions ensuring the existence of the minimum (*Ma*[11]). These results were the subject of a communication at the III UMI Congress, held in Pisa in September 1948.[9] In the case of functionals depending only on second order derivatives (*Ma*[4]), taking the Laplacian of the unknown function as the differential parameter, the minimization problem is set in the class of once continuously differentiable functions, which are assumed to be absolutely continuous along lines parallel to the axes (absolutely continuous according to Tonelli), and such that the pure second derivatives are integrable; for functionals depending on the mixed derivatives (*Ma* [6]), one considers a class of doubly absolutely continuous functions (absolutely continuous according to Vitali).

Studying the papers of Fubini and Beppo Levi on the minimum principle for the Dirichlet integral, Stampacchia realized[10] that, to prove the existence of a minimum for the case of double integrals, in the use of the direct method was not necessary to have uniform convergence; and that the appropriate function space for formulating the problem was not the space of absolutely continuous functions. Thus, with the encouragement of Caccioppoli, he began to examine functions separately continuous with respect to each variables together with the associated notion of quasi uniform convergence and the corresponding compactness criteria (*Ma*[8], [12]), which found applications in the calculus of variations (*Ma*[13]) and in the study of the Dirichlet problem for second order elliptic equations in two variables (*Ma*[14]). In particular, in the paper (*Ma*[13]) published in 1950 in the Giornale di Battaglini, of which Caccioppoli and Miranda were the editing directors, in order to overcome the limitations of Tonelli's theory, Stampacchia introduced a class of functions of the type of Sobolev space, in the two-dimensional case. Thus, independently and using different methods, he found results similar to those of C.B. Morrey on variational problems for multiple integrals,[11] about which he learned much later because of the then existing scientific

[9] G. Stampacchia, *Gli integrali doppi del calcolo delle variazioni in forma ordinaria*, Atti III Congresso UMI, 1948 (1951), p. 110.

[10] Stampacchia 1950, *Ma*[13], p. 171 – Stampacchia 1996-97, p. 31.

[11] C.B. Morrey, *Existence and differentiability theorems for the solution of variational problems for multiple integrals*, Bull. Amer. Math. Soc.,v. 47, 1940, pp. 439-458.

isolation of Italy, due to the war. Stampacchia remarks explicitly that, while his formulation was similar to Morrey's as far as the nature of the functions considered, the treatment was different in so much as Morrey makes use of functions defined up to sets of measure zero, thus precluding an analysis of boundary traces.

Naturally the existence theorems so obtained furnish minimizing functions in the new class, leaving open the regularity problem of the solutions thus found.

The depth and penetration with which these problems were treated and understood are very aptly described in the words with which Ennio De Giorgi remembers a visit to Naples during this period: "Picone was very much interested in problems of the type $\min_{\Omega}\left(\int_{\Omega} g + H_{n-1}(\partial\Omega)\right)$ and thought that Caccioppoli was the right person to find a path to the solution. Hence, he sent me for some days to Naples, where, talking to Caccioppoli, Stampacchia and Carlo Miranda, I could experience all the richness of their ideas from their live voices, much more than as those ideas were stated in their very ingenious writings. In the words of Caccioppoli, Stampacchia and Miranda their personal experiences and the teaching of their masters Picone and Tonelli were integrated, and the spirit of the direct methods of the calculus of variations and, in particular, of the procedure divided into four fundamental steps: relaxation, semicontinuity theorems, representation theorems, regularity theorems, came out clearly."[12]

The years that Stampacchia spent in Pisa and Naples characterize the formation of his personality as an analyst: he was a passionate specialist in calculus of variations and in the theory of partial differential equations, a practitioner and an inspirer of research works of considerable depth and originality of thought. As is well known, his work has contributed notably to the progress of mathematics and the fields of research opened by him are still drawing the attention of the international mathematical community.

THE CHAIR AT THE UNIVERSITY OF GENOVA AND THE ADMITTANCE INTO THE INTERNATIONAL MATHEMATICAL COMMUNITY.

In 1952 Stampacchia came out first in the national competition for a chair of Algebraic and Infinitesimal Mathematical Analysis at the

[12] De Giorgi 1985, p. 185.

University of Palermo[13]; he was nominated Professor on Probation (Professore Straordinario) in the Faculty of Mathematical, Physical and Natural Sciences of the University of Genoa on the 15th of December 1952. In December 1955, he was promoted Full Professor (professore ordinario).[14] He then settled himself with his family in Genova, where his daughters Giulia, in 1955, and Franca, in 1956, were born.

In the years between his habilitation and full professorship, Stampacchia generalized[15] to the case of n variables the class of functions he had introduced for the two dimensional case, and he proved the corresponding compactness criteria (*Ma*[15]). He presented these results at the Congress of UMI held at Taormina in October 1951 (*Ma*[16]). In his communication, he also treated the question of existence of the minimum for multiple integrals depending on the first and second derivatives of the unknown function, leading to the Euler equation being satisfied almost everywhere and thus strengthening the relation between calculus of variations and partial differential equations, a constant theme of all his research. The minimum is attained (*Ma*[17]) in a class of functions having traces on the boundary together with their normal derivative; since it is shown to satisfy the Euler equation, one obtains existence results for a fourth order elliptic equation with assigned boundary values of both the function and its normal derivative. The study of these boundary conditions led Stampacchia to examine the question of the approximability of a function on an assigned surface (*Ma*[18]).

As to the differentiability properties of the minimizing function of multiple integrals depending on the gradient (*Ma*[19]), we recall Stampacchia's results of local integrability of the second derivatives[16] and of analyticity of the solutions, with Hölder-continuous first order derivatives, of regular problems. These regularity results, together with an analysis of variational problems for multiple integrals depending on the first order derivatives of the unknown function, were illustrated in a lecture delivered in Torino in January 1954 (*Ma*[20]). The next step was to study the Euler equations by the direct methods of the calculus of variations in the case of Neumann or mixed boundary conditions

[13] The report of the board of juries, constituted by L. Fantappié, G. Scorza, C. Miranda, S. Cinquini and L. Amerio, is published in Bollettino ufficiale M.P.I., II part, 9 July 1953, n. 28, pp. 2192-2203.

[14] The report of the board of juries, constituted by F. Cecioni, C. Miranda and S. Cinquini, is published in Bollettino ufficiale M.P.I., II part, 26 July 1956, n. 30, pp. 5071-5073.

[15] See Magenes 1978a, pp. 722-725 (XVIII-XXI).

[16] This result was obtained by extending the methods introduced by C. Miranda, *Sui sistemi di tipo ellittico di equazioni lineari a derivate parziali del primo ordine in n variabili indipendenti*, Mem. Accad. Naz. Lincei, s. 8, v. 3, 1952, pp. 85-121.

(*Ma*[23]). Some of the results of this paper, dedicated to Mauro Picone on the occasion of his seventieth birthday, are the subject of a communication presented at the International Congress of Mathematicians held in Amsterdam in September 1954 (*Ma*[22]), while some remarks on the existence and uniqueness of solutions were published in Rendiconti dell'Accademia delle Scienze of Naples (*Ma*[24]).

The variational method was also used by Stampacchia to treat the so called transmission problem, namely, the problem of two equations in two domains having parts of their boundary in common and with a natural condition of matching on the common part of the boundary (*Ma*[27]). This kind of a situation arises in the study of phenomena taking place in a stratified medium. The leading idea is again solving the problem in a weak form, then studying the regularity of the weak solution thus found. A clear exposition of the results on transmission problems for elliptic equations was presented in a talk at the Seminario Matematico of Bari in December 1960 (*Ma*[32]).

Stampacchia's interest in the different classes of functions used in proving existence in weak form can be found in his lectures (*Ma*[28]) at the CIME course on Singular Integrals and Related Questions, held at Varenna in June 1957, where he presented a theory of completion of function spaces following ideas suggested by Aronszajn and Smith.[17]

During the years he spent in Genoa, Stampacchia taught courses in Analisi Matematica, Analisi Superiore, Matematica Superiore (Higher Mathematics), Istituzioni di Matematica II, and Topologia. In addition, during the academic year 1960-61, when he had already moved to Pisa, he also taught a course of Complementi di Matematica (Complements of Mathematics) at the School of Engineering. In February 1960, he was appointed to represent the Faculty of Sciences in the Council of Directors of the newly created Centro di Calcolo Numerico.

He interacted agreably with his colleagues Eugenio G. Togliatti, Enzo Martinelli and Francesco Sbrana. When Martinelli moved to Rome, he favoured the arrival of Enrico Magenes, with whom he had always kept a close relation. He established scientific and friendly relations also with Jaures P. Cecconi, Giulio Aruffo and Emilio Gagliardo; later on, Sergio Campanato also worked with him. During these years, he began to entertain the intense international relations, which were to characterize his future mathematical career. In August 1954, at the conference "Problemi esistenziali e qualitativi per le equazioni differenziali lineari alle derivate parziali" organized by Gaetano Fichera in Trieste,

[17] N. Aronszajn - K.T. Smith, *Functional spaces and functional completion*, Ann. Inst. Fourier, 6, 1956, pp. 125-185.

Stampacchia met Louis Nirenberg, with whom he remained a very close friend until his death. Together with Magenes, he had guests as prestigious as Laurent Schwartz, Henri George Garnir, Antoni Zygmund, Bernard Malgrange and Nachman Aronszajn; he first met Jacques-Louis Lions at Nice, in September 1957, during the Réunion des Mathématiciens d'Expression Latine.

He very often traveled abroad for scientific studies and collaborations. His travels were so frequent that it is impossible to remember all the visits he made, either to attend conferences or to give talks, to innumerable research institutes in Italy and abroad. We shall limit ourselves to mention long-time visits, and only to the most prestigious institutes. Quite often, these visits resulted in significant scientific publications.

As a result of common interests and continued collaboration with Magenes, an exposition of a complete and general survey of the different approaches to boundary value problems for linear elliptic differential equations of arbitrary order was written up in the spring of 1958 (*Ma*[29]). Even more than ten years after its publication, and in spite of notable further developments on the subject matter, this article remained a remarkable instrument for all those who wished to dedicate their studies to these issues. The subject was reconsidered later, in the general lecture (*Ma*[31]) Stampacchia gave at the VI UMI Congress, held in Naples in 1959.

During the mid fifties, the work of Stampacchia on the differentiability properties of the extremals of regular multiple integrals was concerned with different types of questions: on the one hand, existence of square integrable second order derivatives of the solutions to minimum problems, solutions being a priori of very low regularity; on the other hand, analyticity in the case of Hölder continuous first order derivatives (*Ma*[19], [25]). A relation between these results was missing, namely, to assure the analyticity of the minimizing functions whenever the integrand in the regular integral is analytic. Stampacchia stimulated De Giorgi's thoughts on this important problem,[18] which De Giorgi solved in a famous paper presented at the V UMI Congress held in Pavia in October 1955, and published later on, in 1957.[19]

Continuing along the De Giorgi's ideas, Stampacchia considered linear elliptic equations with discontinuous coefficients in spaces of

[18] On this question, in addition to the work of De Giorgi, we recall those of J. Nash and, later, of J. Moser.

[19] E. De Giorgi, *Sulla differenziabilità e la analiticità delle estremali degli integrali multipli regolari*, Mem. Accad. Sci. Torino, s. 3, v. 3, 1957, pp. 25-43. A pre-publication note was published in 1956 in Rendiconti dell'Accademia dei Lincei.

dimension $n > 2$; in the span of about ten years, he obtained results in various directions: L^p estimates for the solutions, the maximum principle, Hölder continuity of the solutions up to the boundary, existence and properties of the Green's function for the Dirichlet problem, characterization of regular points on the boundary.[20] The notable body of scientific publications of the highest international standards on elliptic equations, to which these works of Stampacchia belong, is very well illustrated by Carlo Miranda in a lecture at the VIII UMI Congress,[21] in whose reference list the name of Stampacchia is found among the most eminent specialists in the field.

He first obtained preliminary regularity results (*Ma*[30]), such as boundedness and integrability of solutions for a large class of boundary value problems for second order elliptic equations with bounded and measurable coefficients, imposing only the "cone condition" on the boundary of the domain. In order to explain his technique, he recalled the idea of De Giorgi to obtain the interior regularity of the extremals. Moreover, in the appendix, by comparing some results from the theory of capacity to the Sobolev inequality, he glimpsed a relation between these results and some isoperimetric inequalities. The questions of summability and of boundedness were taken up again in July 1960 (*Ma*[35]), in a lecture at the International Symposium on Linear Spaces held in Jerusalem, refining the technique of truncation with a lemma, by now classical, on decreasing functions. Some limiting cases of summability of the known right hand side were analysed later in (*Ma*[39]), during his visit at the Courant Institute of Mathematical Sciences.

The Hölder continuity up to the boundary of the solution was essentially considered in a memoir dedicated to Giovanni Sansone in the occasion of his seventieth birthday (*Ma*[34]), which had been preannounced by a Comptes Rendus note in February 1960.[22] More precisely, for equations more general than those considered by De Giorgi, and for an "admissible" class of sets Ω introduced to this purpose, Stampacchia proves Hölder continuity in Ω of the solutions of the Dirichelet problem, of the Neumann problem, and of the mixed problem in which the boundary data are discontinous, but the boundary of Ω, the coefficients and the right hand side of the equation are regular. The geometrical conditions ensuring "admissibility" of a set Ω were exposed

[20] See Magenes 1978a, pp. 726-732 (XXII-XXVIII).

[21] C. Miranda, *Progressi e orientamenti della teoria delle equazioni ellittiche negli ultimi quindici anni*, Atti VIII Congresso UMI, 1967 (1968), pp. 23-54.

[22] G. Stampacchia, *Solutions continues de problèmes aux limites elliptiques à données discontinues*. C. R. Acad. Sci. Paris, 250, 1960, pp. 1426-1427.

at the Colloque sur l'Analyse Fonctionnel, held at Louvain in July 1960 (*Ma*[33]).

In 1960 Stampacchia was a visiting professor for one month at the Institut des Hautes Études in Paris, where he gave a brief course under the auspices of the Séminaire Schwartz, during the year devoted to partial differential equations and interpolation. In the published version of this course (*Ma*[36]), he obtained the integrability and the boundedness of solutions of second order elliptic equations with discontinuous data, under less restrictive hypotheses on lower order terms than those considered in earlier works (*Ma*[30], [35]). In the month of November of the same year, he sent a letter of support and solidarity, which was jointly signed by 46 other Italian mathematicians, to Laurent Schwartz, who had been removed from his chair at the École Politechnique.

During his academic career, Stampacchia served as a member in many committees for national competitions for professorial positions in Analysis. Among these, we mention that he was the secretary of the selection committee[23] for a position in Mathematical Analysis at the University of Messina, for which Ennio De Giorgi was first among the selected candidates.

Even after he moved to Genoa, he kept a privileged contact with Naples, both because of his close relationship with his sisters and his other relatives who lived in Naples and because of his relations with the scientific community in Naples, first of all with C. Miranda and Caccioppoli, and also with Federico Cafiero, Donato Greco and Renato Vinciguerra. He was elected a corresponding member of Società Nazionale di Scienze, Lettere ed Arti of Naples in November 1954.

As remarked by Magenes,[24] Stampacchia was always proud of his native Naples and of the family background in which he was born; in fact, he was very happy of being born in such a special place and in an unusual family. These peculiar roots were particularly suited to his personality, rather nonconformist, and certainly contributed to his formation. Part of his characteristic nature was that of a very amiable and easy-going gentlemanliness, a simplicity coupled with the consciousness of his important position in the mathematical community, a very strong sense of criticism and a great sense of humour. Very frank in expressing clearly whatever he thought, he had a very generous and free attitude towards friends and students which was reciprocated with affection. It is worth recalling the testimony of Haïm Brezis, who remembers the beginning of their long and fruitful mathematical relation: "Pendant la

[23] The other members of the board were Giovanni Ricci, Carlo Miranda, Gianfranco Cimmino, and Sandro Faedo.

[24] Magenes 1978a, p. 715 (XI).

préparation de ma thèse j'eu la chance de rencontrer trois maîtres, Félix Browder, Louis Nirenberg et Guido Stampacchia, qui ont donné une ouverture internationale à mon travail. [...] Ma première invitation mathématique est arrivée de l'Université de Pise où enseignait G. Stampacchia. J'avais à peine vingt-trois ans [1967] et j'ai des souvenirs merveilleux de cette visite d'un mois. L'hospitalité de G. Stampacchia était légendaire; il était impossibile de régler une addition, ni même de payer un café, en sa présence. Comme j'ai l'ai dit, mes connaissances en EPD (équations aux dérivées partielles) étaient très fragmentaires et je n'étais même pas familier avec le célèbre principe du maximum; plus précisément, il était enseigné à Paris – en théorie du potentiel – mais sous une forme tellement abstraite et déguisée que le lien avec les EPD s'était perdu. Au lieu d'être surpris de mes lacunes et de me suggérer des lectures, G. Stampacchia s'est chargé lui-même de me l'enseigner en suivant une approche très élégante[25] qu'il avait découverte."[26]

2. BACK TO PISA: THE GOLDEN AGE OF THE MATHEMATICAL INSTITUTE, THE FELTRINELLI PRIZE AND THE PRESIDENCY OF UNIONE MATEMATICA ITALIANA.

While Stampacchia was in Genoa, Sandro Faedo, who had returned definitively to Pisa to the chair left vacant at the death of Tonelli, had already started to do his best to strengthen the Istituto Matematico. Faedo began by bringing in Aldo Andreotti from Turin; together, they tried from 1956 to 1958 to convince Stampacchia to move from Genoa to Pisa, offering him first the chair of Cecioni and then a position at the Scuola Normale.[27] For various reasons, Stampacchia's transfer did not come through; however, the mathematical community in Pisa was enriched by the arrival of Edoardo Vesentini at the University and of Ennio De Giorgi at the Scuola Normale. Thus, in 1960 when a position in Mathematical Analysis at the University of Pisa was available, Stampacchia eventually accepted the offer and moved from Genoa to Pisa with his family. In the same year, Barsotti too returned to Pisa, as Giovanni Prodi and Sergio Campanato did a little later.

[25] See Stampacchia 1963, *Ma* [38], pp. 387-388 – Stampacchia 1996-97, pp. 399-400 and Stampacchia 1965, *Ma* [45], pp. 206-207 – Stampacchia 1996-97, pp. 488-489.

[26] J. Vauthier, *Haïm Brezis un mathématicien juif*, Beauchesne, Paris, 1999, p. 25. Similar feelings are also expressed by David Kinderlehrer (Kinderlehrer 1988).

[27] After the death of Caccioppoli, Stampacchia was contacted for transfer also from the University of Naples.

The presence of this group of highly active and qualified experts in different fields made the scientific atmosphere of the Institute of Mathematics very bright and also led to useful and intense joint collaborative research. Moreover, international relations became very frequent and constructive, with short and extended visits of mathematicians from Pisa to foreign institutions and also by way of the presence of very distinguished visitors from abroad. Among the guests of the famous office with red tapestry, reserved for distinguished visitors at the Institute – at that time located in "La Sapienza", the historical seat of the University of Pisa – and then of the new building at Via Derna, we recall Oscar Zarisky, Hans Lewy, Philip Hartman, Louis Nirenberg, Armand Borel, Joe J. Kohn, Shmuel Agmon, Bernard Malgrange, Donald G. Aronson, Pierre Grisvard, Stanley Kaplan, and Robert Seeley. Thus, the Institute of Mathematics at Pisa became internationally renowned and a pole of attraction and training for research workers both Italian and from abroad, among which Enrico Bombieri stands.

From September 1961 to September 1963, Stampacchia moved with his family to the United States as a Temporary Member of the Courant Institute of Mathematical Sciences of New York University. On this occasion he also visited many other institutions and universities in the United States; in particular, at the University of Minnesota at Minneapolis he established scientific relationships with Walter Littman, Hans Weinberger and James Serrin. In the month of August 1962 he was one of the invited speakers at the International Congress of Mathematicians held in Stockholm; in his talk (*Ma*[37]) he presented the main results obtained up to that time on second order elliptic equations in divergence form with bounded measurable coefficients.

The papers published during his long stay in America are of great relevance. Stampacchia, in collaboration with Littman and Weinberger, obtained a very refined result on regular boundary points for the Dirichlet problem associated to a uniformly elliptic operator with discontinuous coefficients, and he was led to a detailed analysis of the properties of the Green's function and of the capacitory potentials (*Ma*[40]). In March 1963 he completed a paper (*Ma*[38]) strictly related to Calculus of Variations. In this paper[28] he obtained, under suitable hypotheses on the boundary of the domain and on the boundary condition, the existence and regularity of the minimizers of integrals which are only regular or uniformly regular, according to the kind of dependence on the unknown function. These results were obtained by the use of the maximum principle, proved by the truncation method; they can be applied to the problem of minimal surfaces, which could not be treated either by the

[28] See Magenes 1978a, pp. 734-735 (XXX-XXXI).

already mentioned theorem of De Giorgi or by its extensions due to C.B. Morrey and O.A. Ladyzenskaja and N.N. Ura'ltseva.

A new line of research began at this stage,[29] to which S. Campanato and G.N. Meyers had already made contributions: namely, interpolation between $\mathcal{L}^{p,\lambda}$ spaces – which contain for particular values of the parameter λ the Lebesgue spaces of measurable functions whose powers are integrable and the spaces of Hölder continuous functions – and their application to elliptic equations (*Ma*[41]). These results were presented at the VII Congress of UMI held in Genoa at the end of September and October 1963 (*Ma*[42]), while the Dirichlet problem, in the limit case of the space of bounded mean oscillation (BMO) functions of John and Nirenberg, was considered in a section of *Ma*[39].

In the Conference "Convegno Lagrangiano", promoted by the Accademia delle Scienze di Torino in October 1963 in the occasion of the one hundred and fiftieth anniversary of the death of Lagrange, Stampacchia presented a survey (*Ma*[46]) of the main developments of the principle of minimum in the calculus of variations or, more precisely, of the relation between the minima of regular multiple integrals and the boundary value problems for elliptic partial differential equations, thus putting in a historical perspective the most recent results, from Sobolev spaces to trace theorems, from the maximum principle to the regularity of weak solutions.

After returning to Pisa from the United States, Stampacchia continued his study on the interpolation between function spaces, refining, among others, certain properties of inclusion between Morrey spaces (*Ma*[47]). As an application of these results, he and Campanato obtained in a joint paper some L^p estimates for the derivatives of solutions of elliptic equations (*Ma*[50]). Finally, in his talk at the Conference Equadiff II, Differential equations and their applications, held in September 1966 in Bratislava (*Ma*[54]), he presented a panoramic survey of the $\mathcal{L}^{p,\lambda}$ spaces and of their use in interpolation theory and in elliptic equations.

As we have already observed, in his study of variational equations Stampacchia often considered the analysis of potential and capacity theories together; his well known generalization of Lax–Milgram lemma on coercive bilinear forms to convex sets, proved in 1964, can be put in this context (*Ma*[43]). Thus, the theory of variational inequalities was born, driven by the solution given by Gaetano Fichera to the Signorini problem on the elastic equilibrium of a body under unilateral constraints[30]

[29] See Magenes 1978a, pp. 733-734 (XXIX-XXX).

[30] G. Fichera, *Problemi elastostatici con vincoli unilaterali: il problema di Signorini con ambigue condizioni al contorno*, Mem. Accad. Naz. Lincei, s. 8, v. 7, 1964, pp. 91-

and by Stampacchia's work on defining the capacitory potential associated to a non symmetric bilinear form.

In the spring of 1964 he was a visiting professor for a month at Collège de France, on an invitation by Jean Leray. He presented his work on interpolation spaces and he gave a course on second order elliptic equations in divergence form with bounded measurable coefficients, under the auspices of Séminaire Leray sur les équations aux dérivèes partielles. In the published paper based on this course (*Ma*[44]), well known results for the case of the Laplace equation, such as the maximum principle and the properties of the Green function for the Dirichlet problem, were extended to second order elliptic equations in divergence form with bounded measurable coefficients. Also the extension of Harnack inequality by J. Moser,[31] based on the important result of John and Nirenberg on bounded mean oscillation functions, was proved. These topics can be found in an extensive paper published in Annales de l'Institut Fourier in 1965, a paper which was especially devoted to the Dirichlet problem for an elliptic operator in divergence form having discontinuous coefficients and lower order terms (Ma[45]). In this paper, among other things, Stampacchia made use of his result on non symmetric coercive forms to show the existence of the capacitory measure and of the capacitory potential, and to obtain a comparison of the capacities corresponding to different operators. The theory of elliptic equations with divergence structure and with discontinuous coefficients was reviewed in the summer course he gave in the Séminaire de Mathématiques Supérieures at the University of Montreal. A presentation of this topic in its most general form is found in the notes of these lectures published in book form (*Ma*[52]).

Many times at the University of Pisa, Stampacchia was entrusted, in addition to courses in Analysis, with the courses of Analisi Superiore, Metodi Matematici della Fisica (Mathematical Methods in Physics) and Calcoli Numerici e Teoria dei Grafi (Numerical Calculus and Graph Theory); he also taught courses in Equazioni Differenziali and Analisi Superiore at the Scuola Normale. In November 1966, he was elected as the Director of the Istituto di Matematica.

At the Institute, in addition to his already mentioned collaboration with Campanato, Stampacchia worked with other collegues, interested in fields other than Analysis. An important aspect of that period was the weekly Seminar, which all the members of the Institute used to attend

---

140. A preliminary note was published in Rendiconti dell'Accademia Nazionale dei Lincei in 1963.

[31] J. Moser, *On Harnack's theorem for elliptic differential equation*, Comm. Pure Appl. Math., v. 14, 1961, pp. 577-591.

independent of their specific research field. In this context, we recall the collaboration with Andreotti and Vesentini on Carleman estimates for the Laplace-Beltrami equation on complex manifolds. Stampacchia contributed significantly to this work proving an inequality "which highlights the crucial role of the completeness of the metric of the manifold."[32]

Stampacchia and De Giorgi were able to put in evidence a property of minimal surfaces which was not encountered in the case of solutions of elliptic equations in general. Extending a result of Lipman Bers in two dimensions, they showed that a minimal surface which can be represented in Cartesian form on an open set in $\mathfrak{R}^n$ can have singularities in a compact set of zero capacity of order 1 or, equivalently, the singularity set can be a compact set of zero $(n-1)$-dimensional Hausdorff measure (*Ma*[48]).

In collaboration with M.K. Venkatesha Murthy of Tata Institute of Fundamental Research in Bombay, who was a visitor to Pisa several times starting from 1963, Stampacchia considered degenerate elliptic operators, that is, those operators for which the so called ellipticity constant is replaced by a function. This means that the ellipticity constant can depend on the point in the domain and can also vanish on some subset of points, but this function together with its reciprocal satisfies suitable integrability conditions. This leads to the consideration of differential equations in the context of Sobolev spaces with weights and hence to the necessity of recovering in such spaces the relevant properties which allow one to obtain the results known for boundary value problems associated to elliptic equations with discontinuous coefficients (*Ma*[56]).

The results in the papers with De Giorgi[33] and with Murthy (*Ma*[55]) were presented in the Conference "Le equazioni alle derivate parziali", held at Nervi (Genova) in February 1965, of which Stampacchia was one of the organizers.

While Stampacchia was in Chicago as a visiting professor in May 1966, on a proposal of Giovanni Sansone, he was awarded the Feltrinelli Prize for Mathematics, Mechanics and Applications of the Accademia Nazionale dei Lincei. In the motivation for awarding the prize, the committee[34] reviewed the important research activity of Stampacchia and underlined "the vast and ample scientific production", "the importance of

[32] Vesentini 1980, p. 12.

[33] E. De Giorgi - G. Stampacchia, *Sulle singolarità eliminabili delle ipersuperficie minimali*, Atti del Convegno su Le Equazioni alle Derivate Parziali (Nervi, 1965), Edizioni Cremonese, Roma, 1966, pp. 55-58.

[34] The members of the committee were Beniamino Segre, Mauro Picone, Enrico Pistolesi, Giovanni Sansone, Alessandro Terracini and Bruno Finzi.

the results" and "the high esteem and position that these have secured him in the international field".[35] He was appointed as a Corresponding Member of the Accademia dei Lincei in July 1968 with similar motivations. In fact, his frequent visits abroad, his extensive scientific relations in the international field and his vast group of students, direct or indirect, both Italians and foreigners, contributed very much to broaden the horizon of studies and research activities and to enhance the prestige of the Italian School of Mathematics in the scientific world.

The intense scientific activity of those years did not divert Stampacchia's attention from the problems of the teaching of Mathematics and of the formation of teachers, and he personally undertook the work of channeling young people towards mathematical studies. In an article, originally pubblished in Bollettino della Società ex-alunni della Scuola Normale and later reproduced also in the Bollettino dell'UMI,[36] he analysed the genesis of the organization of teaching and of the university curriculum of studies, and made a proposal to reorganize the teaching scheme, including also a three year plan for the future teachers of middle schools. Moreover, he accepted to participate in a pre-University orientation course organized at Erice by the Scuola Normale in September 1966, where he delivered some lectures on the subject "Mathematics as research and as an instrument of scientific and technical enquiry". Stampacchia considered mathematics and, in particular, Mathematical Analysis as a fundamental instrument to study natural phenomena. Like Tonelli, he used to remark that the evolution of a phenomenon is often governed by principles which correspond to a maximum or a minimum of some integral. Accordingly, he assigned a preminent role to Calculus of Variations and to the Theory of Differential Equations in the understanding of problems in Physics and hence in establishing a close relationship between theory and applications. In the preliminary notes of his Erice lectures, one finds remarks expressing his view of Mathematics and his unified vision of theoretical and applied research: "Mathematics, as an expression of human thought, reflects the active will, the contemplative reasoning, the desire for aesthetic perfection. Its fundamental basis consists of logic and intuition, analysis and construction, generalities and individualities. Any development of Mathematics has without doubt its psychological origins in more or less practical requirements or demands, but once it is initiated under circumstances of necessity it acquires a value by itself and transcends the

[35] *Relazione per il conferimento del premio "Antonio Feltrinelli"*, Rendiconti delle adunanze solenni, v. VII (1965-1976), Atti Accad. Naz. Lincei 1976 (1977), p. 143.

[36] G. Stampacchia, *Note sull'insegnamento della matematica*, Boll. UMI, s. 3, v. 21, 1966, pp. 186-190.

limits of its immediate utility. This trend from the applied towards the theoretical science shows itself continuously in history."

Stampacchia appreciated in Mathematics its utility in addition to its intrinsic beauty. He believed that a dynamic relationship between theory and application was of fundamental importance for the development of Mathematics. Quite often, "a theory in Pure Mathematics may become very useful in Applied Mathematics and, conversely, problems suggested by Applied Mathematics can lead to a new theory in Pure Mathematics."[37] He further stated: "In the last three decades Mathematics has gone through a period of profound critical re-examination, trying to recognize its fundamental structure in an abstract manner; at the same time, an increasing number of applied sciences discovered in Mathematics a basic instrument and indicated to it new fields of enquiry. Never before as in this period, has one seen this process of interaction by which, on the one hand Mathematics creates new instruments and new languages for applied sciences, and on the other the latter sciences, with their specific problems, give rise to new areas of research in Mathematics which were unthought of before. Mathematics with its various aspects thus gets inserted as a fundamental fact in cultural, scientific and technical development."[38]

In 1967 Stampacchia was elected, with 239 votes, President of the Unione Matematica Italiana (UMI), of which he was a member since 1948 and a member of the Scientific Committee from 1964. He remained in this office till 1973, when he was re-elected, as per his wish, a member of the Scientific Committee.

As President, he gave the opening addresses at the VIII and at the IX Congresses.[39] In these adresses he identified the role of UMI as that of "preservation and improvement of the level of mathematical research in Italy and of modernization of the teaching"[40] and he expressed his satisfaction for the adequate presence of Italian mathematical activity in the international community. However, he drew the Congress' attention to the necessity of elaborating new policies and a new structure for the development of mathematics, in order to further enlarge the research fields and to think over the teaching methods, starting from the primary school stage.

[37] G. Stampacchia, *Matematica pura e applicata negli sviluppi attuali*, introductory report to the Congress "Rapporti tra ricerca matematica pura ed applicata in Italia", Siena, 27-29 September 1973.

[38] G. Stampacchia, *Discorso*, Atti VIII Congresso UMI, 1967 (1968), p. 19.

[39] The VIII Congress of UMI was held in Trieste from 2 to 7 October 1967. The IX Congress of UMI was held in Bari from 27 September to 3 October 1971.

[40] G. Stampacchia, *Discorso inaugurale*, Atti IX Congresso UMI, 1971 (1974), p. 6.

His activity in UMI was directed, in particular, to the problem of renovating Italian Mathematics "after the sterile closure of dictatorship and war"[41], whose damaging effects he had experienced personally at the beginning of his scientific career. The promotion of research was very dear to Stampacchia. He followed with great interest every initiative which contributed towards this end – for example, the program of visiting professors promoted by the National Committee for Mathematics of the National Research Council, and the CIME (International Mathematical Summer Center), which favoured the insertion of Italians in the international mathematical community, with the help of a program of updating and meetings – and he urged the Ministry of Education to pass the law on the reorganization of the Istituto Nazionale di Alta Matematica, which had been put under a commissioner since 1962. He gave special attention to the journal Bollettino dell'UMI, which, during the period of his Presidency, enhanced its prestige and also enlarged its diffusion.

Being highly sensitive to the problems of the younger generation, he promoted several initiatives in order to direct young people towards mathematics and he was always concerned about the difficulties encountered by graduates in mathematics to obtain adequate jobs. He followed the questions related to the approval of the University Reform Bill by the Parliament and he proposed the institution of a commission to study the problems of teaching mathematics during the first two years of the university courses. Through the Italian Commission for the Teaching of Mathematics, he also took interest in teaching at the secondary school level. Accordingly, he favoured contacts between universities and secondary schools, as well as the diffusion in Italy of modern trends of teaching mathematics in foreign countries.

Even after being fully involved in the activities of UMI and with the University of Pisa, Stampacchia undertook many visits abroad, especially to France and to the United States. Between the years 1965 and 1968 he made numerous visits to Paris, to the University of California at Berkeley, to the University of Minnesota at Minneapolis and to the University of Chicago. The publication of joint papers with Jacques Louis Lions, Philip Hartman, Haïm Brezis and Hans Lewy goes back to the period of these visits.

Starting from this period, his research activity was concentrated on the theory of variational inequalities.[42] Stampacchia took the variational theory of boundary value problems for partial differential equations as

[41] G. Stampacchia, *Discorso*, Atti VIII Congresso UMI, 1967 (1968), p. 18.

[42] For a detailed presentation of the main contributions of G. Stampacchia to the theory of variational inequalities see Lions 1978.

the model: variational inequalities, in fact, represent a very natural generalization of such problems and allow one to consider several questions arising in such different contexts as Mechanics, Physics and Convex Programming.

Continuing the studies he begun in 1964, Stampacchia announced (*Ma*[49]), in collaboration with Lions, the generalization to not necessarily coercive bilinear forms of his first result on this matter published in Comptes Rendus (*Ma*[43]). Associating some of the techniques already used in the study of minimum problems (*Ma*[38]) to the first results obtained for variational inequalities, he analyzed with Hartman the existence, uniqueness and regularity of solutions of the Dirichlet problem for a nonlinear equation with a term depending functionally on the unknown (*Ma*[51]). The result was based on a preliminary theorem where conditions were given for the existence of solutions of a variational inequality related to a monotone operator.[43]

In another important paper with Lions,[44] published in Communications on Pure and Applied Mathematics (*Ma*[53]), Stampacchia re-examined the entire linear theory, studying variational inequalities associated to bilinear forms which are coercive or simply non negative in Hilbert spaces, with applications to elliptic and parabolic operators and to problems with unilateral constraints. The regularity of the solutions of problems with obstacles for a second order linear elliptic operator, and the nature of the set of contact with the obstacle, were studied jointly with Hans Lewy (*Ma*[59]). This paper, published in 1969, was presented at the VIII UMI Congress in 1967 (*Ma*[60]). Finally, in collaboration with Brezis, Stampacchia obtained an abstract regularity theorem for variational inequalities associated to non linear monotone operators, a theorem applicable to a number of examples wherein the convex sets are defined by constraint conditions on the unknown function or on its gradient[45] (*Ma* [57]). The application of the theorem of Brezis-Stampacchia to the case of a convex set defined via an obstacle from above and an obstacle from below was considered in Stampacchia's lecture at the American Mathematical Society Conference, Nonlinear Functional Analysis, held in Chicago in April 1968 (*Ma*[62]).

On an invitation from Aldo Ghizzetti, Stampacchia gave an advanced level course on Variational Inequalities at the NATO Summer School on Theory and Applications of Monotone Operators held in Venice during

[43] A similar result was independently obtained by Felix Browder, *Non linear monotone operators and convex sets in Banach spaces*, Bull. Amer. Math. Soc., 71, 1965, pp. 780-785.

[44] See Lions 1978, pp. 740-741 (XXXVII-XXXVIII).

[45] See Lions 1978, pp. 747-748 (XLIV-XLV).

the summer of 1968. The lecture notes of this course (*Ma*[58]), which are a re-elaboration of the lecture notes for a course delivered in Minneapolis during the spring of the same year, contain a clear exposition of the results on existence and regularity of solutions of variational inequalities obtained up to that time, together with many examples and applications.

3. AT THE UNIVERSITY OF ROME "LA SAPIENZA": THE STUDENT MOVEMENT, THE APPOINTMENT AS DIRECTOR OF THE ISTITUTO PER LE APPLICAZIONI DEL CALCOLO.

Following informal contacts started in February 1967, in November 1968 Stampacchia, the only person whose prestigious name drew complete unanimity among the professors of the then Mathematics Institute, was invited to take the position of Professor of Mathematical Analysis I (second chair) by the Faculty of Sciences of the University of Rome "La Sapienza". He was assigned the course of Istituzioni di Analisi Superiore in the academic year 1968-69, the course of Analisi Superiore in the following year.

In the spring of 1970 he attended the ceremony of dedication to Beppo Levi of the Institute of Mathematics of the University of Rosario, in Argentina; that year he also visited the University of Sussex and the following year the University of Maryland at Baltimore.

It is not difficult to imagine the influence that Stampacchia, who had brought with him a group of Italian as well as foreign young researchers, could have had in Rome both scientifically and also in the field of teaching: there was a lot of hope in this direction. But, instead, because of unfavourable circumstances and conditions, his stay in Rome lasted only two years, after which he accepted with pleasure the invitation to transfer himself to the Scuola Normale at Pisa.

During the winter of 1970 "La Sapienza" was the theater of violent contrasts between the Student Movement and the neo-fascist group of Avanguardia Nazionale, fighting each other to gain control of University politics. There were frequent assemblies, demonstrations and occupations of the buildings; often, the police was called in by academic authorities, to intervene inside the University campus in order to separate the opposing factions or to disperse the demonstrating crowds. In this atmosphere Stampacchia assumed a position against the extreme right student organizations and publicly protested against the fascist demonstrations inside the campus. In a letter sent to the Dean of the Faculty of Sciences, Prof. Montalenti, Stampacchia expressed all his

indignation for the facts he have had to watch, and announced his decision not to go to the Institute of Mathematics until every form of apologetic behaviour ceased. Among other things, he wrote: "Illustrious Dean, it is nearly a year and a half that I have been carrying out my duties as a member of the teaching staff in your Faculty. I have seen episodes of struggle by the Student Movement which I first followed in the hope that it would act as an incentive for the indispensable reforms in the university system; then with a certain amount of mistrust, when these demonstrations became simply a demand for more examinations and less lessons; finally, with a lot of discomfort when a part of the teaching staff, self-defining subordinate, joined this movement, looking for positions of power only alternative to that of the so called barons. ... Unfortunately, while entering the Città Universitaria (University campus) to go to the Institute, I have been forced, for some days, to listen to hymns and to see walls filled with symbols inspired by those harmful ideologies of which our generation very well knows the disastrous consequences. I refer to those ideologies in whose name, incapable, corrupt and violent men rose to power and removed from their teaching positions, and quite often cancelled, the best inteligentsia, lowering the cultural and social standards of Italy and exposing our country to the contempt of the whole world. If I were to continue to teach in this Città Universitaria under these conditions, I would feel responsible for betraying not only my conscience, but also the memory of all those who fought those barbaric ideologies and who worked to lift up Italy from the level to which she had fallen. Therefore I am forced to announce that, until the academic authorities rid the Città Universitaria of the groups of these hooligans, whose objectives are the denial of culture and of any social renovation, it will not be possible for me to go to the Institute of Mathematics."[46]

This decision of not to teach any more at the Institute of Mathematics was referred to in the daily newspapers of Rome as well as in national newspapers, some of which also carried ample sections of the letter to the Dean. The same day two self–styled students of Mathematics sent an insulting letter to Stampacchia, which he always kept among his papers.

In Rome Stampacchia was entrusted with the direction of the Istituto per le Applicazioni del Calcolo (IAC) of the National Research Council (CNR), in December 1968. He took office with the precise aim of strenghtening and updating the research sector of the Institute, by taking on young graduates as research workers, by appointing consultants of high level scientific competence and by encouraging visits of foreign fellowship holders and expert scholars, among whom we find the distinguished names of Lars Hörmander and Hans Lewy.

[46] Paese Sera, March 2nd, 1970, p. 4; Corriere della Sera, March 3th, 1970, p. 7.

With Hans Lewy he studied the question of the regularity of the solution of the obstacle problem for the Laplacian, by considering it as the minimum of the superharmonic functions which respect the constraint (*Ma*[65]) and focused his attention on nonlinear monotone operators defined by means of a vector field, proving, in the case of a Lipschitz continuous obstacle, the existence of a Lipschitz continuous solution which may exhibit further regularization properties (*Ma*[66]). This type of problems contains as a particular case the problem of a minimal surface which lies over an obstacle and assumes fixed boundary values, for which the study of the contact set was later continued by David Kinderlehrer.[47]

Variational inequalities were also the subject of his talks in two international meetings of great relevance. He was one of the invited speakers at the International Conference on Functional Analysis and Related Topics organized in Tokyo by the International Mathematical Union and the Mathematical Society of Japan, in April 1969. In his talk (*Ma*[63]) he gave an exposition of the theorems of existence and of regularity he had obtained with Hartman, Brezis and Hans Lewy, followed by a treatment of several examples. He gave a special lecture (*Ma*[63]) at the International Congress of Mathematicians held in Nice in September 1970, where he presented the main techniques used in problems of variational inequalities, such as the method of penalization for approximating the solution, the use of the lemma of Minty for passing to the limit, and the introduction of pseudo-monotone operators due to Brezis.

The main guiding principle that inspired him in the activities at the Istituto per le Applicazioni del Calcolo was his belief, expressed on several occasions, that the distinction between pure and applied mathematics, being very vague and variable with time, was artificial. In his own words: "To take a right attitude towards mathematics it is necessary to reject a distinction (because it is dangerous) between pure mathematicians and applied mathematicians."[48] He was also very concerned with the influence that powerful industrial groups, increasing in number, could exercise to direct the research work, thus undermining the freedom of science from economic power: "The life of research centers and of advanced teaching institutions is necessary at the present moment, in which Universities seem to be unable to carry out this task.

[47] D. Kinderlehrer, *How a minimal surface leaves an obstacle*, Acta Mathematica 130, 1973, pp. 221-242.

[48] G. Stampacchia, *Matematica pura e applicata negli sviluppi attuali*, introductory report to the Congress "Rapporti tra ricerca pura ed applicata in Italia", Siena, 27-29 September 1973.

The task of research and of preparation of qualified research workers could have been carried out by the National Research Council through some of its institutes. But this task seems now difficult to realize also because a good deal of the funds that the Country has invested for scientific and technological research will end up funding research activities more industry related, thus neglecting fundamental research which would have permitted the formation of real research workers, and not just simple users of national or foreign industrial products."[49]

The experience at the IAC came to an end in January 1971 when, during the time of a visit of three months to Berkeley, a Commissioner was nominated by the President of CNR to substitute him as the Director of the Institute. Stampacchia experienced a great bitterness because of this event and made every effort to remove the shadow that he felt he was under due to his dismissal – even filing a complaint before the Consiglio di Stato (Supreme Administrative Court). This matter ended only in 1975, when Sandro Faedo having become the President of CNR, both parties reached an agreement which Stampacchia, in a letter, considered "satisfactory not so much from the material point of view but rather for my moral condition, since our own past vicissitudes influence the future of every one of us. When one reaches my age, when one thinks of having given to scientific culture as much as one could give, with the impression that in the future one will no more have similar opportunities, at this age it is a pleasure to receive also formal acknowledgements. If instead of this, one suffer an abuse of power, believe me, one feels very depressed."[50]

On reading again Stampacchia speeches, one is often impressed with the extreme lucidity and incredible up-to-dateness of some of his statements: in the introductory report prepared in September 1973 for the Congress "Relations between pure and applied mathematics in Italy", with the aim to denounce the bureaucratization of research and of scientific policies in Italy, he renamed the CNR, National Council for Research, as National Council for Reports, and defined "Comprogresso" ("Comprogress"), the process by which the reforms are often carried out in Italy. He wrote: "One of the fundamental processes put in execution in Italian life can be baptised with the name Comprogress. It is the result of different demands of progress very much heard of, and of their compromise. One of the illustrious victims of Comprogress is the Italian University."

[49] Autograph preparatory manuscript for the inaugural speech for the IX UMI Congress, Bari, 1971.

[50] G. Stampacchia to his lawyer M. Tarello, 29 June 1973, rough draft of an autograph letter.

4. ONCE AGAIN AT PISA: BACK TO THE SCUOLA NORMALE.

Having been invited, with a unanimous vote of the Council of Direction of the Scuola Normale Superiore, including the student representatives, to assume the position of professor for the chair of Analisi Superiore, Stampacchia returned to Pisa starting on November 1st, 1970.

Together with his colleagues De Giorgi and Vesentini, he strove to maintain the international prestige reached by the Scuola Normale and to create a lively scientific atmosphere with frequent visits of foreign mathematicians, among wich we mention the prestigeous names of J. Leray, H. Lewy, J.L. Lions, L. Nirenberg, O. Oleinik, D.G. Aronson, H. Brezis, R. Finn, J. Serrin, M.F. Atiyah, and D. Edmunds. Moreover, he found a new pleasure in teaching: in addition to his course on Analisi Superiore, he taught Analisi Matematica to the students of Physics at the University and then, he also had the charge of the course of Matematica I for the second year students at the Scuola Normale.

We have valuable notes of some of his courses at the Universities of Pisa, of Rome and at the Scuola Normale, notes which have been used partially for his books. We recall, in particular, the notes from his lectures on second order elliptic equations of 1963-64, those of 1967-68 on the theory of ordinary differential equations, which later became the subject of a joint volume with L. C. Piccinini and G. Vidossich (*Ma*[85]), and finally the notes on variational inequalities of 1970-71.

During these years his scientific activity was devoted to variational inequalities. During his stay at Berkeley in 1971, he studied with Alfonso Vignoli, a problem with non Lipschitz continuous obstacle (*Ma*[67]) and in his talk[51] at the Conference on Theory of Ordinary and Partial Differential Equations, held in Dundee, Scotland, in March 1972, he made a survey of the recent results obtained by him and his school. In Pisa, in collaboration with Murthy, he considered an obstacle problem with mixed boundary conditions for a second order elliptic operator (*Ma*[68]). In this paper, which he presented at the International Symposium on Partial Differential Equations and the Geometry of Normed Linear Spaces, held in Jerusalem in June 1972, the regularity of the solution was proved making use preliminarily of the truncation method and then of some nonlinear approximations. Later, in September 1972, in a joint paper with Brezis and Nirenberg, he obtained the extension of some existence results for variational inequalities to the case

[51] G. Stampacchia, *Recent results in the theory of variational inequalities*, Lecture Notes in Math., 280, 1972, pp. 147-153.

of functions of two variables defined in a cartesian product $C \times C$ of a convex set C in a topological vector space, and a minimax principle (*Ma*[71]).

In addition to the theoretical aspects examined so far, Stampacchia had considered also numerical and applied aspects of variational inequalities, both as a consultant of the Istituto per l'Elaborazione dell'Informazione of CNR at Pisa and in some of his publications and internal reports. With Otello Mancino, he studied variational inequalities for monotone operators on convex sets in finite dimensional spaces and gave an algorithm to find the solution of the convex programming problem in a polyhedron (*Ma*[64]).[52] At the Convegno di Analisi Numerica, held in Rome at the Istituto Nazionale di Alta Matematica in January 1972, he examined the problem of the numerical treatment of variational inequalities with an obstacle, approximating the solution with the help of two sequences of solutions of nonlinear equations, one increasing and the other decreasing, thus allowing to estimate the error introduced in the approximation (*Ma*[69]).

A very important feature of variational inequalities associated to elliptic differential operators is their connection with free boundary value problems, a connection put in evidence in Stampacchia's first paper with Hans Lewy (*Ma*[59]) and expounded in his talk at the Conference "Metodi Valutativi della Fisica Matematica" held in Rome at the Accademia dei Lincei in December 1972 (*Ma*[73]). In this talk Stampacchia described the clever trick with which Claudio Baiocchi, in 1971, had transformed the problem of filtration through a porous dam into one of variational inequality.[53] In addition, he took up the study of the stationary irrotational subsonic plane motion of a compressible fluid around a symmetric convex profile, which he had reduced to a free boundary problem associated to a variational inequality in the hodograph plane,[54] in a joint paper with Brezis (*Ma*[72]). This problem had been proposed by the Department of Rational Mechanics of the Politecnico di Torino to the Istituto per le Applicazioni del Calcolo; the interest of the solution method lies in the fact that it allows also the numerical calculation of the solution. Also the Exposé at the Séminaire Goulaouic-Schwartz at the École Polytechnique, in December 1972, was dedicated to questions of this type studied in a joint paper with Brezis. This

[52] The paper Ma [74] collects together the notes of April 1973, taken during a seminar of Stampacchia on this work.

[53] C. Baiocchi, *Su un problema di frontiera libera connesso a questioni di idraulica*, Ann. Mat. Pura Appl., 92, 1972, pp. 107-127. A preliminary note was published in 1971 in Comptes Rendus of the Academie des Sciences of Paris.

[54] See Lions 1978, pp. 794-750 (XLVI-XLVII).

exposition takes up the method of Baiocchi and investigates the motion of a fluid around a convex profile in the incompressible case. This latter study, appropriately developed and completed, was presented at the Conference Sur les Applications de l'Analyse Fonctionnelle aux Problèmes de Mécanique, held in Marseille in September 1975[55], and eventually published in the Archive for Rational Mechanics and Analysis in 1976 (*Ma*[78]).

During these years Stampacchia drew up the project to write a book on variational inequalities, later carried out in collaboration with D. Kinderlehrer (*Ma*[86]).

After leaving the presidency of UMI, Stampacchia occupied himself actively with the task of promoting mathematical research. He accepted the nomination as Director of the Scuola Superiore di Analisi matematica of the International Center of Scientific Culture Ettore Maiorana at Erice. On taking this office, he indicated as his main objective that of encouraging contacts among Italian and foreign specialists in various disciplines and between specialists and young research workers. To this purpose, he was one of the organizers of two courses on the theory, development and recent applications of variational inequalities. The first was held at Erice in March 1975, the second, in 1978, unfortunately turned out to be the first Conference dedicated to his memory. Moreover, he took part as teaching member, in the course on Mathematical and Numerical Methods in Fluid Dynamics, organized by the International Center for Theoretical Physics at Trieste.

There was a forced period of rest in the autumn of 1973, due to serious heart problems; but within a few months he returned to his intense teaching and scientific activities, and also to frequent visits abroad, ignoring the advice of his doctors to entertain a more relaxed and restful life style.

In November 1973 Stampacchia received, with great personal satisfaction, the invitation to give a talk on Hilbert's twenty-third problem, extensions of the Calculus of Variations, at the Mathematical Symposium on Developments Arising from Hilbert's Problems promoted by the American Mathematical Society at De Kalb, Illinois. In his lecture given on 16th May 1974, he described the history of Calculus of Variations starting from the problem of solids with minimum resistance – considered by Newton at the end of the 17th century – to the Dirichlet principle, from the contributions of Beppo Levi, Fubini and Lebesgue in Rendiconti del Circolo Matematico di Palermo to the development of

[55] H. Brezis - G. Stampacchia, *The odograph method in fluid-dynamics in the light of variational inequalities*, in: Applications of methods of functional analysis to problems in mechanics, Lecture Notes in Math., 503, 1976, pp. 239-257.

direct methods. Further he examined the relation with elliptic equations and the extension of variational methods, from the theory of partial differential equations to variational inequalities, from non linear functional analysis to problems of optimal control (*Ma*[79]).

He took the Editorial Directorship of the Science Section of the Annali della Scuola Normale Superiore in 1974. His first task was to include in the editorial committee mathematicians of great reputation from different countries such as S. Agmon, J. Leray, J.L. Lions, L. Nirenberg, in order to raise the journal to high international standards and prestige. At that time he was also a member of the editorial committees of Advances in Mathematics, Applied Mathematics and Optimization, and of Calcolo, and he was a member of the selection committee for invited speakers, presided over by Jean Leray, at the International Congress of Mathematicians held at Moscow (1966) and then at Vancouver (1974).

Stampacchia kept his interest in free boundary problems and considered, in a paper dedicated to the memory of Ivan G. Petrovskii, the motion of a fluid in a porous medium, for instance, water in an earth dam, in a model which could not be reduced to a two dimensional problem (*Ma*[75]). He reduced the three-dimensional problem to a variational inequality, and established the regularity of the solution; furthermore, he showed that the free surface is the graph of a function, leaving open the problem of its regularity, which was later on treated in a general way by Hans W. Alt.[56] In addition to his studies on the regularity of solutions of variational inequalities associated to second order operators recalled so far, Stampacchia considered also operators of fourth order, and established the regularity of solutions of some variational inequalities.[57] As a first step, he examined a one-dimensional model representing the elastoplastic behaviour of a beam, in a paper dedicated to G. Sansone for his eightyfifth birthday (*Ma*[76]); then he studied, in a joint paper with Brezis, a variational inequality for the biharmonic operator in n variables (*Ma*[80]).

In August 1975 he was invited to the Netherlands for the second Scheveningen Conference on New Developments in Differential Equations. There he presented a joint work with Kinderlehrer[58] where they examined a free boundary problem for the Poisson equation in the plane through a variational inequality approach and they proved that the

[56] H.W. Alt, *The fluid flow through porous media. Regularity of the free surface*, Manuscripta math., 21, 1977, pp. 255-272.

[57] Cfr. Lions 1978, pp. 750-751 (XLVII-XLVIII).

[58] G. Stampacchia, *Free boundary problem for Poisson's equation*, New developments in differential equations, 1976, pp. 39-42.

unknown curve, on which conditions of both Dirichlet and Newman types are simultaneously imposed, is regular (*Ma*[77]). Finally, a joint paper with Brezis and Kinderlehrer, which was published posthumously, was once again dedicated to the problem of filtration through a porous dam.[59] In this article, a new formulation was proposed for the problem, whose solution coincides with that of Baiocchi, in the case of a rectangular dam (*Ma*[81]).

Stampacchia also wrote textbooks for university courses and articles for popular scientific publications, such as the ones issued by Istituto Geografico De Agostini and Enciclopedia Einaudi.

His total rejection of any superficially demagogical attitude, his feeling of distrust and his pessimism for the future of the Italian university system, which had already characterized his choices and his speeches, were once more evident towards the end of 1977. At that time, in a renewed climate of violence and uncertainty, he was called upon to preside over a Ministerial Investigating Commission, to evaluate the behaviour of a teacher of Mathematical Analysis and Analytical Geometry at the Faculty of Architecture of the Polytechnic of Milano, where several irregularities had been denounced. The commission, in the report presented to the Minister, interpreted the facts that had taken place as "a symptom of the great uneasiness that one perceives in our universities" after more than ten years of waiting for an organic reform law, and remarked that the legislative measures which had been passed had created "a situation far worse than that which was intended to be remedied". In a handwritten draft, Stampacchia judged the attitude of the entire faculty as "an affront to those who want to defend the principle of the autonomy of the university system, a principle on which the University, its progress and its cultural tradition are based."

He spent a considerable amount of his energies also to make known abroad his research activity, giving seminar talks on the theory of partial differential equations and variational inequalities: he was a visiting professor at the University of Sussex from October to December of 1971 and in the month of Febuary 1976 and he spent a month in Paris at Collège de France between May and June 1976. He also went to the United States for three months from March to May of 1977 as visiting professor at the Courant Institute in New York and at the School of Mathematics of the University of Minnesota at Minneapolis. At last he returned to Paris around the middle of February 1978, as a visiting professor for two months at the University Pierre et Marie Curie, to give a course on partial differential equations.

[59] Cfr. Lions 1978, pp. 751-752 (XLVIII-XLIX).

In Paris he suffered once more a serious heart attack, following which he was admitted to Boucicaut Hospital. Just when the evolution of the illness seemed to be taking such a satisfactory course that he was authorized to return home, Stampacchia expired due to a sudden heart arrest, on 27th April 1978, the same day he was to be discharged from the hospital. He was 56 years old. According to a wish he had expressed many times, he was buried in the British Cemetery in Naples.

On 17th May 1978, the Council of Directors of National and International Schools of the Centro di Cultura Scientifica Ettore Maiorana at Erice decided to honour his memory by dedicating to his name the International School of Mathematics, which hosts the present Conference, and by founding a fellowship to be awarded to a young mathematician to attend the School.

REFERENCES

[1] Cesari L. 1985, *L'opera di Leonida Tonelli e la sua influenza nel pensiero scientifico del secolo*, Convegno celebrativo del centenario della nascita di Mauro Picone e Leonida Tonelli, Atti Convegni Lincei 77 (1986), pp. 42-73.

[2] De Giorgi E. 1978, *La scomparsa a Parigi di Guido Stampacchia*, La Nazione, 4 maggio, p. 4.

[3] De Giorgi E. 1980, *Guido Stampacchia*, Rend. Accad. Naz. Lincei, s. 8, v. 68, pp. 619-625.

[4] De Giorgi E. 1985, *Su alcuni indirizzi di ricerca nel calcolo delle variazioni*, Convegno celebrativo del centenario della nascita di Mauro Picone e Leonida Tonelli, Atti Convegni Lincei 77 (1986), pp. 183-187.

[4] Faedo S. 1985, *Leonida Tonelli e la scuola matematica pisana*, Convegno celebrativo del centenario della nascita di Mauro Picone e Leonida Tonelli, Atti Convegni Lincei 77 (1986), pp. 89-109.

[6] Lions J. L. 1978, *The work of G. Stampacchia in variational inequalities*, Boll. UMI, v. 15-A, n. 3, pp. 736-753 (reprinted in Stampacchia 1996-97, vol. II, pp. XXXIII-L).

[7] Kinderlehrer D. 1988, *Un uomo che pensava agli altri*, L'Unità, 26 aprile, p. 18.

[8] Magenes E. 1978a, *Guido Stampacchia (1922-1978)*, Boll. UMI, v. 15-A, n. 3, pp. 715-736 (reprinted in Stampacchia 1996-97, vol. I, pp. XI-XXXII and vol. II pp. XI-XXXII).

[9] Magenes E. 1978b, *Pubblicazioni scientifiche di Guido Stampacchia*, Boll. UMI, v. 15-A, n. 3, pp. 753-756 (reprinted in Stampacchia 1996-97, vol. I, pp. XXXII-XXXV and vol. II, pp. LI-LIV).

[10] Magenes E. 2000, *Introduzione alla storia di un secolo di matematica alla Normale (1862-1961)*, Boll. Associazione Normalisti, anno III, n. 1, pp. 20-29.

[11] Stampacchia G. 1996-97, *Opere scelte a cura dell'Unione Matematica Italiana con il contributo del Consiglio Nazionale delle Ricerche*, 2 vols., Cremonese, Firenze.

[12] Vesentini E. 1980, *Commemorazione del professor Aldo Andreotti*, Scuola Normale Superiore di Pisa, 2 maggio.

MEMORIES OF GUIDO STAMPACCHIA

L. Nirenberg
Courant Institute, New York, NY, USA

I first met Guido at a conference in 1954 in Trieste, organized by Professor Fichera. This was, essentially, my first contact with Italian analysts working in partial differential equations. Among those were Roberto Conti, Enrico Magenes, Carlo Miranda, Carlo Pucci and Francesco Tricomi. This meeting was the beginning of a long friendship with Italian colleagues and a love affair with Italy. Guido and I became very close friends – to me he was like a brother – our warm friendship continued up to his demise.

It was always a great pleasure to meet him. We would always talk about mathematics, cinema, politics, etc. His enthusiasm and love of mathematics was contagious. He affected everyone who came into contact with him.

It was Guido who pointed out to Ennio De Giorgi the problem of extending to higher dimensions Morrey's regularity result in 2 dimensions for elliptic variational problems. Afterwards Guido extended De Giorgi's results up to the boundary in his beatiful 1966 lecture notes at the University of Montreal. These notes have had a great influence.

In 1958-59 I spent a sabbatical year in Rome and Guido and I organized a seminar which met once a month, over a weekend in Pisa. Over the years, when I visited him and his family in Pisa I was like lo "zio d'America".

Though we discussed mathematics all the time we wrote only one joint paper, together with Brezis, in 1972.

Guido played a very important role in creating scientific contact between Italy and other countries. He visited the Courant Institute several times, for extended periods.

Guido was full of life and humour and I have wonderful memories of our walks, talks and dinners together. He continues to live in the hearts of all who knew him.

IN MEMORY OF GUIDO STAMPACCHIA

C. Sbordone
Dept. of Mathematics and Applications "R. Caccioppoli ", Napoli, Italy

On behalf of the Unione Matematica Italiana I am honored to address a few words on the occasion of the International Conference "Variational Analysis and Applications" in memory of Guido Stampacchia.

Even though the fame of Guido Stampacchia goes beyond the boundary of a national scientific society, I wish to give testimony to his contribution to the developement of the U.M.I.

From 1964 to 1976 he was a member of the Scientific Committee of the UMI. He served as a President from 1967 to 1973.

In that period he dedicated a special care to the Bulletin of the UMI, that started a new series in 1968. The level of the papers highly improved; longer articles were accepted; the referees system was introduced. In 1973 the circulation of the Bulletin increased up to 2400 copies, and the number of the UMI members doubled with respect to the 1967.

He gave talks at many national Congresses of the UMI: in Taormina in 1951, at the age of 29, he was award one of the eight prizes for the best talks of young assistants, in Napoli in 1959 he gave a plenary address.

During his career of a distinguished mathematician he received important recognitions like the invitation to lecture at the International Congresses of Mathematicians in Stocholm in 1962 and in Nice in 1970 and the Feltrinelli Prize from the Accademia dei Lincei in 1966.

Stampacchia was one of the first among Italian Mathematicians who cooperated intensively with foreign colleagues, in particular in France with J.L. Lions and H. Brezis and in US with L. Nirenberg, H. Weinberger, H. Lewy, W. Littman, P. Hartman and D. Kinderlehrer, writing many joint

papers and publishing them in prestigious international Journals (Comm. PAM, CRAS, Ann. Fourier, Acta Math, Archive).

The numerous quotations of his work are evident in the recent annual issues of the Citation Index, where his name appears at least twenty times per year, and this clearly demonstrate that his ideas are still very much alive.

On a personal side I am pleased to acknowledge the influence his papers had on me and my mathematical work and his generous advices during my stay at the Scuola Normale Superiore in 1974-75, a period when there was in Pisa a continuous flow of analysts from many countries due to his and De Giorgi's prestigious personalities.

A selection of his most famous papers has been published into two volumes by the UMI, sponsored by CNR and Scuola Normale Superiore of Pisa, to allow joung mathematicians to know his very important work in Pde's and Calculus of Variations. The idea of publishing these volumes was supported by italian and foreign mathematicians, some of them are present at this memorial meeting.

I take now the opportunity to express my warm congratulations to the winner of the First "Stampacchia Medail", prof. Tristan Riviere.

GUIDO STAMPACCHIA, MY FATHER

G. Stampacchia
Dept. of Neurosciences, University of Pisa, Pisa, Italy

I am grateful, and with me my family, to prof. Franco Giannessi for his invitation to this commemoration, and to prof. Silvia Mazzone for her excellent biography of my father as a man and as a mathematician.

I also wish to thank all the presents and particularly all my father's friends that have spoken before me, tracing a profile of my father, both from the point of view of his scientific work and of his human profile.

I am bringing you in particular the thankfulness of my mother that has not been able to be here in person, but is present here in heart and has recommended me to greet with great affection all my father's –and so also hers- friends.

I must confess that I am in a certain sense feeling uneasy in speaking in front of an audience of "real" scientists. I have had many times , as a medical doctor and researcher in the field of neurology, the occasion to deliver a paper in a scientific meeting, but my audience in those occasions is usually made up of a different sort of "scientists". I cannot avoid to remember the words of my father, when I told him I wanted to become a student in the Medical faculty: "Didn't you say you wanted to enroll in a *scientific* Faculty?". He used to remark that Medicine wasn't what he thought be a science, he saw it with a lot of "witchery" in it. But in effect things have changed since those times, and mathematics has more and more been applied to medical research. Eventually, my father was proud, as I learned in the years, of my medical studies.

Silvia Mazzone writes, quite agreably, that my father was a man of humour and anticonformist attitude. Magenes has spoken of "Naepolitan irony". This aspect of my father's personality was at times not so easy for

me, my brother and sisters (a bunch of four altogether) because we had to face a model very difficult to imitate. Many times my father, when he heard from me something that could sound banal, conformist, what in Italian we call a "frase fatta" (a "stock" or a "set phrase" you would call it in English), told me in a very serious way not to speak *fesserie*, a word much used in Southern Italy to indicate things of no wit and intelligence. His critical spirit determined in us a very strict or better to say a strong education. De Giorgi has said that my father "was never indulgent towards superficiality and laziness", and I believe that that is true not only for my father as a teacher, but also for my father as a father. But he was never grave, or rethorical, or pompous, in the way he behaved but nevertheless showed a deep commitment to seriousness only edulcorated with his ironic wit.

My sister Renata still remembers how, when she finally reached the goal of her degree in Physics, my father, in the United States for work, sent her a congratulation card. The card portrayed a big question mark and he wrote underneath: "Congratulations... and now?".

My father always stimulated us to study with a serious attitude and he wanted good results, that he thought to be in any case due. The fact that we chose to study –at any level, from elementary to University- implied commitment and good results. There were no prices for good qualifications and passing to following classes. He did not, on the other hand, reproached us if we did not achieve good results, but encouraged us. Once I had some "price" for my studies, and I remember well the occasion, because it marked quite a turn in the relation between me and my father. It was when, at my first University exam, I had the maximum vote cum laude. My father bought me a box of chocolates and added some 30 thousand lire as a bonus, quite a lot of money in 1974 Italy. I understood from now then, that my father was beginning to consider me an adult, and started to talk with me as an adult. Unfortunately, due to his premature death, this new period of a Guido-adult Giulia father-daugther relation was bound to last only a few years.

My father was particularly fond of his Neapolitan origins and background. Now he rests, as he desired, in the British Cemetery in Naples, the Protestant Cemetery.

My mother Sara once told me of a love declaration of my father. He said to her: "three things are important for me in my life: 1. Mathematics, 2. the city of Naples, 3. Sara". He liked particularly the food and the Naepolitan dialect. When not in Naples he avoided speaking Naepolitan, and my father spoke Italian without any local accent, but when he was back in Naples he loved to speak Naepolitan. But some word of Naepolitan origin he still used: I remember when I was in junior high school I happened to write, in a composition about my family, that my father, when he came home from work, " *si sparapanzava* on the easy chair". I heard that word used normally

in my family, and I used it in the composition thinking it were a correct Italian word, but my teacher thought it was a mistake and in effect I had to admit there was not such a word in the Italian dictionary. But when I explained it to the teacher, she appreciated such an onomatopoeic word and asked me to keep it in the text, only putting it inside quotation marks.

Neapolitanity, if you pass the word, and his critical and anticonformistic spirit, were the origin of two of his most common words he used. To define a person of mediocre spirit, but quite engaged and convinced of having great cultural prestige, used to say with some benevolence that he was a "*fessacchiotto*". He was much more critical towards people with arrogance, conformist, and speaking with made-up and rethorical phrases. Those were, with no appeal, in a very coloured Naepolitan expression, "*mezze calzette*".

Neapolitan food was also much appreciated. I remember the "*Sara, u'cafè*" request, that meant "please Sara make a good coffee for me". He asked for pizza at meals, and sometimes he liked to cook himself. I remember particularly the tasty *lasagne alla napoletana* or, alternatively, with a meat ball and ricotta sauce.

Naples was also the city of his great *maestro* Caccioppoli, whose photographic portrait was quite in evidence in my father's study, drawing the attention of all visitors.

Although my father was in all sense Neapolitan, he was in the same time very open to the world. He often travelled in Europe, America and Asia also, and opened our house to many foreigners. So we were brought up with people from all the world, because he invited foreigners not only at dinners, but also as guests in our house and as travel companions. I have recently found old photos taken when I was a child and in one of them there was professor Murthy and my family in Assisi.

Being open to the world is one of the teachings my father gave to all of us. We have learned it very well because it given out not only in words but in practice. My father did not state equality between peoples but made us live for periods outside of Italy and brought many persons from all the parts of the world in our house.

Guido Stampacchia didn't love only mathematics, Naples and his wife. He was a man full of interests and he also loved to play. His hobbies were not banal but had to do with creativeness. He loved taking photos and still nowadays my mother has drawers full of photos of us, in the various ages; he loved cinema and he not only had a good time making family movies, but also short documentaries and cartoons. When the Lego brick games was put on the market, my father bought one that had different pieces, bricks, windows and doors. He was very jealous of his constructions, and we were not allowed to touch them. He made a short animation movie that I remember to have seen with a lot of fun and also amazement when I was

child (we are around the Sixties), in which a small house built itself from nothing with Lego bricks. In later times my father's love for the house and garden had satisfaction when in 1966, when he won the Feltrinelli prize, he bought a house in Ronchi of Marina di Massa, where he spent his holidays, that were actually periods of intellectual and professional work. In fact, while us children went to the beach, he often spent days in the garden to think mathematics. He sometimes invited colleagues to work with him in the peace of his garden.

I would like to say some final words on my father as a teacher. Mazzone says about Guido Stampacchia that "he always had towards students an available and generous attitude". In effects I remember that in the Eighties as a Perfezionanda in Neurofisiologia in the Scuola Normale Superiore in Pisa, I made friends with some young mathematicians that had been students with my father, and I was struck from the way they remembered with affection Guido Stampacchia and they expressed me a sort of envy for my having had such a father.

I would like to conclude thanking again the organizers of this ceremony for having invited me to actively participate. I have to confess that this occasion has been for me and my family –mother, brother and sisters- a moment of reflection. We were brought to gather to share the individual memoirs of every one of us concerning the moments and the life spent with Guido Stampacchia, husband and father.

PART 2

CONVERGENCE AND STABILITY OF A REGULARIZATION METHOD FOR MAXIMAL MONOTONE INCLUSIONS AND ITS APPLICATIONS TO CONVEX OPTIMIZATION

Ya. I. Alber,[1] D. Butnariu[2] and G. Kassay[3]
Faculty of Mathematics, Technion - Israel Institute of Technology, Haifa, Israel;[1] Dept. of Mathematics, University of Haifa, Haifa, Israel;[2] Faculty of Mathematics, University Babes-Bolyai, Cluj-Napoca, Romania[3]

Abstract: In this paper we study the stability and convergence of a regularization method for solving inclusions $f \in Ax$, where A is a maximal monotone point-to-set operator from a reflexive smooth Banach space X with the Kadec-Klee property to its dual. We assume that the data A and f involved in the inclusion are given by approximations A^k and f^k converging to A and f, respectively, in the sense of Mosco type topologies. We prove that the sequence $x^k = (A^k + \alpha_k J^\mu)^{-1} f^k$ which results from the regularization process converges weakly and, under some conditions, converges strongly to the minimum norm solution of the inclusion $f \in Ax$, provided that the inclusion is consistent. These results lead to a regularization procedure for perturbed convex optimization problems whose objective functions and feasibility sets are given by approximations. In particular, we obtain a strongly convergent version of the generalized proximal point optimization algorithm which is applicable to problems whose feasibility sets are given by Mosco approximations

Key words: Maximal monotone inclusion, Mosco convergence of sets, regularization method, convex optimization problem, generalized proximal point method for optimization.

2000 Mathematics Subject Classification: Primary: 47J06, 47A52, 90C32; Secondary: 47H14, 90C48, 90C25.

1. INTRODUCTION

Let X be a reflexive, strictly convex and smooth Banach space with the Kadec-Klee property (i.e., such that if a sequence $\{x^k\}_{k\in\mathbb{N}}$ in X converges weakly to some $x\in X$, then $\{x^k\}_{k\in\mathbb{N}}$ converges strongly whenever $\lim_{k\to\infty}\|x^k\|=\|x\|$) and let X^* be the dual of X. Given a maximal monotone mapping $A: X\to 2^{X^*}$ and an element $f\in X^*$, we consider the following problem

$$\text{Find } x\in X \text{ such that } f\in Ax. \tag{1}$$

Problems like (1) are often ill-posed in the sense that they may not have solutions, may have infinitely many solutions and/or small data perturbations may lead to significant distortions of the solution sets. A *regularization technique,* whose basic idea can be traced back to Browder [16] and Cruceanu [23], consists of replacing the original problem (1) by the problem

$$\text{Find } z^\alpha\in X \text{ such that } f\in(A+\alpha J^\mu)z^\alpha, \tag{2}$$

where α is a positive real number and $J^\mu: X\to X^*$ is the duality mapping of gauge μ defined by the equations

$$\langle J^\mu y, y\rangle=\|J^\mu y\|_*\|y\| \text{ and } \|J^\mu y\|_*=\mu(\|y\|), \tag{3}$$

while $\mu:[0,+\infty)\to[0,+\infty)$ is supposed to be continuous, strictly increasing, having $\mu(0)=0$ and $\lim_{t\to\infty}\mu(t)=+\infty$. One does so for several reasons. First, since the mapping $A+\alpha J^\mu$ is surjective and $(A+\alpha J^\mu)^{-1}$ is single valued (cf. [22, Proposition 3.10, p. 165]), the regularized problem (2) has unique solution (even if the inclusion (1) has no solution at all). Second, it follows from [47, p. 129] and [23] that, if $\{\alpha_k\}_{k\in\mathbb{N}}$ is a sequence of positive real numbers and $\lim_{k\to\infty}\alpha_k=0$ then by solving (2) for $\alpha=\alpha_k$ one finds vectors z^{α_k} converging to a solution of (1) provided that this inclusion is consistent. Third, the operator $(A+\alpha J^\mu)^{-1}$ is continuous and, therefore, small perturbations of f will not make the vector z^α be far from the theoretical solution $(A+\alpha J^\mu)^{-1}f$ of (2). In applications it frequently

happens that not only f but also the operator A involved in (1) can be approximated but not precisely computed. This naturally leads to the question whether the regularized inclusion (2) is stable, that is, whether by solving instead of the regularized inclusion (2) a sequence of regularized inclusions

$$f^k \in (A^k + \alpha_k J^\mu)x$$

in which $A^k : X \to 2^{X^*}$ are maximal monotone operators approximating A and f^k approximates f, the sequence of corresponding solutions

$$x^k = \left(A^k + \alpha_k J^\mu\right)^{-1} f^k \tag{4}$$

still converges to a solution of (1) when $\lim_{k\to\infty} \alpha_k = 0$ and the original inclusion (1) is consistent. This question was previously considered by Lavrentev [36] who dealt with it in Hilbert spaces under the assumption that A is linear and positive semidefinite, $\mathrm{Dom}\, A = X$ and $\mu(t) = t/2$. In Alber [1] the problem appears in a more general context but under the assumption that the operator A is defined on the whole Banach space X.

The main purpose of this paper is to show that if the approximations A^k and f^k satisfy some quite mild requirements, then the answer to the question posed above is affirmative, i.e., the sequence $\left\{x^k\right\}_{k\in\mathbb{N}}$ defined by (4) converges strongly to the minimal norm solution of (1) as $\alpha_k \to 0$ and provided that (1) has at least one solution. Subsequently, we prove that the stability results we have obtained for the regularization method presented above apply to the resolution of convex optimization problems with perturbed data and, in particular, to produce a strongly convergent version of a proximal point method.

The stability results proved in this work (see Section 2) do not make additional demands on the data of the original inclusion (1) besides the assumption that A is maximal monotone. The conditions under which we prove those results only concern the quality of the approximations A^k and f^k. They ask that either the Mosco weak upper limit (as defined in [45]) or the weak-strong upper limit (introduced in Subsection 2.1 below) of the sequence of sets $\left\{Graph\left(A^k\right)\right\}_{k\in\mathbb{N}}$ be a subset of $\mathrm{Graph}\left(A\right)$, the later being a somewhat weaker requirement. Also, they ask for a kind of linkage of the approximative data in the form of the boundedness of the sequence

$$\left\{\alpha_k^{-1}\mathrm{dist}_*\left(f^k, A^k v^k\right)\right\}_{k\in\mathbb{N}} \tag{5}$$

for some bounded sequence $\{v^k\}_{k\in\mathbb{N}}$ in X. If approximants A^k and f^k satisfying these conditions exist, then the inclusion (1) is necessarily consistent, the sequence $\{x^k\}_{k\in\mathbb{N}}$ defined by (4) is bounded and its weak accumulation points are solutions of it (see Theorem 2.2 and Corollary 2.3). The main stability results we prove for the proposed regularization scheme are Theorem 2.4 and its Corollary 2.5. They show that if solutions of (1) exist and each of them is the limit of a sequence $\{v^k\}_{k\in\mathbb{N}}$ such that the sequence (5) converges to zero, then the sequence $\{x^k\}_{k\in\mathbb{N}}$ given by (4) converges strongly to the minimal norm solution of (1).

When one has to solve optimization problems like that of finding a vector

$$x^* \in \operatorname{argmin}\{F(x) : g_i(x) \le 0,\ i \in I\}, \tag{6}$$

where the functions F, $g_i : X \to (-\infty, +\infty]$ are convex and lower semicontinuous, perturbations of data are inherent because of imprecise computations and measurements. Since problems like (6) may happen to be ill-posed, replacing the original data F and g_i by approximations F_k and g_i^k may lead to significant distortions of the solution set. In Section 3 we consider (6) and its perturbations in their subgradient inclusion form. We apply the stability results presented in Section 2 for finding out how "good" the approximative data F_k and g_i^k should be in order to ensure that the vectors x^k resulting from the resolution of the regularized perturbed inclusions strongly approximate solutions of (6). Theorem 3.2 answers this question. It shows that for this to happen it is sufficient that the perturbed data would satisfy the conditions (A) and (B) given in Subsection 3.1. Condition (A) asks for sufficiently uniform point-wise convergence of F_k to F. Condition (B) guarantees weak-strong upper convergence of the feasibility sets of the perturbed problems to the feasibility set of the original problem. Proposition 3.6 provides a tool for verifying the validity of condition (B) in the case of optimization problems with affine constraints as well as in the case of some problems of semidefinite programming.

In Section 4 we consider the question whether or under which conditions the generalized proximal point method for optimization which emerged from the works of Martinet [43], [44], Rockafellar [52] and Censor and Zenios [21] can be forced to converge strongly in infinite dimensional Banach spaces. The origin of this question can be traced back to Rockafellar's work [52]. The relevance of the question emerges from the role of the proximal point method in the construction of augmented Lagrangian algorithms (see [53], [18, Chapter 3] and [30]): in this context a better behaved sequence obtained by regularization of the proximal point method may be of use in

order to determine better approximations for a solution of the primal problem. It was shown by Butnariu and Iusem [17] that in smooth uniformly convex Banach spaces the generalized proximal point method converges subsequentially weakly, and sometimes weakly, to solutions of the optimization problem to which it is applied. However, it follows from the work of Güler [28] that the sequences generated by the proximal point method may fail to converge strongly. The generalized proximal point method essentially consists of solving a sequence of perturbed variants of the given convex optimization problem. We apply the results established in Section 3 in order to prove that by regularizing the perturbed problems via the scheme studied in this paper we obtain a sequence $\{(y^k, x^k)\}_{k\in\mathbb{N}}$ in $X \times X$ such that, when the optimization problem is consistent, $\{F(y^k)\}_{k\in\mathbb{N}}$ converges to the optimal value of F and $\{x^k\}_{k\in\mathbb{N}}$ converges strongly to the minimum norm optimal solution of the original optimization problem.

The stability of the regularization scheme represented by (2) was studied before in various settings, but mostly as a way of regularizing variational inequalities involving maximal monotone operators (which, in view of Minty's Theorem, can be also seen as a way of regularizing inclusions involving maximal monotone operators). Mosco [45], [46], Liskovets [39], [40], [41], Ryazantseva [54], Alber and Ryazantseva [6], Alber [2], Alber and Notik [5] have considered the scheme under additional assumptions (not made in our current work) concerning the data A and f (as, for instance, some kind of continuity or that the perturbed operators A^k and A should have the same domains). The stability results they have established usually require Hausdorff metric type convergence conditions for the graphs of A^k. Also under Hausdorff metric type convergence conditions, but with no additional demands on the operator A than its maximal monotonicity, strong convergence of the regularized sequence $\{x^k\}_{k\in\mathbb{N}}$ defined by (4) to the minimal norm solution of (1) was proven by Alber, Butnariu and Ryazantseva in [4]. Recently, weak convergence properties of this regularization scheme were proved by Alber [3] under metric and Mosco type convergence assumptions on the approximants. By contrast, we establish here strong convergence of the regularized sequence $\{x^k\}_{k\in\mathbb{N}}$ by exclusively using variants of Mosco type convergence for the approximants.

The stability of regularization schemes applied to ill-posed problems is a multifaceted topic with multiple applications in various fields as one can see from the monographs of Lions and Magenes [37], Dontchev and Zolezzi [24], Kaplan and Tichatschke [31], Engl, Hanke and Neubauer [27], Showalter [55], and Bonnans and Shapiro [14]. We prove here that the regularization scheme (4) has strong and stable convergence behavior under

undemanding conditions and that it can be applied to a large class of convex optimization problems. An interesting topic for further research is to find out whether and under which conditions this regularization scheme works when applied to other problems like, for instance, differential equations [55], inverse problems [27], linearized abstract equations [14, Section 5.1.3.], etc. which, in many circumstances, can be represented as inclusions involving maximal monotone operators. Convergence of the regularization scheme (4) may happen to be slow (as shown by an example given in [4]). Its rate of convergence seems to depend not only on the properties of A^k and f^k but also on the geometry of the Banach space X in which the problem is set. It is an interesting open problem to evaluate the rate of convergence of the regularization scheme discussed in this work in a way similar to that in which such rates were evaluated for alternative regularization methods by Kaplan and Tichatschke [34], [33], [32] and [42]. Such an evaluation may help decide for which type of problems and in which settings application of the regularization scheme (4) is efficient.

The convergence and the reliability under errors of the generalized proximal point method in finite dimensional spaces was systematically studied along the last decade (see [25], [26] and see [29] for a survey on this topic). In infinite dimensional Hilbert spaces repeated attempts were recently made in order to discover how the problem data should be in order to ensure that the generalized proximal point method converges weakly or strongly under error perturbations (see [8], [9], [15], [30]). Projected subgradient type regularization techniques meant to force strong convergence in Hilbert spaces of Rockafellar's classical proximal point algorithm were discovered by Bauschke and Combettes [12, Corollary 6.2] and Solodov and Svaiter [57]. The regularized generalized proximal point method we propose in Section 4 works in non Hilbertian spaces too. It presents an interesting feature which can be easily observed from Theorem 4.2 and Corollary 4.3: if X is uniformly convex, smooth and separable, then by applying the regularized generalized proximal point method (60) one can reduce resolution of optimization problems in spaces of infinite dimension to solving a sequence of optimization problems in spaces of finite dimension whose solutions will necessarily converge strongly to the minimal norm optimum of the original problem.

2. CONVERGENCE AND STABILITY ANALYSIS FOR MAXIMAL MONOTONE INCLUSIONS

2.1 We start our discussion about the stability of the regularization scheme (2) by recalling (see [45, Definition 1.1]) that a sequence $\{S_k\}_{k\in\mathbb{N}}$ of subsets of X is called *convergent* (*in Mosco sense*) if

$$w-\overline{\lim}\, S_k = s-\underline{\lim}\, S_k,$$

where $s-\underline{\lim}\ S_k$ represents the collection of all $y\in X$ which are limits (in the strong convergence sense) of sequences with the property that $x^k \in S_k$ for all $k\in\mathbb{N}$ and $w-\overline{\lim} S_k$ denotes the collection of all $x\in X$ such that there exists a sequence $\{y^k\}_{k\in\mathbb{N}}$ in X converging weakly to x and with the property that there exists a subsequence $\{S_{i_k}\}_{k\in\mathbb{N}}$ of $\{S_k\}_{k\in\mathbb{N}}$ such that $y^k \in S_{i_k}$ for all $k\in\mathbb{N}$. In this case, the set

$$S := s-\underline{\lim}\, S_k = w-\overline{\lim}\, S_k$$

is called the *limit* of $\{S_k\}_{k\in\mathbb{N}}$ and is denoted $S = \underline{\mathrm{Lim}} S_k$.

By analogy with Mosco's $w-\lim$ we introduce the following notion of limit for sequences of sets contained in $X\times X^*$. This induces a form of graphical convergence for point-to-set mappings from X to X^* which we use in the sequel. For a comprehensive discussion of other notions of convergence of sequences of sets see [13].

Definition. The *weak-strong upper limit of a sequence* $\{U_k\}_{k\in\mathbb{N}}$ of subsets of $X\times X^*$, denoted $ws-\overline{\lim}\ U_k$, is the collection of all pairs $(x,y)\in X\times X^*$ for which there exists a sequence $\{x^k\}_{k\in\mathbb{N}}$ contained in X which converges weakly to x and a sequence $\{y^k\}_{k\in\mathbb{N}}$ contained in X^* which converges strongly to y and such that, for some subsequence $\{U_{i_k}\}_{k\in\mathbb{N}}$ of $\{U_k\}_{k\in\mathbb{N}}$ we have $(x^k,y^k)\in U_{i_k}$ for all $k\in\mathbb{N}$.

It is easy to see that, if $A^k : X\to X^*$, $k\in\mathbb{N}$, is a sequence of point-to-set mappings then the weak-strong upper limit of the sequence $U_k = Graph\left(A^k\right)$, $k\in\mathbb{N}$, is the set U of all pairs $(x,y)\in X\times X^*$ with the

property that there exists a sequence $\{(x^k, y^k)\}_{k\in\mathbb{N}} \subset X \times X^*$ such that $\{x^k\}_{k\in\mathbb{N}}$ converges weakly to x in X, $\{y^k\}_{k\in\mathbb{N}}$ converges strongly to y in X^* and, for some subsequence $\{A^{i_k}\}_{k\in\mathbb{N}}$ of $\{A^k\}_{k\in\mathbb{N}}$ we have

$$y^k \in A^{i_k}(x^k),\ \forall k \in \mathbb{N}.$$

Therefore, in virtue of [11, Proposition 7.1.2.], the graphical upper limit of the sequence $\{A^k\}_{k\in\mathbb{N}}$, $\lim^{\#}_{k\to\infty} A^k$, considered in [11, Definition 7.1.1], the weak-strong upper limit $ws-\overline{\lim}\ Graph(A^k)$ and the Mosco upper limit $w-\overline{\lim}\ Graph(A^k)$ are related by

$$Graph\left(\lim^{\#}_{k\to\infty} A^k\right) \subseteq ws-\overline{\lim}\ Graph(A^k) \subseteq w-\overline{\lim}\ Graph(A^k). \qquad (7)$$

As noted in the Introduction, a goal of this work is to establish convergence and stability of the regularization scheme (4) under undemanding convergence requirements for the approximative data A^k and f^k. As far as we know, the most general result in this respect is that presented in [4, Section 2]. It guarantees convergence and stability of the regularization scheme (4) under the requirement that the maximal monotone operators A^k approximate the maximal monotone operator A in the sense that there exist three functions $a, g, \zeta : \mathbb{R}_+ \to \mathbb{R}_+$, where ζ is strictly increasing and continuous at zero, such that for any $(x,y) \in Graph(A)$ and for any $k \in \mathbb{N}$, there exists a pair $(x^k, y^k) \in Graph(A^k)$ with the property that

$$\|x - x^k\| \le a(\|x\|)k^{-1} \text{ and } \|y - y^k\|_* \le g(\|y\|_*)\zeta(k^{-1}). \qquad (8)$$

Clearly, if this requirement is satisfied, then

$$\mathrm{Graph}(A) \subseteq \mathrm{Graph}\left(\lim^{\flat}_{k\to\infty} A^k\right), \qquad (9)$$

where $\lim^{\flat}_{k\to\infty} A^k$ stands for the graphical lower limit of the sequence $\{A^k\}_{k\in\mathbb{N}}$ (see [11, p. 267]). Since the mappings A^k and A we work with are maximal monotone, Proposition 7.1.7 from [11] applies and, due to (9), it implies that A is exactly the graphical limit of the sequence $\{A^k\}_{k\in\mathbb{N}}$, that is,

$$A = \lim^{b}_{k\to\infty} A^k = \lim^{\#}_{k\to\infty} A^k. \tag{10}$$

In this section we show that convergence and stability of the regularization scheme (4) can be ensured under conditions that are much less demanding than the locally uniform graphical convergence (8). In fact, we prove convergence and stability of the scheme (4) by requiring (see (16) below) less than the graphical convergence (10). This allows us to apply the regularization scheme to a wide class of convex optimization problems as shown in Sections 3 and 4.

All over this paper we denote by $\mu:[0,+\infty)\to[0,+\infty)$ a gauge function with the property that the following limit exists and we have

$$\lim_{t\to\infty}\frac{\mu(t)}{t} > 0. \tag{11}$$

The duality mapping of gauge μ is denoted J^{μ}, as usual.

2.2 The next result shows that, under quite mild conditions concerning the mappings A^k and the vectors f^k, the sequence $\{x^k\}_{k\in N}$ generated in X according to (4) is well defined, bounded and that its weak accumulation points are necessarily solutions of (1).

Theorem. *Let $\{\alpha_k\}_{k\in\mathbb{N}}$ be a bounded sequence of positive real numbers. Suppose that, for each $k\in\mathbb{N}$, the mapping $A^k: X\to 2^{X^*}$ is maximal monotone. Then the following statements are true:*

(*i*) *The sequence $\{x^k\}_{k\in\mathbb{N}}$ given by* (4) *is well defined;*

(*ii*) *If there exists a bounded sequence $\{v^k\}_{k\in\mathbb{N}}$ in X such that the sequence* (5) *is bounded, then the sequence $\{x^k\}_{k\in\mathbb{N}}$ is bounded too and has weak accumulation points;*

(*iii*) *If, in addition to the requirements in* (*ii*), *we have that the sequence $\{\alpha_k\}_{k\in\mathbb{N}}$ converges to zero, the sequence $\{f^k\}_{k\in\mathbb{N}}$ converges weakly to f in X^* and*

$$w-\overline{\lim}\ Graph\left(A^k\right)\subseteq Graph\left(A\right), \tag{12}$$

then the problem (1) *has at least one solution and any weak accumulation point of $\{x^k\}_{k\in\mathbb{N}}$ is a solution of it.*

Proof. Since the mappings A^k are maximal monotone it follows that $A^k + \alpha_k J^\mu$ are surjective and $\left(A^k + \alpha_k J^\mu\right)^{-1}$ are single valued. Hence, the sequence $\left\{x^k\right\}_{k\in\mathbb{N}}$ is well defined. In order to show that this sequence is bounded, observe that, for each $k \in \mathbb{N}$, there exists a function $h^k \in A^k x^k$ such that

$$f^k = h^k + \alpha_k J^\mu x^k. \tag{13}$$

The sets $A^k v^k$ are nonempty because, otherwise, the sequence (5) would be unbounded. Also, these sets are convex and closed. Hence, for each $k \in \mathbb{N}$ there exists $g^k \in A^k v^k$ such that

$$\left\|g^k - f^k\right\|_* = \operatorname{dist}_*(f^k, A^k v^k). \tag{14}$$

Taking into account that A^k is monotone, we deduce

$$\left\langle h^k - g^k, x^k - v^k \right\rangle \geq 0.$$

Hence,

$$\begin{aligned}
\left\langle g^k, x^k - v^k \right\rangle &\leq \left\langle h^k, x^k - v^k \right\rangle = \left\langle f^k - \alpha_k J^\mu x^k, x^k - v^k \right\rangle \\
&= \left\langle f^k, x^k - v^k \right\rangle - \alpha_k \left\langle J^\mu x^k, x^k \right\rangle + \alpha_k \left\langle J^\mu x^k, v^k \right\rangle \\
&= \left\langle f^k, x^k - v^k \right\rangle - \alpha_k \mu\left(\left\|x^k\right\|\right)\left\|x^k\right\| + \alpha_k \left\langle J^\mu x^k, v^k \right\rangle \\
&\leq \left\langle f^k, x^k - v^k \right\rangle - \alpha_k \mu\left(\left\|x^k\right\|\right)\left\|x^k\right\| + \alpha_k \mu\left(\left\|x^k\right\|\right)\left\|v^k\right\|,
\end{aligned}$$

where the first equality follows from (13) and the third equality, as well as the last inequality, follows from (3). By consequence,

$$\begin{aligned}
\alpha_k \mu\left(\left\|x^k\right\|\right)\left(\left\|x^k\right\| - \left\|v^k\right\|\right) &\leq \left\langle f^k - g^k, x^k - v^k \right\rangle \\
&\leq \left\|f^k - g^k\right\|_* \left\|x^k\right\| + \left\|f^k - g^k\right\|_* \left\|v^k\right\|
\end{aligned} \tag{15}$$

for all $k \in \mathbb{N}$. Suppose, by contradiction, that $\{x^k\}_{k\in\mathbb{N}}$ is unbounded. Then, for some subsequence $\{x^{i_k}\}_{k\in\mathbb{N}}$ of it we have $\lim_{k\to\infty}|x^{i_k}| = +\infty$. From (15) we deduce that, for sufficiently large k, we have

$$\frac{1}{\|x^{i_k}\|}\mu\left(\|x^{i_k}\|\right)\left(\|x^{i_k}\| - \|v^{i_k}\|\right) \le \frac{1}{\alpha_{i_k}}\|f^{i_k} - g^{i_k}\|_*\left(1 + \frac{\|v^{i_k}\|}{\|x^{i_k}\|}\right),$$

where, according to (14) and the hypothesis, the sequence $\left\{\alpha_k^{-1}\|f^k - g^k\|_*\right\}_{k\in\mathbb{N}}$ is bounded. Taking on both sides of this inequality the upper limit as $k \to \infty$ and taking into account (11), (14) and the boundedness of $\{v^k\}_{k\in\mathbb{N}}$ one gets that the limit on the left hand side is $+\infty$ while that on the right hand side is finite, that is, a contradiction. This shows that $\{x^k\}_{k\in\mathbb{N}}$ is bounded and, since X is reflexive, $\{x^k\}_{k\in\mathbb{N}}$ has weak accumulation points.

Now, assume that $\{\alpha_k\}_{k\in\mathbb{N}}$ converges to zero, $\{f^k\}_{k\in\mathbb{N}}$ converges weakly to f in X^* and (12) also holds. Observe that the sequence $\left\{\|g^k - f^k\|_*\right\}_{k\in\mathbb{N}}$ converges to zero because $\{\alpha_k\}_{k\in\mathbb{N}}$ converges to zero, $M := \sup_{k\in\mathbb{N}} \alpha_k^{-1}\operatorname{dist}_*(f^k, A^k v^k)$ is finite and

$$\|f^k - g^k\|_* \le \alpha_k M,$$

for all $k \in \mathbb{N}$. Consequently, since $\{f^k\}_{k\in\mathbb{N}}$ converges weakly to f, we deduce that $\{g^k\}_{k\in\mathbb{N}}$ converges weakly to f too. Let v be a weak accumulation point of the sequence $\{v^k\}_{k\in\mathbb{N}}$ and denote by $\{v^{i_k}\}_{k\in\mathbb{N}}$ a subsequence of $\{v^k\}_{k\in\mathbb{N}}$ converging weakly to v. Since for any $k \in \mathbb{N}$ we have $(v^{i_k}, g^{i_k}) \in Graph\left(A^{i_k}\right)$, condition (12) implies that $(v, f) \in Graph(A)$, i.e., v is a solution of (1). Let x be a weak accumulation point of $\{x^k\}_{k\in\mathbb{N}}$ and let $\{x^{j_k}\}_{k\in\mathbb{N}}$ be a subsequence of $\{x^k\}_{k\in\mathbb{N}}$ which converges weakly to x. Note that for any $z \in X$ we have

$$\begin{aligned}\left|\left\langle z, f-h^k\right\rangle\right| &= \left|\left\langle z, f-f^k\right\rangle + \left\langle z, f^k-h^k\right\rangle\right| \\ &= \left|\left\langle z, f-f^k\right\rangle + \alpha_k \left\langle z, J^\mu x^k\right\rangle\right| \\ &\le \left|\left\langle z, f-f^k\right\rangle\right| + \alpha_k \|z\| \left\|J^\mu x^k\right\|_*,\end{aligned}$$

where the last sum converges to zero as $k \to \infty$. This shows that the sequence $\{h^k\}_{k\in\mathbb{N}}$ converges weakly to f. Hence, the sequence $\{(x^{j_k}, h^{j_k})\}_{k\in\mathbb{N}}$ converges weakly to (x, f) in $X \times X^*$. Since we also have that $h^{j_k} \in A^{j_k} x^{j_k}$ for all $k \in \mathbb{N}$, condition (12) implies that $(x, f) \in Graph(A)$, that is, x is a solution of (1).

2.3 Condition (12) involved in Theorem 2.2 is difficult to verify in applications as those discussed in Section 3 below. We show next that this condition can be relaxed at the expense of strenghtening the convergence requirements for $\{f^k\}_{k\in\mathbb{N}}$. Note that in view of (7) condition (16) below is weaker than (12). Precisely, we have the following result:

Corollary. *Let $\{\alpha_k\}_{k\in\mathbb{N}}$ be a sequence of positive real numbers converging to zero. Suppose that, for each $k \in \mathbb{N}$, the mapping $A^k : X \to 2^{X^*}$ is maximal monotone and that there exists a bounded sequence $\{v^k\}_{k\in\mathbb{N}}$ in X such that the sequence* (5) *is bounded. Then the sequence $\{x^k\}_{k\in\mathbb{N}}$ given by* (4) *is well defined, bounded and has weak accumulation points. If, in addition, the sequence $\{f^k\}_{k\in\mathbb{N}}$ converges strongly to f in X^* and*

$$ws - \overline{\lim}\, Graph(A^k) \subseteq Graph(A), \tag{16}$$

then the problem (1) *has solutions and any weak accumulation point of $\{x^k\}_{k\in\mathbb{N}}$ is a solution of it.*

Proof. Well definedness and boundedness of the sequence $\{x^k\}_{k\in\mathbb{N}}$ results from Theorem 2.2. Exactly as in the proof of Theorem 2.2 we deduce that for each $k \in \mathbb{N}$ there exist $h^k \in A^k x^k$ and $g^k \in A^k v^k$ such that (13) and (14) hold. Observe that the sequence $\{g^k\}_{k\in\mathbb{N}}$ converges strongly to f because of (14) and the boundedness of (5). It remain to show that, under the

assumptions that $\{f^k\}_{k\in\mathbb{N}}$ converges strongly to f and (16) holds, any weak accumulation point of $\{x^k\}_{k\in\mathbb{N}}$ is a solution of (1). Let v be a weak accumulation point of the sequence $\{v^k\}_{k\in\mathbb{N}}$ (such a point exists because $\{v^k\}_{k\in\mathbb{N}}$ is bounded and X is reflexive) and denote by $\{v^{i_k}\}_{k\in\mathbb{N}}$ a subsequence of $\{v^k\}_{k\in\mathbb{N}}$ converging weakly to v. Since for all $k\in\mathbb{N}$ we have $(v^{i_k}, g^{i_k}) \in Graph(A^{i_k})$, condition (16) implies that $(v,f)\in Graph(A)$, i.e., v is a solution of (1). Let x be a weak accumulation point of $\{x^k\}_{k\in\mathbb{N}}$ and let $\{x^{j_k}\}_{k\in\mathbb{N}}$ be a subsequence of $\{x^k\}_{k\in\mathbb{N}}$ which converges weakly to x. Note that, according to (13), we have

$$\begin{aligned}\|f-h^k\|_* &\le \|f-f^k\|_* + \|f^k-h^k\|_* \\ &= \|f-f^k\|_* + \alpha_k\|J^\mu x^k\|_*,\end{aligned}$$

where the last sum converges to zero as $k\to\infty$, because $\{x^k\}_{k\in\mathbb{N}}$ is bounded (and, hence, so is $\{J^\mu x^k\}_{k\in\mathbb{N}}$) and the sequence $\{f^k\}_{k\in\mathbb{N}}$ converges to f by hypothesis. Therefore, the sequence $\{h^k\}_{k\in\mathbb{N}}$ converges strongly to f. Since we also have that $h^k\in A^k x^k$ for all $k\in\mathbb{N}$, condition (16) implies that $(x,f)\in Graph(A)$, that is, x is a solution of (1). □

2.4 If problem (1) has only one solution (as happens, for instance, when A is strictly monotone), then Theorem 2.2 guarantees weak convergence of the whole sequence $\{x^k\}_{k\in\mathbb{N}}$. However, in general, we do not know whether the whole sequence $\{x^k\}_{k\in\mathbb{N}}$ converges weakly. The next result shows that not only weak convergence, but also strong convergence of $\{x^k\}_{k\in\mathbb{N}}$ to a solution of (1) can be ensured provided that any element of $A^{-1}f$ (the solution set) is the limit of a sequence $\{v^k\}_{k\in\mathbb{N}}$ satisfying (17) below. In view of the remarks in Subsection 2.1, this result improves upon Theorem 2.2 in [4].

Theorem. *Suppose that problem* (1) *has at least one solution and that the sequence of positive real numbers* $\{\alpha_k\}_{k\in\mathbb{N}}$ *converges to zero. If*

$A^k : X \to 2^{X^*}$, $k \in \mathbb{N}$, are maximal monotone operators with the property (12), *if* $\{f^k\}_{k\in\mathbb{N}}$ *is a sequence converging weakly to* f *in* X^* *and if, for each* $v \in A^{-1}f$, *there exists a sequence* $\{v^k\}_{k\in\mathbb{N}}$ *which converges strongly to* v *in* X *and such that*

$$0 \in s - \underline{\lim}\, \frac{1}{\alpha_k}\left[A^k v^k - f^k\right], \tag{17}$$

then the sequence $\{x^k\}_{k\in\mathbb{N}}$ *given by* (4) *is well defined and converges strongly to the minimal norm solution of problem* (1).

Proof. The assumption that problem (1) has solutions implies, in our current setting, the existence of a bounded sequence $\{v^k\}_{k\in\mathbb{N}}$ as required by Theorem 2.2. Observe that, since (17) holds, the sequence $\{\alpha_k^{-1}\mathrm{dist}_*(f^k, A^k v^k)\}_{k\in\mathbb{N}}$ converges to zero and, therefore, it is bounded. Hence, one can apply Theorem 2.2 in order to deduce well definedness and boundedness of $\{x^k\}_{k\in\mathbb{N}}$ and the fact that any weak accumulation point of it is a solution of (1). Note that, since A is maximal monotone, A^{-1} is maximal monotone too and, therefore, the set $A^{-1}f$, which is exactly the presumed nonempty solution set of problem (1), is convex and closed. The space X is reflexive and strictly convex and, therefore, the nonempty, convex and closed set $A^{-1}f$ contains a unique minimal norm element $\overline{x}$ (the metric projection of 0 onto the set $A^{-1}f$). We show that the only weak accumulation point of $\{x^k\}_{k\in\mathbb{N}}$ is $\overline{x}$. To this end, let $\{x^{i_k}\}_{k\in\mathbb{N}}$ be a subsequence of $\{x^k\}_{k\in\mathbb{N}}$ which converges weakly to some $x \in X$. According to Theorem 2.2, x is necessarily contained in $A^{-1}f$. If $x = 0$, then this is necessarily the minimal norm element of $A^{-1}f$, i.e., $x = \overline{x}$. Suppose that $x \neq 0$. Let v be any other solution of problem (1). By hypothesis, there exists a sequence $\{v^k\}_{k\in\mathbb{N}}$ converging strongly in X to v and such that, for some sequence $\{l^k\}_{k\in\mathbb{N}}$ with $l^k \in A^k v^k$ for each $k \in \mathbb{N}$, we have

$$\lim_{k\to\infty} \frac{1}{\alpha_k}\left(l^k - f^k\right) = 0. \tag{18}$$

Clearly,

$$0 < \|x\| \le \liminf_{k\to\infty} \|x^{i_k}\|$$

and there exists a subsequence $\{x^{j_k}\}_{k\in\mathbb{N}}$ of $\{x^{i_k}\}_{k\in\mathbb{N}}$ such that

$$\liminf_{k\to\infty} \|x^{i_k}\| = \lim_{k\to\infty} \|x^{j_k}\|. \tag{19}$$

The subsequence $\{x^{j_k}\}_{k\in\mathbb{N}}$ is still weakly convergent to x and has

$$0 < \mu(\|x\|) \le \mu\left(\liminf_{k\to\infty} \|x^{i_k}\|\right) = \mu\left(\lim_{k\to\infty} \|x^{j_k}\|\right) = \lim_{k\to\infty} \mu\left(\|x^{j_k}\|\right), \tag{20}$$

because μ is continuous and increasing (as being a gauge function). For each $k \in \mathbb{N}$, let $h^k \in A^k x^k$ be the function for which (13) is satisfied. These functions exist because $\{x^k\}_{k\in\mathbb{N}}$ is well defined. Due to the monotonicity of A^k, we have

$$\begin{aligned} 0 &\le \langle h^k - l^k, x^k - v^k\rangle = \langle f^k - \alpha_k J^\mu x^k - l^k, x^k - v^k\rangle \\ &= \langle f^k - l^k, x^k - v^k\rangle - \alpha_k \langle J^\mu x^k, x^k\rangle + \alpha_k \langle J^\mu x^k, v^k\rangle \\ &\le \langle f^k - l^k, x^k - v^k\rangle - \alpha_k \mu\left(\|x^k\|\right)\|x^k\| + \alpha_k \mu\left(\|x^k\|\right)\|v^k\|, \end{aligned}$$

where the first equality results from (13) and the last inequality follows from (3). This implies

$$\mu\left(\|x^k\|\right)\|x^k\| \le \frac{1}{\alpha_k}\langle f^k - l^k, x^k - v^k\rangle + \mu\left(\|x^k\|\right)\|v^k\|, \tag{21}$$

where the first term of the right hand side converges to zero as $k \to \infty$ because of (18) and because of the boundedness of $\{v^k\}_{k\in\mathbb{N}}$ and $\{x^k\}_{k\in\mathbb{N}}$. Replacing k by j_k in this inequality, we deduce that for k large enough

$$\|x^{j_k}\| \le \frac{1}{\mu\left(\|x^{j_k}\|\right)} \frac{1}{\alpha_k} \langle f^{j_k} - l^{j_k}, x^{j_k} - v^{j_k}\rangle + \|v^{j_k}\|. \tag{22}$$

Letting here $k \to \infty$ we get

$$\|x\| \le \lim_{k\to\infty} \|x^{j_k}\| \le \lim_{k\to\infty} \|v^{j_k}\| = \|v\|,$$

because $\{v^k\}_{k\in\mathbb{N}}$ converges strongly to v and because of (19). Since v is an arbitrarily chosen solution of problem (1), it follows that $x = \overline{x}$. Hence, the sequence $\{x^k\}_{k\in\mathbb{N}}$ converges weakly to $\overline{x}$.

It remains to show that $\{x^k\}_{k\in\mathbb{N}}$ converges strongly. To this end, observe that, since $\{x^k\}_{k\in\mathbb{N}}$ converges weakly to $\overline{x}$ and since X is a space with the Kadec-Klee property, it is sufficient to show that $\{\|x^k\|\}_{k\in\mathbb{N}}$ converges to $\|\overline{x}\|$. In other words, it is sufficient to prove that all convergent subsequences of the bounded sequence $\{\|x^k\|\}_{k\in\mathbb{N}}$ converge to $\|\overline{x}\|$. In order to prove that, let $\{\|x^{p_k}\|\}_{k\in\mathbb{N}}$ be a convergent subsequence of $\{\|x^k\|\}_{k\in\mathbb{N}}$. If $\{\|x^{p_k}\|\}_{k\in\mathbb{N}}$ converges to 0, then

$$0 \le \|\overline{x}\| \le \liminf_{k\to\infty} \|x^k\| \le \lim_{k\to\infty} \|x^{p_k}\| = 0,$$

that is, $\|\overline{x}\| = \lim_{k\to\infty} \|x^{p_k}\| = 0$. Suppose now that

$$\lim_{k\to\infty} \|x^{p_k}\| = \beta > 0.$$

Then, there exists a positive integer k_0 such that, for all integers $k \ge k_0$, we have $\|x^{p_k}\| > 0$. According to (21) this implies that, for $k \ge k_0$, one has

$$\|x^{p_k}\| \le \frac{1}{\mu(\|x^{p_k}\|)} \frac{1}{\alpha_{p_k}} \langle f^{p_k} - l^{p_k}, x^{p_k} - v^{p_k} \rangle + \|v^{p_k}\|.$$

Letting $k \to \infty$ in this inequality we get

$$\|\overline{x}\| \le \liminf_{k\to\infty} \|x^k\| \le \lim_{k\to\infty} \|x^{p_k}\| \le \lim_{k\to\infty} \|v^{p_k}\| = \|v\|.$$

Since v is an arbitrarily chosen solution of problem (1) we can take here $v = \overline{x}$ and obtain $\|\overline{x}\| = \lim_{k\to\infty} \|x^{p_k}\|$. This completes the proof. □

2.5 Similarly to Corollary 2.3 ensuring that the weak accumulation points of $\{x^k\}_{k\in N}$ are solutions of (1), we can use Theorem 2.4 in order to prove strong convergence of $\{x^k\}_{k\in N}$ to a solution of (1) when condition (12) is replaced by the weaker requirement (16) but strenghtening the convergence requirements on $\{f^k\}_{k\in N}$.

Corollary. *Suppose that problem* (1) *has solutions and the sequence of positive real numbers* $\{\alpha_k\}_{k\in\mathbb{N}}$ *converges to zero. If* $A^k : X \to 2^{X^*}$, $k \in \mathbb{N}$, *are maximal monotone operators with the property* (16), *if* $\{f^k\}_{k\in\mathbb{N}}$ *is a sequence converging strongly to* f *in* X^* *and if, for each* $v \in A^{-1}f$, *there exists a sequence* $\{v^k\}_{k\in\mathbb{N}}$ *which converges strongly to* v *in* X *and satisfies* (17), *then the sequence* $\{x^k\}_{k\in\mathbb{N}}$ *given by* (4) *is well defined and converges strongly to the minimal norm solution of problem* (1).

Proof. Well definedness and boundedness of $\{x^k\}_{k\in\mathbb{N}}$ as well as the fact that any weak accumulation point of it is a solution of (1) result from Corollary 2.3. In order to show that $\{x^k\}_{k\in\mathbb{N}}$ converges strongly to the minimal norm solution of the problem one reproduces without modification the arguments made for the same purpose in the proof of Theorem 2.4. □

3. REGULARIZATION OF CONVEX OPTIMIZATION PROBLEMS

3.1 We have noted above that Theorem 2.4 and Corollary 2.5, can be of use in order to prove stability properties of the procedure (4) applied to optimization problems with perturbed data. Such properties are of interest in applications in which the data involved in the optimal solution finding process are affected by computational and/or measurement errors. To make things precise, in what follows $F : X \to (-\infty, +\infty]$ is a lower semicontinuous convex function and Ω is a nonempty, closed convex subset of $\mathrm{Int}(\mathrm{Dom}\, F)$, the interior of the domain of F. We consider the following optimization problem under the assumption that it has at least one solution:

$$(P) \quad \text{Minimize } F(x) \quad \text{subject to} \quad x \in \Omega. \tag{23}$$

It is not difficult to verify that by solving the following inclusion

(P') Find $x \in X$ such that $0 \in Ax$,

where $A: X \to 2^{X^*}$ is the operator defined by

$$A = \partial F + N_{\Omega}, \tag{24}$$

with ∂F denoting the subdifferential of F and $N_{\Omega}: X \to 2^{X^*}$ denoting the normal cone operator associated to Ω, that is,

$$N_{\Omega}(x) = \begin{cases} \{h \in X^* : \langle h, z - x \rangle \le 0, \quad \forall z \in \Omega\} & \text{if } x \in \Omega, \\ \varnothing & \text{otherwise,} \end{cases} \tag{25}$$

one implicity finds solutions of *(P)*. The operators ∂F and N_{Ω} are maximal monotone (cf. [51] by taking into account that N_{Ω} is the subgradient of the indicator function of the set Ω). Consequently, the operator A is maximal monotone too (cf. [50]).

We presume that the function F can not be exactly determined and that, instead, we have a sequence of convex, lower semicontinuous functions $F_k: X \to (-\infty, +\infty]$, $(k \in \mathbb{N})$, such that

$$\text{Dom } F \subseteq \text{Dom } F_k, \quad \forall k \in \mathbb{N}, \tag{26}$$

and which approximates F in the following sense:

Condition (A). *There exists a continuous function* $c: [0, +\infty) \to [0, +\infty)$ *and a sequence of positive real numbers* $\{\delta_k\}_{k \in \mathbb{N}}$ *such that* $\lim_{k \to \infty} \delta_k = 0$ *and*

$$|F_k(x) - F(x)| \le c(\|x\|)\delta_k, \tag{27}$$

whenever $x \in \text{Dom } F$ *and* $k \in \mathbb{N}$.

In real world optimization problems it often happens that the set Ω is defined by a system of inequalities $g_i(x) \le 0$, $i \in I$, where g_i are convex and lower semicontinuous functions on X. The functions g_i may also be hard to precisely evaluate and, then, determining the set Ω (or determining whether a vector belongs to it or not) is done by using some (still convex and lower semicontinuous) approximations g_i^k, $k \in \mathbb{N}$, instead. In other words,

one replaces the set Ω by some nonempty closed convex approximations Ω_k, $k \in \mathbb{N}$, of it. In what follows we assume that

$$\Omega_k \subseteq \operatorname{Int}(\operatorname{Dom} F), \quad \forall k \in \mathbb{N}, \tag{28}$$

and that the closed convex sets Ω_k approximate the set Ω in the following sense:

Condition (B). *The next two requirements are satisfied:*

(i) *For any* $y \in \Omega$ *there exists a sequence* $\{y^k\}_{k\in\mathbb{N}}$ *which converges strongly to* y *in* X *and such that* $y^k \in \Omega_k$ *for all* $k \in \mathbb{N}$;

(ii) *If* $\{z^k\}_{k\in\mathbb{N}}$ *is a sequence in* X *which is weakly convergent and such that for some subsequence* $\{\Omega_{i_k}\}_{k\in\mathbb{N}}$ *of* $\{\Omega_k\}_{k\in\mathbb{N}}$ *we have* $z^k \in \Omega_{i_k}$ *for all* $k \in \mathbb{N}$, *then there exists a sequence* $\{w^k\}_{k\in\mathbb{N}}$ *contained in* Ω *with the property that* $\lim_{k\to\infty} \|z^k - w^k\| = 0$.

Observe that the requirement ($B(i)$) is equivalent to the condition that $\Omega \subseteq s-\underline{\lim}\ \Omega_k$. The requirement ($B(ii)$) implies that $w-\overline{\lim}\ \Omega_k \subseteq \Omega$. Taken together, the requirements ($B(i)$) and ($B(ii)$) imply that $\Omega = Lim\ \Omega_k$. It can be verified that the requirement ($B(ii)$) is satisfied whenever there exists a function $b: X \to [0,+\infty)$ which is bounded on bounded sets, and a sequence of positive real numbers $\{\gamma_k\}_{k\in N}$ converging to zero such that for any $k \in N$ and each $z \in \Omega_k$, we have that $\operatorname{dist}(z,\Omega) < b(z)\gamma_k$. The last condition was repeatedly used in the regularization of variational inequalities involving maximal monotone operators (see [4]).

For each $k \in \mathbb{N}$, we associate to problem (23) the problem

(P_k) Minimize $F_k(x)$ subject to $x \in \Omega_k$,

which can be solved by finding solutions of the inclusion

(P_k') Find $x^k \in X$ such that $0 \in A^k x^k$,

where the operator $A^k : X \to 2^{X^*}$ is defined by

$$A^k := \partial F_k + N_{\Omega_k}, \tag{29}$$

and is also maximan monotone. The question is whether under the conditions $(A),(B)$, (11) and presuming that $\{\alpha_k\}_{k\in\mathbb{N}}$ converges to zero, the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated according to (4) for the operators A^k given by (29) and for $f^k = f = 0, k \in \mathbb{N}$, i.e., the sequence

$$x^k := \left(A^k + \alpha_k J^\mu\right)^{-1}(0.) \tag{30}$$

converges strongly to a solution of problem (P') and, hence, to a solution of the original optimization problem (P). It should be noted that, since by Asplund's Theorem (see, for istance, [22] we have

$$J^\mu x = \partial\phi(\|x\|) \text{ with } \phi(t) := \int_0^t \mu(\tau),$$

determining the vectors x^k defined by (30) amounts to solving the optimization problem

$$(Q_k) \quad \text{Minimize } F_k(x) + \alpha_k\phi(\|x\|) \quad \text{subject to } x \in \Omega_k \tag{31}$$

By contrast to problem (P_k) which may have infinitely many solutions, the problem (Q_k) always has unique solution. Moreover, by choosing $\mu(t) = 2t$ and, thus, $\phi(t) = t^2$, one ensures that the objective function of (Q_k) is strongly convex and, therefore, the problem (Q_k) may be better posed and easier to solve than (P_k).

3.2 We aim now towards giving an answer to the question asked in Subsection 3.1. To this end, when D is a nonempty closed convex subset of X and $x \in X$, we denote by $\text{Proj}_D(x)$ the metric projection of x onto the set D (this exists and it is unique by our hypothesis that the space X is strictly convex and reflexive). The next result shows stability and convergence of the regularization technique when applied to convex optimization problems. For proving it, recall that the objective function F of the problem (P) is assumed to be lower semicontinuous and convex and its domain $\text{Dom}\, F$ has nonempty interior since $\varnothing \neq \Omega \subseteq \text{Int}(\text{Dom}\, F)$ - (see Subsection 3.1). Consequently, F is continuous on $\text{Int}(\text{Dom}\, F)$, for each $x \in \text{Int}(\text{Dom}\, F)$, we have $\partial F(x) \neq \varnothing$ (cf. [48, Proposition 3.2 and Proposition 1.11]) and the right hand sided derivative of F at x, i.e. the function $F^\circ(x,\cdot): X \to \mathbb{R}$ given by

$$F^{\circ}(x,d) := \lim_{t \searrow 0} \frac{F(x+td)-F(x)}{t},$$

is a well defined continuous seminorm on X.

Theorem. *Suppose that conditions* (A) *and* (B) *are satisfied. If there exists a sequence* $\{\alpha_k\}_{k\in\mathbb{N}}$ *of positive real numbers converging to zero such that for each optimal solution* v *of* (P), *there exists a sequence* $\{v^k\}_{k\in\mathbb{N}}$ *with the properties that* $v^k \in \Omega_k$ *for all* $k \in \mathbb{N}$ *and*

$$\lim_{k\to\infty}\|v^k - v\| = 0 = \lim_{k\to\infty} \alpha_k^{-1} \left\| \mathrm{Proj}_{\partial F_k(v^k)+N_{\Omega_k}(v^k)}(0) \right\|_*, \tag{32}$$

then the sequence $\{x^k\}_{k\in\mathbb{N}}$ *given by* (30) *converges strongly to the minimal norm solution of the optimization problem* (P).

Proof. We show that Corollary 2.5 applies to the problems (P') and (P_k'), that is, to the maximal monotone operators A and A^k defined by (24) and (29), respectively, and to the functions $f^k = f = 0$, $(k \in \mathbb{N})$. First, we prove that the condition (16) is satisfied. For this purpose, take $(z,h) \in ws - \lim \mathrm{Graph}\left(A^k\right)$. Then, there exists a sequence $\{z^k\}_{k\in\mathbb{N}}$ converging weakly to z in X and there exists a sequence $\{h^k\}_{k\in\mathbb{N}}$ converging strongly to h in X^* such that for some subsequence $\{A^{i_k}\}_{k\in\mathbb{N}}$ of $\{A^k\}_{k\in\mathbb{N}}$ we have $(z^k,h^k) \in \mathrm{Graph}(A^{i_k})$ for all $k \in \mathbb{N}$. This means that

$$z^k \in \Omega_{i_k} \text{ and } h^k \in \partial F_{i_k}(z^k) + N_{\Omega_{i_k}}(z^k), \quad \forall k \in \mathbb{N},$$

or, equivalently,

$$z^k \in \Omega_{i_k} \text{ and } h^k = \xi^k + \theta^k,$$

with $\xi^k \in \partial F_{i_k}(z^k)$ and $\theta^k \in N_{\Omega_{i_k}}(z^k)$ for all $k \in \mathbb{N}$. We have to show that

$$z \in \Omega \text{ and } h \in \partial F(z) + N_{\Omega}(z). \tag{33}$$

The sequence $\{z^k\}_{k\in\mathbb{N}}$ is weakly convergent to z and $z^k \in \Omega_{i_k}$ for all $k \in \mathbb{N}$. Therefore, according to $(B(ii))$, there exists a sequence $\left\{w^k\right\}_{k\in\mathbb{N}} \subseteq \Omega$ such that $\lim_{k\to\infty}\left\|z^k - w^k\right\| = 0$. Clearly, the sequence $\{w^k\}_{k\in\mathbb{N}}$, converges weakly

to z. Since the set Ω is closed and convex, and therefore weakly closed, we obtain that $z \in \Omega$. In order to complete the proof of (33), let $u \in \Omega$ be fixed. According to $(B(i))$, there exists a sequence $\{u^k\}_{k\in\mathbb{N}}$ which converges strongly to u and such that $u^k \in \Omega_k$ for any $k \in \mathbb{N}$. Since $h^k - \theta^k = \xi^k \in \partial F_{i_k}(z^k)$ we deduce

$$\begin{aligned}\left\langle h^k - \theta^k, u^{i_k} - z^k \right\rangle &\le F_{i_k}(u^{i_k}) - F_{i_k}(z^k) \\ &\le |F_{i_k}(u^{i_k}) - F(u^{i_k})| + |F(z^k) - F_{i_k}(z^k)| \\ &+ F(u^{i_k}) - F(z^k) \\ &\le (c(\|u^{i_k}\|) + c(\|z^k\|))\delta_{i_k} + F(u^{i_k}) - F(z^k),\end{aligned}$$

where the last inequality results from (27). By consequence,

$$\begin{aligned}\left\langle h^k, u^{i_k} - z^k \right\rangle \le (c(\|u^{i_k}\|) + c(\|z^k\|))\delta_{i_k} + F(u^{i_k}) - F(z^k) \\ + \left\langle \theta^k, u^{i_k} - z^k \right\rangle,\end{aligned}$$

where the last term on the right hand side of the inequality is nonpositive because $\theta^k \in N_{\Omega_{i_k}}(z^k)$ and $u^{i_k} \in \Omega_{i_k}$ (see (25)). Thus, for any $k \in \mathbb{N}$, we obtain

$$\left\langle h^k, u^{i_k} - z^k \right\rangle \le (c(\|u^{i_k}\|) + c\|z^k\|))\delta_{i_k} + F(u^{i_k}) - F(z^k). \tag{34}$$

As noted above, the function F is continuous on $\mathrm{Int}(\mathrm{Dom}\, F)$. Hence, the sequence $\{F(u^{i_k})\}_{k\in\mathbb{N}}$ converges to $F(u)$. Since F is also convex, it is weakly lower semicontinuous and, then, we have $F(z) \le \liminf_{k\to\infty} F(z^k)$. Taking lim sup for $k \to \infty$ on both sides of (34), and taking into account that the sequences $\{u^{i_k}\}_{k\in\mathbb{N}}$ and $\{z^k\}_{k\in\mathbb{N}}$ are bounded and that the function c is continuous (see condition (A)), we obtain that

$$\left\langle h, u - z \right\rangle \le F(u) - F(z). \tag{35}$$

Since the latter holds for arbitrary $u \in \Omega$, it implies that $h \in \partial F_\Omega(z)$, where $F_\Omega : X \to (-\infty, +\infty]$ is the lower semicontinuous convex function defined by

$$F_\Omega := F + \iota_\Omega,$$

with ι_Ω standing for the indicator function of the set Ω. As noted above, the function F is continuous on the interior of its domain and, thus, is continuous on $\Omega = \mathrm{Dom}\, F_\Omega$. Hence, applying [48, Proposition 3.23] and observing that $\partial \iota_\Omega = N_\Omega$ (see (25)), we deduce that, for any $x \in X$,

$$\partial F_\Omega(x) = \partial F(x) + \partial \iota_\Omega(x) = \partial F(x) + N_\Omega(x).$$

Consequently,

$$h \in \partial F_\Omega(z) = \partial F(z) + N_\Omega(z)$$

and this completes the proof of (33).

Now observe that, according to (32) and (29), we have that for each solution v of (P) there exists a sequence $\{v^k\}_{k\in\mathbb{N}}$ such that $v^k \in \Omega_k$ for all $k \in \mathbb{N}$ and with the property that

$$\lim_{k\to\infty} \alpha_k^{-1}\, dist_*(0, A^k v^k) = \lim_{k\to\infty} \alpha_k^{-1} \left\| \mathrm{Proj}_{\partial F_k(v^k) + N_{\Omega_k}(v^k)}(0) \right\|_* = 0,$$

that is, condition (17) is also satisfied. □

3.3 Recall (see Subsection 3.1) that we assume that the problem (P) has optimal solutions. By contrast, some or all problems (P_k) may not have optimal solutions. Theorem 3.2 guarantees existence and convergence of $\{x^k\}_{k\in\mathbb{N}}$ to a solution of (P) with no consistency requirements on the problems (P_k). In our circumstances the functions F_k may not have global minimizers either. The following consequence of Theorem 3.2 may be of use for global minimization of F when some of the problems (P_k) have no optimal solutions.

Corollary. *Suppose that conditions (A) and (B) hold. If there exists a sequence* $\{\alpha_k\}_{k\in\mathbb{N}}$ *of positive real numbers converging to zero such that for each optimal solution* v *of* (P), *there exists a sequence* $\{v^k\}_{k\in\mathbb{N}}$ *with the properties that* $v^k \in \Omega_k$ *for all* $k \in N$ *and*

$$\lim_{k\to\infty} \left\| v^k - v \right\| = 0 = \lim_{k\to\infty} \alpha_k^{-1} \left\| \mathrm{Proj}_{\partial F_k(v^k)}(0) \right\|_*, \tag{36}$$

then the sequence $\{x^k\}_{k\in\mathbb{N}}$ *given by* (30) *converges strongly to the minimal norm solution of the optimization problem* (P).

Proof. Note that $0 \in N_{\Omega_k}(v^k)$ for all $k \in \mathbb{N}$ and, therefore, when (36) holds, we have

$$\begin{aligned}\operatorname{dist}_*(0, A^k v^k) &= \inf\left\{\|g+\zeta\|_* : g \in \partial F_k(v^k) \text{ and } \zeta \in N_{\Omega_k}(v^k)\right\} \\ &\le \inf\left\{\|g\|_* : g \in \partial F_k(v^k)\right\} \\ &= \left\|Proj_{\partial F_k(v^k)}(0)\right\|_*.\end{aligned}$$

This implies (32) because of (36). □

3.4 If X is a Hilbert space and the functions F and F_k are differentiable on the interior of $\operatorname{Dom}(F)$, then the condition (36) can be relaxed by taking into account (see [52, Remark 3, p. 890]) that, in this case, we have

$$\left\|\nabla F_k(v^k) + \operatorname{Proj}_{N_{\Omega_k}(v^k)}(-\nabla F_k(v^k))\right\|_* = \left\|\operatorname{Proj}_{T_{\Omega_k}(v^k)}(-\nabla F_k(v^k))\right\|_*, \tag{37}$$

where $T_{\Omega_k}(v^k)$ denotes the tangent cone of Ω_k at the point v^k, that is, the polar cone of $N_{\Omega_k}(v^k)$. Precisely, we have the following result whose proof reproduces without modification the arguments in Theorem 3.2 with the only exception that for showing (17) one uses (37), (39) below, and the equalities

$$\begin{aligned}&\operatorname{dist}_*(0, A^k v^k) = \operatorname{dist}_*(0, \nabla F_k(v^k) + N_{\Omega_k}(v^k)) \\ &= \operatorname{dist}_*(-\nabla F_k(v^k), N_{\Omega_k}(v^k)) \\ &= \left\|\nabla F_k(v^k) + \operatorname{Proj}_{N_{\Omega_k}(v^k)}(-\nabla F_k(v^k))\right\|_*,\end{aligned} \tag{38}$$

where the first is due to the fact that $0 \in N_{\Omega}(v^k)$ and the second follows from (36).

Corollary. *Suppose that X is a Hilbert space and that conditions (A) and (B) hold. If the functions F and F_k, $k \in \mathbb{N}$, are (Gâteaux) differentiable on* $\operatorname{Int}(\operatorname{Dom} F)$ *and if there exists a sequence $\{\alpha_k\}_{k\in\mathbb{N}}$ of positive real numbers converging to zero such that for each optimal solution v of (P), there exists a sequence $\{v^k\}_{k\in\mathbb{N}}$ with the properties that $v^k \in \Omega_k$ for all $k \in \mathbb{N}$ and*

$$\lim_{k\to\infty}\|v^k - v\| = 0 = \lim_{k\to\infty}\alpha_k^{-1}\left\|\operatorname{Proj}_{T_{\Omega_k}(v^k)}(-\nabla F_k(v^k))\right\|_*, \tag{39}$$

then the sequence $\{x^k\}_{k\in\mathbb{N}}$ *given by* (30) *converges strongly to the minimal norm solution of the optimization problem* (P).

3.5 If $\Omega_k = \Omega$ for all $k \in \mathbb{N}$, then condition (B) is, obviously, satisfied. In this case, if there exists a sequence $\{\alpha_k\}_{k\in\mathbb{N}}$ of positive real numbers converging to zero such that, for each solution v of (P), we have

$$\lim_{k\to\infty} \alpha_k^{-1} \left\|\nabla F_k(v) - \nabla F(v)\right\|_* = 0, \tag{40}$$

then (32) holds too. Indeed, if v is a solution of (P), then

$$\left\langle \nabla F(v), u - v \right\rangle = \lim_{t\searrow 0} \frac{F(v + t(u-v)) - F(v)}{t} \geq 0,$$

for any $u \in \Omega$ and this shows that $-\nabla F(v) \in N_\Omega(v)$. Therefore, taking $v^k := v$ for all $k \in \mathbb{N}$ we have

$$\begin{aligned}\left\|\mathrm{Proj}_{\partial F_k(v^k) + N_{\Omega_k}(v^k)}(0)\right\|_* &\leq \left\|\nabla F_k(v) + Proj_{N_\Omega(v)}(-\nabla F_k(v))\right\|_* \\ &= \left\|\mathrm{Proj}_{N_\Omega(v)}(-\nabla F_k(v)) - (-\nabla F_k(v))\right\|_* \\ &\leq \left\|\nabla F_k(v) - \nabla F(v)\right\|_*,\end{aligned}$$

which together with (40) implies (32). Hence, we have the following result:

Corollary. *Suppose that* $\Omega_k = \Omega$ *for all* $k \in \mathbb{N}$ *and condition* (A) *holds. Assume that the functions* F *and* F_k, $k \in \mathbb{N}$, *are (Gâteaux) differentiable on* $\mathrm{Int}(\mathrm{Dom}\, F)$. *If there exists a sequence of positive real numbers* $\{\alpha_k\}_{k\in\mathbb{N}}$ *converging to zero such that for each solution* v *of* (P) *condition* (40) *is satisfied, then the sequence* $\{x^k\}_{k\in\mathbb{N}}$ *given by* (30) *converges strongly to the minimal norm solution of the optimization problem* (P).

3.6 Theorem 3.2 shows that perturbed convex optimization problems can be regularized by the method (4). This naturally leads to the question of whether this regularization technique still works when the set Ω is defined by continuous affine constraints and one has to replace the affine constraints of (P) by approximations which are still continuous and affine. We are going to show that this is indeed the case when Ω satisfies a Robinson type regularity condition. In order to do that, let $\mathcal{L}(X, X^*)$ be the Banach space of all linear continuous operators $L : X \to X^*$ provided with the norm

$$\|L\|_{\diamond} := \sup\{\|Lx\|_* : x \in B_X(0,1)\}, \tag{41}$$

where $B_X(u,r)$ stands for the closed ball of center u and radius r in X. Suppose that L, $L^k \in \mathcal{L}(X,X^*)$, l, $l^k \in X^*$ and let $K \subset X^*$ be a nonempty closed convex cone. Suppose that the sets Ω and Ω_k are defined by

$$\Omega := \{x \in X : L(x) + l \in K\} \tag{42}$$

and

$$\Omega_k := \{x \in X : L^k(x) + l^k \in K\}. \tag{43}$$

The set Ω is called *regular* if the point-to-set mapping $x \to L(x) + l - K$ is regular in the sense of Robinson [49, p. 132], that is,

$$0 \in \mathrm{Int}\{L(x) + l - y : x \in X, y \in K\}. \tag{44}$$

Taking in the next result $K = \{0\}$ one obtains an answer to the question posed above. The fact is that the proposition we prove below is more general and can be also used in order to guarantee validity of condition (B) for some classes of problems of interest in semidefinite programming. Combined with Theorem 3.2 it implies that if the data involved in the constraints of the perturbed problem (P_k) are strong approximations of the data of (P) and if conditions (A) and (32) hold, then the regularization technique (4) can be applied in order to produce strong approximations of the minimal norm solution of (P).

Proposition. *Suppose that* L, $L^k \in L(X,X^*)$ *and* l, $l^k \in X^*$ *for any* $k \in \mathbb{N}$. *Let* $K \subset X^*$ *be a nonempty closed convex cone and consider the problems* (P) *and* (P_k) *with the feasibility sets* Ω *and* Ω_k *defined at* (42) *and* (43), *respectively. If the set* Ω *is regular, if the sequence* $\{L^k\}_{k\in\mathbb{N}}$ *converges strongly to* L *in* $L(X,X^*)$ *and if the sequence* $\{l^k\}_{k\in\mathbb{N}}$ *converges strongly to* l *in* X^*, *then condition* (B) *is satisfied.*

Proof. We first prove $(B(ii))$. For this purpose we apply Corollary 4 in [10, p. 133] to the set $M = K - l$ and to the point-to-set mapping $G : X \to 2^{X^*}$ defined by $G(x) = L(x) - M$. This is possible because of the regularity of Ω (see (44)) which guarantees that $0 \in \mathrm{Int}G(X)$. Hence, by observing that

$\Omega = G^{-1}(0)$, we deduce that for any $x \in \Omega$ there exists a positive real number $\delta(x)$ such that for any $z \in X$ we have

$$\begin{aligned} \operatorname{dist}(z,\Omega) = \operatorname{dist}(z, G^{-1}(0)) &\le (1+\|z-x\|)\frac{1}{\delta(x)}\operatorname{dist}_*(L(z), M) \\ &= (1+\|z-x\|)\frac{1}{\delta(x)}\operatorname{dist}_*(L(z)+l, K). \end{aligned} \tag{45}$$

Now, let $x \in \Omega$ be fixed and let $\delta := \delta(x)$. If $z \in \Omega_k$, then

$$\begin{aligned} \operatorname{dist}_*(L(z)+l, K) &\le \left\| \left(L(z)+l\right) - \left(L^k(z)+l^k\right) \right\|_* \\ &\le \left\| L(z) - L^k(z) \right\|_* + \left\| l - l^k \right\|_* \\ &\le \left\| L - L^k \right\|_\diamond \|z\| + \left\| l - l^k \right\|_* \end{aligned}$$

because of (41). Taking into account (45), it follows that for any $z \in \Omega_k$, we have

$$\begin{aligned} \operatorname{dist}(z,\Omega) &\le \frac{1}{\delta}(1+\|z-x\|)\left[\left\| L - L^k \right\|_\diamond \|z\| + \left\| l - l^k \right\|_* \right] \\ &\le \frac{1}{\delta}(1+\|z\|+\|x\|)(\|z\|+1)\max\left\{ \left\| L - L^k \right\|_\diamond, \left\| l - l^k \right\|_* \right\}. \end{aligned} \tag{46}$$

Consider the bounded on bounded sets function $b : X \to [0, +\infty)$ defined by

$$b(u) := \frac{1}{\delta}(\|u\|+1)(1+\|u\|+\|x\|).$$

By (46) we obtain that, for any $z \in \Omega_k$,

$$\left\| z - \operatorname{Proj}_\Omega(z) \right\| = \operatorname{dist}(z,\Omega) \le b(z)\gamma_k,$$

where $\gamma_k := \max\left\{ \left\| L - L^k \right\|_\diamond, \left\| l - l^k \right\|_* \right\}$ converges to zero as $k \to \infty$. Let $\left\{ z^k \right\}_{k \in \mathbb{N}}$ be a weakly convergent sequence in X such that, for some

subsequence $\{\Omega_{i_k}\}_{k\in\mathbb{N}}$ of $\{\Omega_k\}_{k\in\mathbb{N}}$, we have $z^k \in \Omega_{i_k}$ for all $k \in \mathbb{N}$. According to (47), the vectors $w^k := \mathrm{Proj}_{\Omega}(z^k)$ have the property that

$$\|z^k - w^k\| \le b(z^k)\gamma_k,$$

where, since $\{z^k\}_{k\in\mathbb{N}}$ is bounded, the sequence $\{b(z^k)\}_{k\in\mathbb{N}}$ is bounded too. Hence, $\lim_{k\to\infty}\|z^k - w^k\| = 0$ and ($B(ii)$) is satisfied.

Now we prove that ($B(i)$) is also satisfied. To this end, we consider the function $g: X\times\mathbb{N}\to X^*$ defined by

$$g(x,k)=\begin{cases} L(x)+l & \text{if } k=0,\\ L^{k-1}(x)+l^{k-1} & \text{if } k\ge 1.\end{cases}$$

and the point-to-set mapping $\Gamma: X\times\mathbb{N}\to 2^{X^*}$ defined by

$$\Gamma(x,k)=g(x,k)-K \quad , \quad \forall \quad x\in X \text{ and } k\ge 0.$$

Since Ω is regular (see (44)), we have that $0\in \mathrm{Int}[\mathrm{Im}\Gamma(\cdot,0)]$.

Clearly, $\Omega=\Gamma^{-1}(\cdot,0)(0)$. Let $u^0\in\Omega$. By Theorem 1 in [49], there exists $\eta>0$ such that

$$B_{X^*}(0,\eta)\subseteq g(B_X(u^0,1),0)-K, \tag{48}$$

Since $\|L^k - L\|_\infty$ and $\|l^k - l\|_*$ converge to 0, there exists $k_0\in\mathbb{N}$ such that for any integer $k\ge k_0$ we have

$$\|g(x,0)-g(x,k)\|_* = \|L(x)+l-L^{k-1}(x)-l^{k-1}\|_* \le \frac{\eta}{2}, \quad \forall x\in B_X(u^0,1).$$

This implies that whenever $k\ge k_0$ we have

$$g(B_X(u^0,1),0)\subseteq g\left(B_X(u^0,1),k\right)-K+B_X\left(u^0,\frac{\eta}{2}\right). \tag{49}$$

This and (48) show that the function g satisfies the assumptions in [49, Corollary 2] (with $-K$ instead of K). Consequently, application of this result yields that, for each $x\in\Omega$ and for any integer $k\ge k_0$, the set

$$\Omega_{k+1} = \{x \in X : g(x,k) \in K\}$$

contains the open ball $B_X\left(0, \frac{\eta}{2}\right)$ and that for any $x \in X$ we have

$$\operatorname{dist}(x, \Omega_k) \le \frac{2}{\eta}\left(1 + \left\|x - u^0\right\|\right) \operatorname{dist}_*(0, g(x,k) - K). \tag{50}$$

Note that, if $x \in \Omega$, then $g(x,0) \in K$ and, therefore, we have

$$\begin{aligned}
\operatorname{dist}_*(0, g(x,k) - K) &= \operatorname{dist}_*(g(x,k), K) \\
&\le \left\|g(x,k) - g(x,0)\right\|_* \\
&= \left\|L^{k-1}(x) + l^{k-1} - L(x) - l\right\|_* \\
&\le \left\|L^{k-1} - L\right\|_\diamond \|x\| + \left\|l^{k-1} - l\right\|_* \\
&\le \left(\|x\| + 1\right) \max\left\{\left\|L^{k-1} - L\right\|_\diamond, \left\|l^{k-1} - l\right\|_*\right\}.
\end{aligned}$$

This and (50) implies

$$\operatorname{dist}(x, \Omega_k) \le \frac{2}{\eta}\left(1 + \|x\| + \left\|u^0\right\|\right)\left(\|x\| + 1\right) \max\left\{\left\|L^{k-1} - L\right\|_\diamond, \left\|l^{k-1} - l\right\|_*\right\}, \tag{51}$$

for any $x \in \Omega$. Define the function $a : [0, +\infty) \to [0, +\infty)$ by

$$a(t) = \frac{2}{\eta}\left(1 + \left\|u^0\right\| + t\right)(t+1)$$

and the sequence of nonnegative real numbers

$$\beta_k = \max\left\{\left\|L^{k-1} - L\right\|_\diamond, \left\|l^{k-1} - l\right\|_*\right\} \quad \forall k \in \mathbb{N}.$$

According to (51) we have

$$\left\|x - \operatorname{Proj}_{\Omega_k}(x)\right\| = \operatorname{dist}(x, \Omega_k) \le a(\|x\|)\beta_k,$$

for all integers $k \ge k_0$ and for all $x \in \Omega$. Since, by hypothesis, $\lim_{k \to \infty} \beta_k = 0$, condition ($B(i)$) holds. □

3.7 The implementation of the regularization procedure discussed in this work requires computing vectors x^k defined by (30) with operators A^k given by (29). This implicitly means solving problems like (31). In some circumstances, in the regularization process, one can reduce problems placed in infinite dimensional settings to finite dimensional problems for which many efficient techniques of computing solutions are available. This is typically the case of the problem considered in the following example.

Let $X = \ell^p$ with $p \in (1,\infty)$, $q = p(p-1)^{-1}$ and, then, $X^* = \ell^q$. Let $a \in \ell^q \setminus \{0\}$ and $b^j \in \ell^q_+ \setminus \{0\}$, for all $j \in J$, where J is a nonempty set of indices and ℓ^q_+ stands for the subset of ℓ^q consisting of vectors with nonnegative coordinates. For each $j \in J$, let β_j be a nonnegative real number. Consider (P) to be the following optimization problem in ℓ^p :

$$\text{Minimize } F(x) = \langle a, x \rangle \tag{52}$$

over the set

$$\Omega := \{x \in \ell^p_+ : \langle b^j, x \rangle \le \beta_j,\ j \in J\}. \tag{53}$$

We assume that $a = (a_1, ..., a_i, ...)$ has infinitely many coordinates $a_i \neq 0$ and that the problem (P) has optimal solutions. Whenever u is an element in ℓ^p or in ℓ^q, we denote by $u[k]$ the vector in the same space as u obtained by replacing by zero all coordinates u_i of u with $i > k$. With this notations, for each $k \in \mathbb{N}$, let $\alpha_k := \|a - a[k]\|_*^{1/2}$, and observe that $\{\alpha_k\}_{k\in\mathbb{N}}$ is a sequence of positive real numbers which converge to zero as $k \to \infty$. We associate to problem (52)+(53) the perturbed problems (P_k) given by

$$\text{Minimize } F_k(x) = \langle a[k], x \rangle \text{ over } \Omega. \tag{54}$$

Note that, for each $k \in \mathbb{N}$, the problem (54) has optimal solutions because its objective function F_k is bounded from below on Ω by $F^* = \inf\{F(x) : x \in \Omega\}$. Problem (P) is ill posed and, therefore, even if one can find an optimal solution y^k for each of the essentially finite dimensional linear programming problems (P_k), the sequences $\{y^k\}_{k\in\mathbb{N}}$ may not converge in ℓ^p or, at best, its weak accumulation points (if any) are optimal solutions of (P).

We apply the regularization method (4) to the problems (P) and (P_k) with the function $\mu(t) = t^{p-1}$. It is easy to see that, in this case, determining

the vector x^k defined by (30) reduces to finding the unique optimal solution of the problem

$$\text{Minimize } \langle a[k], x\rangle + \alpha_k \|x\|^p \text{ over } \Omega. \tag{55}$$

Theorem 3.2 applies to problems (P) and (P_k) and guarantees that the sequence $\{x^k\}_{k\in\mathbb{N}}$ converges strongly to the minimal norm solution of (P). Indeed, observe that condition (A) is satisfied because, for any $x \in \ell^p$, we have

$$|F(x) - F_k(x)| \le \|a - a[k]\|_* \|x\| = \alpha_k^2 \|x\|,$$

and condition (B) trivially holds. It remains to prove that (32) holds too, that is, for any optimal solution v of (P) there exists a sequence $\{v^k\}_{k\in\mathbb{N}}$ of vectors in Ω such that

$$\lim_{k\to\infty} \|v^k - v\| = 0 = \lim_{k\to\infty} \alpha_k^{-1} \left\| \text{Proj}_{a[k]+N_\Omega(v^k)}(0) \right\|_*. \tag{56}$$

Take the constant sequence $v^k = v$, $(k \in \mathbb{N})$. Then the first equality in (56) holds and, for any $x \in \Omega$, we have $\langle -a, x - v\rangle = \langle a, v\rangle - \langle a, x\rangle \le 0$, showing that $-a \in N_\Omega(v)$. Thus, for each $k \in \mathbb{N}$, we obtain that

$$\left\| \text{Proj}_{a[k]+N_\Omega(v^k)}(0) \right\|_* \le \|a[k] - a\|_* = \alpha_k^2$$

and this implies the second equality in (56).

Solving problem (55) can be done by finding the unique optimal solution u^k of the following optimization problem in $\mathbb{R}^k$

$$\text{Minimize } \sum_{i=1}^{k} a_i x_i + a_k \sum_{i=1}^{k} |x_i|^p \quad \text{s.t. } \sum_{i=1}^{k} b_i^j x_i \le \beta_j,\ (j \in J),\ x \ge 0 \tag{57}$$

and taking $x^k = (u_1^k, \ldots, u_k^k, 0, \ldots)$. Indeed, for any $x \in \Omega$ we have

$$\begin{aligned} \langle a[k], x\rangle + \alpha_k \|x\|^p &= \langle a[k], x[k]\rangle + \alpha_k \|x\|^p \\ &\ge \langle a[k], x[k]\rangle + \alpha_k \|x[k]\|^p \ge \langle a[k], u^k\rangle + \alpha \|u^k\|^p \end{aligned}$$

where the last inequality holds because $x[k] \in \Omega$ (due to the nonnegativity of the vectors b^j).

4. REGULARIZATION OF A PROXIMAL POINT METHOD

4.1 A question of interest in convex optimization concerns the strong convergence of the *generalized proximal point method* (GPPM for short) which emerged from the works of Martinet [43], [44], Rockafellar [52] and Censor and Zenios [21]. When applied to the consistent problem (P) described in Subsection 3.1 the GPPM produces iterates according to the rule

$$y^0 \in \Omega \text{ and } y^{k+1} := \arg \min\left\{F(x) + \omega_k D_G(x, y^k) : x \in \Omega\right\} \tag{58}$$

with $D_G : \mathrm{Dom}(G) \times \mathrm{Int}(\mathrm{Dom}\, G) \to [0, +\infty)$ defined by

$$D_G(x, y) := G(x) - G(y) - \langle \nabla G(y), x - y \rangle, \quad \forall y \in \mathrm{Int}(\mathrm{Dom}\, G), \tag{59}$$

where $\{\omega_k\}_{k \in \mathbb{N}}$ is a bounded sequence of positive real numbers and $G : X \to (-\infty, +\infty]$ is a *Bregman function* on Ω, that is, a function satisfying the following conditions:

(i) $\Omega \subseteq \mathrm{Int}(\mathrm{Dom}\, G)$;
(ii) G is Fréchet differentiable on $\mathrm{Int}(\mathrm{Dom}\, G)$;
(iii) G is uniformly convex on bounded subsets of Ω;
(iv) For each $x \in \Omega$, the sets

$$R_\alpha^G(x) = \{y \in \Omega : D_G(x, y) \le \alpha\}$$

are bounded for all real numbers $\alpha > 0$.

The sequences $\{y^k\}_{k \in \mathbb{N}}$ generated by the GPPM are well defined, bounded and their weak accumulation points are solutions of (P) - cf. [17]. Weak convergence of these sequences can be ensured only when the Bregman function G has very special properties as, for instance, when ∇G is sequentially weakly-to-weak* continuous on Ω - (see [18, Chapter 3]). Strong convergence may not happen at all even when weak convergence does occur. This is in fact the case of the classical proximal point method for

optimization which is the particular version of GPPM in Hilbert spaces in which $G = \|\cdot\|^2$ (cf. [28]). The conditions under which the GPPM is known to converge strongly (see [52], [35], [7], [17], [20] and the references therein) are quite restrictive and mostly concern the data of (P) [in contrast to those ensuring weak convergence which mostly concern the Bregman function G whose selection can be done from a relatively large pool of known candidates - cf. [18]]. We are going to prove, by applying Theorem 3.2 and its corollaries, that a regularized version of the GPPM produces sequences which behave better than the sequences $\{y^k\}_{k\in\mathbb{N}}$ associated to (P) by (58). By contrast to the regularization method of GPPM proposed in [57] which, in Hilbert spaces, produces strongly convergent sequences whose limits are the projection of their initial points onto the set of optima of (P), the sequences resulting from the regularized version of GPPM proposed here converge strongly to the minimal norm solution of (P).

4.2 From now on we assume that X is an uniformly convex and uniformly smooth Banach space. We are going to show that in this not necessarily Hilbertian setting, by regularizing GPPM following the technique defined in (4), one obtains a procedure which generates sequences converging strongly to optima of (P). To this end, we denote $G(x) = \|x\|^p$ and $\psi(t) = pt^{p-1}$ for some $p \in (1,+\infty)$. Recall (cf. [19]) that G is a Bregman function and that G' is exactly the duality mapping J^{ψ}. We denote by S the presumed nonempty set of optimal solutions of the problem (P) described in Subsection 3.1.

Theorem. *Let* $\{\Omega_k\}_{k\in\mathbb{N}}$ *be a sequence of closed convex sets contained in* Ω *such that* $(B(i))$ *is satisfied and*

(a) $S \cap \left(\bigcap_{k=0}^{\infty} \Omega_k\right) \neq \varnothing$;

(b) $\operatorname{Lim}\ (S \cap \Omega_k) = S$.

If $\{\alpha_k\}_{k\in\mathbb{N}}$ *and* $\{\omega_k\}_{k\in\mathbb{N}}$ *are sequences of positive real numbers such that the first converges to zero and the second has the property that* $\lim_{k\to\infty}(\omega_{k+1}/\alpha_k) = 0$, *then, for any initial point* $y^0 \in \Omega_0$, *the sequences* $\{x^k\}_{k\in\mathbb{N}}$ *and* $\{y^k\}_{k\in\mathbb{N}}$ *generated according to the rule*

$$y^k = \arg\ \min\left\{F(x) + \omega_k D_G(x, y^{k-1}) : x \in \Omega_k\right\},$$

$$x^k = \left[\partial F + N_{\Omega_k} + \omega_k(J^{\psi} - J^{\psi} y^k) + \alpha_k J^{\mu}\right]^{-1}(0), \tag{60}$$

are well defined and have the following properties:

(*i*) *The sequence* $\{y^k\}_{k\in\mathbb{N}}$ *is bounded, the sequence* $\{F(y^k)\}_{k\in\mathbb{N}}$ *converges and*

$$\lim_{k\to\infty} F(y^k) = \inf\{F(y) : y \in \Omega\}; \tag{61}$$

(*ii*) *The sequence* $\{x^k\}_{k\in\mathbb{N}}$ *converges strongly to a solution of* (P).

Proof. For each $k \in \mathbb{N}$, define the functions $E_k, H_k : X \to (-\infty, +\infty]$ by

$$E_k(x) = F(x) + \iota_{\Omega_k}(x),$$

and

$$H_k(x) = E_k(x) + \omega_k D_G(x, y^{k-1}),$$

where ι_{Ω_k} stands for the indicator function of the set Ω_k. The functions E_k and H_k are lower semicontinuous, convex and bounded from below. According to [18, Proposition 3.1.5] applied to them we deduce that, for any integer $k \geq 1$, the vector

$$y^k = \arg\ \min\{H_k(y) : y \in \Omega_k\}.$$

exists and is well defined. Note that this is exactly the vector y^k given by (60) and, thus, the sequence $\{y^k\}_{k\in\mathbb{N}}$ is well defined. Let $\overline{z} \in S \cap \left(\bigcap_{k=0}^{\infty} \Omega_k\right)$. Observe that, for each positive integer k, the vector $\overline{z}$ is also a minimizer of the function E_k over Ω_k. An argument similar to that in the proof of [18, Proposition 3.1.6] applied to the functions E_k and H_k shows that

$$D_G(\overline{z}, y^{k-1}) - D_G(\overline{z}, y^k) - D_G(y^k, y^{k-1}) \geq \frac{1}{\omega_k}\left[E_k(y^k) - E_k(\overline{z})\right],$$

for all integers $k \geq 1$. Thus, if ω is a positive upper bound of the bounded sequence $\{\omega_k\}_{k\in N}$, we get

$$D_G(\overline{z}, y^{k-1}) - D_G(\overline{z}, y^k) - D_G(y^k, y^{k-1}) \geq \frac{1}{\omega}\left[F(y^k) - F(\overline{z})\right] \geq 0, \tag{62}$$

for all integers $k \geq 1$, because $y^k \in \Omega_k \subseteq \Omega$ and $\overline{z}$ is a solution of (P). From (62) it can be easily seen that the sequence $\{D_G(\overline{z}, y^k)\}_{k\in\mathbb{N}}$ is nonincreasing, hence, convergent, and, consequently, that the sequence $\{D_G(y^k, y^{k-1})\}_{k\in N}$ converges to zero. These and (62) imply that $\lim_{k\to\infty}[F(y^k) - F(\overline{z})] = 0$. Hence, (61) is proved. Boundedness of $\{y^k\}_{k\in\mathbb{N}}$ follows from the fact that $\{D_G(\overline{z}, y^k)\}_{k\in\mathbb{N}}$ is bounded by $\alpha = D_G(\overline{z}, y^0)$ and, then, all y^k are contained in the set $R_\alpha^G(\overline{z})$ which is bounded because, as noted above, G is a Bregman function. Hence, the proof of (i) is complete.

In order to prove (ii), we apply Theorem 3.2 to the problem (P) given at (23) and to the problems (P_k) with the functions $F_k : X \to (-\infty, +\infty]$ given by

$$F_k(x) := F(x) + \omega_{k+1} D_G(x, y^k),$$

where, for each nonnegative integer k, the vector y^k is defined by (60), that is,

$$y^k = \arg\ \min\{F_{k-1}(x) : x \in \Omega_k\}.$$

Note that F_k is convex, lower semicontinuous and has $\mathrm{Dom}\, F_k = \mathrm{Dom}\, F$. Also, by Asplund's Theorem which shows that ∇G is exactly the duality mapping J^ψ, we obtain

$$\partial F_k(x) := \partial F(x) + \omega_{k+1}(J^\psi x - J^\psi y^k).$$

We associate to each function F_k the maximal monotone operator A^k defined by (29). Observe that the vectors x^k defined by (60) are exactly those given by (30) for this specific operator A^k and, thus, it is well defined. We show next that the operators A^k have the properties required by the hypothesis of Theorem 3.2.

Let $x \in \mathrm{Dom}\, F$. Then, for any $k \in \mathbb{N}$, we have

$$\begin{aligned} F_k(x) - F(x) &= \omega_{k+1} D_G(x, y^k) \\ &= \omega_{k+1}\left(\|x\|^p - \|y^k\|^p - \langle J^\psi y^k, x - y^k\rangle\right) \\ &= \omega_{k+1}\left(\|x\|^p + (p-1)\|y^k\|^p - \langle J^\psi y^k, x\rangle\right) \\ &\le \omega_{k+1}\left(\|x\|^p + (p-1)\|y^k\|^p + \|J^\psi y^k\|_* \|x\|\right) \\ &= \omega_{k+1}\left(\|x\|^p + (p-1)\|y^k\|^p + p\|y^k\|^{p-1}\|x\|\right). \end{aligned}$$

The sequence $\{y^k\}_{k\in\mathbb{N}}$ is bounded as shown above. Let M be a positive upper bound of the sequence $\{\|y^k\|\}_{k\in N}$. Define the continuous function $c:[0,+\infty)\to[0,+\infty)$ by

$$c(t) = t^p + pM^{p-1}t + (p-1)M^{p-1}.$$

Hence, for each $k\in\mathbb{N}$ we have

$$|F_k(x) - F(x)| \le \omega_{k+1}\left[\|x\|^p + (p-1)M^p + pM^{p-1}\|x\|\right] = \omega_{k+1}c(\|x\|),$$

showing that condition (A) is satisfied with $\delta_k = \omega_{k+1}$, $k\in\mathbb{N}$. Condition $(B(ii))$ holds in our case because, by hypothesis, all Ω_k are subsets of Ω. So, condition (B) is satisfied.

It remains to show that for any solution $v\in\Omega$ of (P) there exists a sequence $\{v^k\}_{k\in\mathbb{N}}$, which has $v^k\in\Omega_k$ for all $k\in\mathbb{N}$, and satisfies (32). To this end, note that each $S_k := \Omega_k \cap S$ is a nonempty, closed and convex subset of X. By (b), we have

$$S = \operatorname{Lim} S_k. \tag{63}$$

Also, there exists a sequence $\{w^k\}_{k\in N}$ such that, for each $k\in\mathbb{N}$, $w^k\in\Omega_k$ and $\lim_{k\to\infty}\|w^k - v\| = 0$. Let $v^k = \operatorname{Proj}_{S_k}(w^k)$. Observe that, according to [24, p. 40], the space X is an E-space because it is uniformly convex. Therefore, Theorem 10 in [24, p. 49] combined with (63) imply that

$$\lim_{k\to\infty} v^k = \lim_{k\to\infty} \operatorname{Proj}_{S_k}(w^k) = \operatorname{Proj}_S(v) = v. \tag{64}$$

Now, observe that each $v^k \in S_k$ and, hence, is a minimizer of E_k over X, that is, $0 \in \partial E_k(v^k)$. Similarly to F, the functions F_k are continuous on the interiors of their respective domains and, therefore, they are continuous on Ω_k (see (26) and (28)). By consequence, we have

$$\partial E_k(v^k) = \partial F(v^k) + \partial \iota_{\Omega_k}(v^k) = \partial F(v^k) + N_{\Omega_k}(v^k). \tag{65}$$

From (65) we obtain

$$0 \in \partial F(v^k) + N_{\Omega_k}(v^k). \tag{66}$$

For each $k \in \mathbb{N}$, we have

$$\begin{aligned}\left\|\mathrm{Proj}_{\partial F_k(v^k)+N_{\Omega_k}(v^k)}(0)\right\|_* &= \mathrm{dist}_*\left(0, \partial F(v^k) + N_{\Omega_k}(v^k) + \omega_{k+1}\left(J^\psi v^k - J^\psi y^k\right)\right)\\ &= \mathrm{dist}_*\left(\omega_{k+1}\left(J^\psi y^k - J^\psi v^k\right), \partial F(v^k) + N_{\Omega_k}(v^k)\right)\\ &\le \omega_{k+1}\left\|J^\psi y^k - J^\psi v^k\right\|_*,\end{aligned}$$

where the last inequality follows from (66). By consequence, taking into account (3), we obtain

$$\begin{aligned}\left\|\mathrm{Proj}_{\partial F_k(v^k)+N_{\Omega_k}(v^k)}(0)\right\|_* &\le \omega_{k+1}\left\|J^\psi v^k - J^\psi y^k\right\|_*\\ &\le \omega_{k+1}\left(\left\|J^\psi v^k\right\|_* + \left\|J^\psi y^k\right\|_*\right)\\ &= p\omega_{k+1}\left(\left\|v^k\right\|^{p-1} + \left\|y^k\right\|^{p-1}\right),\end{aligned}$$

and, thus, we have

$$\left\|\mathrm{Proj}_{\partial F_k(v^k)+N_{\Omega_k}(v^k)}(0)\right\|_* \le p\omega_{k+1}\left(\left\|v^k\right\|^{p-1} + M^{p-1}\right), \quad \forall k \in \mathbb{N}. \tag{67}$$

Let N be an upper bound of the sequence $\left\{\left\|v^k\right\|\right\}_{k\in\mathbb{N}}$ and denote

$$q := p\left(N^{p-1} + M^{p-1}\right).$$

Then, by (67), we have

$$\alpha_k^{-1}\left\|\mathrm{Proj}_{\partial F_k(v^k)+N_{\Omega_k}(v^k)}(0)\right\|_* \le q\frac{\omega_{k+1}}{\alpha_k}, \forall k \in \mathbb{N}. \tag{68}$$

This and (64) imply (32) by hypothesis. According to (68), condition (32) holds too. These show that Theorem 3.2 is applicable to F and to the functions F_k. In turn, Theorem 3.2 implies that the sequence $\{x^k\}_{k\in\mathbb{N}}$ converges strongly to the minimal norm minimizer of F over Ω. □

4.3 Verifying the conditions (a) and (b) of Theorem 4.2 may be difficult. In some circumstances the following consequence of Theorem 4.2 may be of use. For instance, if X has a countable system of generators $\{e^k\}_{k\in\mathbb{N}}$ and the problem (P) is unconstrained (i.e., $\Omega = X$), then using the next result with the sets

$$\Omega_k = \overline{\mathrm{aff}}\{e^i : 0 \le i \le k\},$$

which necessarily satisfy condition (c) below, one reduces the resolution of (P) to solving a sequence of unconstrained problems in spaces of finite dimension whose solutions x^k will necessarily converge strongly to an optimum of (P).

Corollary. *Let* $\{\Omega_k\}_{k\in\mathbb{N}}$ *be a sequence of closed convex subsets of* Ω *such that* $(B(i))$ *is satisfied and one of the following conditions hold:*

(c) $S \subseteq \bigcup_{k=0}^{\infty}\Omega_k$ *and* $\Omega_k \subseteq \Omega_{k+1}$ *for all* $k \in \mathbb{N}$;

(d) $\mathrm{Int}\, S \neq \varnothing$ *and* $S \cap \Omega_k \neq \varnothing$ *for all* $k \in \mathbb{N}$.

If $\{\alpha_k\}_{k\in\mathbb{N}}$ *and* $\{\omega_k\}_{k\in\mathbb{N}}$ *are sequences of positive real numbers such that the first converges to zero and the second has the property that* $\lim_{k\to\infty}(\omega_{k+1}/\alpha_k) = 0$, *then, for any initial point* $y^0 \in \Omega$, *the sequences* $\{x^k\}_{k\in\mathbb{N}}$ *and* $\{y^k\}_{k\in\mathbb{N}}$ *generated according to the rule* (60) *are well defined and have the properties* (i) *and* (ii) *from Theorem 4.2.*

Proof. Suppose that condition (c) holds. Take any $\overline{z} \in S$. There exists a number $k_0 \in \mathbb{N}$ such that $\overline{z} \in \Omega_{k_0}$. Denote $\Omega_0' = \bigcup_{k=0}^{k_0}\Omega_k$ and $\Omega_k' = \Omega_{k_0+k}$ for $k \ge 1$. Applying Lemma 1.2 from [45] we deduce that $\mathrm{Lim}(\Omega_k' \cap S) = S$. Applying Theorem 4.2 to the sets Ω_k' we obtain the result. Now, assume that (d) holds. Then, Lemma 1.4 from [45] guarantees that condition (b) of Theorem 4.2 is satisfied and we can apply that proposition in this case too.

□

4.4 The previous results in this section deal with the case that the sets Ω_k are contained in Ω. If F has bounded level sets

$$L_F(\alpha) := \{x \in X : F(x) \leq \alpha\},$$

then the regularized proximal point method (60) is also stable under outer approximations of Ω.

Proposition. *Let* $\{\Omega_k\}_{k\in\mathbb{N}}$ *be a sequence of closed convex sets contained in* $\mathrm{Int}(\mathrm{Dom}\, F)$ *such that condition* (B) *is satisfied,* $\Omega = \bigcap_{k=0}^{\infty} \Omega_k$ *and* $\Omega_{k+1} \subseteq \Omega_k$ *for all* $k \in \mathbb{N}$. *Let* $\{\alpha_k\}_{k\in\mathbb{N}}$ *be a sequence of positive real numbers converging to zero. Suppose that for each number* $\alpha \geq 0$ *the level set* $L_F(\alpha)$ *of the objective function* F *is bounded. Then, for any initial point* $y^0 \in \Omega$, *the sequences* $\{x^k\}_{k\in\mathbb{N}}$ *and* $\{y^k\}_{k\in\mathbb{N}}$ *generated according to the rule* (60) *where the positive numbers* ω_k *are chosen such that* $\lim_{k\to\infty}(\omega_{k+1}/\alpha_k) = 0$ *and for some positive number* K,

$$\omega_k D_G(y^0, y^{k-1}) \leq K, \tag{69}$$

are well defined and have the properties (i) *and* (ii) *from Theorem 4.2.*

Proof. We use the notations F_k, E_k and H_k introduced in the proof of Theorem 4.2. Observe that, since X is reflexive and the level sets of F are bounded, for any $k \in \mathbb{N}$, there exists

$$z^k \in \arg\ \min\{F(x) : x \in \Omega_k\}.$$

According to [18, Proposition 3.1.5] applied to E_k and H_k we deduce that, for any integer $k \geq 1$, the vector

$$y^k = \arg\ \min\{H_k(y) : y \in \Omega_k\}.$$

exists. Let $z \in S$ and observe that, for any $k \in \mathbb{N}$,

$$F(z^k) \leq F(z^{k+1}) \leq F(z), \tag{70}$$

because $\Omega \subseteq \Omega_k$ and $\Omega_{k+1} \subseteq \Omega_k$. This shows that all z^k belong to $L_F(F(z))$ and, therefore, the sequence $\{z^k\}_{k\in\mathbb{N}}$ is bounded. Since X is reflexive, the bounded sequence $\{z^k\}_{k\in\mathbb{N}}$ contains a weakly convergent subsequence $\{z^{i_k}\}_{k\in\mathbb{N}}$. Let z' be the weak limit of $\{z^{i_k}\}_{k\in\mathbb{N}}$. According to [45, Lemma 1.3], we have that

$$\Omega = \bigcap_{k=0}^{\infty} \Omega_k = \operatorname{Lim} \Omega_k$$

and this implies that $z' \in \Omega$. Hence, by taking (70) into account we get

$$F(z^k) \le F(z) \le F(z').$$

Since F is lower semicontinuous and the sequence $\{F(z^k)\}_{k\in\mathbb{N}}$ is nondecreasing this implies

$$F(z') \le \lim_{k\to\infty} F(z^{i_k}) = \lim_{k\to\infty} F(z^k) \le F(z) \le F(z'),$$

that is, $F(z') = F(z)$ showing that $\{F(z^k)\}_{k\in\mathbb{N}}$ converges to the minimal value of F over Ω. Now observe that according to (60) and (69) we have

$$\begin{aligned} F(y^k) &\le F(y^k) + \omega_k D_G(y^k, y^{k-1}) \\ &\le F(y^0) + \omega_k D_G(y^0, y^{k-1}) \\ &\le F(y^0) + K \end{aligned}$$

because $y^0 \in \Omega \subseteq \Omega_k$. This implies that the sequence $\{y^k\}_{k\in\mathbb{N}}$ is bounded because it is contained in $L_F(F(y^0) + K)$. Also according to (60) we have

$$\begin{aligned} &0 \le F(y^k) - F(z^k) \le \omega_k \left[D_G(z^k, y^{k-1}) - D_G(y^k, y^{k-1}) \right] \\ &= \omega_k \left[\|z^k\|^p - \|y^k\|^p + \left\langle J^{\psi} y^{k-1}, y^k - z^k \right\rangle \right], \end{aligned} \tag{71}$$

where the quantity between the square brackets is bounded because both sequences $\{y^k\}_{k\in\mathbb{N}}$ and $\{z^k\}_{k\in\mathbb{N}}$ are bounded as shown above. Note that $\{\omega_k\}_{k\in\mathbb{N}}$ converges to zero. Hence, by (71), we obtain that

$$\lim_{k\to\infty}\left[F(y^k)-F(z^k)\right]=0,$$

and this proves (i).

For proving (ii) one reproduces without modifications the arguments made for the same purpose in the proof of Theorem 4.2 and keeping in mind that in the current circumstances we have that $S_k := S\cap\Omega_k = S$ for all $k\in\mathbb{N}$. □

ACKNOWLEDGMENTS

The work of Yakov Alber was supported in part by the KAMEA Program of the Israeli Ministry of Absorption. Dan Butnariu gratefully acknowledges the support of the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities (Grant 592/00). The work of Gábor Kassay was partially supported by a grant of the foundation "Sapientia" from Cluj-Napoca, Romania.

REFERENCES

[1] Alber Y., The solution of nonlinear equations with monotone operators in Banach spaces (Russian), *Siberian Mathematical Journal*, **16** (1975), 1-8.

[2] Alber Y., The regularization method for variational inequalities with nonsmooth unbounded operators in Banach space, *Applied Mathematical Letters*, **6** (1993), 63-68.

[3] Alber Y., Stability of the proximal projection algorithm for nonsmooth convex optimization problems with perturbed constraint sets, *preprint* 2001.

[4] Alber Y., Butnariu D. and Ryazantseva, I., Regularization methods for ill-posed inclusions and variational inequalities with domain perturbations, *Journal of Nonlinear and Convex Analysis*, **2** (2001), 53-79.

[5] Alber Y. and Notik A.I., Perturbed unstable variational inequalities with unbounded operators on approximately given sets, *Set-Valued Analysis*, **1** (1993), 393-402.

[6] Alber Y. and Ryazantseva I., Variational inequalities with discontinuous mappings, *Soviet Mathematics Doklady*, ,**25** (1982), 206-210.

[7] Alber Y., Burachik R. and Iusem A.N., A proximal point method for nonsmooth convex optimization problems in Banach spaces, *Abstract and Applied Analysis*, **2** (1997), 100-120.

[8] Alexandre P., The perturbed generalized Tikhonov's algorithm, *Serdica Mathematical Journal*, **25** (1999), 91-102.

[9] Alexandre P., Nguyen V.H. and Tossings P., The perturbed generalized proximal algorithm, *Mathematical Modelling and Numerical Analysis*, **32** (1998), 223-253.

[10] Aubin J.-P. and Ekeland, Applied Nonlinear Analysis, *John-Wiley and Sons*, New York, 1984.

[11] Aubin J.-P. and Frankowska H., Set-Valued Analysis, *Birkhäuser*, Boston, 1990.

[12] Bauschke H.H. and Combettes P.L., A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces, *Mathematics of Operation Research*, **26** (2000), 248-264.

[13] Beer G., Topologies on Closed and Closed Convex Sets, *Kluwer Academic Publishers*, Dordrecht, 1993.

[14] Bonnans J.F. and Shapiro A., Perturbation Analysis of Optimization Problems, *Springer Verlag*, 2000.

[15] Brohe M. and Tossings P., Perturbed proximal point algorithm with nonquadratic kernel, *Serdica Mathematical Journal*, **26** (2000), 177-206.

[16] Browder F.E., Existence and approximation of solutions for nonlinear variational inequalities, *Proceedings of the National Academy of Science USA*, **56** (1966), 1080-1086.

[17] Butnariu D. and Iusem A.N., On a proximal point method of optimization in Banach spaces, *Numerical Functional Analysis and Optimization*, **18**, (1998), 723-744.

[18] Butnariu D. and Iusem A.N., Totally Convex Functions for Fixed Point Computation and Infinite Dimensional Optimization, *Kluwer Academic Publishers*, Dordrecht, 2000.

[19] Butnariu D., Iusem A.N. and Resmerita E., Total convexity for the powers of the norm in uniformly convex Banach spaces, *Journal of Convex Analysis* **7** (2000), 319-334.

[20] Butnariu D., Iusem A.N. and Zalinescu C., On uniform convexity, total convexity and the convergence of a proximal point and an outer Bregman projection method in Banach spaces, *Journal of Convex Analysis*, to appear.

[21] Censor Y. and Zenios S., Proximal minimization algorithm with D-functions, *Journal of Optimization Theory and Applications*, **73** (1992) 451-464.

[22] Cioranescu I., Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems, *Kluwer Academic Publishers*, Dordrecht, 1990.

[23] Cruceanu S., Regularization pour les problèmes à operateurs monotones et la mèthode de Galerkine, *Commentationes Mathematicae Universitatis Carolinae* 12 (1971), 1-13.

[24] Dontchev A.L. and Zolezzi T., Well-Posed Optimization Problems, *Springer Verlag*, Berlin, 1991.

[25] Eckstein J., Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, *Mathematics of Operation Research*, **18** (1993), 202-226.

[26] Eckstein J., Approximate iterations in Bregman-function-based proximal algorithms, *Mathematical Programming* **83** (1998), 113-123.

[27] Engl H.W., Hanke M. and Neubauer A., Regularization of Inverse Problems, *Kluwer Academic Publishers*, Dordrecht, 1996.

[28] Güler O., On the convergence of the proximal point algorithm for convex minimization, *SIAM Journal on Control and Optimization*, **29** (1991), 403-419.

[29] Iusem A., Augmented Lagrangian methods and proximal point methods for convex optimization, *Investigacion Operativa* **8** (1999), 11-49.

[30] Iusem A.N. and Otero R.G., Inexact version of proximal point and augmented Lagrangian algorithms in Banach spaces, *Numerical Functional Analysis and Optimization* **22** (2001), 609-640.

[31] Kaplan A. and Tichatschke R., Stable Methods for Ill-Posed Variational Problems, *Akademie Verlag*, 1994.

[32] Kaplan A. and Tichatschke R., Proximal point approach and approximation of variational inequalities, *SIAM Journal on Control and Optimization*, **39** (2000), 1136-1159.

[33] Kaplan A. and Tichatschke R., Proximal methods for variational inequalities with set-valued monotone operators, In: "From Convexity to Nonconvexity" (Edited by R.P. Gilbert, P.D. Panagiotopoulos and P.M. Pardalos), pp. 345-361, *Kluwer Academic Publishers*, Dordrecht, 2001.

[34] Kaplan A. and Tichatschke R., A general view on proximal point methods to variational inequalities in Hilbert spaces - iterative regularization and approximation, *Journal of Nonlinear and Convex Analysis* **2** (2001), 305-332.

[35] Kassay G., The proximal point algorithm for reflexive Banach spaces, *Studia Universitatis "Babes-Bolyai" - Sectia Mathematica* **30** (1985), 9-17.

[36] Lavrentev M.M., On Some Ill-Posed Problems of Mathematical Physics (Russian), *Nauka*, Novosibirsk, 1962.

[37] Lions J.L. and Magenes E., Problèmes aux limites non homogènes et applications, Vol. 1, *Dunod*, Paris, 1968.

[38] Liskovets O.A., Variational Methods for Solving Unstable Problems (Russian), *Nauka i Tekhnika,* Minsk, 1981.

[39] Liskovets O.A., Regularization of problems with discontinuous monotone arbitrarily perturbed operators, *Soviet Mathematics Doklady* **28** (1983), 324-327.

[40] Liskovets O.A., Discrete convergence of elements and operators for ill-posed problems with a monotone operator, *Soviet Mathematics Doklady* **31** (1985), 202-206.

[41] Liskovets O.A., External approximations for the regularization of monotone variational inequalities, *Soviet Mathematics Doklady* **36** (1988), 220-224.

[42] Liu F. and Nashed M.Z., Regularization of nonlinear ill-posed variational inequalities and convergence rates, *Set Valued Analysis* **6** (1998), 313-344.

[43] Martinet B., Régularization d'inéquations variationelles par approximations successive, *Revue Française de Informatique at Recherche Opérationelle* **2** (1970), 154-159.

[44] Martinet B., Algorithmes pour la résolution de problèmes d'optimisation et minimax, Thèse d'état, *Université de Grenoble*, Grenoble, France, 1972.

[45] Mosco U., Convergence of convex sets and solutions of variational inequalities, *Advances in Mathematics* **3** (1969), 510-585.

[46] Mosco U., Perturbations of variational inequalities, in: "Nonlinear Functional Analysis", Proceedings of the Symposium in Pure Mathematics, 18, Part 1, pp. 182-194, *American Mathematical Society*, Providence, R.I., 1970.

[47] Pascali D. and Sburlan S., Nonlinear Mappings of Monotone Type, *Stijthoff&Noordhoff International Publishers*, Alphen aan den Rijn, The Netherlands, 1978.

[48] Phelps R.R., Convex Functions, Monotone Operators and Differentiability, 2nd Edition, *Springer Verlag*, Berlin, 1993.

[49] Robinson S.M., Regularity and stability for convex multivalued functions, *Mathematics of Operations Research* **1** (1976), 130-143

[50] Rockafellar R.T., On the maximality of sums of nonlinear monotone operators, *Transactions of the American Mathematical Society* **149** (1970), 75-88.

[51] Rockafellar R.T., On the maximal monotonicity of subdifferential mappings, *Pacific Journal of Mathematics* **33** (1970), 209-216.

[52] Rockafellar R.T., Monotone operators and the proximal point algorithm, *SIAM Journal on Control and Optimization* **14** (1976), 877-898.

[53] Rockafellar R.T., Augmented Lagrangians and applications of the proximal point algorithm in convex programming, *Mathematics of Operation Research* **1** (1976), 97-116.

[54] Ryazantseva I., Variational inequalities with monotone operators on sets which are specified approximately, *USSR Computational Mathematics and Mathematical Physics* **24** (1984), 194-197.

[55] Showalter R.E., Monotone Operators in Banach Space and Nonlinear Partial Differential Equations, *American Mathematical Society*, 1997.

[56] Solodov M.V. and Svaiter B.F., An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions, *Mathematics of Operation Research* **25** (2000), 214-230.

[57] Solodov M.V. and Svaiter B.F., Forcing strong convergence of proximal point iterations in a Hilbert space, *Mathematical Programming* **87** (2000), 189-202.

[58] Yamada I., The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings, In: "Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications" (D. Butnariu, Y. Censor and S. Reich, Editors), *Elsevier Science Publishers*, Amsterdam, The Netherlands, 2001, pp. 473-504.

PARTITIONABLE MIXED VARIATIONAL INEQUALITIES

E. Allevi,[1] A. Gnudi,[1] I.V. Konnov[2] and E.O. Mazurkevich[2]
Dept. of Mathematics, Statistics, Computer Science and Applications, Bergamo University, Bergamo, Italy;[1] Dept. of Applied Mathematics, Kazan University, Kazan, Russia[2]

Abstract: Two recent papers [1] and [2] have presented existence and uniqueness results for solutions of mixed variational inequality problems involving P-mappings and convex and separable but not necessarily differentiable functions where the feasible set is defined by box type constraints. In this paper we generalise these results for the case where the subspaces constituting the initial space are not real lines.

Key words: Mixed variational inequalities, nondifferentiable functions, product sets, order monotonicity, existence and uniqueness results.

1. INTRODUCTION

Many equilibrium problems arising in Mathematical Physics, Economics, Operations Research and other fields possess a partitionable structure which enable one to essentially weaken the conditions for existence and uniqueness results of solutions and for convergence of solution methods. Usually, such results are based on order monotonicity type assumptions, however, they are restricted with the case where subspaces are one-dimensional; see e.g. [3]–[5]. In two recent papers [1] and [2], several existence and uniqueness results for solutions of such problems involving P type mappings and convex and separable but not necessarily differentiable functions have been established.

In this work, we consider extensions of order monotonicity concepts for mappings to the case where subspaces need not be real lines with applications to the mixed variational inequality problems.

Let M be an index set $M = \{1,...,m\}$. We consider a fixed partition of the real Euclidean space R^n associated to M, i.e.

$$R^n = \prod_{s \in M} R^{n_s}, \tag{1}$$

hence for each $x \in R^n$ we have $x = (x_s \mid s \in M)$, where $x_s \in R^{n_s}$. Let K_s be a nonempty closed convex set in R^{n_s} for every $s \in M$, and let

$$K = \prod_{s \in M} K_s.$$

Let $Q : K \to R^n$ be a continuous mapping. Then we can consider also the partition $(Q_s \mid s \in M)$ of Q associated to M such that $Q_s : K \to R^{n_s}$ for $s \in M$. Let $f : K \to R$ be a function of the form $f(x) = \sum_{s \in M} f_s(x_s)$ where $f_s : K_s \to R$ be a convex function for each $s \in M$. We consider the partitionable *mixed variational inequality problem* (MVI for short) of the form: Find $x^* = (x_s^* \mid s \in M) \in K$ such that

$$\sum_{s \in M} \left(\left\langle Q_s(x^*), x_s - x_s^* \right\rangle + f_s(x_s) - f_s(x_s^*) \right) \geq 0 \quad \forall x_s \in K_s, \forall s \in M. \tag{2}$$

Note that MVI involves a continuous mapping and a convex, but not necessarily differentiable function. Taking into account the partition of the problem associated to M, we see that MVI (2) can be equivalently rewritten in the more standard form: Find $x^* \in K$ such that

$$\left\langle Q(x^*), x - x^* \right\rangle + f(x) - f(x^*) \geq 0 \quad \forall x \in K,$$

where $f(x) = \sum_{s \in M} f_s(x_s)$. We intend to present existence and uniqueness results of solutions for this problem under extended order monotonicity type assumptions on the cost mapping.

2. THEORETICAL BACKGROUND

So, we consider MVI (2) under the following standing assumptions.

(A1) *$Q: K \to R^n$ is a mapping with the partition $(Q_s \mid s \in M)$ associated to M such that $Q_s : K \to R^{n_s}, s \in M$ are continuous mappings.*

(A2) *$f: K \to R$ is of the form $f(x) = \sum_{s \in M} f_s(x_s)$ where $f_s : K_s \to R$ is a convex continuous function for every $s \in M$.*

(A3) *K is of the form*

$$K = \prod_{s \in M} K_s,$$

where K_s is a convex and closed subset of R^{n_s} for every $s \in M$.

Note that K is obviously convex and closed; in the case where $K_s = R_+^{n_s}$ for all $s \in M$, we obtain $K = R_+^n$, hence MVI (2) involves complementarity problems. First we give an equivalence result for MVI (2).

Proposition 1. *The following assertions are equivalent:*

(i) $x^* = (x_s^* \mid s \in M)$ is a solution to (2);

(ii) it holds that $x^* = (x_s^* \mid s \in M) \in K$ and

$$\langle Q_s(x^*), x_s - x_s^* \rangle + f_s(x_s) - f_s(x_s^*) \geq 0 \quad \forall x_s \in K_s, \forall s \in M; \tag{3}$$

(iii) it holds that $x^* = (x_s^* \mid s \in M) \in K$ and

$$\exists g_s^* \in \partial f_s(x_s^*) : \langle Q_s(x^*), x_s - x_s^* \rangle + \langle g_s^*, x_s - x_s^* \rangle \geq 0 \quad \forall x_s \in K_s, \forall s \in M. \tag{4}$$

Proof. It is clear that (iii) implies (ii) and (ii) implies (i). Conversely, let x^* solve (2) and there exist an index l and a point $y_l \in K_l$ such that

$$\langle Q_l(x^*), y_l - x_l^* \rangle + f_l(y_l) - f_l(x_l^*) < 0.$$

Set $\tilde{x} = (x_1^*, \ldots, x_{l-1}^*, y_l, x_{l+1}^*, \ldots, x_n^*) \in K$, then we have

$$\sum_{s \in M} \langle Q_s(x^*), \tilde{x}_s - x_s^* \rangle + f_s(\tilde{x}_s) - f_s(x_s^*) = \langle Q_l(x^*), y_l - x_l^* \rangle + f_l(y_l) - f_l(x_l^*) < 0,$$

which is a contradiction. Hence, (i) implies (ii). Next, if x^* solves (3), then it is a solution to the following convex programming problem:

$$\min_{x_s \in K_s} \sum_{s \in M} (\langle Q_s(x^*), x_s \rangle + f_s(x_s)).$$

It is clear that (4) represents necessary and sufficient optimality conditions for these problems. Therefore, (ii) implies (iii) and the proof is complete. □

Definition 1. Let M be an index set such that (1) holds, and let $F : R^n \to R^n$ be a mapping with the partition $(F_s \mid s \in M)$ associated to M. Then the mapping F is said to be

(a) a $P_0(M)$*-mapping*, if for all $x, y \in R^n$, $x \neq y$, there exists an index $s \in M$ such that $x_s \neq y_s$ and

$$\langle x_s - y_s, F_s(x) - F_s(y) \rangle \geq 0.$$

(b) a $P(M)$*-mapping*, if

$$\max_{s \in M} \langle x_s - y_s, F_s(x) - F_s(y) \rangle > 0 \text{ for all } x, y \in R^n, x \neq y;$$

(c) a *strict* $P(M)$*-mapping*, if there exists $\gamma > 0$ such that $F - \gamma I_n$ is a $P(M)$-mapping, where I_n is the identity map on R^n;

(d) a *uniform* $P(M)$*-mapping*, if

$$\max_{s \in M} \langle x_s - y_s, F_s(x) - F_s(y) \rangle \geq \mu \|x - y\|^2 \text{ for all } x, y \in R^n,$$

for some constant $\mu > 0$;

It is clear that each $P(M)$-mapping is a $P_0(M)$-mapping, each strict $P(M)$-mapping is a $P(M)$-mapping, and that each uniform $P(M)$-mapping is a strict $P(M)$-mapping. Moreover, each monotone (respectively, strictly, strongly monotone) mapping is a $P_0(M)$ (respectively, $P(M)$, uniform $P(M)$) - mapping.

Now we give an example of nonmonotone $P_0(M)$- and $P(M)$-mappings.

Example 1. Let us consider the mapping $Q : R^n \to R^n$ with the partition $(Q_s \mid s \in M)$ associated to M, such that (1) holds, $x = (x_s \mid s \in M)$, and

$$Q_s(x) = \begin{cases} F_s(x_s) + A_{sm}x_m & \text{if} \quad s \neq m, \\ -\sum_{i=1}^{m-1} \lambda_i A_{im}^T x_i + F_m(x_m) & \text{if} \quad s = m; \end{cases}$$

where $F_i : R^{n_i} \to R^{n_i}$ is a mapping, $\lambda_i > 0$, and A_{im} is an arbitrary $n_i \times n_m$ matrix for each $i \in M$. Suppose that the mappings F_i are monotone. Then Q may be nonmonotone since $\lambda_i \neq 1$ in general. However, it is clear that Q is a $P_0(M)$-mapping. Similarly, if the mappings F_i are strictly monotone, then Q is a $P(M)$-mapping.

3. PROPERTIES OF GAP FUNCTIONS

Let us consider the function

$$\varphi_\alpha(x) = \max_{y \in K} \sum_{s \in M} \Phi_s^\alpha(x, y_s) = \sum_{s \in M} \max_{y_s \in K_s} \Phi_s^\alpha(x, y_s) \tag{5}$$

where

$$\Phi_s^\alpha(x, y_s) = \langle Q_s(x), x_s - y_s \rangle - 0.5\alpha \|x_s - y_s\|^2 + f_s(x_s) - f_s(y_s),$$

$s \in M$, $\alpha > 0$. The function $\Phi_s^\alpha(x, \cdot)$ is strongly concave, hence, there exist unique solutions to each inner problem in (5), i.e., there exist elements $y_s^\alpha(x) \in K_s$ such that

$$\max_{y_s \in K_s} \Phi_s^\alpha(x, y_s) = \Phi_s^\alpha(x, y_s^\alpha(x)), \quad s \in M.$$

Set $y^\alpha(x) = (y_1^\alpha(x), ..., y_m^\alpha(x))^T$. It is clear that $y^\alpha(x)$ solves the problem

$$\max_{y \in K} \sum_{s \in M} (\langle Q_s(x), x_s - y_s \rangle - 0.5\alpha \|x_s - y_s\|^2 + f_s(x_s) - f_s(y_s)) \tag{6}$$

for any fixed x. Taking Proposition 1 as a basis, we obtain the following equivalence result.

Lemma 1. *The following relations are equivalent*:

(i) $\overline{y} = y^\alpha(x)$;

(ii) $\overline{y} \in K$ and

$\forall s \in M,\ z_s \in K_s\ \exists \overline{g}_s \in \partial f_s(\overline{y}_s)$ such that

$$\sum_{s \in M} (\langle Q_s(x) + \alpha(\overline{y}_s - x_s), z_s - \overline{y}_s \rangle + \langle \overline{g}_s, z_s - \overline{y}_s \rangle) \geq 0 \tag{7}$$

(iii) $\overline{y} \in K$ *and*

$$\sum_{s \in M} (\langle Q_s(x) + \alpha(\overline{y}_s - x_s), z_s - \overline{y}_s \rangle + \\ + f_s(z_s) - f_s(\overline{y}_s)) \geq 0 \quad \forall z_s \in K_s, s \in M; \tag{8}$$

(iv) $\overline{y} \in K$ *and*

$$\langle Q_s(x) + \alpha(\overline{y}_s - x_s), z_s - \overline{y}_s \rangle + f_s(z_s) - f_s(\overline{y}_s) \geq 0 \quad \forall z_s \in K_s, s \in M. \tag{9}$$

Proof. Obviously, (7) represents the necessary and sufficient condition of optimality for problem (6), e.g. see [7, Theorem 27.4]. Therefore, (i) is equivalent to (ii). The equivalence of (ii), (iii) and (iv) follows from Proposition 1. □

We will show that φ_α is a gap function for the initial MVI (2).

Proposition 2. *It holds that*

(i) $\varphi_\alpha(x) \geq 0$ *for all* $x \in K$ *;*

(ii) the following assertions are equivalent:

(a) $\varphi_\alpha(x^*) = 0$ *and* $x^* \in K$ *,*

(b) x^* *solves (1),*

(c) $x^* = y^\alpha(x^*)$ *.*

Proof. Assertion (i) follows directly from the definition. In case (ii), if x^* solves (2), then

$$\sum_{s \in M} (\langle Q_s(x^*), x_s^* - y_s \rangle - 0.5\alpha \|x_s^* - y_s\|^2 + f_s(x_s^*) - f_s(y_s)) \leq \\ \leq \sum_{s \in M} (\langle Q_s(x^*), x_s^* - y_s \rangle + f_s(x_s^*) - f_s(y_s)) \leq 0$$

for all $y_s \in K_s, s \in M$, hence $\varphi_\alpha(x^*) = 0$ due to (i). Next, using (8) with $z = x = x^*$ and $\overline{y} = y^\alpha(x^*)$ gives

$$\varphi_\alpha(x^*) \geq \sum_{s\in M}\left(\left\langle Q_s(x^*), x_s^* - y_s^\alpha(x^*)\right\rangle - 0.5\alpha\left\|x^* - y_s^\alpha(x^*)\right\|^2 + f_s(x_s^*) - f_s(y_s^\alpha(x^*))\right) \geq$$

$$\geq 0.5\alpha\left\|x^* - y^\alpha(x^*)\right\|^2.$$

Hence, $\varphi_\alpha(x^*) = 0$ now implies $x^* = y^\alpha(x^*)$. Next, if $x^* = y^\alpha(x^*)$, then (8) yields (1) and the proof is complete. □

Proposition 3. *The mapping* $x \mapsto y^\alpha(x)$ *is continuous.*

Proof. Take arbitrary points $x', x'' \in R^n$ and set $y' = y^\alpha(x')$ and $y'' = y^\alpha(x'')$. Adding (8) with $x = x'$, $\overline{y} = y'$, $z = y''$ and (8) with $x = x''$, $\overline{y} = y''$, $z = y'$ gives

$$\sum_{s\in M}\left(\left\langle Q_s(x') - Q_s(x''), y_s'' - y_s'\right\rangle + \alpha\left\langle y_s' - y_s'', y_s'' - y_s'\right\rangle - \alpha\left\langle x_s' - x_s'', y_s'' - y_s'\right\rangle\right) \geq 0.$$

It follows that

$$\alpha\left\|y'' - y'\right\|^2 \leq \left\|Q(x') - Q(x'')\right\|\left\|y'' - y'\right\| + \alpha\left\|x'' - x'\right\|\left\|y'' - y'\right\|,$$

or equivalently,

$$\left\|y'' - y'\right\| \leq \alpha^{-1}\left\|Q(x') - Q(x'')\right\| + \left\|x'' - x'\right\|.$$

Since Q is continuous, this inequality implies that $x \mapsto y^\alpha(x)$ is continuous, as desired. □

So, under the blanket assumptions, MVI (2) reduces to the problem of minimizing the function φ_α over K, i.e., it is equivalent to the optimization problem

$$\min_{x\in K} \varphi_\alpha(x). \tag{10}$$

4. GENERAL EXISTENCE AND UNIQUENESS RESULTS

In this section, we establish existence and uniqueness results for MVI (2). We first consider the case where Q possesses $P(M)$ type properties.

Proposition 4. *Suppose K is a bounded set. Then MVI (2) has a solution.*

Proof. It was shown in Proposition 3 that the mapping $x \mapsto y^{\alpha}(x)$ is continuous. Applying Brouwer's fixed point theorem, we conclude that there exists $x^* = y^{\alpha}(x^*)$. Using Proposition 2, we deduce that x^* is a solution to MVI (2). □

Proposition 5. *Let Q be a $P(M)$-mapping. Then MVI (2) has at most one solution.*

Proof. Suppose for contradiction that there exist x' and x'', $x' \neq x''$, which are solutions to MVI (2). By Proposition 1, for each $s \in M$ we have

$$\langle Q_s(x'), x_s'' - x_s' \rangle + f_s(x_s'') - f_s(x_s') \geq 0$$

and

$$\langle Q_s(x''), x_s' - x_s'' \rangle + f_s(x_s') - f_s(x_s'')) \geq 0.$$

Adding these inequalities yields

$$\langle Q_s(x') - Q_s(x''), x_s'' - x_s' \rangle \geq 0,$$

which is a contradiction.

Combining both the propositions yields the following result.

Corollary 1. *Let Q be a $P(M)$-mapping and let K be a bounded set. Then MVI (2) has a unique solution.*

Now we present an existence and uniqueness result for the unbounded case.

Theorem 1. *Let Q be a strict P(M)-mapping. Then MVI (2) has a unique solution.*

Proof. Due to Proposition 5, it suffices to show that MVI (2) is solvable. Clearly, if K is bounded, then MVI (2) is solvable due to Proposition 4. Therefore, we have to consider the unbounded case. Fix a point $z=(z_s \mid s \in M)$. For a number $r>0$ we set

$$B_s(z_s,r)=\{x_s \in R_s^n \mid \|x_s - z_s\| \le r\}$$

for each $s \in M$. Let x^r denote a unique solution of the problem (2) over the set

$$K_r=\{x \in R^n \mid x_s \in K_s \bigcap B_s(z_s,r), s \in M\}.$$

By Proposition 1, we have

$$\exists g_s^r \in \partial f_s(x_s^r): \left\langle Q_s(x^r), y_s - x_s^r \right\rangle + \left\langle g_s^r, y_s - x_s^r \right\rangle \ge 0 \tag{11}$$

for all $y_s \in K_s \bigcap B_s(z_s,r)$, $s \in M$.

We now proceed to show that $\|x_s - z_s\| < r$, $s \in M$ for $r>0$ large enough. Assume for contradiction that $\|x^r - z\| \to \infty$ as $r \to \infty$. Choose an arbitrary sequence $\{r_k\} \to \infty$ and set $y^k = x^{r_k}$. Choose the index set $J=\{s \mid \|y_s^k\| \to \infty \text{ as } k \to \infty\}$. Letting

$$\tilde{z}_s^k = \begin{cases} y_s^k & \text{if } s \notin J, \\ z_s & \text{if } s \in J; \end{cases}$$

we have

$$\begin{aligned} &\left\langle y_{s_k}^k - \tilde{z}_{s_k}^k, Q_{s_k}(y^k) - Q_{s_k}(\tilde{z}^k) - \gamma(y_{s_k}^k - \tilde{z}_{s_k}^k) \right\rangle = \\ &= \max_{s \in M} \left\langle y_s^k - \tilde{z}_s^k, Q_s(y^k) - Q_s(\tilde{z}^k) - \gamma(y_s^k - \tilde{z}_s^k) \right\rangle > 0. \end{aligned} \tag{12}$$

Since the set M={ 1, ..., m } is finite, without loss of generality we can supposes that s_k is fixed, i.e. $s_k = l$. Note that $l \in J$ due to (12). Taking into account the monotonicity of ∂f_l and the assumptions of Theorem 1, we have

$$\left\langle y_l^k - z_l, Q_l(y^k) - Q_l(\tilde{z}^k) \right\rangle + \left\langle y_l^k - z_l, g_l^k - g_l \right\rangle \ge \gamma \|y_l^k - z_l\|^2$$

for all $g_l^k \in \partial f_l(y^k)$ and $g_l \in \partial f_l(z_l)$, or equivalently,

$$\left\langle Q_l(y^k), y_l^k - z_l \right\rangle + \left\langle g_l^k, y_l^k - z_l \right\rangle \geq \gamma \left\| y_l^k - z_l \right\|^2 - \left\langle z_l - y_l^k, Q_l(\tilde{z}^k) + g_l \right\rangle.$$

Since $\{\tilde{z}^k\}$ is bounded, we must have $\left\| Q_l(\tilde{z}^k) \right\| \leq C$ and $\left\| g_l \right\| < C$. Hence, it holds that $\left\| y_l^k - z_l \right\| \to \infty$ and

$$\gamma \left\| y_l^k - z_l \right\|^2 - \left| \left\langle -y_l^k, Q_l(\tilde{z}^k) + g_l \right\rangle \right| \to +\infty,$$

hence that

$$\left\langle Q_l(y^k), z_l - y_l^k \right\rangle + \left\langle g_l^k, z_l - y_l^k \right\rangle < 0$$

for k large enough, which contradicts (12). Thus, there exists a number k' such that $\left\| y_s^k - z_s \right\| < r_k$, $s \in M$; if $k \geq k'$. It follows that, for any $x_s \in K_s$, there is $\varepsilon > 0$ such that

$$x_s^r + \varepsilon(x_s - x_s^r) \in K_s \bigcap B_s(z_s, r), \quad s \in M; \text{ if } r \geq r_{k'}.$$

Applying now (12) with $y_s = x_s^r + \varepsilon(x_s - x_s^r)$ gives

$$\exists g_s^r \in \partial f_s(x^r) : \left\langle Q_s(x^r), x_s^r + \varepsilon(x_s - x_s^r) - x_s^r \right\rangle + \left\langle g_s^r, x_s^r + \varepsilon(x_s - x_s^r) - x_s^r \right\rangle \geq 0,$$

or equivalently,

$$\exists g_s^r \in \partial f_s(x^r) : \left\langle Q_s(x^r), x_s - x_s^r \right\rangle + \left\langle g_s^r, x_s - x_s^r \right\rangle \geq 0$$

$s \in M$. Due to Proposition 1, it means that x^* is a solution to MVI (2). The proof is complete. □

So, joint existence and uniqueness results require for Q to be at least $P(M)$. Now we intend to obtain similar results under weaker assumptions on Q. We give an additional relationship between $P_0(M)$ and strict $P(M)$-mappings.

Proposition 6. *If $F : K \to R^n$ is a $P_0(M)$-mapping, then, for any $\varepsilon > 0$, $F + \varepsilon I_n$ is a strict $P(M)$-mapping.*

Proof. First we show that $F^{(\varepsilon)} = F + \varepsilon I_n$ is a $P(M)$-mapping for each $\varepsilon > 0$. Choose $x', x'' \in U$, $x' \neq x''$, set $I = \{s \mid x'_s \neq x''_s\}$ and fix $\varepsilon > 0$. Since F is a $P_0(M)$-mapping, there exists an index $k \in I$ such that

$$\langle F_k(x') - F_k(x''), x'_k - x''_s \rangle = \max_{s \in M} \langle F_s(x') - F_s(x''), x'_s - x''_s \rangle.$$

Then, by definition,

$$\langle F_k(x') - F_k(x''), x'_k - x''_k \rangle \geq 0, \; x'_k \neq x''_k,$$

and

$$\varepsilon \langle x'_k - x''_k, x'_k - x''_k \rangle > 0.$$

Adding these inequalities yields

$$\langle F_k^{(\varepsilon)}(x') - F_k^{(\varepsilon)}(x''), x'_k - x''_k \rangle > 0.$$

Hence, $F^{(\varepsilon)}$ is a $P(M)$-mapping. Since $F^{(\varepsilon'')} = F^{(\varepsilon')} - (\varepsilon' - \varepsilon'')I_n =$ $= F + \varepsilon'' I_n$ is a $P(M)$-mapping, if $0 < \varepsilon'' < \varepsilon'$, we conclude that $F^{(\varepsilon)}$ is a strict $P(M)$-mapping. □

Thus, if (A1)–(A3) hold and is a $P_0(M)$-mapping, we can consider the regularized problem with the cost mapping $Q^{(\varepsilon)} = Q + \varepsilon I_n$. From Proposition 6 it follows that $Q^{(\varepsilon)}$ is a strict $P(M)$-mapping for each $\varepsilon > 0$, hence, using Theorem 1, we conclude that each regularized MVI, which approximates the initial MVI (2), will have a unique solution.

At the same time, replacing the (strict) $P(M)$ property of Q with (strong) strict convexity of f_s, we can obtain similar results in the case where Q is a $P_0(M)$-mapping. It also means that we can apply the same regularization approach to the functions f_s in the general case.

Theorem 2. *Let Q be a $P_0(M)$-mapping and let f_s be strictly convex for $s \in M$. Then MVI (2) has at most one solution.*

Proof. Suppose for contradiction that there exist x' and x'', $x' \neq x''$, which are solutions to MVI(2). By Proposition 1, we have

$$\exists g'_s \in \partial f_s(x'_s) : \langle Q_s(x'), x''_s - x'_s \rangle + \langle g'_s, x''_s - x'_s \rangle \geq 0$$

and

$$\exists g_s'' \in \partial f_s(x_s''): \langle Q_s(x''), x_s' - x_s'' \rangle + \langle g_s'', x_s' - x_s'' \rangle \geq 0$$

$s \in M$. Adding these inequalities yields

$$\langle Q_s(x') - Q_s(x''), x_s'' - x_s' \rangle + \langle g_s' - g_s'', x_s'' - x_s' \rangle \geq 0, \tag{13}$$

$s \in M$. For brevity, set $I = \{s \mid x_s' \neq x_s''\}$. Since Q is a $P_0(M)$-mapping, there exists an index $k \in I$ such that

$$\langle Q_k(x') - Q_k(x''), x_k' - x_k'' \rangle = \max_{s \in M} \langle Q_s(x') - Q_s(x''), x_s' - x_s'' \rangle.$$

Then, by definition, $\langle Q_k(x') - Q_k(x''), x_k' - x_k'' \rangle \geq 0$. Due to (13) we now obtain

$$\langle g_k' - g_k'', x_k'' - x_k' \rangle \geq 0,$$

which is a contradiction, since f_k is strictly convex, i.e., ∂f_k is strictly monotone. □

Again, combining Theorem 2 and Proposition 4 yields the following result immediately.

Corollary 2. *In addition to the assumptions of Theorem 2, suppose K is a bounded set. Then MVI (2) has a unique solution.*

We now present an existence and uniqueness result on unbounded sets under the $P_0(M)$ condition. This result can be viewed as a counterpart of that in Theorem 1.

Theorem 3. *Let Q be a $P_0(M)$-mapping and let f_s be a strongly convex function for each $s \in M$. Then MVI (2) has a unique solution.*

Proof. By Proposition 1, the initial problem is equivalent to the following VI: Find $x^* \in K$ such that

$$\exists g_s^* \in \partial f_s(x_s^*): \langle Q_s(x^*) + \varepsilon x_s^*, x_s - x_s^* \rangle +$$
$$+ \langle g_s^* - \varepsilon x_s^*, x_s - x_s^* \rangle \geq 0 \quad \forall x_s \in K_s, s \in M; \tag{14}$$

which can be rewritten equivalently as

$$\exists t_s^* \in \partial\psi_s(x_s^*) : \left\langle F_s^{(\varepsilon)}(x^*), x_s - x_s^* \right\rangle + \left\langle t_s^*, x_s - x_s^* \right\rangle \geq 0 \quad \forall x_s \in K_s,\ s \in M; \ (15)$$

where $F_s^{(\varepsilon)}(x) = Q_s(x) + \varepsilon x_s$ and $\psi_s(\sigma) = f_s(\sigma) - \varepsilon\sigma^2/2$. Again, on account of Proposition 1, problem (15) is equivalent to the MVI: Find $x^* \in K$ such that

$$\sum_{s \in M} (\left\langle F_s^{(\varepsilon)} x^*), x_s - x_s^* \right\rangle + \psi_s(x_s) - \psi_s(x_s^*)) \geq 0 \quad \forall x_s \in K_s,\ s \in M. \qquad (16)$$

From Proposition 6 it follows that $F^{(\varepsilon)}$ is a strict $P(M)$-mapping for every $\varepsilon > 0$. We will show that each ψ_s is a convex function for some $\varepsilon > 0$. Since f_s is strongly convex, we see that for all x_s', x_s'' and $t_s' \in \partial\psi_s(x_s')$, $t_s'' \in \partial\psi_s(x_s'')$, we have

$$\left\langle t_s' - t_s'', x_s' - x_s'' \right\rangle = \left\langle g_s' - g_s'', x_s' - x_s'' \right\rangle - \varepsilon \left\langle x'_s - x''_s, x'_s - x''_s \right\rangle$$

for some $g_s' \in \partial f_s(x_s')$, $g_s'' \in \partial f_s(x_s'')$, hence

$$\left\langle t_s' - t_s'', x_s' - x_s'' \right\rangle \geq \tau(x_s' - x_s'')^2 - \varepsilon(x_s' - x_s'')^2 \geq 0$$

if $\varepsilon < \tau$, where τ is the smallest constant of strong monotonicity of ∂f_k (strong convexity of f_k). So, ψ_s is a convex function if $0 < \varepsilon < \tau$.

By Theorem 1, problem (16) has a unique solution. However, problem (16) is equivalent to (15), i.e., it is equivalent to (14). This completes the proof. □

REFERENCES

[1] I.V. Konnov, *Properties of gap functions for mixed variational inequalities*, Siberian Journal of Numerical Mathematics, **3** (2000), 259–270.

[2] I.V. Konnov and E.O. Volotskaya, *Mixed variational inequalities and ecomonic equilibrium problems*, Journal of Applied Mathematics, **6** (2002), 289–314.

[3] J.J. Moré, *Classes of functions and feasibility conditions in nonlinear complementarity problems*, Mathematical Programming, **6** (1974), 327 – 338.

[4] C. Kanzow and M. Fukushima, *Theoretical and numerical investigation of the D-gap function for box constrained variational inequalities*, Mathematical Programming, **83** (1998), 55 – 87.

[5] F. Facchinei and C. Kanzow, *Beyond monotonicity in regularization methods for nonlinear complementarity problems*, SIAM Journal on Control and Optimization, **37** (1999), 1150 – 1161.

[6] F. Facchinei and J.S. PANG, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag, Berlin, 2003 (two volumes).
[7] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, (1970).

IRREDUCIBILITY OF THE TRANSITION SEMIGROUP ASSOCIATED WITH THE TWO PHASE STEFAN PROBLEM

Viorel Barbu[1]* and Giuseppe Da Prato[2]**
University of Iasi, Iasi, Romania[1]; Scuola Normale Superiore, Pisa, Italy[2]

Abstract: We prove that the transition semigroup associated with the two phase Stefan problem is irreducible. The proof relies on a general result of approximate controllability for maximal monotone systems, see [1].

Mathematics Subject Classification AMS : 35K35, 35R15, 60H15.

Key words: Stochastic Stefan problem, invariant measures, transition semigroup, irreducibility.

1. INTRODUCTION

Let $\mathcal{O}$ be a bounded open domain of $\mathbb{R}^n$, $n \in \mathbb{N}$, with a smooth boundary $\partial\mathcal{O}$.

We are concerned with the following stochastic differential equation,

* This was done during the stay in Scuola Normale Superiore di Pisa.
** Partially supported by the Italian National Project MURST "Equazioni di Kolmogorov."

$$\begin{cases} dX - \Delta\beta(X)dt = \sqrt{Q}\, dW_t \quad \text{in } (0,+\infty)\times\mathcal{O}, \\ \beta(X) = 0 \quad \text{on } (0,+\infty)\times\partial\mathcal{O}, \\ X(0,\xi) = x(\xi) \quad \text{in } \mathcal{O}, \end{cases} \tag{1.1}$$

where $\sqrt{Q}\, dW_t$ is a coloured noise of covariance Q with $Q \in L(L^2(\mathcal{O}))$ symmetric, nonnegative and of trace class, defined in some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $L^2(\mathcal{O})$. We shall assume further that

$$Q \in L(H^{-1}(\mathcal{O}), L^2(\mathcal{O})). \tag{1.2}$$

A typical example of operator Q is $Q = (-A_0)^{-\sigma}$ with $\sigma > n/2$, so that Tr $Q < +\infty$, where $A_0 = \Delta, D(A_0) = H^2(\mathcal{O}) \cap H_0^1(\mathcal{O}))$.

Finally, $\beta : \mathbb{R} \to \mathbb{R}$ is a continuous nondecreasing function, such that

$$\beta(0) = 0, \quad \lim_{r\to\pm\infty} \beta(r) = \pm\infty \tag{1.3}$$

and there exists $\delta > 0$ such that

$$(\beta(r) - \beta(r_1))(r - r_1) \geq \delta(\beta(r) - \beta(r_1)^2, \quad r, r_1 \in \mathbb{R}. \tag{1.4}$$

A typical example is

$$\beta(r) = \begin{cases} \alpha_1 r \quad \text{for } r \leq 0, \\ 0 \quad \text{for } 0 < r \leq \rho, \\ \alpha_2(r - \rho) \quad \text{for } r > 0, \end{cases} \tag{1.5}$$

(where α_1, α_2, ρ are given positive numbers) which reduces problem (1.1) to the two phase Stefan problem studied in [2].

Under the above assumptions one proves in [2] that for any $x \in H = H^{-1}(\mathcal{O})$ problem (1.1) has a unique generalized solution

$$X(\cdot, x) \in C_w([0,T]; L^2(\Omega, H^{-1}(\mathcal{O})).$$

Moreover, the corresponding transition semigroup

$$P_t\varphi(x) = \mathbb{E}[\varphi(X(t,x))], \quad t \geq 0, x \in K,$$

defined for every φ bounded and Borel, has an invariant measure ν with the support in

$$\{x \in L^2(\mathcal{O}) : \beta(X) \in H_0^1(\mathcal{O})\}.$$

We recall that

$$\int_H P_t\varphi(x)\nu(dx) = \int_H \varphi(x)\nu(dx), \quad \varphi \in C_b(H). \tag{1.6}$$

Here we shall study the *irreducibility* of P_t. We recall that P_t is said to be irreducible if for all $T > 0, r > 0, x_0, x_1 \in H$ one has

$$P_T\chi_{B(x_1,r)}(x_0) > 0,$$

where $B(x_1,r) = \{x \in H : |x - x_1| < r\}$. Equivalently,

$$\mathbb{P}\big(|X(T,x_0) - x_1|_{-1} \geq r\big) < 1 \quad \text{for all } T > 0, r > 0, x_0, x_1 \in H. \tag{1.7}$$

Irreducibility is an important property for the transition semigroup P_t because it implies that the measure ν is *full*. In fact, from (1.6) it follows, for any $x \in H$, $r, t > 0$, that

$$\nu(B(x,r)) = \int_H P_t\chi_{B(x_1,r)}(x)\nu(dx) > 0.$$

Our main result is the following.

Theorem 1.1 *Under assumptions* (1.2), (1.3), (1.4) *the transition semigroup* P_t *is irreducible.*

Theorem 1.1 will be proven in Sect.3. The main ingredient is an approximate controllability result for the deterministic equation

$$\begin{cases} \dfrac{dy}{dt} - \Delta\beta(y) = \sqrt{Q}\,u \quad \text{in } (0,+\infty)\times\mathcal{O}, \\ \\ \beta(y) \in H_0^1(\mathcal{O}), \\ \\ y(0,\xi) = x_0(\xi) \quad \text{in } \mathcal{O}, \end{cases} \tag{1.8}$$

which will be proved in a general framework in Sect.2.

We shall use the following notations.

$H = H^{-1}(\mathcal{O})$ with the norm $|\cdot|_{-1}$ and scalar product $(\cdot,\cdot)_{-1}$. $H_0^1(\mathcal{O})$ and $H^2(\mathcal{O})$ are standard Sobolev spaces.

$B_b(H)$ is the Banach space of all real bounded mappings in H endowed with the sup norm

$$\|\varphi\|_0 = \sup_{x\in H} |\varphi(x)|$$

and $C_b(H)$ the closed subspace of all uniformly continuous and bounded mappings.

Moreover, we set

$$A_0 = \Delta, D(A_0) = H^2(\mathcal{O}) \cap H_0^1(\mathcal{O}),$$

and

$$\begin{cases} Ay = -\Delta y, \quad y \in D(A) \cap H, \\ \\ D(A) = \{y \in L^2(\mathcal{O}) : \beta(y) \in H_0^1(\mathcal{O})\}. \end{cases} \tag{1.9}$$

We recall that A is m–accretive (maximal monotone) in $H\times H$ (see e.g. [1], [3]).

Finally, $C_W([0,T];L^2(\Omega,H^{-1}(\mathcal{O}))$ is the space of all stochastic processes which are square mean continuous and adapted to W.

2. APPROXIMATE CONTROLLABILITY

Let A be a nonlinear, multivalued operator on a Hilbert space H. Let U be another Hilbert space and $B: U \to H$ a linear operator. We denote by $|\cdot|$ and $(\cdot,\cdot)$ the norm and the scalar product of H and U.

We shall assume that

Hypothesis 2.1

(i) $A: D(A) \subset H \to H$ *is quasi–m–accretive, i.e. there exists* $\gamma > 0$ *such that* $A + \gamma I$ *is* m*–accretive (equivalently, maximal monotone) in* $H \times H$.

(ii) $B \in L(U,H)$ *and Ker* $B^* = \{0\}$, *where* B^* *is the adjoint of* B.

We are concerned with the controlled equation

$$\begin{cases} y'(t) + A(y(t)) \ni Bu(t), \quad t \in [0,T] \\ \\ y(0) = y_0 \in H. \end{cases} \tag{2.1}$$

It is well known (see e.g. [1]) that for each $y_0 \in \overline{D(A)}$ and any $u \in L^2(0,T;U)$ the Cauchy problem (2.1) has a unique "mild" solution $y = y^u(t, x_0) \in C([0,T];H)$. If $A = \partial\varphi$ is the subdifferential of a lower semicontinuous, convex function $\varphi: H \to (-\infty, +\infty]$ then (see [3], [1]):

$$y^u \in W^{1,2}([0,T];H) \text{ and } Ay^u \in L^2([0,T];H).$$

Proposition 2.2 below amounts to say that under above assumptions, system (2.1) is approximately controllable.

Proposition 2.2 *Let* $y_0, y_1 \in \overline{D(A)}$. *Then* $\forall \varepsilon > 0$, $\exists u \in C([0,T];U)$, *such that*

$$|y^u(T) - y_1| \leq \varepsilon. \tag{2.2}$$

We shall prove Proposition 2.2 in two steps.

Lemma 2.3 *Let* $y_0, y_1 \in D(A)$. *Then* $\exists v \in L^\infty(0,T;H)$, *such that the mild solution* $z \in C([0,T];H)$ *to problem*

$$\begin{cases} z'(t) + A(z(t)) \ni v(t), \quad \text{a.e. } t \in [0,T] \\ \\ z(0) = y_0 \in H, \end{cases} \tag{2.3}$$

satisfies also

$$z \in W^{1,\infty}([0,T];H) \tag{2.4}$$

$$z(T) = y_1. \tag{2.5}$$

Proof. Consider the nonlinear mapping

$$F(z) = \rho \operatorname{sgn}(z - y_1),$$

where sgn is the multivalud mapping

$$\operatorname{sgn} z = \begin{cases} \dfrac{z}{|z|}, & \text{if } z \neq 0, \\ \\ \{z : |z| \leq 1\}, & \text{if } z = 0. \end{cases}$$

It is not difficult to show that the operator $A + F$ is quasi–m–accretive as well and consequently the Cauchy problem

$$\begin{cases} z'(t) + A(z(t)) + \rho \operatorname{sgn}(z(t) - y_1) \ni 0, \quad \text{a.e. } t \in [0,T]. \\ \\ z(0) = y_0 \in H, \end{cases} \tag{2.6}$$

has a unique strong solution $z \in W^{1,\infty}([0,T];H)$. (As a matter of fact, z is right differentiable and $(Az(t) + v(t))^0$ (the minimal section of $Az(t) + v(t)$) is continuous from the right on $[0,T)$.

Now, multiplying the first equation in (2.6) by $z(t) - y_1$ and integrating over $(0,t)$, we get

$$\frac{1}{2}\frac{d}{dt}|z(t)-y_1|^2+(u(t),z(t)-y_1)+\rho|z(t)-y_1|=0,$$

where $u(t)\in A(z(t))$. But, by the quasi accretivity of A, we have:

$$\begin{aligned}(u(t),z(t)-y_1) &= (u(t)-A^0y_1,z(t)-y_1)+(A^0y_1,z(t)-y_1)\\ &\geq -\gamma|z(t)-y_1|^2+(A^0y_1,z(t)-y_1),\end{aligned}$$

so that

$$\frac{1}{2}\frac{d}{dt}|z(t)-y_1|^2+\rho|z(t)-y_1|\leq\gamma|z(t)-y_1|^2+|A^0(y_1)||z(t)-y_1|.$$

This yields

$$\frac{d}{dt}\left(e^{-\gamma t}|z(t)-y_1|\right)+(\rho-|A^0(y_1)|)e^{-\gamma t}\leq 0$$

i.e.

$$|z(t)-y_1|e^{-\gamma t}+\frac{1}{\gamma}(\rho-|A^0(y_1)|)(1-e^{-\gamma t})\leq|y_0-y_1|,\quad t\geq 0. \tag{2.7}$$

for $\rho>|A_0(y_1)|$. By (2.7) we see that $|z(t)-y_1|=0$ for

$$t\geq T_0=-\frac{1}{\gamma}\log\left(1-\frac{\gamma|y_0-y_1|}{\rho-|A^0(y_1)|}\right).$$

For ρ sufficiently large T_0 can be taken just equal to T apriori fixed. (We note that if $\gamma=1$ then $T_0=\frac{|y_0-y_1|}{\rho-|A^0(y_1)|}$).

Then $v(t)=-\rho\,\mathrm{sign}\,(z(t)-y_1)$ (or more precisely its single valued section) from equation (2.6) is the desired controller.

We note that $v\in L^\infty(0,T;H)$. Indeed $|v(t)|\leq\rho$ for all $t\in[0,T]$, $y'(t)\in-Ay(t)+v(t)$ by maximal monotonicity of $A+\gamma I$ and it follows that v is measurable.

Next by continuity of the map $y_0\to z^v(t,y_0)$ the exact controllability results extends to all $y_0\in\overline{D(A)}$. Finally, by density it extends to all $y_1\in\overline{D(A)}$ as claimed. □

Lemma 2.4 *The set*

$$\{v = Bu : u \in C([0,T];U)\}$$

is dense in $L^2(0,T;H)$.

Proof. It is immediate that $\{v = Bu : u \in L^2(0,T;U)\}$ is dense in $L^2(0,T;H)$ because otherwise there exists $\eta \in L^2(0,T;H)$, not identically zero, such that

$$\int_0^T (Bu(t),\eta(t))dt = 0, \quad u \in L^2(0,T;U).$$

This yields $\int_0^T |B^*\eta|^2\, dt = 0$. Since Ker $B^* = \{0\}$ we infer that $\eta = 0$. Since $C([0,T];U)$ is dense in $L^2(0,T;U)$ the desired result follows by the continuous dependence of y^u with respect to u. □

Proof of Proposition 2.2. Nothing remains to do except to combine Lemma 2.3 with Lemma 2.4. By Lemma 2.3 there is $v \in L^2(0,T;H)$, such that $y(T) = y_1$ and

$$\begin{cases} y'(t) + A(y(t)) \ni v(t), t \in (0,T) \\ \\ \qquad y(0) = y_0 \in H. \end{cases} \tag{2.8}$$

On the other hand, by Lemma 2.4 for each $\varepsilon > 0$ there is $u_\varepsilon \in C([0,T];U)$ such that

$$\|Bu_\varepsilon - v\|_{L^2(0,T;H)} \le \varepsilon.$$

Let $y_\varepsilon \in C([0,T];H)$ be the solution to

$$\begin{cases} y_\varepsilon'(t) + A(y_\varepsilon(t)) \ni Bu_\varepsilon(t), \quad t \in [0,T], \\ \\ \qquad y(0) = y_0 \in H. \end{cases} \tag{2.9}$$

Subtracting (2.8) and (2.9) we get from monotonicity of A that

$$|y_\varepsilon(t)-y(t)|\le\int_0^t|(Bu_\varepsilon(s)-v(s))|\,ds\le\varepsilon T^{1/2},\quad t\in[0,T].$$

Hence $|y_\varepsilon(T)-y_1|\le\varepsilon T^{1/2}$ as claimed. This completes the proof. ∎

3. PROOF OF THEOREM 1.1

We shall consider the deterministic equation (1.8) and apply Proposition 2.2 where $H=H^{-1}(\mathcal{O})$, $U=H^{-1}(\mathcal{O})$, $B=\sqrt{Q}$ and $A:D(A)\subset H\to H$ is defined by (1.9).

Consequently, for each $\varepsilon>0$, $T>0$ and all $x_0,x_1\in H^{-1}(\mathcal{O})$ there is $u\in C([0,T];H^{-1}(\mathcal{O}))$, such that

$$|y(T)-x_1|_{-1}<\varepsilon. \tag{3.1}$$

We note that by assumption (1.2) we have

$$\sqrt{Q}\,u\in L^2(0,T;L^2(\mathcal{O}))\cap C([0,T];H^{-1}(\mathcal{O})).$$

Next substracting equations (1.1) and (1.9) we get

$$X(t,x_0)-y(t,x_0)-\int_0^t\Delta(\beta(X(s,x_0))-\beta(y(s,x_0)))ds=\sqrt{Q}\,(W(t)-u(t)). \tag{3.2}$$

We set

$$Z(t)=\int_0^t(\beta(X(s,x_0))-\beta(y(s,x_0)))ds$$

so that

$$X(t,x_0)-y(t,x_0)-\Delta Z(t)=\sqrt{Q}\,(W(t)-u(t)). \tag{3.3}$$

Multiplying both sides by $Z'(t)=\beta(X(t,x_0))-\beta(y(t,x_0))$ and integrating in t and ξ yields

$$\int_0^t (X(s,x_0)-y(s,x_0),\beta(X(s,x_0))-\beta(y(s,x_0)))ds-\int_0^t (\Delta Z(s),Z'(s))_{L^2(\mathcal{O})}ds$$

$$=\int_0^t (\sqrt{Q}\,(W(s)-u(s)),\beta(X(s,x_0))-\beta(y(s,x_0)))_{L^2(\mathcal{O})}ds.$$

Since

$$-\int_0^t (\Delta Z(s),Z'(s))_{L^2(\mathcal{O})}ds=\left\|\frac{1}{2}Z(t)\right\|^2_{H_0^1(\mathcal{O})},$$

we get

$$\int_0^t (X(s,x_0)-y(s,x_0),\beta(X(s,x_0))-\beta(y(s,x_0)))_{L^2(\mathcal{O})}ds+\frac{1}{2}\|Z(t)\|^2_{H_0^1(\mathcal{O})}$$

$$=\int_0^t (\sqrt{Q}\,(W(s)-u(s)),\beta(X(s,x_0))-\beta(y(s,x_0)))_{H_0^1(\mathcal{O})}ds.$$

Taking into account (1.4) we find

$$\delta\|\beta(X(\cdot,x_0))-\beta(y(\cdot,x_0))\|^2_{L^2(0,t;L^2(\mathcal{O}))}+\frac{1}{2}\|Z(t)\|^2_{H_0^1(\mathcal{O})}$$

$$\le\left\|\sqrt{Q}\,W-\sqrt{Q}\,u\right\|^2_{L^2(0,t;L^2(\mathcal{O}))}\|\beta(X(\cdot,x_0))-\beta(y(\cdot,x_0))\|^2_{L^2(0,t;L^2(\mathcal{O}))},$$

which, by a standard device, yields

$$\|\beta(X(\cdot,x_0))-\beta(y(\cdot,x_0))\|^2_{L^2(0,t;L^2(\mathcal{O}))}+\|Z(t)\|^2_{H_0^1(\mathcal{O})}$$

$$\le C\left\|\sqrt{Q}\,W-\sqrt{Q}\,u\right\|^2_{L^2(0,t;L^2(\mathcal{O}))} \tag{3.4}$$

for a suitable constant C. Now, by (3.3) we have for $t=T$

$$|X(T,x_0)-y(T,x_0)|_{-1}\le\|Z(T)\|_{H_0^1(\mathcal{O})}+\left\|\sqrt{Q}\,W(T)-\sqrt{Q}\,u(T)\right\|_{-1}$$

and, taking into account (3.4) we get for $t = T$

$$|X(T,x_0) - y(T,x_0)|_{-1} \leq \left\|\sqrt{Q}\,W - \sqrt{Q}\,u\right\|^2_{L^2(0,t;L^2(\mathcal{O}))}$$

$$+|\sqrt{Q}\,W(T) - \sqrt{Q}\,u(T)|_{-1}.$$

We choose x_1 as in (3.1) and get therefore

$$|X(T,x_0) - x_1|_{-1} \leq |y(T,x_0) - x_1|_{-1}$$

$$+C_2\left\|\sqrt{Q}\,W - \sqrt{Q}\,u\right\|^2_{L^2(0,T;L^2(\mathcal{O}))} + |\sqrt{Q}\,W(T) - \sqrt{Q}\,u(T)|_{-1}$$

$$\leq \varepsilon + \left\|\sqrt{Q}\,W - \sqrt{Q}\,u\right\|^2_{L^2(0,T;L^2(\mathcal{O}))} + |\sqrt{Q}\,W(T) - \sqrt{Q}\,u(T)|_{-1}.$$

Therefore for any $r > 0$

$$\mathbb{P}\left(|X(T,x_0) - x_1|_{-1} \geq r\right)$$

$$\leq \mathbb{P}\left(\left\|\sqrt{Q}\,W - \sqrt{Q}\,u\right\|^2_{L^2(0,T;L^2(\mathcal{O}))} + |\sqrt{Q}\,W(T) - \sqrt{Q}\,u(T)|_{-1} \geq r - \varepsilon\right).$$

It is clear that the random variable $(\sqrt{Q}\,W(\cdot), \sqrt{Q}\,W(T))$ in

$$\Lambda := L^2(0,T;L^2(\mathcal{O})) \times H^{-1}(\mathcal{O})$$

is Gaussian. We claim that $(\sqrt{Q}\,W(\cdot), \sqrt{Q}\,W(T))$ is nondegenerate. This will imply that $\mathbb{P}\left(|X(T,x_0) - x_1|_{-1} \geq r\right) < 1$ as required.

Let us prove the claim. Denote by $\begin{pmatrix}\varphi\\ x\end{pmatrix}$ the generic element of Λ and by $\mathcal{Q}$ the covariance operator of the Gaussian random variable $(\sqrt{Q}\,W(\cdot), \sqrt{Q}\,W(T))$. Then we have for any $\begin{pmatrix}\varphi\\ x\end{pmatrix} \in \Lambda$

$$\left\langle \mathcal{Q}\begin{pmatrix}\varphi\\ x\end{pmatrix}, \begin{pmatrix}\varphi\\ x\end{pmatrix}\right\rangle = \mathbb{E}\left[\left|\left\langle \begin{pmatrix}\sqrt{Q}\,W(\cdot)\\ \sqrt{Q}\,W(T)\end{pmatrix}, \begin{pmatrix}\varphi\\ x\end{pmatrix}\right\rangle\right|^2\right]$$

$$= \mathbb{E}\left[\left(\int_0^T (\sqrt{Q}\,W(s), \varphi(s))ds + (\sqrt{Q}\,W(T), x)\right)^2\right] \tag{3.5}$$

$$= \mathbb{E}[\int_0^T \int_0^T (\sqrt{Q}\,W(t), \varphi(t))\,(\sqrt{Q}\,W(s), \varphi(s))dtds$$

$$+2(\sqrt{Q}\,W(T), x)\int_0^T (\sqrt{Q}\,W(s), \varphi(s))ds + |(\sqrt{Q}\,W(t), x)|^2].$$

Since

$$\mathbb{E}\left[(\sqrt{Q}\,W(t), W(s))\right] = \min\{t,s\}\,\mathrm{Tr}\,Q, \quad t,s \geq 0,$$

and

$$\mathbb{E}\left[(\sqrt{Q}\,W(T), x)\,(\sqrt{Q}\,W(s), \varphi(s))\right] = \min\{T,s\}(Q\varphi(s), x), \quad t,s \geq 0,$$

we obtain by (3.5), that

$$\left\langle \mathcal{Q}\begin{pmatrix}\varphi\\ x\end{pmatrix}, \begin{pmatrix}\varphi\\ x\end{pmatrix}\right\rangle = \mathrm{Tr}\,Q \int_0^T \int_0^T \min\{t,s\}\,(\varphi(t), \varphi(s))dtds$$
$$+2\int_0^T s(Q\varphi(s), x)ds + T(Qx, x). \tag{3.6}$$

Assume now that there exists $\begin{pmatrix}\varphi\\ x\end{pmatrix} \in \Lambda$ such that $\mathcal{Q}\begin{pmatrix}\varphi\\ x\end{pmatrix} = 0$. We have to show that $\begin{pmatrix}\varphi\\ x\end{pmatrix} = 0$. In fact, by (3.6) we obtain

$$\mathrm{Tr}\,Q \int_0^T \min\{t,s\}\,\varphi(s)ds + tQx = 0 \tag{3.7}$$

and

$$\int_0^T sQ\varphi(s)ds + Qx = 0. \tag{3.8}$$

Eliminating x from (3.6) and (3.7) yields

$$\mathrm{Tr}\, Q \int_0^T \min\{t,s\}\varphi(s)ds = t\int_0^T sQ\varphi(s)ds,$$

that it is equivalent to

$$\mathrm{Tr}\, Q \int_0^t s\varphi(s)ds + t\mathrm{Tr}; Q\int_t^T \varphi(s)ds = t\int_0^T sQ\varphi(s)ds.$$

Differentiating with respect to t yields

$$\mathrm{Tr}\, Q \int_t^T \varphi(s)ds = \int_0^T sQ\varphi(s)ds,$$

which implies $\varphi = 0$ and consequently by (3.8) $x = 0$ since Ker $Q = \{0\}$. The proof is complete. □

REFERENCES

[1] V. Barbu, "Analysis and control of nonlinear infinite dimensional systems", Academic Press, San Diego, 1993.

[2] V. Barbu and G. Da Prato, "The two phase stochastic Stefan problem", Probab. Theory Relat. Fields, 124, 544–560, 2002.

[3] H. Brézis, "Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations", Contributions to Nonlinear Functional Analysis, E. Zarantonello, ed., Academic Press, New York, 1971.

ON SOME BOUNDARY VALUE PROBLEMS FOR FLOWS WITH SHEAR DEPENDENT VISCOSITY

H. Beirão da Veiga
Dept. of Applied Mathematics " U. Dini," University of Pisa, Pisa, Italy

Abstract: This notes concern the Navier-Stokes equations with gradient dependent viscosity and slip (or non-slip) type boundary conditions. Regularity up to the boundary still presents many open problems. In the sequel we present some regularity results for weak solutions to the Ladyzhenskaya model in the half space $\mathbb{R}^n_+$. See Theorems 3.1 and 3.2. Complete proofs of these results are done, and will appear in the forthcoming paper [6].

1. INTRODUCTION

The Navier-Stokes equations with shear dependent viscosity has been studied in the last half century by a great number of researchers. A typical example is the *Ladyzhenskaya model*

$$\begin{cases} \dfrac{\partial u}{\partial t} + u \cdot \nabla u - \nabla \cdot T(u,\pi) = f \\ \nabla \cdot u = 0, \end{cases} \tag{1.1}$$

where T denotes the stress tensor

$$T = -\pi I + \nu_T(u)\mathcal{D}u. \tag{1.2}$$

Here,

$$\begin{aligned} &\mathcal{D}u = \nabla u + \nabla u^T, \\ &\nu_T(u) = \nu_0 + \nu_1 |\mathcal{D}u|^{p-2}, \end{aligned} \tag{1.3}$$

and ν_0, ν_1 are strictly positive constants.

Note that (1.2) satisfies the Stokes Principle, see [36]. See also the reference [31] page 231.

For $p = n = 3$, the system (1.1) is the classical Smagorinsky turbulence model, see [34]. See also [14] and references therein.

From the mathematical viewpoint, the crucial characteristic of models like (1.3) is the growth of the *convex* potential $|\mathcal{D}u|^p$ near infinity (and, to a minor extent, near zero). This leads us to show the main points by considering the classical, and more representative case (1.3), rather than risk hiding ideas and methods in a more general setting.

The first mathematical studies on the above kind of equations go back to O.A. Ladyzenskaya in a series of remarkable contributions. See [17], [18], [19] and [20]. Similar results were obtained by J.-L. Lions for models in which $\nabla u + \nabla u^T$ is essentially replaced by ∇u. See [23] and [24], Chap.2, n.5.

Other fundamental existence, uniqueness and regularity results for Ladyzhenskaya type models, under the non-slip boundary condition (1.5), can be found in [25] and references therein. Without any claim of completeness, we also refer to [1], [9], [10], [21], [25], [26],[28], [30], and to the references given by these authors.

Theoretical contributions (contrary to applied results) mostly concern the homogeneous boundary condition $u = 0$. However, many other boundary conditions are crucial in applications as, for instance, the following nonhomogeneous slip type boundary condition

$$\begin{cases} (u \cdot n)_{|\Gamma} = 0 \\ \beta u_\tau + \underline{\tau}(u)_{|\Gamma} = b(x), \end{cases} \tag{1.4}$$

that will be considered in the sequel together with the non-slip boundary condition

$$u_{|\Gamma} = 0. \tag{1.5}$$

In (1.4) $\underline{n}$ is the unit outward normal to the domain's boundary $\Gamma, \beta \geq 0$ is a given constant and $b(x)$ is a given tangential vector field. We denote by $\underline{t} = T \cdot \underline{n}$ the normal component of the tensor T, by $u_\tau = u - (u \cdot \underline{n})\underline{n}$ the tangential component of u and by $\underline{\tau}$ the tangential component of $\underline{t}$

$$\underline{\tau}(u) = \underline{t} - (\underline{t} \cdot \underline{n})\underline{n}. \tag{1.6}$$

The first deep mathematical study of this type of boundary conditions was done by V.A. Solonnikov and V.E. Ščadilov in reference [35].

For results and applications of boundary conditions like (1.4) see, for instance, [3], [4], [5], [8], [12], [15], [16], [22], [27], [29], [32], [35] [37], and references therein. See also [31], page 240, for a discussion of this subject.

We are interested in strong regularity results, *up to the boundary*, of weak solutions. The really *new obstacles* to face arise due to the interaction between the nonlinear terms containing $\nabla u + \nabla u^T$ and the boundary conditions. We concentrate our attention on this new point, by considering the following stationary problem in $\mathbb{R}^n_+$:

$$\begin{cases} -\nu_0 \nabla \cdot \left(\nabla u + \nabla u^T\right) - \\ \nu_1 \nabla \cdot \left(\left| \nabla u + \nabla u^T \right|^{p-2} \left(\nabla u + \nabla u^T\right)\right) + \nabla \pi = f, \\ \nabla \cdot u = 0. \end{cases} \tag{1.7}$$

Similar, but stronger, results hold for solutions to the simplest Lions model

$$\begin{cases} -\nu_0 \Delta u - \nu_1 \nabla \cdot (|\nabla u|^{p-2} \nabla u) + \nabla \pi = f(x), \\ \nabla \cdot u = 0, \end{cases} \tag{1.8}$$

The full non-linear evolution problem is studied in the forthcoming paper [7]. See the Remark 3.1 below.

2. WEAK SOLUTIONS. KNOWN RESULTS AND NOTATION

Let us now introduce the functional setting used in the following.

We set $p' = p/(p-1)$ for each $p \in]1, +\infty[$. If X is a Banach space we denote by X' its strong dual space. We use the same notation for functional spaces and norms for both scalar and vector fields. The symbol $\|\cdot\|_p$ denotes the canonical norm in $L^p(\mathbb{R}^n_+)$, and $\|\cdot\|$ that in $L^2(\mathbb{R}^n_+)$. In general, "integer norms", as well as "integer Sobolev spaces", relate to $\mathbb{R}^n_+$, and "fractional norms" concern the boundary $\Gamma = \mathbb{R}^{n-1}$. For instance, $\|\cdot\|_{1/2} = \|\cdot\|_{1/2,\Gamma}$, and $H^{1/2} = H^{1/2}(\mathbb{R}^{n-1})$.

We define $D^1 := D^{1,2}(\mathbb{R}^n_+)$ as the completion of $C_0^\infty(\overline{\mathbb{R}^n_+})$ (or $C_0^k(\overline{\mathbb{R}^n_+})$, $k \geq 1$) with respect to the norm $\|\nabla v\|$. Moreover, D^1_0 is the completion of $C_0^\infty(\mathbb{R}^n_+)$ with respect to $\|\nabla v\|$. It is well-known that

$$D^1 = \{v : v \in L^r, \nabla v \in L^2\}, \tag{2.1}$$

where $1/r = 1/2 - 1/n$. In particular, the norms $\|\nabla v\|$ and $\|\nabla v\| + \|v\|_{L^r}$ are equivalent in D^1 and in D^1_0.

Since the restriction to a bounded set B of any function in D^1 belongs to the Sobolev space $H^1(B)$, it follows that its trace on the boundary $\mathbb{R}^{n-1}$ is (locally) well defined as an element of $H^{\frac{1}{2}}$. Trace spaces in $\mathbb{R}^{n-1}$ may be studied, in a convenient way, by resorting to the Fourier transform. The trace space of D^1 is denoted here by $D^{1/2} = D^{1/2}(\mathbb{R}^{n-1})$. Actually, it is the completion of $C_0^\infty(\mathbb{R}^{n-1})$ with respect to the norm induced in $\mathbb{R}^{n-1}$ by the norm $\|\nabla v\|$ in $C_0^\infty(\mathbb{R}^n_+)$. It consists of functions (distributions) that have a "half derivative" in $L^2(\mathbb{R}^{n-1})$ (in the usual Fourier-transform sense) and that, actually, belong to $L^s(\mathbb{R}^{n-1})$, where s is given by the Sobolev embedding exponent

$$\frac{1}{s} = \frac{1}{2} - \frac{1/2}{n-1}. \tag{2.2}$$

See [13], Theorem II.3 and Definition II.1. See also [2], [11], [33] and references.

We set $D^{-1/2} = (D^{1/2})'$. Norms in $D^{1/2}$ and $D^{-1/2}$ are denoted respectively by $[.]_{1/2}$ and $[.]_{-1/2}$. Note that, by (2.2), one has $L^{s'} \subset D^{-1/2}$ where $s' = 2(n-1)/n$.

We define

$$D^1_\tau = \{v \in D^1 : v_n = 0 \text{ on } \Gamma\} \qquad \text{and} \qquad D^1_0 = \{v \in D^1 : v = 0 \text{ on } \Gamma\}.$$

V_2 denotes the space

$$V_2 = \left\{ v \in D_\tau^1 : \nabla \cdot v = 0 \text{ in } \mathbb{R}_+^n \right\} \tag{2.3}$$

if the boundary value problem under consideration is (1.4), and denotes the space

$$V_2 = \left\{ v \in D_0^1 : \nabla \cdot v = 0 \text{ on } \mathbb{R}_+^n \right\} \tag{2.4}$$

if the boundary value problem under consideration is (1.5). The above subspaces of D^1 are endowed with the norm $\|\nabla u\|$. Moreover, $[\cdot]_{-1}$ denotes the strong norm in the dual space $(V_2)'$.

We set

$$V = \left\{ v \in V_2 : \|\mathcal{D}v\|_p < \infty \right\},$$

endowed with the norm

$$\|v\|_V = \|\nabla v\|_2 + \|\mathcal{D}v\|_p .$$

It should be remarked that, by appealing to inequalities of Korn's type, we can verify that $V = \left\{ v \in V_2 : \|\nabla v\|_p < \infty \right\}$ and also that $\|\nabla v\|_2 + \|\mathcal{D}v\|_p$ and $\|\nabla v\|_2 + \|\nabla v\|_p$ are equivalent norms in V.

Weak solutions exist under the assumptions

$$f \in (V_2)', \tag{2.5}$$

and, concerning the tangential vector field b,

$$b \in D^{-\frac{1}{2}}(\mathbb{R}^{n-1}). \tag{2.6}$$

Note that (2.5) holds if $f \in L^{r'}$, and (2.6) holds if $b \in L^{s'}(\mathbb{R}^{n-1})$.

The following definition is well know.

Definition 2.1. *We say that u is a* weak solution *to problem* (1.7), (1.4) *if* $u \in V$ *satisfies*

$$\frac{1}{2}\int_\Omega \nu_T(u)\mathcal{D}u \cdot \mathcal{D}v dx + \beta \int_\Gamma u \cdot v d\Gamma = \int_\Omega f \cdot v dx + \int_\Gamma b \cdot v d\Gamma, \tag{2.7}$$

for all $v \in V$.

If we consider the Dirichlet boundary value problem (1.5), this definition applies as well, by dropping in (2.7) the terms with β and b.

By defining $< Au, v >$, for each pair $u, v \in V$, as the left hand side of (2.7), the operator $A : V \to V'$ satisfies the assumptions in the Theorems 2.1 and 2.2, Chap.2, Sect.2, [24]. This shows existence and uniqueness of the weak solution.

By replacing v by u in equation (2.8) one gets

$$\nu_0 \|\nabla u\|^2 + \nu_1 \|\mathcal{D}u\|_p^p + \beta \|u\|_\Gamma^2 = < b, u >_\Gamma + < f, u >_\Omega, \tag{2.8}$$

where the symbols $< \cdot, \cdot >$ denote "duality pairings" and the trace of u on the boundary is denoted simply by u. Note that the left hand side of equation (2.8) is just $< Au, u >$. This shows that the assumption (2.2) in the above Theorem 2.1, reference [24], holds.

From (2.8) there readily follows the basic estimate

$$\frac{\nu_0^2}{2} \|\nabla u\|^2 + \nu_0 \nu_1 \|\mathcal{D}u\|_p^p + \beta \|u\|_\Gamma^2 \le c_n ([f]_{-1}^2 + [b]_{-\frac{1}{2}}^2), \tag{2.9}$$

where the constant c_n depends only on n.

By restriction of (2.7) to divergence-free test-functions v with compact support in $\mathbb{R}_+^n$ there follows the existence of a distribution π (determined up to a constant) such that

$$\nabla \pi = -\nabla \cdot \left[\nu_0 \nabla u + \nu_1 \, | \, \mathcal{D}u \, |^{p-2} \, \mathcal{D}u \right] + f. \tag{2.10}$$

Equation (2.10) shows that the first equation (1.7) holds in the distributional sense.

In the sequel

$$B_R^+ = \{ x : | \, x \, | < R, x_n > 0 \},$$

and $| \, B_R^+ \, |$ denotes the Lebesgue measure of B_R^+.

We end this section by introducing some more notation.

We denote by $D^2 u$ the set of all the second derivatives of u. The meaning of expressions like $\|D^2 u\|$ is clear. The symbol $D_*^2 u$ may denote

any of the second order derivatives $\partial^2 u_j/\partial x_i \partial x_k$ except for the derivatives $\partial^2 u_j/\partial x_n^2$, if $j < n$. Moreover,

$$|D_*^2 u|^2 := \left|\frac{\partial^2 u_n}{\partial x_n^2}\right|^2 + \sum_{\substack{i,j,k=1 \\ (i,k)\neq(n,n)}}^{n} \left|\frac{\partial^2 u_j}{\partial x_i \partial x_k}\right|^2.$$

Similarly, ∇^* may denote any first order partial derivative, except for $\partial/\partial x_n$.

We set

$$\|\cdot\|_{\alpha,R} = \|\cdot\|_{L^\alpha(B_R^+)}.$$

Finally,

$$\|f,b\|^2 = \|f\|^2 + [b]_{1/2}^2$$

and

$$[f,b]^2 = [f]_{-1}^2 + [b]_{-\frac{1}{2}}^2.$$

3. NEW RESULTS.

Now we state the two main theorems. We set, for each $q > 1$,

$$\begin{aligned} K_q = c_n R\left(|B_R^+|^{\frac{1}{q}-\frac{1}{2}}[f,b] + \nu_1 |B_R^+|^{\frac{1}{q}-\frac{1}{p}} \|\mathcal{D}u\|_p^{p-1}\right) + \\ c_n\left(|B_R^+|^{\frac{1}{r}-\frac{1}{2}} + (p-1)\left(\frac{\nu_1}{\nu_0}\right)^{\frac{1}{2}} |B_R^+|^{\frac{1}{q}-\frac{1}{p}} \|\mathcal{D}u\|_p^{\frac{p-2}{2}}\right)\|f,b\|. \end{aligned} \tag{3.1}$$

Theorem 3.1. *Assume that* $2 < p$ *and that*

$$\begin{cases} f \in L^2\left(\mathbb{R}^n_+\right) \\ b \in D^{\frac{1}{2}}\left(\mathbb{R}^{n-1}\right). \end{cases} \tag{3.2}$$

Let u, π *be the weak solution to problem* (1.7) *under one of the boundary conditions* (1.4) *or* (1.5). *Then the derivatives* $D^2_* u$ *belong to* $L^2(\mathbb{R}^n_+)$ *and satisfy the estimate*

$$\nu_0 \left\| D^2_* u \right\| + (\nu_0 \nu_1)^{\frac{1}{2}} \left\| \mathcal{D}u \right|^{\frac{p-2}{2}} \nabla^* \mathcal{D}u \Big\| \le c_n \left\| f, b \right\|. \tag{3.3}$$

On the other hand,

$$D^2 u, \nabla^* \pi \in L^{p'}_{loc}(\overline{\mathbb{R}^n_+}),$$

where

$$p' = \frac{p}{p-1}.$$

In particular, if $p < \frac{n}{n-2}$, *then* $u \in C^{0,\alpha}_{loc}(\overline{\mathbb{R}^n_+})$ *where* $\alpha = \frac{n-(n-2)p}{p}$.

More precisely, for each $R > 0$,

$$\frac{1}{p-1} \left\| \nabla^* \pi \right\|_{p',R} + \nu_0 \left\| D^2 u \right\|_{p',R} \le \mathcal{K}, \tag{3.4}$$

where

$$\begin{aligned} \mathcal{K} = {} & c_n R \left(|B^+_R|^{\frac{1}{p'} - \frac{1}{2}} [f, b] + \nu_1 \left\| \mathcal{D}u \right\|^{p-1}_p \right) + \\ & c_n \left(|B^+_R|^{\frac{1}{p'} - \frac{1}{2}} + (p-1) \left(\frac{\nu_1}{\nu_0} \right)^{\frac{1}{2}} \left\| \mathcal{D}u \right\|^{\frac{p-2}{2}}_p \right) \left\| f, b \right\|. \end{aligned} \tag{3.5}$$

Moreover, if $p < 4$,

$$\frac{\partial \pi}{\partial x_n} \in L^{\bar{p}}_{loc}(\overline{\mathbb{R}^n_+}),$$

where

$$\bar{p} = \frac{2p}{3p-4}. \tag{3.6}$$

Recall that $\|\mathcal{D}u\|_p$ *satisfies the estimate* (2.9) .

The above theorem may be improved. Merely for convenience assume that $n=3$. For brevity, $\nu_0 = \nu_1 = 1$ and C_R depends on $|B_R^+|$.

Theorem 3.2. *Assume that* $n=3$, $\nu_0 = \nu_1 = 1$, *and*

$$2 \le p \le 3.$$

Let f, b, u *and* π *be as in Theorem 3.1. Then, in addition to the results stated in this last theorem, one has*

$$D^2 u, \nabla^* \pi \in L^l_{loc}(\overline{\mathbb{R}^n_+}),$$

where $l = 3(4-p)/(5-p)$.
In particular, $u \in C^{0,\alpha}_{loc}(\overline{\mathbb{R}^n_+})$, *where* $\alpha = \frac{3-p}{4-p}$.

More precisely, for each $R > 0$,

$$\|\nabla^* \pi\|_{l,R} + \|D^2 u\|_{l,R} \le K_1 + c_p \|f, b\|^{\frac{2}{4-p}}. \tag{3.7}$$

Finally,

$$\left\|\frac{\partial \pi}{\partial x_n}\right\|_{m,R} \le C_R\left(\|\mathcal{D}u\|_p + \|\mathcal{D}u\|_p^{p-1} + \|f\| + \|f\|^{\frac{2}{4-p}}\right),$$

where $m = 6(4-p)/(8-p)$.

Remark 3.1. Assume that $n=3$. Among the other results, in reference [7] we show that (under natural regularity hypotheses on the data) the solutions to the initial-boundary value problem (1.1), (1.4) (or (1.5)) in a regular bounded open set Ω satisfy

$$u \in L^2\left(0,T;W^{2,p'}(\Omega)\right) \qquad \text{if} \qquad p \in]\,2+\frac{2}{5},4[$$

and

$$u \in L^{4-p}\left(0,T;W^{2,p'}(\Omega)\right) \qquad \text{if} \qquad p \in]\,2+\frac{2}{5},3[\,.$$

For the Stokes evolution problem (drop the term $(u\cdot\nabla)u$*) the above results hold for each* $p \geq 2$.

The linear case, $p=2$, is well studied; see [35], [4] and [5]. Nevertheless, it is significant that, in this particular case, the statements and estimates established in Theorems 3.1 and 3.2 coincide with the classical results.

REFERENCES

[1] H. AMANN, *Stability of the rest state of a viscous incompressible fluid*, Arch. Rat. Mech. Anal., **126** (1944), 231-242.

[2] A. AVANTAGGIATI, *Spazi di Sobolev con peso ed alcune applicazioni*, J.Fluid Mech., **30** (1967), 197-207.

[3] G.J. BEAVERS, D.D. JOSEPH, *Boundary conditions of a naturally permeable wall*, J.Fluid Mech., **30** (1967), 197-207.

[4] H. BEIRÃO DA VEIGA, *Regularity of solutions to a nonhomogeneous boundary value problem for general Stokes systems in* $\mathbb{R}^n_+$, to appear in Math. Annalen.

[5] H. BEIRÃO DA VEIGA, *Regularity for Stokes and generalized Stokes systems under nonhomogeneous slip type boundary conditions*, to appear in Advances in Diff. Equations.

[6] H. BEIRÃO DA VEIGA, *On the regularity of flows with Ladyzhenskaya shear dependent viscosity and slip or non-slip boundary conditions equations*, to appear in Comm. Pure Appl. Math.

[7] H. BEIRÃO DA VEIGA, in preparation.

[8] C. CONCA, *On the application of the homogenization theory to a class of problems arising in fluid mechanics*, J. Math. Pures Appl., **64** (1985), 31-75.

[9] J. Frehse, J. Málek, and M. Steinhauer, *An existence result for fluids with shear dependent viscosity-steady flows*, Nonlinear Analysis. TMA., **30** (1997), 3041-3049.

[10] H. Fujita, *Remarks on the Stokes flow under slip and leak boundary conditions of friction type*, in "Topics in Mathematical Fluid Mechanics", Quaderni di Matematica, Vol.**10**, Napoli 2002, 73-94.

[11] G.P. Galdi, *An Introduction to the Mathematical Theory of the Navier-Stokes Equations: Vol.I: Linearized Steady Problems*, Springer Tracts in Natural Philosophy,**38**,Second corrected printing, Springer-Verlag, 1998.

[12] G.P. Galdi, W. Layton, *Approximation of the larger eddies in fluid motion: A model for space filtered flow*, Math. Models and Meth. in Appl. Sciences, **3** (2000), 343-350.

[13] B. Hanouzet, *Espaces de Sobolev avec poids. Application au problème de Dirichlet dans un demi espace*, Rend. Sem. Mat. Univ. Padova,**46** (1971), 227-272.

[14] T.J.R. Hughes, L. Mazzei, and A.A. Oberai, *The multiscale formulation of large eddy simulation: Decay of homogeneous isotropic turbulence*, Physics of Fluids, **13** (2001), 505-512.

[15] V. John, *Slip with friction and penetration with resistence boundary conditions for the Navier-Stokes equations-numerical tests and aspects of the implementations*, J. Comp. Appl. Math., to appear.

[16] V. John, *Large eddy simulation of turbulent incompressible flows. Analytical and numerical results for a class of LES models*, Habilitationsschrift, Otto-von-Guericke-Universitat Magdeburg, April 2002.

[17] O.A. Ladysenskaya, *On nonlinear problems of continuum mechanics*, Proc. Int. Congr. Math. (Moscow, 1966). Nauka, Moscow, 1968, p.p.560-573; English transl. in Amer.Math. Soc. Transl.(2), **70** (1968).

[18] O.A. Ladyzenskaya, *Sur de nouvelles équations dans la dynamique des fluides visqueux et leurs résolution globale*, Troudi Math. Inst. Steklov, **CII** (1967), 85-104.

[19] O.A. Ladysenskaya, *Sur des modifications des équations de Navier-Stokes pour des grand gradients de vitesses*, Séminaire Inst. Steklov, **7** (1968), 126-154.

[20] O.A. Ladysenskaya, *The Mathematical Theory of Viscous Incompressible Flow*, Gordon and Breach, New-York, 2° édition, 1969.

[21] O.A. Ladyzenskaya, G.A. Seregin, *On regularity of solutions to two-dimensional equations of the dynamics of fluids with nonlinear viscosity*, Zap. Nauch. Sem. Pt. Odel. Mat. Inst., **259** (1999), 145-166.

[22] A. Liakos, *Discretization of the Navier-Stokes equations with slip boundary condition*, Num. Meth. for Partial Diff. Eq.,**1** (2001), 1-18.

[23] J.-L. Lions, *Sur cértaines équations paraboliques non linéaires*, Bull. Soc. Math. France, **93** (1965), 155-175.

[24] J.-L. Lions, *Quelques Méthodes de Résolution des Problèmes aux Limites Non Linéaires*, Dunod, Paris, 1969.

[25] J. Malek, J. Nečas, and M. Ružička, *On weak solutions to a class of non-Newtonian incompressilble fluids in bounded three-dimensional domains: the case* $p \geq 2$, Advances in Diff. Equations, **6** (2001), 257-302.

[26] J. Malek,K.R. Rajagopal, and M. Ružička, *Existence and regulaity of solutions and stability of the rest state for fluids with shear dependent viscosity*, Math. Models Methods Appl. Sci., **6**, (1995), 789-812.

[27] C. Pare's, *Existence, uniqueness and regularity of solutions of the equations of a turbulence model for incompressible fluids*, Appl. Analysis, **43** (1992), 245-296.

[28] M. Ružičza, *A note on steady flow of fluids with shear dependent viscosity*, Nonlinear Analysis. TMA., **30** (1997), 3029-3039.

[29] H. Saito, L.E. Scriven, *Study of the coating flow by the finite elemente method*, J. Comput. Phys., **42** (1981), 53-76.

[30] G.A. Seregin,*Interior regularity for solutions to the modified Navier-Stokes equations*, J.Math. Fluid Mech., **1** (1999), 235-281.

[31] J. Serrin, *Mathematical Principles of Classical Fluid Mechanics*, in *Encyclopedia of Physics* VIII, p.p. 125-263, Springer-Verlag, Berlin, 1959.

[32] J. Silliman, L.E. Scriven, *Separating flow near a static contact line: slip at a wall and shape of a free surface*, J.Comput. Physics, **34** (1980), 287-313.

[33] C.G. Simader, H. Sohr, *The Dirichlet problem for the Laplacian in Bounded and unbounded domains*, Pitman Research Notes in Mathematics Series., Longman Scientific and Technical, **360**, 1997.

[34] J.S. Smagorinsky,*General circulation experiments with the primitive equations. I. The basic experiment*, Mon. Weather Rev., **91** (1963), 99-164.

[35] V.A. Solonnikov and V.E. Ščadilov, *On a boundary value problem for a stationary system of Navier-Stokes equations*, Proc. Steklov Inst. Math.,**125** (1973), 186-199.
[36] G. Stokes, Trans. Cambridge Phil. Soc., **8**, 287 (1845), 75-129.
[37] R. Verfürth, *Finite element approximation of incompressible Navier-Stokes equations with slip boundary conditions*, Numer. Math., **50** (1987), 697-721.

HOMOGENIZATION OF SYSTEMS OF PARTIAL DIFFERENTIAL EQUATIONS

A. Bensoussan
University Paris Dauphine, Paris, France

1. INTRODUCTION

In this paper, we consider the class of systems of nonlinear partial differential equations, which has been lengthily studied by Prof J. FREHSE and the A., with application to stochastic differential games with N players. In particular, we refer to the book, A. BENSOUSSAN, J. FREHSE [1]. The regularity theory is instrumental to prove the existence of equilibriums in noncooperative games. The objective in this paper is to show that regularity theory is also extremely useful for obtaining the limit of problems with small parameters, like in homogenization. The methods used for scalar equations cannot extend, and the regularity results become instrumental

2. STATEMENT OF THE PROBLEM AND RESULTS

2.1 Notation

We consider a family of matrices $a^{\varepsilon}(x)$ satisfying

$$a^{\varepsilon}(x) \text{ is measurable on } R^n \tag{2.1}$$

$$a^{\varepsilon}(x)\xi.\xi \geq \alpha \mid \xi \mid^2, \forall \xi \in R^n, \alpha > 0 \tag{2.2}$$

$$(a^{\varepsilon})^{-1}(x)\xi.\xi \geq \alpha_0 \mid \xi \mid^2, \forall \xi \in R^n, \alpha_0 > 0. \tag{2.3}$$

We shall say that a^{ε} belongs to the class $M(\alpha, \alpha_0)$. Let Ω be a smooth bounded domain of R^n. Following the theory of abstract homogenization introduced by L. TARTAR [6] and F. MURAT, L. TARTAR [5], we shall consider the following properties. There exists a sequence of vectors $v^{\varepsilon}(x) \in R^n$, such that

$$v^{\varepsilon} \in (H^1(\Omega))^n \text{ and } v^{\varepsilon} \to x \text{ in } (H^1(\Omega))^n \text{ weakly} \tag{2.4}$$

$$a^{\varepsilon}(Dv^{\varepsilon})^* \to a \text{ in } (L^2(\Omega))^{n\times n} \text{ weakly} \tag{2.5}$$

$$\operatorname{div}(a^{\varepsilon}(Dv^{\varepsilon})^*) \to \operatorname{div} a \text{ in } (H^{-1}(\Omega))^n \text{ strongly} \tag{2.6}$$

Note that

$$(Dv^{\varepsilon})_{ij} = \frac{\partial v_i^{\varepsilon}}{\partial x_j}.$$

It is a classical result that a belongs to $M(\alpha, \alpha_0)$.

We also assume, for technical reasons,

$$\| Dv^{\varepsilon}(x) \| \leq B, x \in \Omega \tag{2.7}$$

We next consider N Hamiltonians $H^{\varepsilon,\nu}(x,s,\xi)$, where $s \in R^N, \xi \in R^{N\times n}, \nu = 1, \ldots, N$, such that

$$H^{\varepsilon,\nu}(x,s,\xi)\mid_{\{\xi^{\nu}=0\}} - H^{\varepsilon,\nu}(x,s,\xi)\{\xi^{\nu} = 0, s^{\nu} = 0\} \geq \beta^{\nu} s^{\nu}, \text{ if } s^{\nu} > 0 \tag{2.8}$$

$$H^{\varepsilon,\nu}(x,s,\xi)\mid_{\{\xi^{\nu}=0\}} - H^{\varepsilon,\nu}(x,s,\xi)\{\xi^{\nu} = 0, s^{\nu} = 0\} \leq \beta^{\nu} s^{\nu}, \text{ if } s^{\nu} < 0 \tag{2.9}$$

$$\mid H^{\varepsilon,\nu}(x,s,\xi)\mid_{\{s^{\nu}=0,\xi^{\nu}=0\}} \mid \leq M^{\nu} \tag{2.10}$$

Writing H^{ε} for the vector $H^{\varepsilon,\nu}$, we also assume

$$|H^{\varepsilon}(x,s,\xi)-H^{\varepsilon}(x,s',\xi)|\leq \varpi(|s-s'|)(1+|\xi|^2) \tag{2.11}$$

and

$$|H^{\varepsilon}(x,s,\xi)-H^{\varepsilon}(x,s,\xi')|\leq \gamma|\xi-\xi'|(1+|\xi|+|\xi'|+|s|^{\frac{1}{2}}) \tag{2.12}$$

where $\varpi: R^+ \to R^+$ is continuous, increasing, $\varpi(0)=0$. Next we assume a special growth assumption on the Hamiltonians

$$\begin{gathered}|H^{\varepsilon,\nu}(x,s,\xi)|\leq K^{\nu}|\xi||\xi^{\nu}|+\sum_{\{\mu=1\}}^{\nu} K_{\mu}^{\nu}|\xi^{\mu}|^2+k^{\nu}(x),\nu=1,\ \ldots,N-1 \\ |H^{\varepsilon,N}(x,s,\xi)|\leq K^{N}|\xi|^2+k^{N}(x)\end{gathered} \tag{2.13}$$

where

$$K^{\nu},K_{\mu}^{\nu} \text{ are positive constants}, k^{\nu}\geq 0\in L^q(\Omega), q>\frac{n}{2}. \tag{2.14}$$

We consider the system of elliptic equations

$$-\mathrm{div}(a^{\varepsilon}(x)Du^{\varepsilon,\nu})+H^{\varepsilon,\nu}(x,u^{\varepsilon},Du^{\varepsilon})=0, x\in\Omega, u^{\varepsilon,\nu}|_{\partial\Omega}=0 \tag{2.15}$$

where u^{ε} denotes the vector of components $u^{\varepsilon,\nu}$. The functions $u^{\varepsilon,\nu}$ belong to

$$u^{\varepsilon,\nu}\in H_0^1(\Omega)\times L^{\infty}(\Omega) \tag{2.16}$$

In our following development estimates will be proven, which will be uniform in ε, so we shall assume the existence of u^{ε} so that (2. 15), (2. 16) hold. We can refer to A. BENSOUSSAN, J. FREHSE [2].

Remark 2.1. *There is an additional degree of freedom, related to the ordering of equations in writing the system. Let* Γ *be an* $N\times N$ *matrix, which is invertible. To* Γ *we associate the transform of* H^{ε} *denoted* H_{Γ}^{ε}, *defined as follows*

$$H_{\Gamma}^{\varepsilon}(x,s,\xi)=\Gamma H^{\varepsilon}(x,\Gamma^{-1}s,\Gamma^{-1}\xi). \tag{2.17}$$

Setting

$$z^\varepsilon = \Gamma u^\varepsilon \tag{2.18}$$

then z^ε *is the solution of (2.15),(2.16), with* H^ε *replaced by* H^ε_Γ. *We shall need that (2.8),(2.9),(2.10) hold for some transform* H^ε_Γ, *with* Γ *satisfying the Maximum Principle, which means*

$$\Gamma s \geq 0 \Rightarrow s \geq 0 \tag{2.19}$$

and that (2.11), (2.12), (2.13) hold for another transfom H^ε_Γ, *not necessarily the same, in particular with* Γ *not satisfying the Maximum Principle. We shall neeed (2.8),(2.9),(2.10) to prove that* z^ε *is bounded. Since* Γ *satisfies the Maximum Principle and is invertible, this implies that* u^ε *is bounded. This being achieved, another transformation, not necessarily satisfying the Maximum Principle preserves the* L^∞ *bound. It permits to show* C^δ *estimates, which are also valid for* u^ε. *In the statement of results, this flexibility will be implicit.*

2.2 Statement of Results

Our objective is to prove the following

Theorem 2.1. *We make the assumptions (2.1), (2.2), (2.3), (2.4), (2.5), (2.6), (2.7), (2.8), (2.9), (2.10), (2.11), (2.12), (2.13), (2.14). For the assumptions (2.8) to (2.14), we take into account Remark 1. Let* u^ε *be a solution of the system (2.15), then*

$$\| u^{\varepsilon,\nu} \|_{L^\infty(\Omega)} \leq C, \| u^{\varepsilon,\nu} \|_{H^1_0(\Omega)} \leq C \tag{2.20}$$

If we pick a subsequence, still denoted $u^{\varepsilon,\nu}$, *such that then*

$$Du^{\varepsilon,\nu} - (Dv^\varepsilon)^* Du^\nu \to 0 \text{ in } L^2(\Omega). \tag{2.21}$$

Moreover there exist Hamiltonians $H^\nu(x,s,\xi)$ *satisfying assumptions (2. 8) to (2. 14), with possibly different constants, such that* u *the vector of components* u^ν *satisfies the equations*

$$-\mathrm{div}(a(x)Du^\nu) + H^\nu(x,u,Du) = 0, x \in \Omega, u^\nu|_{\partial\Omega} = 0 \tag{2.22}$$

3. A PRIORI ESTIMATES

3.1 Preliminaries

We note that the solution of (2. 15) has the full regularity, namely $u^{\varepsilon,\nu} \in W^{2,p}(\Omega)$, in particular $u^{\varepsilon,\nu} \in C^1(\bar{\Omega})$. We prove first

Lemma 3.1. *We have the estimates*

$$|u^{\varepsilon,\nu}(x)| \leq \frac{M^\nu}{\beta^\nu} \tag{3.23}$$

PROOF:
Note first that, from (2.8), (2.10) one has

$$H^{\varepsilon,\nu}(x,s,\xi)|_{\{\xi^\nu=0\}} \geq \beta^\nu s^\nu - M^\nu, \text{ if } s^\nu > 0 \tag{3.24}$$

and from (2.9), (2.10)

$$H^{\varepsilon,\nu}(x,s,\xi)|_{\{\xi^\nu=0\}} \leq \beta^\nu s^\nu + M^\nu, \text{ if } s^\nu < 0. \tag{3.25}$$

The function $u^{\varepsilon,\nu}(x)$ being continuous in $\bar{\Omega}$ attains its maximum in x^ε (we omit to write the dependence in ν). Suppose the maximum is strictly positive, then $x^\varepsilon \in \Omega$. From the Maximum Principle, we have

$$H^{\varepsilon,\nu}(x,u^\varepsilon(x^\varepsilon),Du^\varepsilon(x^\varepsilon))|_{\{\xi^\nu=0\}} \leq 0 \tag{3.26}$$

so, using (3.24), we deduce

$$u^{\varepsilon,\nu}(x^\varepsilon) \leq \frac{M^\nu}{\beta^\nu} \tag{3.27}$$

if $u^{\varepsilon,\nu}(x^\varepsilon) > 0$, and this inequality is obvious if $u^{\varepsilon,\nu}(x^\varepsilon) \leq 0$. A similar inequality is proven for the minimum, using this time (3.25). The result (3. 23) is thus obtained. □

We shall now make use of the growth assumptions (2. 13),(2.14). We first notice that we can write

$$H^{\varepsilon,\nu}(x,s,\xi)=Q^{\varepsilon,\nu}(x,s,\xi).\xi^{\nu}+H_0^{\varepsilon,\nu}(x,s,\xi) \tag{3.28}$$

with the properties

$$\begin{gathered} Q^{\varepsilon,\nu}, \text{ measurable, continuous in } s,\xi, \text{ for } \xi^{\nu}\neq 0 \\ |Q^{\varepsilon,\nu}(x,s,\xi)|\leq K^{\nu}\,|\xi|, \nu=1,\ \ldots,N-1 \\ Q^{\varepsilon,N}=Q^{\varepsilon,N-1} \end{gathered} \tag{3.29}$$

and

$$|H_0^{\varepsilon,\nu}(x,s,\xi)|\leq \sum_{\{\mu=1\}}^{\nu} K_{\mu}^{\nu}\,|\xi^{\mu}|^2+k^{\nu}(x), \nu=1,\ \ldots,N \tag{3.30}$$

where the constants not yet defined are K_{μ}^{N} defined as follows

$$K_{\mu}^{N}=K^{N}+\frac{1}{2}K^{N-1}, \mu=1,\ \ldots,N-1; K_{N}^{N}=K^{N}+K^{N-1}. \tag{3.31}$$

Indeed, we set, for $\nu=1,\ \ldots,N-1$

$$\sigma^{\varepsilon,\nu}(x,s,\xi)=\frac{H^{\varepsilon,\nu}(x,s,\xi)}{K^{\nu}\,|\xi|\,|\xi^{\nu}|+\sum_{\{\mu=1\}}^{\nu}K_{\mu}^{\nu}\,|\xi^{\mu}|^2+k^{\nu}(x)} \tag{3.32}$$

and successively

$$Q^{\varepsilon,\nu}(x,s,\xi)=K^{\nu}\sigma^{\varepsilon,\nu}(x,s,\xi)\,|\xi|\,\frac{\xi^{\nu}}{|\xi^{\nu}|} \tag{3.33}$$

for $\nu=1,\ \ldots,N-1$, and $Q^{N}=Q^{N-1}$. Then we set

$$H_0^{\varepsilon,\nu}(x,s,\xi)=H^{\varepsilon,\nu}(x,s,\xi)-Q^{\varepsilon,\nu}(x,s,\xi).\xi^{\nu} \tag{3.34}$$

then, it is easy to check that (3.28), (3.29), (3.30), (3.31) are verified.

We then proceed with a fundamental inequality. For simplicity at this stage, we shall omit to write explicitly ε, since all estimates will be uniform with respect to ε. We call

$$\rho = \max_{\nu} \frac{M^{\nu}}{\beta^{\nu}} \tag{3.35}$$

which is an L^{∞} bound for the solution of (2. 15), which we call temporarily u^{ν}, without ε. To any solution $u = (\ldots, u^{\nu}, \ldots)$, we associate a constant vector c, such that

$$\| c \| \leq \rho \tag{3.36}$$

and we write

$$\tilde{u} = u - c.$$

Let also

$$\psi \geq 0, \psi \in H^1 \cap L^{\infty}(\Omega), \psi |_{\partial\Omega} = 0 \text{ if } c \neq 0. \tag{3.37}$$

We introduce the notation

$$\beta(x) = \exp x - x - 1 \tag{3.38}$$

and the map $X(s) : R^N \to R^N$ defined backwards by the formulas

$$\begin{aligned} X^N(s) &= \exp[\beta(\gamma^N s^N) + \beta(-\gamma^N s^N)] \\ X^{\nu}(s) &= \exp[\beta(\gamma^{\nu} s^{\nu}) + \beta(-\gamma^{\nu} s^{\nu}) + X^{\nu+1}(s)], \nu = 1, \ldots, N-1 \end{aligned} \tag{3.39}$$

where γ^{ν} are positive constants ans $s = (s^1, \ldots, s^N)$. We note the formula

$$\frac{\partial X^{\nu}}{\partial s^{\mu}} = \begin{cases} 0 \text{ if } \mu < \nu \\ \gamma^{\mu} X^{\nu} \ldots X^{\mu} (\beta'(\gamma^{\mu} s^{\mu}) - \beta'(-\gamma^{\mu} s^{\mu})) \text{ if } \mu \geq \nu \end{cases} \tag{3.40}$$

We call

$$X(x) = X(\tilde{u}(x)) \tag{3.41}$$

hence clearly

$$DX^{\nu} = \sum_{\{\mu=\nu\}}^{N} \gamma^{\mu} X^{\nu} \ldots X^{\mu}(\beta'(\gamma^{\mu}\tilde{u}^{\mu}) - \beta'(-\gamma^{\mu}\tilde{u}^{\mu}))Du^{\mu} \tag{3.42}$$

from which we deduce the estimates

$$\begin{aligned} &|DX| \le c(\rho)|\tilde{u}||Du| \\ &0 \le X(x) - X_0 \le c(\rho)|\tilde{u}|^2 \end{aligned} \tag{3.43}$$

where in the sequel, $c(\rho)$ denotes a constant depending only of ρ (this assumes that the constants γ^{μ} depend only of ρ), and X_0 is the value of $X(s)$ for $s = 0$. We have

$$X_0 \ge 1.$$

We state the

Proposition 3.1. We assume (2.1), (2.2), (2.3), (2.13), (2.14), a solution u of (2. 15), bounded by ρ. There exist constants $\gamma^{\nu}(\rho)$, $c(\rho)$ such that, for any constant vector c satisfying (3. 36), and any ψ such that (3. 37) holds, one has

$$\int_{\Omega} a_{ij} \frac{\partial X^{1}}{\partial x_j} \frac{\partial \psi}{\partial x_i} dx + \alpha \int_{\Omega} \psi |Du|^2 dx \le c(\rho) \int_{\Omega} \psi \sum_{\{\nu=1\}}^{N} k^{\nu} dx \tag{3.44}$$

PROOF:
We take as a test function in (2.15)

$$v^{\nu} = \psi\gamma^{\nu}(\beta'(\gamma^{\nu}\tilde{u}^{\nu}) - \beta'(-\gamma^{\nu}\tilde{u}^{\nu}))\prod_{\mu=1}^{\nu} X^{\mu}$$

then

$$\int_{\Omega} -\operatorname{div}(a(x)Du^{\nu})v^{\nu}\, dx = \int_{\Omega} a_{ij} \frac{\partial u^{\nu}}{\partial x_j} \frac{\partial \psi}{\partial x_i} \gamma^{\nu}(\beta'(\gamma^{\nu}\tilde{u}^{\nu}) - \beta'(-\gamma^{\nu}\tilde{u}^{\nu}))\prod_{\mu=1}^{\nu} X^{\mu}\, dx +$$

$$+ \int_{\Omega} a_{ij} \frac{\partial u^{\nu}}{\partial x_j} \psi\gamma^{\nu 2}(\beta''(\gamma^{\nu}\tilde{u}^{\nu}) + \beta''(-\gamma^{\nu}\tilde{u}^{\nu})) \frac{\partial u^{\nu}}{\partial x_i} \prod_{\mu=1}^{\nu} X^{\mu}\, dx +$$

$$+\int_\Omega a_{ij}\frac{\partial u^\nu}{\partial x_j}\psi\gamma^\nu(\beta'(\gamma^\nu\tilde{u}^\nu)-\beta'(-\gamma^\nu\tilde{u}^\nu))\frac{\partial}{\partial x_i}\prod_{\mu=1}^{\nu}X^\mu\,dx$$

$$=I+II+III.$$

Then, one checks easily that

$$I=\int_\Omega a_{ij}\frac{\partial X^1}{\partial x_j}\frac{\partial\psi}{\partial x_i}dx$$

$$II\geq\alpha\int_\Omega\psi\,|Du^\nu|^2\,{\gamma^\nu}^2(\exp\gamma^\nu\tilde{u}^\nu+\exp-\gamma^\nu\tilde{u}^\nu)\prod_{\mu=1}^{\nu}X^\mu\,dx$$

$$III=\int_\Omega a_{ij}\frac{\partial F^\nu}{\partial x_j}\frac{\partial F^\nu}{\partial x_i}\psi\prod_{\mu=1}^{\nu}X^\mu\,dx$$

where

$$F^\nu=\log X^\nu.$$

Next, we have

$$\sum_{\nu=1}^{N-1}\int_\Omega Q^\nu v^\nu\,dx=\sum_{\nu=1}^{N-1}\int_\Omega\psi Q^\nu(DF^\nu-DX^{\nu+1})\prod_{\mu=1}^{\nu}X^\mu\,dx$$

$$=\sum_{\nu=1}^{N-1}\int_\Omega\psi(Q^\nu-Q^{\nu-1})DF^\nu\prod_{\mu=1}^{\nu}X^\mu\,dx-\int_\Omega\psi Q^{N-1}DF^N\prod_{\mu=1}^{N}X^\mu\,dx$$

where we have set $Q^0=0$. Since $Q^{N-1}=Q^N$, it follows that

$$\sum_{\nu=1}^{N}\int_\Omega Q^\nu v^\nu\,dx=\sum_{\nu=1}^{N-1}\int_\Omega\psi\tilde{Q}^\nu DF^\nu\prod_{\mu=1}^{\nu}X^\mu\,dx$$

where

$$\tilde{Q}^\nu=Q^\nu-Q^{\nu-1},\nu=1,\cdots,N-1.$$

Collecting results and performing additional majorations we obtain

$$\int_\Omega a_{ij}\frac{\partial X^1}{\partial x_j}\frac{\partial \psi}{\partial x_i}dx+$$

$$+\alpha\sum_{\nu=1}^{N}\int_\Omega \psi\,|Du^\nu|^2\,{\gamma^\nu}^2(\exp\gamma^\nu\tilde{u}^\nu+\exp-\gamma^\nu\tilde{u}^\nu)\prod_{\mu=1}^{\nu}X^\mu\,dx$$

$$+\sum_{\nu=1}^{N}\int_\Omega \psi H_0^\nu\gamma^\nu(\exp\gamma^\nu\tilde{u}^\nu-\exp-\gamma^\nu\tilde{u}^\nu)\prod_{\mu=1}^{\nu}X^\mu\,dx\le\frac{1}{4}\sum_{\nu=1}^{N-1}\int_\Omega\psi(\frac{a+a^*}{2})^{-1}\tilde{Q}^\nu.\tilde{Q}^\nu\prod_{\mu=1}^{\nu}X^\mu\,dx.$$

Thanks to the properties (3.30), (3.31), it follows that

$$\int_\Omega a_{ij}\frac{\partial X^1}{\partial x_j}\frac{\partial \psi}{\partial x_i}dx+$$

$$+\alpha\sum_{\nu=1}^{N}\int_\Omega \psi\,|Du^\nu|^2\,{\gamma^\nu}^2(\exp\gamma^\nu\tilde{u}^\nu+\exp-\gamma^\nu\tilde{u}^\nu)\prod_{\mu=1}^{\nu}X^\mu\,dx\le$$

$$\sum_{\nu=1}^{N}\int_\Omega \psi\,|Du^\nu|^2\,[\frac{1}{4\alpha}\sum_{\sigma=1}^{N-1}(K^\sigma+K^{\sigma-1})^2$$

$$+\sum_{\sigma=\nu}^{N}\gamma^\sigma K_\nu^\sigma\,|\exp\gamma^\sigma\tilde{u}^\sigma-\exp-\gamma^\sigma\tilde{u}^\sigma|]\prod_{\mu=1}^{\sigma}X^\mu\,dx+$$

$$+\sum_{\nu=1}^{N}\int_\Omega\psi\gamma^\nu k^\nu\,|\exp\gamma^\nu\tilde{u}^\nu-\exp-\gamma^\nu\tilde{u}^\nu|\prod_{\mu=1}^{\nu}X^\mu\,dx.$$

Suppose the constants $\gamma^\nu(\rho)$ are chosen so that

$$\alpha{\gamma^\nu}^2-2\gamma^\nu K_\nu^\nu\ge\frac{1}{4\alpha}[\sum_{\sigma=1}^{\nu}(K^\sigma+K^{\sigma-1})^2+$$

$$+\sum_{\sigma=\nu+1}^{N-1}(K^\sigma+K^{\sigma-1})^2\prod_{\mu=\nu+1}^{\sigma}X^\mu]+ \tag{3.45}$$

$$+\sum_{\sigma=\nu+1}^{N}\gamma^\sigma K_\nu^\sigma\,|\exp\gamma^\sigma\tilde{u}^\sigma-\exp-\gamma^\sigma\tilde{u}^\sigma|\prod_{\mu=\nu+1}^{\sigma}X^\mu$$

and $c(\rho)$ is such that

$$\gamma^\nu\,|\exp\gamma^\nu\tilde{u}^\nu-\exp-\gamma^\nu\tilde{u}^\nu|\prod_{\mu=1}^{\nu}X^\mu\le c(\rho) \tag{3.46}$$

then the result (3.44) follows. The constants $\gamma^{\nu}(\rho)$ can be defined by the relations (3.45) backwards, observing that X^{μ} can be majorized by a number depending only on $\rho, \gamma^{\mu}, \ldots, \gamma^{N}$.

The proof has been completed. □

3.2 Estimates

We begin by stating the following result concerning the H_0^1 estimates

Proposition 3.2. *We have the estimate*

$$\alpha \int_{\Omega} | Du^{\varepsilon} |^2 dx \leq c(\rho) \tag{3.47}$$

PROOF :
One just pick $c = 0$ and $\psi = 1$ in (3. 44). The result follows immediately. □

We then proceed with the Hölder estimate, which is essential in the case of systems

Proposition 3.3 *For* $\delta < \delta_0 = 1 - \frac{n}{2q}$, *one has the estimate*

$$| u^{\varepsilon,\nu} |_{C^{\delta}} \leq c_{\delta}(\rho) \tag{3.48}$$

We begin by introducing the Green function, with respect to a point $x_0 \in \Omega$. Let Q be a ball such that $\bar{\Omega} \subset Q$. The Green function is the solution $G = G^{x_0}$ of the equation

$$\int_{Q} aD\phi.DG\, dx = \phi(x_0), \forall \phi \in C_0^{\infty}(Q) \tag{3.49}$$

Moreover, G satisfies the estimates

$$c_0 | x - x_0 |^{2-n} \leq G(x) \leq c_1 | x - x_0 |^{2-n} \tag{3.50}$$

for all x in a neighborhood of x_0, whose closure is contained in Q. In particular, (3. 50) holds for $x \in \bar{\Omega}$. The constants c_0, c_1 depend only on α, α_0, therefore they do not depend on ε, whereas G depends on ε.

The next ingredient is the cut-off function. Let $\tau(x)$ be a smooth function such that $0 \leq \tau \leq 1$, and

$$\tau(x) = 1, \forall x \text{ such that } |x| \leq 1, \tau(x) = 1, \forall x \text{ such that } |x| \geq 2.$$

We define

$$\tau_R(x) = \tau(\frac{x - x_0}{R})$$

and we denote by $B_R = B_R(x_0)$ the ball of center x_0 and of radius R. We assume $R \leq R_0$. An essential element in the proof of Proposition 3 is the following

Lemma 3.2. *We have the inequality*

$$\int_{B_R} |Du^\varepsilon|^2 |x - x_0|^{2-n} dx \leq C \int_{B_{4R} - B_R} |Du^\varepsilon|^2 |x - x_0|^{2-n} dx + CR^\beta \quad (3.51)$$

for all $R \leq R_0$ *and* $\beta \leq \beta_0 = 2 - \frac{n}{q}$, *with* C *depending only on* ρ.

PROOF of LEMMA 3.2:
We apply (3.44) with

$$\psi = G\tau_R^2$$

and

$$c = c^R = \begin{cases} 0 \text{ if } B_{2R} \cap (R^n - \Omega) \neq \varnothing \\ \dfrac{1}{|B_{2R} - B_R|} \displaystyle\int_{B_{2R} - B_R} u\, dx \text{ if } B_{2R} \subset \Omega \end{cases} \quad (3.52)$$

We can also consider that u is extended outside Ω with the value 0.

We first notice that

$$\alpha \int_\Omega G\tau_R^2 |Du|^2 dx \geq \alpha c_0 \int_{B_R} |Du|^2 |x - x_0|^{2-n} dx. \quad (3.53)$$

Next

$$c(\rho) \int_\Omega G\tau_R^2 \sum_{\{\nu=1\}}^N k^\nu dx \leq c(\rho) \left(\int_{B_{2R} - B_R} G^{q'} dx \right)^{\frac{1}{q'}}$$

$$\leq c_1 c(\rho)\left(\int_{B_{2R}-B_R} | x - x_0 |^{(2-n)q'} \, dx \right)^{\frac{1}{q'}}$$

hence

$$c(\rho) \int_\Omega G\tau_R^2 \sum_{\{\nu=1\}}^{N} k^\nu \, dx \leq c(\rho) R^{2-\frac{n}{q}} \tag{3.54}$$

where, of course the constant $c(\rho)$ is generic. We turn to the main term in (3.44)

$$\int_\Omega a_{ij} \frac{\partial X^1}{\partial x_j} \frac{\partial G\tau_R^2}{\partial x_i} dx = I + II$$

with

$$I = 2 \int_\Omega G\tau_R a_{ij} \frac{\partial X^1}{\partial x_j} \frac{\partial \tau_R}{\partial x_i} dx$$

and

$$II = \int_\Omega \tau_R^2 a_{ij} \frac{\partial X^1}{\partial x_j} \frac{\partial G}{\partial x_i} dx.$$

Then, we can write

$$II = \int_\Omega a_{ij} \frac{\partial ((X^1 - X_0^1)\tau_R^2)}{\partial x_j} \frac{\partial G}{\partial x_i} dx - 2 \int_\Omega \tau_R (X^1 - X_0^1) a_{ij} \frac{\partial \tau_R}{\partial x_j} \frac{\partial G}{\partial x_i} dx$$

and from the definition of the Green function, see (3. 49)

$$II \geq -2 \int_\Omega \tau_R (X^1 - X_0^1) a_{ij} \frac{\partial \tau_R}{\partial x_j} \frac{\partial G}{\partial x_i} dx.$$

Making use of (3. 43), and performing easy majorations we obtain

$$I + II \geq -c(\rho)[\int_{(B_{2R}-B_R)\cap\Omega} \frac{|u-c^R|^2}{R^2} G\,dx + \int_{B_{2R}-B_R} |Du|^2\, G\,dx + III]$$

where

$$III = \int_{(B_{2R}-B_R)\cap\Omega} G^{-1}\,|DG|^2 |u-c^R|^2\ \tau_R^2\,dx.$$

Note that

$$\int_{(B_{2R}-B_R)\cap\Omega} \frac{|u-c^R|^2}{R^2} G\,dx \leq c_1 \int_{(B_{2R}-B_R)\cap\Omega} \frac{|u-c^R|^2}{R^2} |x-x_0|^{2-n}\,dx$$
$$\leq c_1 R^{2-n} \int_{(B_{2R}-B_R)\cap\Omega} \frac{|u-c^R|^2}{R^2} dx$$

and using Poincaré's inequality, we obtain

$$\leq CR^{2-n} \int_{B_{2R}-B_R} |Du|^2\,dx \leq C \int_{B_{2R}-B_R} |Du|^2 |x-x_0|^{2-n}\,dx.$$

Therefore we have proven

$$I + II \geq -c(\rho)[\int_{B_{2R}-B_R} |Du|^2 |x-x_0|^{2-n}\,dx + III].$$

To estimate III, one introduces a new cut-off function, defined as follows

$$\chi = 0 \text{ if } |x| \leq \frac{1}{2}$$
$$\chi = \tau \text{ if } |x| \geq 1$$

and χ smooth, $0 \leq \chi \leq \tau$. We set

$$\chi_R = \chi(\frac{x-x_0}{R})$$

and note that

$$\chi_R = \tau_R, \text{ outside } B_R.$$

We take in (3.49)

$$\phi = G^{-\frac{1}{2}} \mid u - c^R \mid^2 \chi_R^2$$

noting that $\phi(x_0) = 0$. We obtain the relation

$$\frac{1}{2}\int aDG.DGG^{-\frac{3}{2}} \mid u - c^R \mid^2 \chi_R^2 \, dx = \int aD(\mid u - c^R \mid^2 \chi_R^2).DGG^{-\frac{1}{2}} \, dx \quad (3.55)$$

Using now the system (2.15), testing with $(u^\nu - c^{\nu,R})G^{\frac{1}{2}}\chi_R^2$, it follows

$$\int aDu^\nu .Du^\nu G^{\frac{1}{2}}\chi_R^2 \, dx + \frac{1}{2}\int aDu^\nu (u^\nu - c^{\nu,R})G^{-\frac{1}{2}}DG\chi_R^2 \, dx +$$

$$2\int aDu^\nu (u^\nu - c^{\nu,R})D\chi_R G^{\frac{1}{2}} \, dx + \int H^\nu (u^\nu - c^{\nu,R})G^{\frac{1}{2}}\chi_R^2 \, dx = 0.$$

Hence

$$\int aD(\mid u - c^R \mid^2 \chi_R^2).DGG^{-\frac{1}{2}} \, dx \le 2\int aD\chi_R .DGG^{-\frac{1}{2}}\chi_R \mid u - c^R \mid^2 -$$

$$-8\int aDu^\nu (u^\nu - c^{\nu,R})D\chi_R G^{\frac{1}{2}} \, dx + 4\int H^\nu (u^\nu - c^{\nu,R})G^{\frac{1}{2}}\chi_R^2 \, dx$$

Using the quadratic growth of H, one checks easily that

$$\int aD(\mid u - c^R \mid^2 \chi_R^2).DGG^{-\frac{1}{2}} \, dx \le C\delta \int \mid DG \mid^2 G^{-\frac{3}{2}} \mid u - c^R \mid^2 \chi_R^2 \, dx +$$

$$+\frac{C}{\delta}\int_{(B_{2R} - B_{\frac{R}{2}})\cap\Omega} \frac{\mid u - c^R \mid^2}{R^2} G^{\frac{1}{2}} \, dx +$$

$$+\int_{B_{2R} - B_{\frac{R}{2}}} \mid Du \mid^2 G^{\frac{1}{2}} \, dx + CR^{1-n+\frac{n}{q}}$$

where δ is aribtrarily small. Combining with (3.55)we obtain

$$\int \mid DG \mid^2 G^{-\frac{3}{2}} \mid u - c^R \mid^2 \chi_R^2 \, dx \le C\int_{B_{2R} - B_{\frac{R}{2}}} \mid Du \mid^2 G^{\frac{1}{2}} \, dx +$$

$$+C\int_{(B_{2R}-B_{\frac{R}{2}})\cap\Omega}\frac{|u-c^R|^2}{R^2}G^{\frac{1}{2}}\,dx+CR^{1-n+\frac{n}{q}}.$$

Finally , we can assert that

$$III\le C\int_{B_{2R}-B_{\frac{R}{2}}}|Du|^2|x-x_0|^{2-n}\,dx+CR^{2-\frac{n}{q}}.$$

Combining results, and changing R by $2R$, the result (3. 51) is obtained. This concludes the proof of Lemma 2. □

PROOF of PROPOSITION 3.3
Proceeding as for Lemma 2, with $\psi=G$ and $c=0$, one obtains

$$\int|Du|^2|x-x_0|^{2-n}\,dx\le C \tag{3.56}$$

We can then use the hole filling technique of Widman (see K.O. WIDMAN [7]) to obtain

$$\int_{B_R}|Du|^2|x-x_0|^{2-n}\,dx\le C_\beta R^\beta,\beta<\beta_0=2-\frac{n}{q} \tag{3.57}$$

and the result (3. 48) follows from the classical result of MORREY [4], with $\delta=\frac{\beta}{2}$. □

4. PROOF OF THEOREM 2.1

4.1 Strong Convergence

From PROPOSITIONS 3.2 and 3.3, we deduce that we can extract a subsequence such that

$$u^{\varepsilon,\nu}\to u^\nu \text{ in } H_0^1(\Omega) \text{ weakly and in } C^0(\bar{\Omega}) \tag{4.58}$$

Note also that $H^{\varepsilon,\nu}(x,u^\varepsilon,Du^\varepsilon)$ remains bounded in $L^1(\Omega)$ and in $H^{-1}(\Omega)$ and thus we can assume that

$$H^{\varepsilon,\nu} \to \lambda^{\nu} \text{ in } H^{-1}(\Omega) \text{ weakly and in } (C^0(\bar{\Omega}))^* \text{ weak star} \tag{4.59}$$

Let us consider functions ϕ^{ν} such that

$$\begin{aligned} &\phi^{\nu} \in C^2(\bar{\Omega}) \\ &\phi^{\nu}|_{\partial\Omega} = 0, D\phi^{\nu}|_{\partial\Omega} = 0 \end{aligned} \tag{4.60}$$

and set

$$\phi^{\varepsilon,\nu} = \phi^{\nu} + D\phi^{\nu}(v^{\varepsilon} - x).$$

Note that

$$D\phi^{\varepsilon,\nu} = (Dv^{\varepsilon})^* D\phi^{\nu} + D^2\phi^{\nu}(v^{\varepsilon} - x).$$

From the assumptions (2.4),(2.5),(2.6), we then deduce

$$\begin{aligned} \phi^{\varepsilon,\nu} &\to \phi^{\nu} \text{ in } H_0^1(\Omega) \text{ weakly} \\ D\phi^{\varepsilon,\nu} - (Dv^{\varepsilon})^* D\phi^{\nu} &\to 0 \text{ in } (L^2(\Omega))^n \text{ weakly} \\ a^{\varepsilon} D\phi^{\varepsilon,\nu} &\to aD\phi^{\nu} \text{ in } (L^2(\Omega))^n \text{ weakly} \\ \text{div} a^{\varepsilon} D\phi^{\varepsilon,\nu} &\to \text{div} aD\phi^{\nu} \text{ in } H^{-1}(\Omega) \text{ strongly} \end{aligned} \tag{4.61}$$

From (2.4), (2.7), we can assert that

$$v^{\varepsilon} \to x \text{ in } C^0(\bar{\Omega})$$

hence also

$$\phi^{\varepsilon,\nu} \to \phi^{\nu} \text{ in } C^0(\bar{\Omega}) \tag{4.62}$$

We then state the Lemma

Lemma 4.1. *We have the property*

$$\alpha \limsup \int \sum_{\nu} | Du^{\varepsilon,\nu} - (Dv^{\varepsilon})^* D\phi^{\nu} |^2 dx \leq$$
$$-\sum_{\nu} \int aD\phi^{\nu}.D(u^{\nu} - \phi^{\nu}) dx - \sum_{\nu} < \lambda^{\nu}, u^{\nu} - \phi^{\nu} > \tag{4.63}$$

PROOF:
Consider (2.15), which we test with $u^{\varepsilon,\nu} - \phi^{\varepsilon,\nu}$. We deduce

$$\int a^{\varepsilon} D(u^{\varepsilon,\nu} - \phi^{\varepsilon,\nu}).D(u^{\varepsilon,\nu} - \phi^{\varepsilon,\nu}) dx +$$
$$- \int \operatorname{div}(a^{\varepsilon} D\phi^{\varepsilon,\nu})(u^{\varepsilon,\nu} - \phi^{\varepsilon,\nu}) dx + \int H^{\varepsilon,\nu}(u^{\varepsilon,\nu} - \phi^{\varepsilon,\nu}) dx = 0.$$

Using (4.61), (4.62) together with (4.59), we obtain

$$\alpha \limsup \int \sum_{\nu} | Du^{\varepsilon,\nu} - D\phi^{\varepsilon,\nu} |^2 dx \leq$$
$$\sum_{\nu} \int \operatorname{div}(aD\phi^{\nu})(u^{\nu} - \phi^{\nu}) dx - \sum_{\nu} < \lambda^{\nu}, u^{\nu} - \phi^{\nu} > .$$

Taking into account the second property (4.61), we obtain (4.63). □

Now, we can assert that (4.63) holds also for $\phi^{\nu} \in H_0^1(\Omega)$, since the second derivative has disappeared from the formula. Taking then $\phi^{\nu} = u^{\nu}$ we deduce

Proposition 4.1. *We have the property*

$$Du^{\varepsilon,\nu} - (Dv^{\varepsilon})^* Du^{\nu} \to 0 \text{ in } (L^2(\Omega))^n \tag{4.64}$$

4.2 Construction of the Limit Hamiltonian

Consider the sequence $H^{\varepsilon,\nu}(x,s,(Dv^{\varepsilon})^*\xi)$. From (2. 8) to ((2. 12), we deduce

$$| H^{\varepsilon}(x,s,(Dv^{\varepsilon})^*\xi) | \leq M + \varpi(|s|)(1 + B^2 |\xi|^2) + \gamma B |\xi| (1 + B|\xi|) \tag{4.65}$$

where

$$M = (\sum_{\nu} (M^{\nu})^2)^{\frac{1}{2}}.$$

We follow the approach of BOCCARDO-MURAT [3]. Let Z be a countable dense subset of $R^{N(n+1)}$, containing 0 and dense subsets of the subspaces $s^{\nu} = 0$. There exists a subsequence such that

$$H^{\varepsilon,\nu}(x,s,(Dv^{\varepsilon})^{*}\xi) \to H^{\nu}(x,s,\xi) \text{ weakly in } L^2(\Omega), \forall s,\xi \in Z \quad (4.66)$$

One checks easily that the properties (2.8) to ((2.14) are satisfied for $H^{\nu}(x,s,\xi)$ on Z, with possibly different constants. These estimates, which imply a uniform continuity in Z for bounded sets, permit to extend the definition of $H^{\nu}(x,s,\xi)$ to any pair s^{ν},ξ^{ν}, and (2. 8) to ((2. 14) are satisfied. Moreover one has the property

$$H^{\varepsilon,\nu}(x,s,(Dv^{\varepsilon})^{*}\xi) \to H^{\nu}(x,s,\xi) \text{ weakly in } L^2(\Omega), \forall s,\xi \qquad (4.67)$$

We then asset the following result

Lemma 4.2. *Let* $\phi^{\nu} \in L^{\infty}(\Omega)$ *and* $\Gamma^{\nu} \in (L^2(\Omega))^n$. *Let* $z^{\varepsilon,\nu}$ *be functions bounded in* $L^{\infty}(\Omega)$, *which converge in* $L^2(\Omega)$ *to functions* z^{ν}. *Then one has the convergence property*

$$\int (H^{\varepsilon,\nu}(x,\phi(x),(Dv^{\varepsilon})^{*}\Gamma(x)) - H^{\nu}(x,\phi(x),\Gamma(x)))z^{\varepsilon,\nu}(x)\,dx \to 0 \qquad (4.68)$$

PROOF:
From (4.67), one can assert that (4.68) holds whenever ϕ^{ν},Γ^{ν} are bounded step functions. We can then consider approximations $\phi^{k,\nu},\Gamma^{k,\nu}$ such that

$$\phi^{k,\nu} \to \phi^{\nu} \text{ a.e. and } \| \phi^{k,\nu} \| \le C$$
$$\Gamma^{k,\nu} \to \Gamma^{\nu} \text{ in } (L^2(\Omega))^n \text{ and a.e.}$$

$\phi^{k,\nu},\Gamma^{k,\nu}$ being step functions. We write

$$\int (H^{\varepsilon}(x,\phi(x),(Dv^{\varepsilon})^{*}\Gamma(x)) - H(x,\phi(x),\Gamma(x)))z^{\varepsilon,\nu}(x)\,dx$$
$$= \int (H^{\varepsilon}(x,\phi(x),(Dv^{\varepsilon})^{*}\Gamma(x)) - H^{\varepsilon}(x,\phi^{k}(x),\Gamma^{k}(x)))z^{\varepsilon,\nu}(x)\,dx$$
$$+ \int (H^{\varepsilon}(x,\phi^{k}(x),\Gamma^{k}(x)) - H(x,\phi^{k}(x),\Gamma^{k}(x)))z^{\varepsilon,\nu}(x)\,dx$$

$$+\int (H(x,\phi^k(x),\Gamma^k(x))-H(x,\phi(x),\Gamma(x)))z^{\varepsilon,\nu}(x)\,dx$$
$$=I+II+III.$$

From uniform estimes, we can check that I, III are bounded by $o(k)$ independent of ε, and $o(k)\to 0$, as $k\to\infty$. Moreover, for fixed k, $II\to 0$ as $\varepsilon\to 0$. The result follows. □

4.3 End of Proof

We can now complete the proof of Theorem 1. From (4.64) and the uniform estimates (2.11), (2.12) it follows that

$$H^{\varepsilon,\nu}(x,u^\varepsilon,Du^\varepsilon)-H^{\varepsilon,\nu}(x,u,(Dv^\varepsilon)^*Du)\to 0 \text{ in } L^1(\Omega) \tag{4.69}$$

Moreover from Lemma 4.2, we have

$$H^{\varepsilon,\nu}(x,u,(Dv^\varepsilon)^*Du)-H^\nu(x,u,Du)\to 0 \text{ in } L^1(\Omega) \text{ weakly} \tag{4.70}$$

Therefore we deduce

$$H^{\varepsilon,\nu}(x,u^\varepsilon,Du^\varepsilon)-H^\nu(x,u,Du)\to 0 \text{ in } L^1(\Omega) \text{ weakly} \tag{4.71}$$

Hence, we have

$$\lambda^\nu = H^\nu(x,u,Du)$$

From 4.64, we have

$$a^\varepsilon Du^{\varepsilon,\nu}-a^\varepsilon(Dv^\varepsilon)^*Du^\nu\to 0 \text{ in } (L^2(\Omega))^n \tag{4.72}$$

hence

$$\operatorname{div} a^\varepsilon Du^{\varepsilon,\nu}-\operatorname{div}(a^\varepsilon(Dv^\varepsilon)^*Du^\nu)\to 0 \text{ in } H^{-1}(\Omega) \text{ weakly} \tag{4.73}$$

But, from the assumption (2.5) it follows

$$\operatorname{div}(a^\varepsilon(Dv^\varepsilon)^*Du^\nu)\to \operatorname{div}(aDu^\nu) \text{ in } H^{-1}(\Omega) \text{ weakly} \tag{4.74}$$

hence finally

$$\mathrm{div} a^{\varepsilon} Du^{\varepsilon,\nu} \to \mathrm{div}(aDu^{\nu}) \text{ in } H^{-1}(\Omega) \text{ weakly} \tag{4.75}$$

The proof has been completed. □

REFERENCES

[1] A. BENSOUSSAN, J. FREHSE, Regularity Results for Nonlinear Elliptic Systems and Applications , Springer-Verlag, Berlin, 2002.

[2] A. BENSOUSSAN, J. FREHSE, Smooth Solutions of Systems of Quasilinear Equations, in volume in memory of J.L. LIONS, dec 2001

[3] L. BOCCARDO, F. MURAT, Homogéneisation de problèmes quasi-linéaires, Atti del Convegno "Studio di problemi limite della analisi funzionale", Bressanone, (1982) , 13-51, Pitagora Editrice, Bologna

[4] C.B. MORREY Jr, Multiple Integrals in the Calculus of Variations , Springer-Verlag, Berlin, 1966.

[5] F. MURAT, L. TARTAR, Calcul des variations et homogénéisation , Eyrolles, coll DER-EDF, Lecture Notes, Paris

[6] L. TARTAR, Cours Peccot, Collège de France

[7] K.O. WIDMAN, Hölder continuity of solutions od elliptic equations, Manuscripta Math 5 (1971), 299- 308

ABOUT THE DUALITY GAP IN VECTOR OPTIMIZATION

G. Bigi[1] and M. Pappalardo[2]
Dept. of Computer Sciences, University of Pisa, Pisa, Italy;[1] Dept. of Applied Mathematics, University of Pisa, Pisa, Italy[2]

Abstract: Since a vector program has not just an optimal value but a set of optimal ones, the analysis of duality gap requires at least the comparison between two sets of vector optimal values. Relying only on a weak duality property, the situations that can occur are analysed in detail and some concepts of duality gap are proposed. Some numerical examples are also provided.

Key words: vector optimization, duality gap

1. INTRODUCTION

Duality is one of the most important topics in optimization both from a theoretical and algorithmic point of view. In scalar optimization, one generally looks for a dual problem in such a way that the difference between the optimal values is non–negative, small and possibly zero. This difference is called duality gap. However, such a definition cannot be applied to vector optimization easily, since a vector program has not just an optimal value but a set of optimal ones. A first attempt to analyse the vector case appeared in [2] but it was based on scalarization techniques and the considered duality gap was between scalar problems. Though a large number of papers dealing with duality for vector optimization have been published, to the best of our knowledge, studies about the duality gap have not been carried out. The first difficulty to overcome is the definition of duality gap itself: the aim of this note is to address possible answers to this question and not to present new duality results. Therefore, we review some of the vector optimization duality

schemes developed in literature and we propose some concepts of duality gap, which match the known results.

2. DUALITY SCHEMES

Throughout all the paper we consider the following vector optimization problem as the primal problem:

$$(P) \quad \min_C f(x) \quad \text{subject to } x \in X_0, -g(x) \in Q$$

where $f : \mathbb{R}^n \to \mathbb{R}^\ell$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are vector–valued functions, $X_0 \subseteq \mathbb{R}^n$ is any subset and $Q \subseteq \mathbb{R}^m$ is a convex cone; let X denote the feasible region, i.e. $X = \{x \in X_0 : -g(x) \in Q\}$. The notation $\min_C$ marks vector minimum: we recall that $\overline{x} \in X$ is a *vector minimum point* of problem (P) if $f(\overline{x})$ is a *minimal element* of $f(X)$ with respect to the partial order induced by the convex, pointed cone $C \subseteq \mathbb{R}^\ell$ with $0 \in C$, i.e. if there is no feasible x such that $f(\overline{x}) - f(x) \in C \setminus \{0\}$. Moreover, the minimal elements of $f(X)$ will be referred to as optimal values of (P) and $\min_C f(X)$ will denote the set of all the optimal values. Analogous definitions can be introduced for maximization problems.

Many dual problems for (P) have been introduced in different ways. The most studied approaches rely on vector–valued Lagrangians, starting from the pioneering paper by Tanino and Sawaragi [8]. The Lagrangian $L(x, \Lambda) := f(x) + \Lambda g(x)$, where Λ is a matrix of multipliers, has been employed in that paper to introduce the following dual problem:

$$(D_L) \quad \max_C d_L(\Lambda) \quad \text{subject to } \Lambda \in Y_L := \{\overline{\Lambda} \in \mathbb{R}^{\ell \times m} : \overline{\Lambda}(Q) \subseteq C\}$$

where $d_L(\Lambda) := \min_C \{L(x, \Lambda) : x \in X_0\}$ is the dual mapping. It is worth stressing that d_L is set–valued; therefore, (D_L) is actually a set–valued optimization problem and it simply borrows the optimality concept of the vector case in the most natural way, that is $\overline{\Lambda} \in Y_L$ is a maximum point of (D_L) if there exists $y \in d_L(\overline{\Lambda})$ such that $y \in \max_C d_L(Y_L)$, where $d_L(Y_L)$ denotes the union of the sets $d_L(\Lambda)$ over all $\Lambda \in Y_L$.

Another Lagrangian type approach relies on the real–valued Lagrangian function $\ell(x, \theta, \lambda) = \theta \cdot f(x) + \lambda \cdot g(x)$. It has originally been proposed by Jahn [5] as a generalization of the duality approach developed by Isermann [4] for linear vector optimization problems:

$$(D_J) \quad \max_C d_J(\theta, \lambda) \quad \text{subject to } (\theta, \lambda) \in C^+ \times Q^*$$

where

$$d_J(\theta,\lambda) := \{v \in \mathbb{R}^\ell : \theta \cdot v \le \inf\{\ell(x,\theta,\lambda) : x \in X_0\}\}.$$

Thus, the dual mapping is set–valued also in this scheme but it does not involve any vector optimality concept. The dual variables θ and λ are bounded to take their values in the strict dual cone of C, i.e.

$$C^+ := \{\theta \in \mathbb{R}^\ell : \theta \cdot c > 0 \ \ \forall c \in C \setminus \{0\}\}$$

and in the dual cone of Q, i.e. $Q^* := \{\lambda \in \mathbb{R}^m : \lambda \cdot q \ge 0 \ \ \forall q \in Q\}$.

Also Wolfe type duality has been widely studied for vector optimization [3,10,11]. In this approach the constraints are just the Kuhn–Tucker necessary conditions for problem (P): supposing for the sake of simplicity that f and g are differentiable functions and X_0 is an open set, Wolfe dual is the following maximization problem:

$$(D_W) \quad \begin{array}{c} \max_C f(u) + (\lambda \cdot g(u))c \\ \text{subject to} \\ \theta \cdot \nabla f(u) + \lambda \cdot \nabla g(u) = 0, \\ u \in \mathbb{R}^n, \theta \in C^+, \lambda \in Q^*, \theta \cdot c = 1 \end{array}$$

where $c \in \operatorname{int} C$ is a fixed vector. The dual mapping is not set–valued but it is worth stressing that unlike the previous schemes this one requires convexity assumptions on f and g even to achieve weak duality results.

Actually, also some other duality schemes have been developed, among which we just recall the Mond–Weir variant of Wolfe duality [10], conjugate duality [9] and the scheme based on monotone nonlinear multipliers [6]. Since we aim to analyse concepts of duality gap for all schemes in a unifed way, we will just consider the following generic form of dual problem:

$$(D) \quad \max_C d(y) \quad \text{subject to } y \in Y$$

where Y is the set of the feasible dual variables and the dual mapping d takes values in $\mathbb{R}^\ell$ and may be set–valued.

3. DUALITY GAP

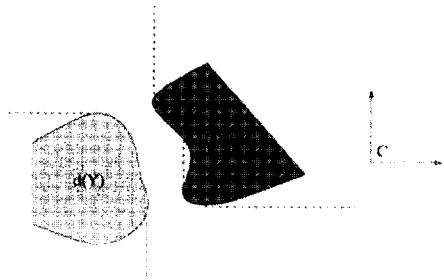
All the duality schemes we recalled in the previous section satisfy a weak duality property, that can be written as

$$d(Y) \cap [f(X) + (C \setminus \{0\})] = \varnothing \tag{1}$$

or equivalently

$$f(X)\cap[d(Y)-(C\setminus\{0\})]=\varnothing \tag{2}$$

and means that no value of the dual problem (D) is greater (with respect to the partial order induced by C) than any value of the primal problem (P) or equivalently that no value of (P) is smaller than any value of (D).



If this weak duality property holds, then it is easy to check that $f(\overline{x})\in d(\overline{y})$ implies that $\overline{x}$ is a minimum point of (P) and $\overline{y}$ is a maximum point of (D) and $f(\overline{x})$ is an optimal value both for (P) and (D) (see [4,5,6,8]). Therefore, whenever the images have nonempty intersection, i.e.

$$d(Y)\cap f(X)\neq\varnothing, \tag{3}$$

the two problems possess at least a common optimal value. It is worth noting that (3) is actually equivalent to the condition

$$\max_C d(Y)\cap\min_C f(X)\neq\varnothing.$$

In order to check whether (3) holds or to give a numerical measure of its failure, we can consider the quantity

$$\Delta:=\inf\{\|v-u\|: v\in d(Y),u\in f(X)\}.$$

In fact, $\Delta=0$ if (3) holds. Unfortunately, the vice versa does not hold, unless suitable compactness assumptions on $f(X)$ and $d(Y)$ are made. Thus, Δ be may be zero even if no common value exists. However, $\Delta=0$ means by definition that the two objective functions achieve values whose difference has arbitrarily small norm; notice that in scalar optimization (i.e. $\ell=1$) this means exactly that the optimal values coincide but in the vector framework it is not so.

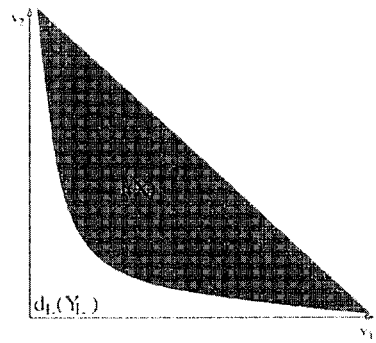
Example 3.1. Consider (P) with $n=2$, $\ell=2$, $m=1$, $X_0=\mathbb{R}^2_+$, $Q=\mathbb{R}_+$, the functions $f(x_1,x_2)=(x_1,x_2)$ and $g(x_1,x_2)=1-x_1x_2$ and $C=\mathbb{R}^2_+$ as the ordering cone. The set of the optimal values of (P) is

$$\min_C f(X)=\{v\in\mathbb{R}^2_+ : v_1v_2=1\}.$$

Let us examine the Lagrangian dual (D_L). Easy calculations show that

$$d_L(\Lambda)=\begin{cases}\{(\Lambda_{11},\Lambda_{21})\} & \text{if } \Lambda_{11}\Lambda_{21}=0\\ \varnothing & \text{if } \Lambda_{11}\Lambda_{21}\neq 0\end{cases}$$

and thus $d_L(Y_L)=\{v\in\mathbb{R}^2_+ : v_1v_2=0\}$. Therefore, we have $\Delta=0$ even though (3) does not hold; this is possible since the images $f(X)$ and $d_L(Y_L)$ are not compact sets.



Generally, the results known in literature as strong duality relations are different from (3). In fact, they state that under suitable assumptions (typically convexity requirements on f and g and Slater constraint qualification) for any minimum point $\overline{x}\in X$ of (P) there exists $\overline{y}\in Y$ such that $f(\overline{x})\in d(\overline{y})$ (see [5,7,8]); that is, not only (3) is satisfied but also the stronger conditon

$$\min_C f(X)\subseteq d(Y) \tag{4}$$

or equivalently

$$\min_C f(X)\subseteq \max_C d(Y)$$

holds. In order to check whether (4) holds or to give a numerical measure of its failure, we can consider the quantity

$$\Delta^P := \sup\{\inf\{\| v-u \| : v \in d(Y)\} : u \in \min_C f(X)\}.$$

Since (4) implies (3), we have $\Delta \le \Delta^P$ and the inequality may be strict.

Example 3.2. Consider (P) with $n=2$, $\ell=2$, $m=1$, $X_0=[0,\pi/2]\times[0,2]$, $Q=\mathbb{R}_+$, the functions $f(x_1,x_2)=(x_2\cos x_1, x_2\sin x_1)$ and $g(x_1,x_2)=1-x_2$ and $C=\mathbb{R}^2_+$ as the ordering cone. The set of the optimal values of (P) is

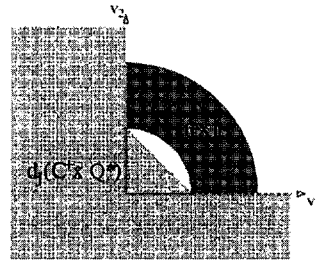
$$\min_C f(X)=\{v\in\mathbb{R}^2_+ : v_1^2+v_2^2=1\}.$$

Let us examine the Lagrangian dual (D_J). Standard optimization techniques provide

$$\inf\{\ell(x,\theta,\lambda) : x\in X_0\}=\begin{cases}\lambda & \text{if } \min\{\theta_1,\theta_2\}\ge\lambda\\ 2\min\{\theta_1,\theta_2\}-\lambda & \text{if } \min\{\theta_1,\theta_2\}\le\lambda.\end{cases} \tag{5}$$

Choosing $\overline{\theta}_1=\overline{\theta}_2=\lambda$, we get $d_J(\overline{\theta},\overline{\lambda})=\{v\in\mathbb{R}^2_+ : v_1+v_2\le 1\}$ and thus $\Delta=0$ since $\{(1,0),(0,1)\}\subseteq d_J(C^+\times Q^*)\cap f(X)$. This inclusion is actually an equality since (5) allows to check easily that

$$d_J(C^+\times Q^*)=(\mathbb{R}^2\setminus\mathbb{R}^2_+)\cup\{v\in\mathbb{R}^2_+ : v_1+v_2\le 1\}$$



which implies $\Delta^P>0$; precisely, we have $\Delta^P=\|(\sqrt{2}/2,\sqrt{2}/2)-(1/2,1/2)\|=$ $=(2-\sqrt{2})/2$. Notice that f is not convex and therefore not all the standard assumptions to achieve the strong duality relation (4) hold.

Obviously, $\Delta^P=0$ if (4) holds while the vice versa does not hold unless $d(Y)$ is a closed set. However, $\Delta^P=0$ means that any optimal value of (P) is the limit of a sequence of values of (D) even when (D) has no optimal values as in the following example.

Example 3.3. Consider (P) with $n=2$, $\ell=2$, $m=1$, $X_0=\mathbb{R}^2$, $Q=\mathbb{R}_+$, the convex functions $f(x_1,x_2)=(x_1,x_2)$ and $g(x_1,x_2)=(x_1+x_2)^2$ and $C=\mathbb{R}^2_+$ as the ordering cone. The set of the optimal values of (P) is

$$\min_C f(X)=\{v\in\mathbb{R}^2 : v_2=-v_1\}.$$

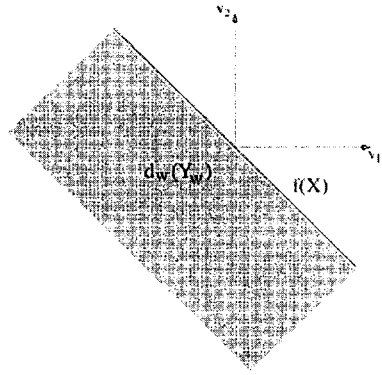
Let us examine the Wolfe dual (D_W), which turns out to be

$$\begin{aligned}&\max_C(u_1+\lambda(u_1+u_2)^2, u_2+\lambda(u_1+u_2)^2)\\ &\qquad\text{subject to}\\ &\theta_1+2\lambda(u_1+u_2)=0,\\ &\theta_2+2\lambda(u_1+u_2)=0,\\ &\theta_1+\theta_2=1, \lambda\geq 0, \theta_1>0, \theta_2>0.\end{aligned}$$

The constraints imply not only $\lambda\neq 0$ and $(u_1+u_2)\neq 0$ but also $\theta_1=\theta_2=1/2$ and $\lambda=-1/4(u_1+u_2)$. Therefore, (D_W) reduces to the maximization problem

$$\begin{aligned}&\max_C(3u_1-u_2, 3u_2-u_1)\\ &\qquad\text{subject to}\\ &u_1+u_2<0.\end{aligned}$$

Relying on a linear transformation of coordinates, it is easy to check that the set of the values of (D_W) is $d_W(Y_W)=\{v\in\mathbb{R}^2 : v_2<v_1\}$.



Therefore, we have $\Delta=\Delta^P=0$ though neither (3) nor (4) holds; this is possible since (P) does not satisty Slater constraint qualification.

Exchanging the roles of the primal and dual problem, it is reasonable to consider also the following duality relation

$$\max_C d(Y) \subseteq f(X). \tag{6}$$

Results which guarantee that this relation holds are known also as converse duality. We can consider the quantity

$$\Delta^D := \sup \{\inf \{\| v-u \| : u \in f(X)\} : v \in \max_C d(Y)\}$$

to check whether or not (6) holds. Obviously, $\Delta^D = 0$ if (6) holds while the vice versa does not hold unless $f(X)$ is a closed set. Moreover, $\Delta \le \Delta^D$ while no relationship between Δ^P and Δ^D readily follows from the definitions.

Actually, not many converse duality results have been proved for vector optimization (see [4] for the linear case and [5] for the duality scheme involving (P) and (D_J)); furthermore, as far as we know, no converse duality results have been obtained yet without actually supposing C to be a closed cone.

Example 3.4. Consider (P) with $n=2$, $\ell=2$, $m=1$, $X_0=[0,2]\times[0,2]$, $Q=\mathbb{R}_+$, the linear functions $f(x_1,x_2)=(x_1,x_2)$ and $g(x_1,x_2)=1-x_1$ and $C=\operatorname{int}\mathbb{R}^2_+ \cup \{0\}$ as the ordering cone. The set of the optimal values of (P) is

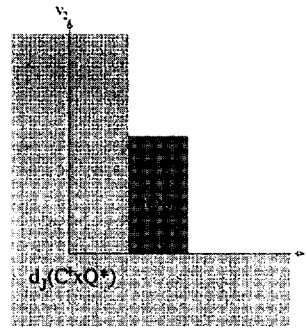
$$\min_C f(X) = ([1,2]\times\{0\}) \cup (\{1\}\times[0,2]).$$

Let us examine the Lagrangian dual (D_J). It is easy to prove

$$\inf\{\ell(x,\theta,\lambda) : x \in X_0\} = \begin{cases} \lambda & \text{if } \theta_1 \ge \lambda \\ 2\theta_1 - \lambda & \text{if } \theta_1 \le \lambda. \end{cases} \tag{7}$$

Notice that $C^+ = \mathbb{R}^2_+ \setminus \{0\}$; therefore, any dual variables such that $\theta_1 = \lambda = 0$ yield $d_J(\theta,\lambda) = \mathbb{R}\times\mathbb{R}_-$ while those variables such that $\theta_1 = \lambda > 0$ and $\theta_2 = 0$ yield $d_J(\theta,\lambda) = (-\infty,1]\times\mathbb{R}$. Actually, (7) allows to check that

$$d_J(C^+ \times Q^*) = \{v \in \mathbb{R}^2 : y_1 \le 1 \text{ or } y_2 \le 0\}.$$



Therefore, we have $\Delta = \Delta^P = 0$ but $\Delta^D = +\infty$. Though (P) is a linear vector optimization problem, no converse duality relation holds: the only assumption of the converse duality theorems presented in [4,5] which is not fulfilled is the closedness of C. In fact, just considering the problems (P) and (D_J) of this eaxmple with $C = \mathbb{R}^2_+$, we get $\Delta = \Delta^P = \Delta^D = 0$ in accordance with [4,5].

Remark 3.1. Actually, many duality results have originally been developed through scalarization techniques, considering just proper minima and proper maxima. It is worth stressing that (3) does not guarantee that the common values are properly optimal, since the definition of proper optimality is not related to any fixed partial order. Obviously, concepts of duality gaps could be introduced also in this framework but relying only on the relationships between the set of proper minima of (P) and proper maxima of (D).

4. OPEN PROBLEMS

The quantities Δ, Δ^P and Δ^D are three concepts of duality gap, which seem adequate for vector optimization: they have been introduced in accordance with the known duality results as a numerical measure to test whether or not they hold. Furthermore, notice that in the case of scalar optimization they all collapse to the well–known concept of duality gap, since we have $\Delta = \Delta^P = \Delta^D$.

We did not prove any new result since this was not the aim of this paper; however, we want to address some research directions on this topic, which we believe to be interesting. Is there any relationship between Δ^P and Δ^D? Are there further relationships between Δ, Δ^P and Δ^D? For instance, if $\Delta > 0$ with Δ^P and Δ^D being both finite, is it true that they are all equal? How can these quantities be estimated, relying for instance on the lack of convexity of a function (see the results in [1] for scalar optimization) and/or on other numerical measures of the failure of those conditions, which

guarantee a zero duality gap? Are there any relationships between these concepts of duality gap and the duality gaps achieved considering scalar optimization reformulations of the original vector problem or considering scalarization methods such as the weighting and the ε –constraint ones?

REFERENCES

[1] Aubin, J.P. and Ekeland, I. (1976), "Estimates of the Duality Gap in Nonconvex Optimization", *Mathematics of Operations Research*, Vol. 1, pp. 225-245.

[2] Di Guglielmo, F. (1977), "Nonconvex Duality in Multiobjective Optimization", *Mathematics of Operations Research*, Vol. 2, pp. 285-291.

[3] Egudo, R.R. (1989), "Efficiency and Generalized Convex Duality for Multiobjective Programs", *Journal of Mathematical Analysis and Applications*, Vol. 138, pp. 84-94.

[4] Isermann, H. (1978), "On Some Relations Between a Dual Pair of Multiple Objective Linear Programs", *Zeitschrift für Operations Research*, Vol. 22, pp. 33-41.

[5] Jahn, J. (1983), "Duality in Vector Optimization", *Mathematical Programming*, Vol. 25, pp. 343-353.

[6] Luc, D.T. and Jahn J. (1991), "Axiomatic Approach to Duality in Optimization", *Numerical Functional Analysis and Optimization*, Vol. 13, pp. 305-326.

[7] Sawaragi, Y., Nakayama, H. and Tanino, T. (1985), *Theory of Multiobjective Optimization*, Academic Press.

[8] Tanino, T. Sawaragi, Y. (1979), "Duality Theory in Multiobjective Programming", *Journal of Optimization Theory and Applications*, Vol. 27, pp. 509-529.

[9] Tanino, T. and Sawaragi, Y. (1980), "Conjugate Maps and Duality in Multiobjective Programming", *Journal of Optimization Theory and Applications*, Vol. 31, pp. 473-499.

[10] Weir, T. and Mond, B. (1988), "Pre-invex Functions in Multiple Objective Optimization", *Journal of Mathematical Analysis and Applications*, Vol. 136, pp. 29-38.

[11] Weir, T., Mond, B. and Craven, B.D. (1987), "Weak Minimization and Duality", *Numerical Functional Analysis and Optimization*, Vol. 9, pp. 181-192.

SEPARATION OF CONVEX CONES AND EXTREMAL PROBLEMS

V.Boltyanski
CIMAT, Guanajuato, Mexico

Abstract: In 1958 the author proved the *Maximum Principle* [2]. B. Pshenichni wrote that the proof was sensational, using topology to obtain a result of variational calculus. Later the author worked out the *Tent Method* [3] as a general way to solve extremal problems. In fact, main ideas of the Method were contained in [2]. We give here a short survey of the Tent Method and the idea of the proof of the Maximum Principle. *AMS* 1991 *Math. Subject Classification.* Primary 15A15; 52A20; Secondary 15A18; 52B12.

Key words: optimization, variational calculus, maximum principle.

1. CLASSICAL CALCULUS OF VARIATIONS

We formulate problems of the classical Calculus of Variations in terms of controlled objects. Note that the connection between variational problems and controlled objects was discovered by Graves [11] (see also the survey [12]).

Consider the controlled object

$$\dot{x} = f(x,u), \qquad x = (x^1,...,x^n)^T \in \mathbb{R}^n, \quad u = (u^1,...,u^r)^T \in U, \tag{1}$$

where the function $f(x,u) = (f^1(x,u),...,f^n(x,u))^T$ is smooth, and $U \subset \mathbb{R}^r$ is a given *resource set.* Every piecewise continuous function $u(t), 0 \le t \le t_1$, with values in U, is an *admissible control.* We always assume that

$u(\tau+0)=u(\tau)$ for $\tau<t_1$ and, moreover, $u(t_1-0)=u(t_1)$. For every *initial state* $x_0\in\mathbb{R}^n$ there is a unique *trajectory* $x(t), 0\le t\le t_1$, with $x(0)=x_0$ that corresponds to the control. The pair $(u(t),x(t))$, $0\le t\le t_1$, will be denoted as an admissible process for the controlled object.

The *Lagrange optimization problem* requires to find an admissible control such that the corresponding trajectory starting from x_0 satisfies the terminal inclusion $x(t_1)\in\Omega_1$ and minimizes the integral functional

$$J=\int_0^{t_1} f^0(x(t),u(t))\,dt.$$

Here $\Omega_1\in\mathbb{R}^n$ is a given *terminal set* and $f^0(x,u)$ is a given positive integrable function. The terminal time t_1 is not fixed in advance.

If $f^0(x,u)\equiv 1$, then $J=t_1$, and we obtain the *time-optimal control problem*: to find an admissible control that transfers x_0 to Ω_1 in the shortest time.

The *Mayer problem* requires to find an admissible control minimizing the value $g(x(t_1))$ of a given smooth function $g(x)$ at the terminal point $x(t_1)\in\Omega_1$.

Finally, the *Bolza problem* is a combination of Lagrange's and Mayer's ones.

The four problems are *equivalent*, i.e., every one of them can be reduced to another one with a suitable change of variables (see the nice monograph [1]).

There are two main classical necessary conditions of optimality: the *Lagrange Multiplier Rule* and the *Weierstrass Theorem*.

In the middle of XX century M.Hestenes deduced the *Maximum Principle* from the Weierstrass theorem. We formulate the Principle for the time-optimization problem, assuming that the terminal set Ω_1 is a smooth manifold. Let us introduce an auxiliary vector $\psi=(\psi_1,\dots,\psi_n)$, the *Hamiltonian*

$$H(\psi,x,u)=\langle\psi,f(x,u)\rangle=\sum_{i=1}^{n}\psi_i f^i(x,u), \tag{2}$$

and the *conjugate system* corresponding to the process $x(t),u(t), 0\le t\le t_1$:

$$\dot\psi=-\frac{\partial H}{\partial x},\quad \text{i.e.,}\quad \dot\psi_j=-\frac{\partial H(\psi,x(t),u(t))}{\partial x^j},\quad j=1,\dots,n. \tag{3}$$

Hestenes' Maximum Principle affirms that, if the process is time-optimal, then there is a nontrivial solution $\psi(t)$ of the conjugate system, such that

(*i*) for $0 \le t \le t_1$ the following *maximum condition* holds:

$$H(\psi(t),x(t),u(t)) = \max_{u \in U} H(\psi(t),x(t),u); \tag{4}$$

(*ii*) $H(\psi(t),x(t),u(t)) = \text{const} \ge 0$;
(*iii*) $\psi(t_1)$ is orthogonal to M_1 at the terminal point $x(t_1)$.

Hestenes obtained this result in the framework of the *classical Calculus of Variations*, i.e., $f(x,u)$ is smooth, U is an *open* set in $\mathbb{R}^r$, and the optimization is considered in the *local sense*, i.e., $u(t)$ is optimal among the controls u satisfying $\| u(t) - u \| < \varepsilon$ for all t (for more details, see [12]).

2. NON-CLASSICAL CALCULUS OF VARIATIONS

In 1953 Feldbaum for the first time solved a *non-classical* time-optimization problem. His ideology came from the *theory of remote control*. The main problem of the theory was to ensure stability of the closed loop system (controlled plant + regulator). In the linear case stability holds if and only if the roots of the characteristic equation (poles) belong to the left complex half-plane. The larger of the absolute values of the real part of the poles shows the faster convergence of the system state to the origin. That was the initial understanding of "time-optimality". Then the researchers paid attention to "bang-bang" control lows. The fundamental difference from the linear control low is that the state $x = 0$ can be reached in a finite time. How to reach the origin *in the shortest time* is the time-optimality problem.

The first statement of the problem and some initial results in this direction are connected with the name of A. Feldbaum, who is undoubtedly a pioneer of the mathematical theory of optimal control. In [7] he investigated the system

$$\dot{x}^1 = x^2, \dot{x}^2 = u, \qquad (x^1, x^2)^T \in \mathbb{R}^2, \quad -1 \le u \le 1, \tag{5}$$

assuming that the terminal set Ω_1 consists of the only point $(0,0)^T \in \mathbb{R}^2$. The problem is *non-classical*, because the resource set $-1 \le u \le 1$ is *closed* (not open) in $\mathbb{R}^1$. Feldbaum often said in his talks that for engineering problems it is important to consider variational problems with *closed* resource sets.

Feldbaum proved that every time-optimal control for (5) takes only the values $u = \pm 1$ and has no more than one switching, i.e., no more than two intervals of constancy. Fig. 1 shows the synthesis of optimal trajectories.

Generalizing this theorem, Feldbaum established his *n-Interval Theorem.* In his talks and articles Feldbaum formulated the *synthesis problem* [8].

Being based on Feldbaum's ideology on *closed* resource sets and on some linear examples [5], L. Pontryagin conjectured that for *closed* resource set the Maximum Principle is a *local, sufficient condition* for the time-optimality:

PONTRYAGIN's MAXIMUM PRINCIPLE. *Let* $u(t), x(t), 0 \le t \le t_1$, *be a process of the controlled object* (1) *with a closed resource set* $U \subset \mathbb{R}^r$ *and a fixed end-condition* $x(t_1) = x_1$. *Assume that there is a nontrivial solution* $\psi(t)$ *of the conjugate system* (3) *such that* $H \ge 0$ *and that along the process the maximum condition* (4) *holds. Then the process is time-optimal* (*in local sense*).

This conjecture (though later proved to be, in general, incorrect) was the only but the most important contribution of L.Pontryagin in the development of the Maximum Principle. His hypothesis was very essential, since it signified the passing to the *non-classical Calculus of Variations* that ignores the openness of the resource set. Note that Pontryagin formulated his hypothesis as a *sufficient* condition under the influence of *Legendre's sufficient condition.*

In 1957 Gamkrelidze [9] proved that, for linear controlled objects with convex polyhedral resource set (under a "general position condition"), the Maximum Principle is a *necessary and sufficient condition* of time-optimality. And in 1958 the following theorem was proved [2] which affirms that the Maximum Principle is a *global, necessary condition* of time-optimality (in the non-linear case this condition is *not sufficient*, contradicting Pontryagin's hypothesis):

Maximum principle. *Let* $u(t), x(t), 0 \le t \le t_1$, *be an admissible process of the controlled object* (1) *where the resource set* U *is a Hausdorff topological space. The right-hand side is assumed to be continuous with respect to* x,u *and smooth with respect to* x. *For time-optimality of the process it is necessary that there exists a nontrivial solution* $\psi(t)$ *of the conjugate system* (3) *such that along the process* $H \ge 0$ *and the maximum condition* (4) *is satisfied.*

Today there are several dozens of different versions of the Maximum Principle (see [13] and Chapter I in [6]). That group of results is the kernel of the modern non-classical Calculus of Variations. A non-classical *sufficient* condition of optimality (as the union of the Maximum Principle, the Dynamic Programming in a revised form, and Feldbaum's idea of synthesis) was recently given in [4].

3. TENT METHOD

The remaining part of the article is, in some sense, a continuation of the historical article [10]. Indeed, in [10] it is shown, how L.Pontryagin and R.Gamkrelidze came to the statement of the Maximum Principle, using arguments close to the classical Lagrange's Multiplying Rule and Legendre's sufficient condition. As to the proof of the Maximum Principle for non-linear case, in [10] is written that "there was no real progress until Boltyanski introduced the needle variation of the control". But the needle-shaped variations don't work themselves; they are used in some geometrical and topological environments. Here we give a description of these environments.

Consider the classical Lagrange *conditional extremal problem*: to find the minimum of a function $g(x), x \in \mathbb{R}^n$, under the constraints $f_i(x) = 0$ for $i = 1,...,s$, i.e., to find the minimum of g on the set $\Sigma = \Omega_1 \cap ... \cap \Omega_s$ where $\Omega_i = \{x : f_i(x) = 0\}$.

Theorem 1. *A point* $x_1 \in \Sigma$ *is a minimizer of the function g on* Σ *if and only if* $\Omega_0 \cap \Omega_1 \cap ... \cap \Omega_s = \{x_1\}$ *where* $\Omega_0 = \{x : g(x) < g(x_1)\} \cup \{x_1\}$.

Theorem 1 leads us the following general problem:

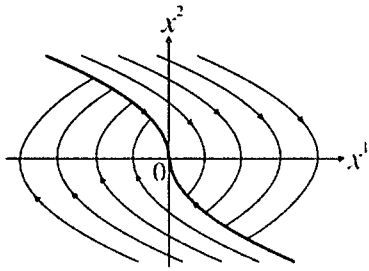


Fig. 1

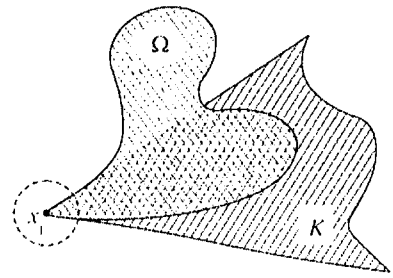


Fig. 2

Abstract intersection problem. *There are sets* $\Omega_0, \Omega_1, ..., \Omega_s$ *in* R^n *with a common point* x_1. *Find a condition under which the intersection* $\Omega_0 \cap \Omega_1 \cap ... \cap \Omega_s$ *consists only of the point* x_1.

This problem includes a wide category of extremal problems. The *Tent Method* is a tool to solve the problem. The idea is to replace each set $\Omega_i, i = 0,1,...,s,$ by its "linear approximation" (*tent* as we say in the sequel, see Fig. 2) in order to pass from $\Omega_0 \cap \Omega_1 \cap ... \cap \Omega_s = \{x_1\}$ to a simpler condition on tents. For example, if $\Omega \subset \mathbb{R}^n$ is a smooth manifold and $x_1 \in \Omega$, then the *tangential plane* of Ω at x_1 is a tent of Ω at x_1. We restrict

ourselves to this intuitive description (for exact definition of tents and relative topics see [3]).

Definition 1. Closed, convex cones $K_0, K_1, ..., K_s \subset \mathbb{R}^n$ with common apex x_1 are said to be *separable* if there exists a hyperplane $\Gamma \subset \mathbb{R}^n$ passing through x_1 that separates one of the cones from the intersection of others, i.e., for an index $i \in \{0,1,...,s\}$ the cone K_i is situated in Π_1 and the intersection of other cones is situated in Π_2 where Π_1 and Π_2 are two closed half-spaces defined by Γ (see Fig. 3 for $i = 0$).

Definition 2. Let $K \subset \mathbb{R}^n$ be a closed cone with apex a. A vector $y \in \mathbb{R}^n$ is said to be a *dual vector of* K if $\langle y, x - a \rangle \le 0$ for all $x \in K$.

The following two theorems [3] form a kernel of the Tent Method.

Theorem 2. *For separability of convex cones* $K_0, K_1, ..., K_s$ *in* $\mathbb{R}^n$ *it is necessary and sufficient that there exist dual vectors* $a_0, a_1, ..., a_s$ *of the cones, at least one of which is nonzero, such that* $a_0 + a_1 + ... + a_s = 0$.

Theorem 3. *Let* $\Omega_0, \Omega_1, ..., \Omega_s$ *be sets in* $\mathbb{R}^n$ *with a common point* x_0, *and* $K_0, K_1, ..., K_s$ *be tents of the sets at the point* x_0. *Assume that at least one of the tents is distinct from a plane. If* $K_0, K_1, ..., K_s$ *are not separable, then there exists a point* $x' \in \Omega_0 \cap \Omega_1 \cap ... \cap \Omega_s$ *distinct from* x_0. *In other words, separability is a necessary condition for* $\Omega_0 \cap \Omega_1 \cap ... \cap \Omega_s = \{x_1\}$.

We show how the Tent Method works in the above Lagrange problem. By Theorem 1, the equality $\Omega_0 \cap \Omega_1 \cap ... \cap \Omega_s = \{x_1\}$ is a necessary (and sufficient) condition that g takes its minimal value at x_1. Remark that the half-space $K_0 = \{x : \langle \operatorname{grad} g(x_1), x - x_1 \rangle \le 0\}$ is the tent of Ω_0 at the point x_1, and this tent is distinct from its affine hall in $\mathbb{R}^n$. Theorems 2 and 3 imply that the tents $K_0, K_1, ..., K_s$ are separable, i.e., there are dual vectors $a_0, a_1, ..., a_s$ not all equal to 0 with $a_0 + a_1 + ... + a_s = 0$. Here $K_1, ..., K_s$ are the tangential hyperplanes of the manifolds $\Omega_1, ..., \Omega_s$, i.e., $a_i = \lambda_i \operatorname{grad} f_i(x_1), i = 1, ..., s$. By definition of K_0, we have $a_0 = \lambda_0 \operatorname{grad} g(x_1)$ with $\lambda_0 \ge 0$. Thus

$$\lambda_0 \operatorname{grad} g(x_1) + \lambda_1 \operatorname{grad} f_1(x_1) + ... + \lambda_s \operatorname{grad} f_s(x_1) = 0. \qquad (6)$$

Supposing that the vectors $\operatorname{grad} f_i(x_1), i = 1, ..., s,$ are linearly independent, we obtain $\lambda_0 \ne 0$. By homogeneity we may suppose $\lambda_0 = 1$, and (6) gives us Lagrange's necessary condition of extremum.

4. A SHORT PROOF OF THE MAXIMUM PRINCIPLE

In conclusion we outline the proof of the Maximum Principle for Mayer's optimization problem defined in section (1) (for more details, see chapter I in [6]):

Maximum principle. *Let* $u(t),x(t),0\le t\le t_1$, *be an admissible process with* $x(0)=x_0$. *If the process solves the Mayer optimization problem, then there exists a solution* $\psi(t)$ *of conjugate system* (3) *such that* $x(t),u(t),\psi(t)$ *satisfy maximum condition and the transversality condition:* $H(\psi(t_1),x(t_1),u(t_1))=0$ *and there is a number* $\lambda\ge 0$ *such that* $\psi(t_1)+\lambda\,\mathrm{grad}\,g(x(t_1))\perp\Omega_1$ *at the point* $x(t_1)$, *the vector* $\psi(t_1)$ *being distinct from* 0 *if* $\lambda=0$.

To prove this theorem, denote by Ω_2 the *controllability region*, i.e., the set of all points which can be reached, starting from the initial point x_0. Then the problem is to minimize $g(x)$ on $\Omega_1\cap\Omega_2$. Denoting by Ω_0 the set as in the above Lagrange problem, we again have to solve the Abstract intersection problem: $\Omega_0\cap\Omega_1\cap\Omega_2=\{x_1\}$.

First we describe a tent of Ω_2 at the point $x_1=x(t_1)$ [2]. Let $\tau<t_1$ and $u\in U$. Consider the solution $\xi(t)$ of the *variational system of equations*

$$\dot{\xi}^k=\sum_{i=1}^{n}\frac{\partial f^k(x(t),u(t))}{\partial x^i}\xi^i,\ \tau\le t\le t_1,$$

with the initial condition $\xi(\tau)=f(x(\tau),u)-f(x(\tau),u(\tau))$.

Then $\Delta(\tau,u)=\xi(t_1)$ is said to be the *deviation vector* corresponding to τ and u. By Q denote the closed convex cone generated by all deviation vectors and by K the vector sum of Q and the line through the origin that is parallel to the vector $f(x(t_1),u(t_1))$. Then $K(P)=x_1+K$ is a tent of Ω_2 at x_1 for the considered admissible process $P=\{u(t),x(t),0\le t\le t_1\}$.

Indeed, consider the process $u_\varepsilon(t),x_\varepsilon(t),0\le t\le t_1$, with $x_\varepsilon(0)=x_0$ where $u_\varepsilon(t)$ is the following *needle-shaped variation* [2] of the control $u(t)$:

$$u_\varepsilon(t)=\begin{cases}u(t) & \text{for } t<\tau,\\ u & \text{for } \tau\le t<\tau+\varepsilon,\\ u(t) & \text{for } t\ge\tau+\varepsilon.\end{cases}$$

Then $x_\varepsilon(t_1)=x(t_1)+\varepsilon\Delta(\tau,u)+o(\varepsilon)$ where $\Delta(\tau,u)$ is the deviation vector as above, (see Fig.4). Since $x_\varepsilon(t_1)\in\Omega_2$, the deviation vector $\Delta(\tau,u)$ is a *tangential vector* of Ω_2. Moreover, the sum of deviation vectors with

positive coefficients also is a tangential vector of Ω_2 [2]. This means that $K(P)$ is a tent of Ω_2 at x_1.

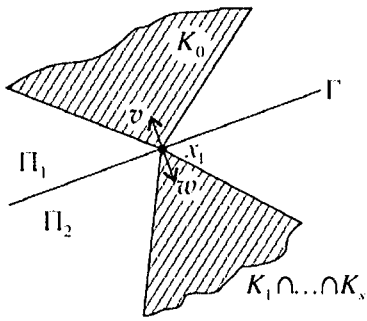


Fig. 3

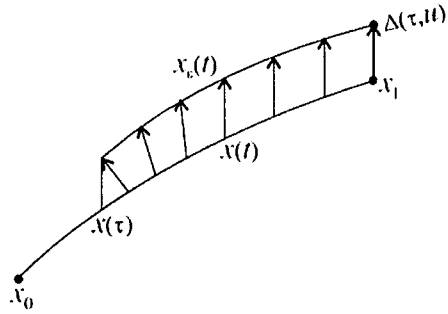


Fig. 4

According to Theorem 1, we conclude that at optimum the final state $x_1 = x(t_1)$ satisfies the condition $\Omega_0 \cap \Omega_1 \cap \Omega_2 = \{x_1\}$. Since Ω_1 is a smooth manifold, its tangential plane K_1 at x_1 is its tent at x_1. Furthermore, the cone K_2 constructed above is a tent of Ω_2 at the point x_1. Finally, the half-space $K_0 = \{x : \langle \operatorname{grad} g(x_1), x - x_0 \rangle \le 0\}$ is the tent of Ω_0 at the point x_1, and hence every its dual vector a_0 has the form $a_0 = \lambda \operatorname{grad} g(x_1)$ where $\lambda \ge 0$.

By Theorem 3, there exist dual vectors a_1, a_2 and a number $\lambda \ge 0$ such that

$$\lambda \operatorname{grad} g(x_1) + a_1 + a_2 = 0 \tag{7}$$

where at least one of the vectors $\lambda \operatorname{grad} g(x_1), a_1, a_2$ is distinct from zero. Since $a_1 \perp \Omega_1$ at the point x_1, the necessary condition (7) can be formulated in the following form: there exist a number $\lambda \ge 0$ and a dual vector a_2 of the cone K_2 such that $\lambda \operatorname{grad} g(x_1) + a_2 \perp \Omega_1$ at the point x_1 and $a_2 \ne 0$ when $\lambda = 0$.

Denote by $\psi(t)$ the solution of the conjugate system with $\psi(t_1) = a_2$. Furthermore, consider the solution of the variational system with the initial condition $\xi(\tau) = f(x(\tau), u) - f(x(\tau), u(\tau))$. The scalar product $\langle \psi(t), \xi(t) \rangle$ keeps a constant value for $\tau \le t \le t_1$. Consequently

$$\langle \psi(\tau), \xi(\tau) \rangle = \langle \psi(t_1), \xi(t_1) \rangle = \langle a_2, \xi(t_1) \rangle \le 0.$$

Thus $\langle \psi(\tau), \xi(\tau) \rangle = \langle \psi(\tau), f(x(\tau), u) - f(x(\tau), u(\tau)) \rangle \le 0$. In other words, $H(\psi(\tau), x(\tau), u) \le H(\psi(\tau), x(\tau), u(\tau))$, i.e., the maximum condition holds.

It remains to establish the transversality condition. Since

$$x_1 + f(x(t_1),u(t_1)) \in K(P) \quad \text{and} \quad x_1 - f(x(t_1),u(t_1)) \in K(P),$$

we have $\langle a_2, f(x(t_1),u(t_1))\rangle = 0$, i.e., $H(\psi(t_1),x(t_1),u(t_1)) = 0$. Furthermore, there is a number $\lambda \geq 0$ such that $\lambda \operatorname{grad} g(x_1) + a_2 \perp \Omega_1$ at the point $x(t_1)$, where $\lambda \neq 0$ if the vector $a_2 = \psi(t_1)$ vanishes.

REFERENCES

[1] G.A. Bliss. "Lectures on the Calculus of Variations". Chicago University Press. Chicago-London-Toronto, 1946.

[2] V.G. Boltyanski. "The maximum principle in the theory of optimal processes" (in Russian). *Doklady Akad. Nauk SSSR* 119 (1958), no. 6, 1070-1073.

[3] V.Boltyanski: "The tent method in the theory of extremal problems" (In Russian). *Uspehi Mat. Nauk* 30 (1975), 3-65.

[4] V.G. Boltyanski. "Sufficient Conditions for Lagrange", Mayer, and Bolza Optimization Problems. *Mathematical Problems in Engineering* 7 (2001), 177-203.

[5] V.G. Boltyanski, R.V. Gamkrelidze, and L.S. Pontryagin. "On the theory of optimal processes" (in Russian). *Doklady Akad. Nauk SSSR* 110 (1956), no. 1, 7-10.

[6] V. Boltyanski, H. Martini, and V. Soltan. "Geometric Methods and Optimization Problems". Kluwer Academic Publishers. Dordrecht - Boston - London, 1999. viii + 429 pp.

[7] A.A. Feldbaum. "Optimal processes in systems of automatic control" (in Russian). *Avtomatika i Telemehanika* 14 (1953), no. 6, 712-728.

[8] A.A. Feldbaum. "On synthesis of optimal systems with the help of phase space" (in Russian). *Avtomatika i Telemehanika* 16 (1955), no. 2, 129-149.

[9] R.V. Gamkrelidze. "On the theory of optimal processes in linear systems" (in Russian). *Doklady AN SSSR* 116 (1957), no. 1, 9-11.

[10] R.V. Gamkrelidze. "Discovery of the Maximum Principle". *Journal of Dynamical and Control Systems* 5, no.4, 1999, 437 - 451.

[11] L.M. Graves. "A transformation of the problem of Lagrange in the Calculus of Variations". *Transactions of the American Math. Soc.* 35 (1933), 675-682.

[12] E. McShane. "The calculus of variations from the beginning through optimal control theory". *SIAM Journal of Control and Optimization,* 27(5),1989, 916-939.

[13] Pontryagin L., Boltyanski V., Gamkrelidze R., and Mishchenko E. Selected works. Vol 4. "The mathematical theory of optimal processes". Classics of Soviet Mathematics. Gordon & Breach Science Publications, New York, 1986. xxiv + 360 pp.

INFINITELY MANY SOLUTIONS FOR THE DIRICHLET PROBLEM VIA A VARIATIONAL PRINCIPLE OF RICCERI

F.Cammaroto*, A.Chinnì and B. Di Bella
Dept. of Mathematics, University of Messina, Sant'Agata, Messina, Italy

Abstract: Using a recent variational principle of B. Ricceri, we present some results of existence of infinitely many solutions for the Dirichlet problem involving the p-Laplacian.

1. INTRODUCTION

The aim of this note is to investigate the following autonomous Dirichlet problem

$$\begin{cases} -\Delta_p u = f(u) & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases} \qquad (D_{n,p})$$

where Ω is a bounded open subset of the euclidean space $(\mathbb{R}^n, |\cdot|)$ with boundary of class C^1, $p > n$, $\Delta_p u = \operatorname{div}(|\nabla u|^{p-2}\nabla u)$ and $f : \mathbb{R} \to \mathbb{R}$ is a continuous function having a suitable oscillating behaviour. Let us recall that a weak solution of $(D_{n,p})$ is any $u \in W_0^{1,p}(\Omega)$ such that

* Corresponding author. Because of surprising coicidence of names within the same Department, we have to point out the author was born on August 4, 1968.

$$\int_\Omega |\nabla u(x)|^{p-2}\nabla u(x)\nabla v(x)\,dx - \int_\Omega f(u(x))v(x)\,dx = 0$$

for each $v \in W_0^{1,p}(\Omega)$.

The existence of infinitely many solutions of the problem $(D_{n,p})$ has been studied extensively. Many results have been obtained usually under sublinearity or superlinearity conditions at 0 and at $+\infty$ of function f (see, for instance, [3]). More rarely, multiplicity of solutions has been investigated when f has an oscillating behaviour, we refer to [4], [5] and [7].

In our results, we make use of a recent general variational principle obtained by B. Ricceri in [6]. The following result is a direct consequence of Theorem 2.5 of [6].

Theorem 1.1 *Let X be a reflexive real Banach space, and let Φ, $\Psi : X \to \mathbb{R}$ be two sequentially weakly lower semicontinuous and Gâteaux differentiable functionals. Assume also that Ψ is (strongly) continuous and satisfies $\lim_{\|x\|\to+\infty} \Psi(x) = +\infty$. For each $r > \inf_X \Psi$, put*

$$\varphi(r) = \inf_{x\in\Psi^{-1}(]-\infty,r[)} \frac{\Phi(x) - \inf_{\overline{(\Psi^{-1}(]-\infty,r[))_w}} \Phi}{r - \Psi(x)},$$

where $\overline{(\Psi^{-1}(]-\infty,r[))_w}$ is the closure of $\Psi^{-1}(]-\infty,r[)$ in the weak topology. Fixed $\lambda \in \mathbb{R}$, then

(a) *if $\{r_n\}_{n\in\mathbb{N}}$ is a real sequence with $\lim_{n\to\infty} r_n = +\infty$ such that $\varphi(r_n) < \lambda$, for each $n \in \mathbb{N}$, the following alternative holds: either $\Phi + \lambda\Psi$ has a global minimum, or there exists a sequence $\{x_n\}$ of critical points of $\Phi + \lambda\Psi$ such that $\lim_{n\to\infty} \Psi(x_n) = +\infty$.*

(b) *if $\{s_n\}_{n\in\mathbb{N}}$ is a real sequence with $\lim_{n\to\infty} s_n = (\inf_X \Psi)^+$ such that $\varphi(s_n) < \lambda$, for each $n \in \mathbb{N}$, the following alternative holds: either there exists a global minimum of Ψ which is a local minimum of $\Phi + \lambda\Psi$,* or there exists a sequence $\{x_n\}$ *of pairwise distinct critical points of $\Phi + \lambda\Psi$, with $\lim_{n\to\infty} \Psi(x_n) = \inf_X \Psi$, which weakly converges to a global minimum of Ψ.*

Throughout the sequel, $f : \mathbb{R} \to \mathbb{R}$ is a continuous function such that $f(x) = 0$ for each $x \in]-\infty, 0]$ and $F : \mathbb{R} \to \mathbb{R}$ is the function defined by setting

$$F(\xi)=\int_0^{\xi} f(t)dt$$

for each $\xi \in \mathbb{R}$.

We shall consider the Sobolev space $W_0^{1,p}(\Omega)$ endowed with the norm

$$\|u\| := \left(\int_\Omega |\nabla u(x)|^p \, dx\right)^{1/p}.$$

We recall that there exists a constant $c>0$ such that

$$\sup_{x\in\Omega} |u(x)| \le c\|u\| \tag{1}$$

for each $u \in W_0^{1,p}(\Omega)$. Moreover we put $\omega := \dfrac{\pi^{n/2}}{\frac{n}{2}\Gamma(\frac{n}{2})}$ the measure of the n-dimensional unit ball.

2. RESULTS

Our first result guarantees that the problem $(D_{n,p})$ has infinitely many weak solutions that form an unbounded set in $W_0^{1,p}(\Omega)$.

Theorem 2.1 *Assume that, for each* $\xi \in \mathbb{R}$, $F(\xi) \ge 0$. *Moreover suppose that there exist* $x_0 \in \Omega$, *a positive number* $\delta \le d(x_0, \partial\Omega)$ *and four real sequences* $\{r_k\}_{k\in\mathbb{N}}$, $\{\gamma_k\}_{k\in\mathbb{N}}$, $\{\varepsilon_k\}_{k\in\mathbb{N}}$, $\{\xi_k\}_{k\in\mathbb{N}}$ *with* $\lim_{k\to\infty} r_k = +\infty$, $0 < \gamma_k \le \operatorname{dist}(x_0, \partial\Omega)$, $\varepsilon_k \in]0, \gamma_k[$ *and* $\xi_k \in]0, +\infty[$ *for all* $k \in \mathbb{N}$, *such that*

(i) $F(\xi_k) = \max\limits_{[0, c r_k^{\frac{1}{p}}]} F$ for each $k \in \mathbb{N}$;

(ii) $\xi_k < (\gamma_k - \varepsilon_k)\left(\dfrac{r_k}{\omega(\gamma_k^n - \varepsilon_k^n)}\right)^{\frac{1}{p}}$ for each $k \in \mathbb{N}$;

(iii) $F(\xi_k) < \dfrac{1}{p(|\Omega| - \omega\varepsilon_k^n)}\left[r_k - \dfrac{\omega\xi_k^p}{(\gamma_k - \varepsilon_k)^p}(\gamma_k^n - \varepsilon_k^n)\right]$ for each $k \in \mathbb{N}$,

(iv) $\limsup_{\xi\to+\infty} \dfrac{F(\xi)}{\xi^p} > \dfrac{2^p}{p\delta^p}(2^n - 1)$.

Then, the problem ($D_{n,p}$) has infinitely many weak solutions that form an unbounded set in $W_0^{1,p}(\Omega)$.

Proof. Let us apply Theorem 1.1. To this end choose $X = W_0^{1,p}(\Omega)$ and for each $u \in X$, put

$$\Phi(u) = -\int_\Omega \left(\int_0^{u(x)} f(t)dt \right) dx$$

and

$$\Psi(u) = \|u\|^p .$$

It is well known that the critical points in X of the functional $\Phi + \frac{1}{p}\Psi$ are precisely the weak solutions of problem ($D_{n,p}$). Clearly, the functionals Φ and Ψ are Gâteaux differentiable and sequentially weakly lower semicontinuous; moreover Ψ is obviously (strong) continuous and coercive.

In our case the function φ of Theorem 1.1 is defined by setting

$$\varphi(r) = \inf_{\|u\|^p < r} \frac{\sup_{\|v\|^p \le r} \int_\Omega F(v(x))dx - \int_\Omega F(u(x))dx}{r - \|u\|^p}$$

for each $r \in]0, +\infty[$.

Now we wish to prove that $\varphi(r_k) < \frac{1}{p}$ for each $k \in \mathbb{N}$. To this aim, it sufficies to prove that, for each $k \in \mathbb{N}$, there exists a function $u_k \in X$, with $\|u_k\|^p < r_k$, such that

$$\sup_{\|v\|^p \le r_k} \int_\Omega F(v(x))dx - \int_\Omega F(u_k(x))dx < \frac{1}{p}(r_k - \|u_k\|^p).$$

Fix $k \in \mathbb{N}$ and consider the function $u_k \in X$ defined by setting

$$u_k(x)=\begin{cases} 0 & \text{if } x\in\Omega\setminus B(x_0,\gamma_k) \\ \xi_k & \text{if } x\in B(x_0,\varepsilon_k) \\ \dfrac{\xi_k}{\gamma_k-\varepsilon_k}(\gamma_k-|x-x_0|) & \text{if } x\in B(x_0,\gamma_k)\setminus B(x_0,\varepsilon_k) \end{cases}$$

Obviously

$$\|u_k\|^p=\int_\Omega|\nabla u_k(x)|^p\,dx=\int_{B(x_0,\gamma_k)\setminus B(x_0,\varepsilon_k)}\frac{\xi_k^p}{(\gamma_k-\varepsilon_k)^p}dx=$$

$$=\frac{\xi_k^p}{(\gamma_k-\varepsilon_k)^p}(|B(x_0,\gamma_k)|-|B(x_0,\varepsilon_k)|)=\frac{\xi_k^p}{(\gamma_k-\varepsilon_k)^p}\frac{\omega(\gamma_k^n-\varepsilon_k^n)}{\omega}<r_k$$

thanks to hypothesis (ii).

In view of (1), for each $v\in X$, with $\|v\|^p\le r_k$, one has

$$\sup_{x\in\Omega}|v(x)|\le cr_k^{\frac{1}{p}}$$

and, thanks to (i), it follows that

$$F(v(x))\le\max_{[0,cr_k^{\frac{1}{p}}]}F=F(\xi_k)$$

for each $x\in\Omega$. Hence, using (iii), we get

$$\sup_{\|v\|^p\le r_k}\int_\Omega F(v(x))dx-\int_\Omega F(u_k(x))dx\le F(\xi_k)|\Omega|-\int_{B(x_0,\varepsilon_k)}F(\xi_k)dx=$$

$$=F(\xi_k)|\Omega|-F(\xi_k)|B(x_0,\varepsilon_k)|=F(\xi_k)[|\Omega|-\omega\varepsilon_k^n]$$

$$<\frac{1}{p}(r_k-\|u_k\|^p).$$

Bearing in mind that $\lim_{k\to\infty}r_k=+\infty$, the previous inequality assures that the hypothesis of part (a) of Theorem 1.1 is satisfied. Thus, by that result, it follows that either the functional $\Phi+\frac{1}{p}\Psi$ has a global minimum, or there

exists a sequence $\{u_k\}_{k\in\mathbb{N}}$ of solutions of problem $(D_{n,p})$ such that $\lim_{k\to\infty}\|u_k\|=+\infty$.

The other step is to verify that the functional $\Phi+\frac{1}{p}\Psi$ has no global minimum.

By (iv) we can choose a constant $h>\frac{2^p}{p\delta^p}(2^n-1)$ such that, for each $k\in\mathbb{N}$, one has

$$\sup_{\eta\geq k}\frac{F(\eta)}{\eta^p}>h$$

and so there exists $\eta_k\geq k$ such that

$$\frac{F(\eta_k)}{\eta_k^p}>h.$$

Now, if we consider a function $w_k\in X$ defined by setting

$$w_k(x)=\begin{cases}0 & \text{if } x\in\Omega\setminus B(x_0,\delta)\\ \eta_k & \text{if } x\in B(x_0,\frac{\delta}{2})\\ \frac{2\eta_k}{\delta}(\delta-|x-x_0|) & \text{if } x\in B(x_0,\delta)\setminus B(x_0,\frac{\delta}{2})\end{cases}$$

one has

$$\Phi(w_k)+\frac{1}{p}\Psi(w_k)=-\int_\Omega F(w_k(x))dx+\frac{1}{p}\|w_k\|^p\leq-\int_{B(x_0,\frac{\delta}{2})}F(\eta_k)dx+$$

$$+\frac{2^p\eta_k^p\omega\delta^n}{p\delta^p}(1-\frac{1}{2^n})=\omega\delta^n\left[2^p\frac{\eta_k^p}{p\delta^p}(1-\frac{1}{2^n})-\frac{F(\eta_k)}{2^n}\right]<$$

$$<\omega\delta^n\left[2^p\frac{\eta_k^p}{p\delta^p}(1-\frac{1}{2^n})-h\frac{\eta_k^p}{2^n}\right]=\frac{\omega\delta^n\eta_k^p}{2^n}\left[\frac{2^p}{p\delta^p}(2^n-1)-h\right].$$

Since $h > \frac{2^p}{p\delta^p}(2^n - 1)$, it forces $\lim_{k\to\infty} \eta_k^p \left[\frac{2^p}{p\delta^p}(2^n - 1) - h \right] = -\infty$ and so the previous inequality shows that the functional $\Phi + \frac{1}{p}\Psi$ is not bounded from below and then it has no global minimum.

Therefore, Theorem 1.1 assures that there is a sequence $\{v_k\}_{k\in\mathbb{N}} \subseteq X$ of critical points of $\Phi + \frac{1}{p}\Psi$ such that $\lim_{k\to\infty} \|v_k\| = +\infty$. As previously observed, every function v_k is a weak solution of $(D_{n,p})$ and this completes the proof. □

A possible function that verifies Theorem 2.1 is the following

Example 2.1 Let $\Omega = B(0,r)$ the open ball of $\mathbb{R}^2$, $p = 3$ and $F : \mathbb{R} \to \mathbb{R}$ be the function defined by setting

$$F(x) = \begin{cases} 0 & \text{if } x \in]-\infty, 0[\\ A(x) & \text{if } x \in [0, e] \\ B_k(x) & \text{if } x \in]e^{8k-7}, e^{8k-3}] \\ C_k(x) & \text{if } x \in]e^{8k-3}, e^{8k+1}] \end{cases}$$

with $k \in \mathbb{N}$,

$$A(x) = a(-2x^3 + 3e\,x^2)$$

$$B_k(x) = \frac{a}{(e^4-1)^3}(2x^3 - 3e^{8k-7}(e^4+1)x^2 + 6e^{16k-10}x + e^{24k-13}(e^4-3))$$

$$C_k(x) = \frac{a}{(e^4-1)^3}(-2e^{12}x^3 + 3e^{8k+9}(e^4+1)x^2 - 6e^{16k+10}x + e^{24k+3}(3e^4-1))$$

where a is a real number such that

$$\frac{64}{r^3} < a < \frac{8e^{12}}{45\pi^4 r^3} - \frac{64}{15r^3}.$$

This function satisfies all assumptions of Theorem 2.1 taking $x_0 = 0$, $\delta = \frac{r}{2}$, $r_k = \frac{e^{24k-9}}{2\pi^3 r}$, $\gamma_k = \frac{r}{2}$, $\varepsilon_k = \frac{r}{4}$ and $\xi_k = e^{8k-7}$; in particular, the choice of a makes true hypotheses (iii) and (iv).

It is interesting to note that, in this case, one has

$$a \le \limsup_{\xi \to +\infty} \frac{F(\xi)}{\xi^3} < +\infty.$$

For $n = 1$ and $p = 2$, taking $\Omega =]0,1[$ (in this case $c = \frac{1}{2}$), $x_0 = \frac{1}{2}$, $\delta = \frac{1}{2}$ and $\gamma_k = \frac{1}{2}$ for each $k \in \mathbb{N}$ Theorem 2.1 gives the following result.

Theorem 2.2 (Theorem 2.1 of [1]) *Assume that, for each* $\xi \in \mathbb{R}$, $F(\xi) \ge 0$ *and that there exist three real sequences* $\{r_k\}_{k\in\mathbb{R}}$, $\{\varepsilon_k\}_{k\in\mathbb{N}}$, $\{\xi_k\}_{k\in\mathbb{N}}$ *with* $\lim_{k\to\infty} r_k = +\infty$, $\{\varepsilon_k : k \in \mathbb{N}\} \subseteq]0, \frac{1}{2}[$ *and* $\{\xi_k : k \in \mathbb{N}\} \subseteq]0, +\infty[$, *such that*

(i) $F(\xi_k) = \max_{[0,\frac{\sqrt{r_k}}{2}]} F$;

(ii) $\xi_k < \sqrt{\frac{r_k \varepsilon_k}{2}}$ for each $k \in \mathrm{IN}$;

(iii) $F(\xi_k) < \frac{1}{4\varepsilon_k}\left(r_k - 2\frac{\xi_k^2}{\varepsilon_k}\right)$ for each $k \in \mathrm{IN}$;

(iv) $\limsup_{\xi\to+\infty} \frac{F(\xi)}{\xi^2} > 8$.

Then, the problem

$$\begin{cases} -u'' = f(u) & \text{in }]0,1[\\ u(0) = u(1) = 0 \end{cases} \qquad (D_{1,2})$$

has infinitely many classical solutions that form an unbounded set in $W_0^{1,2}(]0,1[)$.

Remark 2.1 Observe that, for assuring that the functional $\Phi+\frac{1}{p}\Psi$ has no global minimum, in the proof of Theorem 2.1 (and similar in Theorem 2.2) we guarantee that the functional is not bounded from below. It would be interesting to find some conditions such that $\Phi+\frac{1}{p}\Psi$ is bounded from below but without a global minimum.

An explicit example of function F that fits all the hypotheses of Theorem 2.2 is the following.

Example 2.2 Let a a real number such that

$$4<a<4e^{\frac{\pi}{2}}-8:=\sigma.$$

Let $F:\mathbb{R}\to\mathbb{R}$ be the function defined by setting

$$F(x)=\begin{cases} 0 & \text{if } x\in]-\infty,0] \\ ax^2\left(\sin\left(\ln x^2\right)+1\right) & \text{if } x\in]0,+\infty[\end{cases}$$

The function F assumes its local maxima in $x=e^{\frac{\pi}{2}+k\pi}$, for each $k\in\mathbb{Z}$ and its local minima in $x=e^{\frac{3}{4}\pi+k\pi}$, for each $k\in\mathbb{Z}$. Moreover it satisfies all the hypotheses of Theorem 2.2

To justify this assertion we choose, for each $k\in\mathbb{N}$ $(k\geq 0)$, $r_k=4e^{\frac{3}{2}\pi+2k\pi}$, $\xi_k=e^{\frac{\pi}{2}+k\pi}$ and $\varepsilon_k=\frac{1}{4}$. It is easy to prove that (i), (ii), (iii) and (iv) are satisfied: in particular, the choice of a makes true hypotheses (iii) and (iv).

Note that, in this case, one has

$$\limsup_{\xi\to+\infty}\frac{F(\xi)}{\xi^2}=\limsup_{\xi\to+\infty}\ a\left(\sin\left(\ln x^2\right)+1\right)=2a<+\infty$$

and

$$\liminf_{\xi\to+\infty}\frac{F(\xi)}{\xi^2}=\liminf_{\xi\to+\infty}a\left(\sin\left(\ln x^2\right)+1\right)=0.$$

With a slight modification on function F (see Example 2.1 of [1]) it is possible to obtain

$$\liminf_{\xi\to+\infty}\frac{F(\xi)}{\xi^2}>0.$$

The next results are two simpler but less general form of Theorem 2.1. With similar arguments utilized in the proof of Theorem 2.1, making use of part (a) of Theorem 1.1, we obtain:

Theorem 2.3 *Assume that, for each* $\xi\in\mathbb{R}$, $F(\xi)\geq 0$. *Moreover, suppose that there exist two real sequences* $\{a_k\}_{k\in\mathbb{N}}$ *and* $\{b_k\}_{k\in\mathbb{N}}$ *in* $]0,+\infty[$ *with* $a_k<b_k$, $\lim_{k\to\infty}b_k=+\infty$, *such that*

(i) $\lim_{k\to\infty}\frac{b_k}{a_k}=+\infty$;

(ii) $\max_{[a_k,b_k]}f\leq 0$ *for all* $k\in\mathbb{N}$;

(iii) $$\frac{2^p}{p(\sup_{x\in\Omega}\operatorname{dist}(x,\partial\Omega))^p}(2^n-1)<\limsup_{\xi\to+\infty}\frac{F(\xi)}{\xi^p}<+\infty.$$

Then, problem $(D_{n,p})$ *admits an unbounded sequence of non-negative weak solutions in* $W_0^{1,p}(\Omega)$.

Likewise, applying part (b) of Theorem 1.1, we get

Theorem 2.4. *Assume that, for each* $\xi\in\mathbb{R}$, $F(\xi)\geq 0$. *Moreover, suppose that there exist two real sequences* $\{a_k\}_{k\in\mathbb{N}}$ *and* $\{b_k\}_{k\in\mathbb{N}}$ *in* $]0,+\infty[$ *with* $a_k<b_k$, $\lim_{k\to\infty}b_k=0$, *such that*

(j) $\lim_{k\to\infty}\frac{b_k}{a_k}=+\infty$;

(jj) $\max_{[a_k,b_k]}f\leq 0$ *for all* $k\in\mathbb{N}$;

(jjj) $$\frac{2^p}{p(\sup_{x\in\Omega}\operatorname{dist}(x,\partial\Omega))^p}(2^n-1)<\limsup_{\xi\to 0^+}\frac{F(\xi)}{\xi^p}<+\infty.$$

Then, problem $(D_{n,p})$ *admits a sequence of nonzero weak solutions which strongly converges to* 0 *in* $W_0^{1,p}(\Omega)$.

Remark 2.2 Observe that, in the mere condition:

$$0<\limsup_{\xi\to+\infty}\frac{F(\xi)}{\xi^p}<+\infty$$

we can apply Theorem 2.3 and 2.4 taking Ω sufficiently large.

An explicit example of function which satisfies all the assumptions of Theorem 2.3 is the following.

Example 2.3 Let Ω be a bounded open subset of $\mathbb{R}^n$ with boundary of class C^1 and $p > n$. Let $f : \mathbb{R} \to \mathbb{R}$ the function defined by setting

$$f(\xi) = \sum_{k=1}^{\infty} \frac{2Lh_k}{k!} \operatorname{dist}(\xi, \mathbb{R} \setminus [k!k, (k+1)!]) \ ,$$

for each $\xi \in \mathbb{R}$, where

$$L > \frac{2^p}{p(\sup_{x \in \Omega} \operatorname{dist}(x, \partial\Omega))^p} (2^n - 1)$$

and

$$h_k = 2(k!)^{p-1}[(k+1)^p - 1]$$

for each $k \in \mathbb{N}$. A more explicit expression of f is

$$f(\xi) = \begin{cases} 0 & \text{if} \quad \xi \in \mathbb{R} \setminus \bigcup_{k \in \mathbb{N}} [k!k, (k+1)!] \\ \dfrac{2Lh_k}{k!} \min\{\xi - k!k, (k+1)! - \xi\} & \text{if} \quad \xi \in [k!k, (k+1)!], \ k \in \mathbb{N} \end{cases}$$

By choosing, for each $k \in \mathbb{N}$,

$$a_k = k!$$
$$b_k = k!k$$

the hypotheses of Theorem 2.3 are satisfied and one has

$$\limsup_{\xi \to +\infty} \frac{F(\xi)}{\xi^p} = L.$$

In fact

$$\frac{F(a_k)}{a_k^p} = \frac{L}{2(k!)^p} \sum_{i=1}^{k-1} ((i+1)! - i!i) h_i =$$

$$= \frac{L}{(k!)^p} \sum_{i=1}^{k-1} [((i+1)!)^p - (i!)^p)] = L \frac{(k!)^p - 1}{(k!)^p}.$$

On the other side, for each $\xi \in [b_k, b_{k+1}]$, one has

$$\frac{F(\xi)}{\xi^p} \leq \frac{F(a_{k+1})}{b_k^p} = L \frac{((k+1)!)^p - 1}{(k!)^p k^p}.$$

In a similar way it is possible to obtain an example of function satisfying Theorem 2.4.

Note, in particular, the following corollary of Theorem 2.3.

Corollary 2.1 *Let* $\{a_k\}_{k\in\mathbb{N}}$ *and* $\{b_k\}_{k\in\mathbb{N}}$ *two real sequences in* $]0,+\infty[$ *with* $a_k < b_k$, $\lim_{k\to\infty} b_k = +\infty$, *such that* $\lim_{k\to\infty} \frac{b_k}{a_k} = +\infty$. *Moreover, let* $g \in C^1(\mathbb{R})$ *such that* $\inf_{\xi\in\mathbb{R}} g(\xi) \geq 0$ *and*

$$\frac{2^p}{(\sup_{x\in\Omega} \operatorname{dist}(x, \partial\Omega))^p} (2^n - 1) < \limsup_{\xi\to+\infty} g(\xi) < +\infty$$

and

$$max_{\xi\in[a_k,b_k]} \left[g(\xi) + \frac{\xi}{p} g'(\xi) \right] \leq 0$$

for each $k \in \mathbb{N}$. *Then the problem*

$$\begin{cases} -\Delta_p u = \frac{1}{p} u^{p-1} (pg(u) + ug'(u)) & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases} \tag{2}$$

admits an unbounded sequence of non-negative weak solutions in $W_0^{1,p}(\Omega)$.

A similar result can be obtained using Theorem 2.4.

3. CONCLUDING REMARKS

Now we wish to recall some other results existing in literature concerning the existence of infinitely many solutions for the problem $(D_{n,p})$

The following result comes directly from a recent theorem obtained by J. Saint-Raymond (Theorem 3.1 of [7]).

Theorem 3.1 *Let* $f:\mathbb{R}\to\mathbb{R}$ *be a continuous function and let* $F:\mathbb{R}\to\mathbb{R}$ *the function defined by setting*

$$F(\xi)=\int_0^{\xi} f(t)dt$$

for each $\xi\in\mathbb{R}$. *Assume that*

(1) *there exists* $M>0$ *such that, for every* $\rho>0$, *there exists* $t>0$ *satisfying* $F(t)\geq\rho(1+t^2)$ *and* $F(s)\geq -MF(t)$ *for each* $s\in[0,t]$;

(2) $\sup\{t\in\mathbb{R} : f(t)<0\}=+\infty$;

(3) $\inf\{t\in\mathbb{R}.\ f(t)>0\}<0$.

Then there are unboundedly (infinitely) many solutions of the problem : $(D_{1,2})$.

We wish to emphasize that Theorem 3.1 cannot be applied to the function of the Example 2.2. In fact the hypotheses (1) and (3) of Theorem 3.1 are surely not satisfied; if we consider the function F of Example 2.2 it is easy to observe that $\inf\{t\in\mathbb{R} : f(t)>0\}=0$. Moreover, for each $t>0$, one has

$$F(t)\leq a(1+b)t^2<3at^2<3\sigma(1+t^2).$$

This means that, in particular, hypothesis (2) cannot be satisfied when $\rho\geq 3\sigma$.

Another comparison we wish to make is with a recent result of Korman and Li (see [2]). Before dealing with this result we should state first a definition.

Definition 3.1 Let $f:\mathbb{R}\to\mathbb{R}$ be a function. We say that f satisfies Schaaf-Schmitt condition if there exist two monotone sequences $\{x_n\}_{n\in\mathbb{N}}$ and $\{y_n\}_{n\in\mathbb{N}}$ with $\lim_{n\to\infty}x_n=\lim_{n\to\infty}y_n=+\infty$ such that

$$F(x_n)-F(x)\geq 0 \quad \text{for all } 0\leq x<x_n$$

$$F(y_n) - F(x) \leq 0 \quad \text{for all } 0 \leq x < y_n$$

where F is, as usual, the integral function of f.

Now, making use of Theorem 2 of [2] with opportune choices it is easy to obtain the following theorem that assures the existence of infinitely many solutions of problem $(D_{1,2})$.

Theorem 3.2 *Assume that* $f \in C^2([0,+\infty[)$ *satisfies* $f(0)=0$, $f'(0)>0$, $f(t)>0$ *for each* $t>0$ *and* $\lim_{t\to+\infty} \frac{f(t)}{t} = 0$. *Moreover the function* $\frac{1}{\pi^2} f(t) - t$ *satisfies the Schaaf-Schmitt condition. Under these assumptions the problem* $(D_{1,2})$ *admits infinitely many solutions.*

In this result the strong hypothesis is the existence of the $\lim_{t\to+\infty} \frac{f(t)}{t}$ and Example 2.2 shows a case in which, even if $\frac{1}{\pi^2} f(t) - t$ satisfies the Schaaf-Schmitt condition, $\lim_{t\to+\infty} \frac{f(t)}{t}$ doesn't exist, making the previous result not appliable.

Other recent results in which infinitely many solutions of the problem $(D_{n,p})$ are assured is contained in [4] and [5].

In [4] Omari and Zanolin obtain the following result

Theorem 3.3 (Corollary 1.2 of [4]) *Assume that*

$$\liminf_{\xi\to+\infty} \frac{F(\xi)}{\xi^p} = 0 \quad \text{and} \quad \limsup_{\xi\to+\infty} \frac{F(\xi)}{\xi^p} = +\infty \tag{3}$$

then problem $(D_{n,p})$ *has a sequence* $\{u_n\}_{n\in\mathbb{N}}$ *of positive solutions in* $W_0^{1,p}(\Omega)$ *with* $\max_{\overline{\Omega}} u_n \to +\infty$.

In [5] the same authors replace the conditions (3) at $+\infty$ by similar ones at 0, in order to produce arbitrarily small positive solutions of problem $(D_{n,p})$. Namely, the following holds

Theorem 3.4 *Assume that*

$$\liminf_{\xi\to 0^+}\frac{F(\xi)}{\xi^p}=0 \quad \text{and} \quad \limsup_{\xi\to 0^+}\frac{F(\xi)}{\xi^p}=+\infty \tag{4}$$

then problem $(D_{n,p})$ *has a sequence* $\{u_n\}_{n\in\mathbb{N}}$ *of positive solutions in* $W_0^{1,p}(\Omega)$ *with* $\max_{\bar{\Omega}} u_n$ *decreasing to zero and* $\frac{1}{p}\int_\Omega |\nabla u_n(x)|^p\,dx-\int_\Omega F(u_n(x))dx$ *increasing to zero.*

We note that in these results it is requested that $\limsup_{\xi\to+\infty}\frac{F(\xi)}{\xi^p}=+\infty$ and $\limsup_{\xi\to 0^+}\frac{F(\xi)}{\xi^p}=+\infty$; these are stronger requests with respect of hypothesis (iii) of Theorem 2.3 and (jjj) of Theorem 2.4. Moreover, in our results, nothing is said about the behaviour of $\liminf_{\xi\to+\infty}\frac{F(\xi)}{\xi^p}$ and $\liminf_{\xi\to 0^+}\frac{F(\xi)}{\xi^p}$. In fact, as we have already observed, the function F of Example 2.1 of [1] doesn't satisfy any of the (3) and the function F of Example 2.1 doesn't satisfy $\limsup_{\xi\to+\infty}\frac{F(\xi)}{\xi^3}=+\infty$.

REFERENCES

[1] F. Cammaroto and A. Chinnì. *Infinitely many solutions for a two points boundary value problem.* Far East Journal of Mathematical Sciences, 11, n.1:41-51, 2003.

[2] P. Korman and Y. LI. *Infinitely many solutions at a resonance.* Nonlinear Differential Equations, pages 105–111, 2000.

[3] G.B. Li and H.S. Zhou. *Multiple solutions to p-Laplacian problems with asymptotic nonlinearity as* u^{p-1} *at infinity.* J. London Math. Soc., **65**, n. 2:123–138, 2002.

[4] P. Omari and F. Zanolin. *Infinitely many solutions of a quasilinear elliptic problem with an oscillatory potential.* Commun. in Partial Differential Equations, **21** :721–733, 1996.

[5] P. Omari and F. Zanolin. *An elliptic problem with arbitrarily small positive solutions.* Nonlinear Differential Equations, Electron. J. Diff. Eqns., Conf. **05**: 301–308, 2000.

[6] B. Ricceri. *A general variational principle and some of its applications.* J. Comput. Appl. Math., **113**: 401–410, 2000.

[7] J. Saint Raymond. *On the multiplicity of the solutions of the equation* $-\Delta u=\lambda f(u)$. J. Differential Equations, **180**: 65–88, 2002.

A DENSITY RESULT ON THE SPACE VMO_ω

A.O. Caruso and M.S. Fanciullo
Dept. of Mathematics and Computer Sciences, University of Catania, Catania, Italy

Abstract: In Carnot–Carathéodory metric spaces related to a family of free Hörmander vector fields $X_1,...,X_q$, we prove that the space C^∞ is locally dense in VMO_ω with respect to BMO_ω norm.

Key words: VMO spaces, spaces of homogeneous type, Carnot–Carathéodory metric

1. SOME PRELIMINARIES

In [4] we introduced the space of the functions with bounded mean oscillation defined using cubes on a space of homogeneous type. We proved that this space is equivalent to the classical space BMO defined using balls. Then we proved that, as in the euclidean case (see [13]), in Carnot–Carathéodory metric spaces the space of the C^∞ functions is locally dense in the space VMO of the functions with vanishing mean oscillation with respect to the BMO norm. In this note we give the definition of the space BMO_ω^C of the "multipliers" of BMO on spaces of homogeneous type. BMO_ω^C was introduced by Spanne and it is the set of functions for which the mean oscillation behaves as $|\log r|^{-1}$ over the cubes with diameter r (for definition in euclidean setting see [14], [1]). Then, in an obvious way, we introduce the space VMO_ω^C and in Carnot–Carathéodory metric spaces related to a family of free Hörmander vector fields, we prove that C^∞ is locally dense in VMO_ω^C with respect to BMO_ω^C norm. We remark that these density results are fundamental to solve regularity problems for equations and systems with discontinuous coefficients (see [1], [3], [5], [9], [10]).

Let us begin by giving some basic definitions.

A quasimetric d on a set S is a function $d: S \times S \to [0, +\infty[$ with the following properties

$d(x,y) = 0$ if and only if $x = y$;

$d(x,y) = d(y,x) \quad \forall \quad x, y \in S$;

$d(x,y) \le A_0 \left[d(x,z) + d(z,y) \right] \quad \forall x, y, z \in S$.

Now we give the definition of space of homogeneous type:

Definition 1.1. *A space of homogeneous type (S,d,μ) is a set S with a quasimetric d and a measure μ on S such that, for all $x \in S$ and $r > 0$ it results $0 < \mu(B(x,r)) < +\infty$ and the following doubling property holds*

$$\mu(B(x,2r)) \le A_1 \, \mu(B(x,r)). \tag{D}$$

The number $Q = \log_2 A_1$ (A_1 is the infimum satisfying (1)) is called the homogeneous dimension of the space (S,d,μ). To have more details on spaces of homegeneous type we refer the reader to [8].

Definition 1.2. *A Borel measure μ on a quasimetric space is said to be Ahlfors regular of dimension Q if there exist two positive constants a and A such that for all $x \in S$ and $r > 0$ it results*

$$a r^Q \le \mu(B(x,r)) \le A r^Q. \tag{A}$$

Let (S,d,μ) be a space of homogeneous type with μ Ahlfors regular measure. If $f \in L^1(\Omega)$, $\Omega \subseteq S$, we denote by f_Ω the integral average $\fint_\Omega f d\mu = \frac{1}{\mu(\Omega)} \int_\Omega f d\mu$.

We consider the non–decreasing function $\omega(r) = 1/\ln \frac{eR_0}{r}$ for $r < R_0$ and $\omega(r) = 1$ for $r \ge R_0$. Then we can give the following definitions (see [2]).

Definition 1.3. *BMO_ω is the set of equivalence of functions f (with finite integral on bounded sets), modulo additive constants, such that*

$$\|f\|_{BMO_\omega} = \sup_{\substack{x \in S \\ r>0}} \frac{1}{\omega(r)} \fint_{B(x,r)} | f - f_{B(x,r)} | \, d\mu < +\infty.$$

BMO_ω is a Banach space with the above norm.

Definition 1.4. *A function $f \in BMO_\omega(S)$ belongs to the space $VMO_\omega(S)$ if*

$$\sup_{\substack{x\in S \\ 0<r\le a}} \frac{1}{\omega(r)} \fint_{B(x,r)} |f - f_{B(x,r)}|\, d\mu \to 0 \quad \text{as} \quad a \to 0.$$

In the space of homogeneous type (S,d,μ) we can give the definition of dyadic cubes (see [6] and [7]) and, then the definition of cubes (see [4]). Indeed in [6] and in [7] the following facts have been proved.

Theorem 1.1. *For all integers k there exist a numerable set I_k and a family of subsets $Q_\alpha^k \subseteq S$, $\alpha \in I_k$, such that*

(1) $\mu(S \setminus \bigcup_\alpha Q_\alpha^k) = 0 \ \ \forall k \in \mathbb{Z}$;

(2) *for any α, β, k, l with $l \ge k$, either $Q_\beta^l \subseteq Q_\alpha^k$ or $Q_\beta^l \cap Q_\alpha^k = \varnothing$;*

(3) *for each Q_α^{k+1} there exists exactly one Q_β^k (parent of Q_α^{k+1}) such that $Q_\alpha^{k+1} \subseteq Q_\beta^k$;*

(4) *for each Q_α^k there exists at least one Q_β^{k+1} (child of Q_α^k) such that $Q_\beta^{k+1} \subseteq Q_\alpha^k$.*

These subsets are called *dyadic cubes* since they are the analogous of the euclidean dyadic cubes. Now we give the definition of cubes.

Definition 1.5. *Let Q and Q' be two dyadic cubes. We say that Q' is 1–step contiguous to Q if $\partial Q' \cap \partial Q \neq \varnothing$. Moreover we say that Q' is k–step contiguous ($k \ge 2$) to Q if Q' is 1–step contiguous to some $(k-1)$–step dyadic cube contiguous to Q.*

Definition 1.6. *We call cube either a dyadic cube or the union of a given dyadic cube with its contiguous cubes of the same generation up to some step $k \ge 1$.*

We denote by $d(Q)$ the diameter of a generic cube Q. Then we can introduce the class of the multipliers of *BMO* using cubes.

Definition 1.7. *BMO_ω^C is the set of equivalence of functions f (with finite integral on bounded sets), modulo additive constants, such that*

$$\|f\|_{BMO_\omega^C} = \sup_Q \frac{1}{\omega(d(Q))} \fint_Q |f - f_Q|\, d\mu < +\infty. \ f$$

Definition 1.8. *A function $f \in BMO_\omega^C(S)$ belongs to the space $VMO_\omega^C(S)$ if*

$$M_0(f) := \lim_{a \to 0^+} M_a(f) = 0,$$

where

$$M_a(f) := \sup_{d(Q) \le a} \frac{1}{\omega(d(Q))} \fint_Q | f - f_Q | \, d\mu.$$

It is possible to prove the following theorem.

Proposition 1.1. *Let* (S,d,μ) *be a space of homogeneous type with* μ *Ahlfors regular measure. Then there exists a positive constant* C *such that*

$$\frac{1}{C} \| \cdot \|_{BMO_\omega(S)} \le \| \cdot \|_{BMO_\omega^C(S)} \le C \| \cdot \|_{BMO_\omega(S)},$$

and

$$VMO_\omega(S) = VMO_\omega^C(S).$$

Proof. It is similar to the proof of the Theorem 2.2 and Theorem 2.3 in [4].

□

2. THE DENSITY RESULT

Now we introduce some particular spaces of homogeneous type in which we prove our density result.

Given q smooth real vector fields $X_1, X_2, ..., X_q$ on a bounded domain Ω in $\mathbb{R}^N$, a Lipschitz continuous curve $\gamma : [0,T] \to \Omega$ is said to be $X-$*subunit* if there exists a measurable vector function $h = (h_1, ..., h_q) : [0,T] \to \mathbb{R}^q$ such that $\dot{\gamma}(t) = \sum_{i=1}^q h_i(t) X_i(\gamma(t))$ for a.e. $t \in [0,T]$ and $\|h\|_\infty \le 1$. Set

$$d_X(x,y) = \inf \{ T \ge 0 : \exists \gamma, X - \text{subunit curve, such that } \gamma(0) = x, \gamma(T) = y \},$$

we have that d_X is a metric in Ω, usually called the Carnot–Carathéodory distance associated to the system $X = (X_1, X_2, ..., X_q)$ (see [11]).

Denoted by $X_\alpha = \left[X_{\alpha_1}, \left[X_{\alpha_2}, ..., \left[X_{\alpha_{d-1}}, X_{\alpha_d}\right]...\right]\right]$ a commutator, where $\alpha = (\alpha_1, \alpha_2, \ ..., \alpha_{d-1}, \alpha_d)$ is a multi–index with $|\alpha| = d$, we say that the system X satisfies Hörmander condition of step s at some point $\xi_0 \in \Omega$ if $\{X_\alpha(\xi_0)\}_{|\alpha| \le s}$ spans $\mathbb{R}^N$ as vector space. The vector fields are free of step s if $N = \dim g(q,s)$ where $g(q,s)$ is the free Lie algebra of step s on q generators.

In $\mathbb{R}^N$ the euclidean topology and the $C-C$ metric one are the same, nevertheless the two metrics are not equivalent. Moreover, Lebesgue measure is locally doubling with respect to d_X; indeed, for any bounded set $E \Subset \Omega$ there exists $R > 0$ such that $\mathcal{L}^N(B) \approx r^Q$ for any $C-C$ ball B centered in $c \in E$ with radius $0 < r \le R$. In order to work with these spaces, the following theorem (due to Rothschild and Stein, [12]), is crucial:

Theorem 2.1. *Let $X = (X_1, X_2, ..., X_q)$ be a system of q real C^∞ vector fields on an open set $\Omega \subseteq \mathbb{R}^N$ satisfying Hörmander condition of step s and free up to the same order at $\xi_0 \in \Omega$. Then, there exist open neighborhoods U of 0 and $W \Subset V$ of ξ_0 such that, for any $\xi \in V$, the mapping $U \ni y \to \eta = \exp\left(\sum_{|\alpha| \le N} y_\alpha X_\alpha\right)\xi \in V$ is invertible, and calling $y = \Theta_\xi(\eta)$ its inverse, it results:*

a) $\Theta_{\xi|_W}$ *is a diffeomorphhism onto the image for every* $\xi \in V$;

b) $U \subseteq \Theta_\xi(V)$ *for every* $\xi \in W$;

c) $\Theta : V \times V \to \mathbb{R}^N$ *defined by* $\Theta(\xi,\eta) := \Theta_\xi(\eta)$ *is* $C^\infty(V \times V)$.

We will assume $(V, d_X, \mathcal{L}^N)$ as our space of homogeneous type in order to prove the density result. From the properties of the function $\Theta(\xi,\eta)$ we can construct the convolution f_ε of a function f in the space $(V, d_X, \mathcal{L}^N)$ (for more details we refer the reader to [4]).

As in [4], for the function f_ε it is possible to prove the following lemma.

Lemma 2.1. *If $f \in BMO_\omega(V)$ then $f_\varepsilon \in BMO_\omega(W)$, moreover, for $\varepsilon > 0$ sufficiently small it results*

$$\|f_\varepsilon\|_{BMO_\omega(W)} \le c\|f\|_{BMO_\omega(V)}$$

where c is an absolute constant.

Hence we prove our main result.

Theorem 2.2. *There exists a positive constant A such that for all $f \in BMO_\omega(V)$ there exists a function $g \in C^\infty(W)$ such that $\|f-g\|_{BMO^C_\omega(W)} \le A M_a(f)$. In particular, if $f \in VMO_\omega(V)$, then there exists a sequence $\{f_n\}$ in $C^\infty(W)$ such that $f_n \to f$ in $BMO_\omega(W)$.*

Proof. Fix $a>0$ and l such that $M_a(f) \le l$. Taken a suitable $\bar{k}$, let h be the step function such that h takes value $f_{Q_\alpha^{\bar{k}}}$ in $Q_\alpha^{\bar{k}}$. Now we estimate $\|f-h\|_{BMO^C_\omega(W)}$. Let Q be a cube: we can assume that Q is union of dyadic cubes of generation k. Take $k' \ge \max\{k,\bar{k}\}$: then $Q = \cup_{\alpha=1}^m Q_\alpha^{k'}$. Since $d(Q) \ge d(Q_\alpha^{k'})$ for $\alpha = 1,2,...,m$ and by monotonicity of the function ω, it results for an absolute constant $\tilde{c}$

$$\frac{1}{\omega(d(Q))} \fint_Q |f-h-(f-h)_Q| \, d\xi \le \frac{2}{\omega(d(Q))} \fint_Q |f-h| \, d\xi =$$
$$= \frac{2}{\omega(d(Q))\,|Q|} \sum_{\alpha=1}^m \int_{Q_\alpha^{k'}} |f - f_{Q_\alpha^{k'}}| \, d\xi \le \tilde{c}\, l.$$

As in [4], for a suitable $r>0$, we can construct the function h_ε. Take $\xi \in W$, then ξ belongs to some $Q_\alpha^{\bar{k}}$. There exists only an absolute number $\overline{C}$ of dyadic cubes $Q_\alpha^{\bar{k}}$ such that $Q_\alpha^{\bar{k}} \cap B(\xi,r) \ne \varnothing$, from these cubes we can construct a cube Q' such that $d(Q') < a$. Since $\omega(\mathrm{d}(Q')) \le 1$, we have

$$|f_{Q_\alpha^{\bar{k}}} - f_{Q'}| \le \fint_{Q_\alpha^{\bar{k}}} |f - f_{Q'}| \, d\xi \le \frac{|Q'|}{|Q_\alpha^{\bar{k}}|} \fint_{Q'} |f - f_{Q'}| \, d\xi \le c\, l,$$

where the constant c depends only on $\overline{C}$, a_0 and the homogeneous dimension. Then, arguing as in the proof of Theorem 2.2 and Theorem 2.3 in [4] we have the thesis. □

REFERENCES

[1] P. Acquistapace, *On BMO regularity for linear elliptic systems*, Ann. Mat. Pura Appl., (4) 161 (1992) 231-269.

[2] M. Bramanti and L. Brandolini, *Estimates of BMO type for singular integrals on spaces of homogeneous type and applications to hypoelliptic PDES*, to appear on Revista Matematica Iberoamericana.

[3] A.O. Caruso, *Interior $S^{1,p}_{X,loc}$ estimates for variational hypoelliptic operators with coefficients locally in VMO_X* , preprint (2003).

[4] A.O. Caruso and M.S. Fanciullo, *BMO on spaces of homogeneous type: a density result*, preprint (2003).

[5] F. Chiarenza, M.Frasca and P.Longo, *$W^{2,p}$-solvability of the Dirichlet problem for non divergence elliptic equations with VMO coefficients*, Trans. Amer. Math. Soc., 336 , 1, (1993), 841-853.

[6] M. Christ, *Lectures on singular integral operators*, Conference Board of the Mathematical Sciences, Regional Conference Series in Mathematics, 77 (1990).

[7] M. Christ, *A T(b) Theorem with remarks on analytic capacity and the Cauchy integral*, Colloq. Math., **LX/LXI**, 2, (1990), 601-628.

[8] R. Coifman and G.Weiss, *Analyse Harmonique Non-Commutative sur Certains Espaces Homogenes*, Lectures Notes in Mathematics, 242, Springer–Verlag (1971).

[9] G.Di Fazio, *L^p estimates for Divergence Form Elliptic Equations with Discontinuous Coefficients*, Boll. U.M.I. (7) 10-A, (1996), 409-420.

[10] G.Di Fazio and M.S. Fanciullo, *BMO regularity for elliptic systems in Carnot–Carathéodory spaces*, Comm. on Applied Nonlinear Analysis, 10 (2003), n.2, 81-95

[11] M. Gromov, *Carnot–Carathéodory spaces seen from within*, in *Sub-Riemannian Geometry*, Progress in Mathematics 144 , ed. by A.Bellaïche and J.Risler, Birkhäuser (1996).

[12] L.P. Rothschild and E.M. Stein, *Hypoelliptic differential operators and nilpotent groups*, Acta Math., 137 (1976), 247-320.

[13] D. Sarason, *Functions of vanishing mean oscillation*, Trans. Amer. Math. Soc., 207 (1975), 391-405.

[14] S. Spanne,*Some functions spaces defined using the mean oscillation over cubes*, Ann. Sc. Norm. Sup. Pisa (3), 19 (1965), 593-608.

LINEAR COMPLEMENTARITY SINCE 1978

Richard W. Cottle
Dept. of Operation Research, Stanford University, Stanford, CA., USA

Abstract: We survey developments on the Linear Complementarity Problem (LCP) since 1978, the year in which the International School of Mathematics on *Variational Inequalities and Complementarity Problems* took place at the 'Ettore Majorana' Centre of Scientific Culture in Erice, Sicily. This report will touch on matrix classes and the existence of solutions, complexity, degeneracy resolution, algorithms, software products, applications and generalizations of the LCP.

1. INTRODUCTION

The proceedings of the International School of Mathematics on Variational Inequalities and Complementarity Problems [16] contains a reasonable summary of what was known about the linear complementarity problem (LCP) in the year 1978. Of the 25 papers in that volume, only [9] deals exclusively, albeit briefly, with the LCP. Lemke's more extensive paper [71] has much to say about the LCP as well as the broader topic of Constructive Approximation Methods (CAM) by which he meant the reliance on algorithms rather than fixed-point theorems of the Brouwer or Kakutani type.

In general, a finite-dimensional complementarity problem is expressed in terms of a closed convex cone $K \subset \mathbb{R}^n$ and a mapping $F : \mathbb{R}^n \to \mathbb{R}^n$. One seeks a vector satisfying the conditions

$$x \in K, \quad F(x) \in K^* \text{ (the polar of } K) \text{ and } \langle x, F(x) \rangle = 0 . \tag{1}$$

This formulation and its equivalence with the variational inequality

$$\text{Find } x \in K \text{ such that } \langle F(x), v - x \rangle \geq 0, \forall v \in K \tag{2}$$

were established by Karamardian [60], [61]. An important distinction between these problems is that the variational inequality can be posed with respect to *any* closed convex set, not just a cone, wereas the complementarity problem is *always* defined relative to a closed convex cone.

During the last 25 years, the field of optimization (mathematical programming) has seen considerable growth in research activity and the number of publications dealing with theory , algorithms, and applications of complementarity and related subjects. The author's own informal survey suggests that roughly 10% of papers in the journal *Mathematical Programming* fall into this broad category. Scores of doctoral theses and numerous monographs have been written on complementarity. Among the latter, we single out [81], [18], [56], [57], and [33]. In the addition to these contributions to the literature, there now exists computer software for complementarity problems incorporated within commercial optimisation packages.

The standard *linear complentarity problem* (LPC) corresponds to the case where K is the nonnegative orthant R^n_+ and F is an affine transformation $x \mapsto q + Mx$ of R^n into itself. It is plain to see that the $n \times n$ matrix M and the n-vector q determine the LCP. For this reason we often use the pair (q, M) as a notation for the problem (1) where F and K are as just described.

The aim of this expository paper is to sketch *some* of the progress made on the LCP since 1978. A thorough treatment of the subject would be inappropriate on this occasion. As a consequence, many valuable contributions had to be omitted. Section 2 discusses some contributions from the traditional topic of matrix classes with particular emphasis on the important questions of the existence and possible uniqueness of solutions. Section 3 is about complexity issues, algorithmic developments and computer software for "processing" linear complementarity problems. Some applications of the LCP are discussed in Section 4, and a variety of generalizations of the standard LCP are reviewed in Section 5.

2. EXISTENCE OF SOLUTIONS: THE ROLE OF MATRIX CLASSES

With *any* affine transformation of R^n into itself there corresponds a linear complementarity problem, but with the data so freely chosen, there is no reason to expect the LCP to have a solution. From the mathematical

standpoint, it is amusing to study the existence question "abstractly," and much of this kind of thing ha been done. The applicability of complementarity problems to satisfying first-optimality conditions in mathematical programming and to satisfying equilibrium conditions in economics and engineering justifies the attention given to finding conditions on the problem data that guarantee the existence of their solutions. When these conditions are in harmony with algorithms intended to produce solutions, these investigations are all the more fruitful. This section will highlight some contributions to the identification of interesting matrix classes that shed light on the question of existence. For this discussion, we follow the notational system used in [18].

Given an $n \times n$ matrix M there is an associated cone $K(M)$ which can be defined as the set of all $q \in R^n$ such that the LCP (q,M) has a solution. This set can also be described as the union of all *complementary cones* induced by M. That is,

$$K(M) = \bigcup_{\alpha} \operatorname{pos} C_M(\alpha) \tag{3}$$

where $\alpha \subseteq \{1,...,\}$ and $C_M(\alpha)$ is the $n \times n$ matrix defined as follows:

$$\left(C_M(\alpha)\right)_{\bullet i} = \begin{cases} -M_{\bullet i} & \text{if } i \in \alpha \\ I_{\bullet i} & \text{otherwise} \end{cases}. \tag{4}$$

A vector x is said to be *feasible* for the LCP (q,M) if $x \geq 0$ and $q + Mx \geq 0$. The set of all feasible solutions for (q,M) is denoted $\operatorname{FEA}(q,M)$, whereas $\operatorname{SOL}(q,M)$ denotes the set of all feasible solutions such that $x^{\mathrm{T}}(q+Mx) = \langle x, q+Mx \rangle = 0$.

At an early stage in the development of the subject, Parsons [84] and Murty [79] introduced the matrix classes $\mathbf{Q}_0$ and $\mathbf{Q}$. The definitions of these classes are as follows:

$$\mathbf{Q}_0 = \bigcup_{n=1}^{\infty} \left\{ M \in R^{n \times n} : \operatorname{FEA}(q,M) \neq \emptyset \Rightarrow \operatorname{SOL}(q,M) \neq \emptyset \right\} \tag{5}$$

$$\mathbf{Q} = \bigcup_{n=1}^{\infty} \left\{ M \in R^{n \times n} : \operatorname{SOL}(q,M) \neq \emptyset \text{ for all } q \in R^n \right\}. \tag{6}$$

Efforts to find useful, nontrivial characterizations of the classes $\mathbf{Q}_0$ and $\mathbf{Q}$ have not been successful. See [9, pages 99-100] in this regard. Nevertheless, it is clear that $M \in \mathbf{Q}$ only if M belongs to the matrix class $\mathbf{S}$, which is to say the linear inequality system $Mx > 0$, $x > 0$ has a solution. Whether or not $M \in \mathbf{S}$ is a question that can be answered by linear programming. The issue of characterizing $\mathbf{Q}_0$ and $\mathbf{Q}$ boils down to that of characterizing $\mathbf{Q}_0$ alone since $\mathbf{Q} = \mathbf{S} \cap \mathbf{Q}_0$. As noted in [9], Eaves [27] showed that $M \in \mathbf{Q}_0$ if and only if $K(M)$ is convex. Unfortunately, this is not an easily tested conditions.

The bulk of research on these questions is aimed at identifying sub classes of $\mathbf{Q}_0$. This trend was well underway in 1978. The main matrix classes of the time were

	PSD	positive semidefinite (not necessarily symmetric)
★	**PD**	positive definite (not necessarily symmetric)
★	**P**	positive principal minors
★	**E**	strictly semimonotone
★	**R**	regular
	$\mathbf{CP}^+$	copositive-plus
★	**SCP**	strictly copositive
	A	adequate
	Z	nonpositive off-diagonal elements

Classes indicated by ★ also belong to $\mathbf{Q}$. All these classes are discussed in [18].

One notable class defined by the intersection of two classes listed above is $\mathbf{K} = \mathbf{P} \cap \mathbf{Z}$. This remarkable matrix class has many applications and a rich theory which includes about 50 equivalent definitions, most of which can be found in Berman and Plemmons [3]. For others see [18].

In some cases, membership in a matrix class defined with absolutely no reference to the LCP is equivalent to a property of the LCP. The class **P** is an example (the property being existence and uniqueness of a solution to *every* LCP formed with the matrix). The class **K** is another (here the property is the existence and uniqueness for every LCP of a solution that is also the least element of the feasible region).

These results were well known even before 1978. Since then, some new matrix classes have found. One of these was identified by Cottle, Pang, and Venkateswaran [19] and shown to have intimate connections with properties of the LCP. Because of their strong association with both new and old algorithms for the LCP, we shall devote a disproportionate amount of attention to some developments in this topic.

These matrices are called *sufficient*[1] and the class of all such matrices is denoted **SU**. This class is actually the intersection of two others: the *row sufficient matrices* (**RSU**) and the *column sufficient matrices* (**CSU**). For the person who thinks of n-vectors as *columns*, the primary definition needed here would be that of column sufficiency.

Definition 1. An $n \times n$ matrix is *column sufficient* if for all $x \in R^n$

$$x_i(Mx)_i \leq 0 \text{ for all } i = 1,\ldots,n \quad \Rightarrow \quad x_i(Mx)_i = 0 \text{ for all } i = 1,\ldots,n.$$

A matrix is *row sufficient* if its transpose is column sufficient. A matrix that is both row and column sufficient it simply called sufficient.

It is not hard to show that the class **SU** contains the union of **P** and **PSD**. (As noted above, we do not assume the symmetry of the matrix M in discussing positive semidefiniteness. Instead, for M to be **PSD** we require $x^{\mathrm{T}}Mx = \sum_i x_i(Mx)_i \geq 0$ for all $x \in R^n$.) It even contains **A** as well as the direct sum of the classes **A**, **P** and **PSD**.

The definitions just given make no direct reference to the LCP, and yet such a connect exists and is an intimate one. In fact, each of the classes can be defined in terms of a property of the LCP. We state these as theorems.

Theorem 1. The $n \times n$ matrix M is column sufficient if and only if for every $q \in M$ the (possibly empty) solution set of (q,M) is convex.

We remark that $\mathbf{CSU} \not\subset \mathbf{Q}_0$. For instance, any 2×2 matrix with one zero column and one positive column is column sufficient but not a $\mathbf{Q}_0$-matrix. The story is different for **RSU**, however.

The fact that $\mathbf{RSU} \subset \mathbf{Q}_0$ follows from

Theorem 2. The $n \times n$ matrix M is row sufficient if and only if for each $q \in R^n$, if (x,u) is a Karush-Kuhn-Tucker pair for the quadratic program

$$\begin{aligned} &\text{minimize } x^{\mathrm{T}}(q + Mx) \\ &\text{subject to } q + Mx \geq 0 \\ &\qquad\qquad x \geq 0, \end{aligned} \tag{7}$$

then x solves the LCP (q,M).

[1] This name for the matrix class **SU** was chosen partly in jest as a parody of the name "adequate" for the matrix class **A** introduced by Ingleton [54].

Thus, when $M \in \mathbf{SU}$, every feasible LCP (q,M) has a *nonempty* convex solution set. Moreover, since every row or column sufficient matrix must have nonnegative principal minors (and thus belong to the matrix class $\mathbf{P}_0$), the difference between any two solutions of an LCP formed with such a matrix is zero. Hence $(q+Mx)$ is *constant* for all x solving (q,M).

The class **SU** nicely unifies the classes **A**, **P** and **PSD** in other ways. For example, it is "invariant" under principal pivoting. This means that, if $M_{\alpha\alpha}$ is a nonsingular principal submatrix of a sufficient matrix M, then the corresponding principal pivot transform matrix M' is also sufficient. More explicitly, if

$$M = \begin{bmatrix} M_{\alpha\alpha} & M_{\alpha\bar{\alpha}} \\ M_{\bar{\alpha}\alpha} & M_{\bar{\alpha}\bar{\alpha}} \end{bmatrix} \in \mathbf{SU}$$

then

$$M' = \begin{bmatrix} M'_{\alpha\alpha} & M'_{\alpha\bar{\alpha}} \\ M'_{\bar{\alpha}\alpha} & M'_{\bar{\alpha}\bar{\alpha}} \end{bmatrix} := \begin{bmatrix} M_{\alpha\alpha}^{-1} & -M_{\alpha\alpha}^{-1}M_{\alpha\bar{\alpha}} \\ M_{\bar{\alpha}\alpha}M_{\alpha\alpha}^{-1} & M_{\bar{\alpha}\bar{\alpha}} - M_{\bar{\alpha}\alpha}M_{\alpha\alpha}^{-1}M_{\alpha\bar{\alpha}} \end{bmatrix} \in \mathbf{SU}$$

The question of checking a real square matrix for membership in **CSU** is addressed in [17], [45], and [94]. We have no polynomial test for this property, but the ones we have are at least finite. One of these tests is somewhat interesting. It is based on the fact that checking the column sufficiency of 2×2 matrices is quite easy.

Lemma 1. The matrix $M \in R^{2\times 2}$ is column sufficient if and only if the following two conditions are satisfied:

(i) $M \in \mathbf{P}_0$;
(ii) no principal pivot transform or principal rearrangement of M has the form

$$\begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}, \quad b \neq 0.$$

Definition 2. If $M \in R^{n\times n}$ and $1 \leq k \leq n$ we say that M is *column sufficient of order k* if every $k\times k$ principal submatrix of M is column sufficient.

In [17] this definition and the notion of principal pivot transformation are combined to yield the following criterion for column sufficiency. It happens to be a generalization of a theorem due to Parsons [84] regarding **P**-matrices.

Theorem 3. A matrix $M \in R^{n\times n}$ is column sufficient if and only if every principal pivot transform of M is column sufficient of order 2.

The above lemma can be used to check the column sufficiency of the 2×2 matrices mentioned in the theorem. Väliaho [94] gave some additional criteria for (row and column) sufficiency. He closes with the observation, "all the above tests are combinatorially explosive and thus practicable for small matrices only". This is not altogether surprising, for Coxson [21] showed that testing for membership on **P** is co-NP-complete.

Let us now turn to another class of matrices. This one was introduced in Section 3.2 of *A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems* by Kojima, Megiddo, Noma and Yoshise [67]. The class is called $\mathbf{P}_*(k)$ where the parameter K is a nonnegative real number. The condition that a member of this class is required to satisfy is

$$(1+4\kappa)\sum_{i\in I_+(x)} x_i(Mx)_i + \sum_{i\in I_-(x)} x_i(Mx)_i \geq 0 \quad \forall x \in R^n \tag{8}$$

where

$$I_+(x)=\{i : x_i(Mx)_i > 0\} \text{ and } I_-(x)=\{i : x_i(Mx)_i < 0\} \tag{9}$$

Notice that $\mathbf{P}_*(0)$ = **PSD** and that if $\kappa_1 \leq \kappa_2$, then $\mathbf{P}_*(\kappa_1)\subseteq \mathbf{P}_*(\kappa_2)$. The smallest value of κ for which $M \in \mathbf{P}_*(\kappa)$ is denoted $\hat{\kappa}(M)$. The scalar is $\hat{\kappa}(M)$ measures the smallest "boost" that $x^{\mathrm{T}}Mx$ needs to become a nonnegative-valued function. Accordingly, $\hat{\kappa}(M)$ is called the handicap of M.

What we have here *really* is a family of matrix classes (one for each κ). Relative to this family Kojima et al. defined the class

$$\mathbf{P}_* = \bigcup_{\kappa\geq 0} \mathbf{P}_*(\kappa). \tag{10}$$

They show that for each $\kappa \geq 0$ the class $\mathbf{P}_*(\kappa)$ is a subclass of **CSU** and hence is a subclass of $\mathbf{P}_0$.

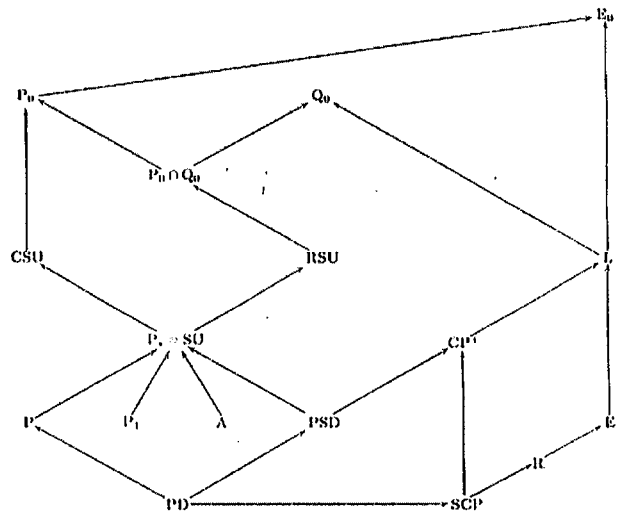
Recall our observation that **CSU** is not a subclass of $\mathbf{Q}_0$. Despite this, they observe [67, p. 39] that $\mathbf{P}_*$ is a subclass of $\mathbf{Q}_0$. *This inclusion leads one to suspect that there must be more to the class* $\mathbf{P}_*$ *than meets the eye, and indeed this is just the case.*

In [45] it is shown that $\mathbf{P}_* \subset \mathbf{SU}$ and that, for 2×2 matrices, the reverse inclusion holds as well. This led to the conjecture (also stated in [45]) that $\mathbf{P}_* \supset \mathbf{SU}$ and to the question [13] "Are $\mathbf{P}_*$-matrices just sufficient?"

An answer to this question was not long in coming. Hannu Väliaho's paper [95], with the crisp title "$\mathbf{P}_*$-matrices are just sufficient", provided an affirmative answer and did so with a beautiful proof. Notice that the equality of $\mathbf{P}_*$ and $\mathbf{SU}$ explains why $\mathbf{P}_*$-matrices belong to $\mathbf{Q}_0$: they are row sufficient as well as column sufficient.

In a subsequent paper [96], Väliaho introduced a method for computing the handicap of a (sufficient) matrix. The calculation is not a simple matter. In addition, Väliaho showed that the handicap is the same for a matrix and its transpose. He conjectured that the handicap of a matrix is a continuous function of its elements. It is also true that if M' is a principal pivot transform of M, then $\hat{\kappa}(M)=\hat{\kappa}(M')$, whereas if $M_{\alpha\alpha}$ is a principal submatrix of M, then $\hat{\kappa}(M_{\alpha\alpha})\le\hat{\kappa}(M)$.

The figure below indicates some inclusions among most of the matrix classes discussed so far. It is worth noting that the five classes along the bottom edge of this figure are all subclasses of $\mathbf{Q}$.



Inclusions among matrix classes. All are strict.

In the interest of clarity, some other inclusions have been omitted from this figure. For example, $\mathbf{P}$ is a subclass of $\mathbf{E}$ which is also known [10] to be the class of completely-$\mathbf{Q}$ matrices, i.e., those for which the matrix and all of its principal submatrices belong to $\mathbf{Q}$. Moreover, Eaves [27] showed that $\mathbf{L}$ contains $\mathbf{P}$ and $\mathbf{A}$; in his Ph.D. thesis, Stone [89] showed that $\mathbf{L}$ contains $\mathbf{P}_1$.

3. ALGORITHMS, COMPLEXITY AND COMPUTING

A few years before the period under discussion here, Klee and Minty [66] stunned the mathematical programming world by demonstrating that solving a linear programming with the standard simplex algorithm can take an exponential number of pivot steps. Soon, variants of the class of problems exhibited by Klee and Minty turned up in the world of linear complementarity, and it was shown that all the known pivoting algorithms of the day could require an exponential number of iterations to obtain a solution. Results of this sort can be found in [34], [80], [40], [4]. Some unification of these problems is available in [11] where connections Hamiltonian paths, puzzles and Gary code are established.

Such exponential behavior of the simplex algorithm is at variance with its typical behavior.

Decades of experience has shown that the number of iterations required to solve a linear program with an m-rowed constraint matrix is a small multiple of m. Seeking an understanding of the behavior of the simplex algorithm several researchers concentrated on *probabilistic analysis*. (See Todd [92, p. 422] for some discussion and citations of the relevant literature.) In some of these studies, the discussion was facilitated by the expression of the optimality conditions for linear programming as an LCP. The keen interest in this line of research on linear programming tended to increase the optimization community's interest in the LCP as such.

The search for a polynomial-time linear programming algorithm was on. Credit for the first to be discovered goes to L. Khachiyan [64], [65]. His ellipsoid method [64] was guaranteed to solve a linear program in polynomial time, but although it was a major achievement, the method proved disappointing because its typical behavior was much like its worst-case behavior.

And then, in the year 1984, N.K. Karmarkar [63] advanced a linear programming algorithm that differed from the simplex method by taking steps through the interior of the feasible region rather than by traversing edges of the polyhedron. There are now many algorithms of this type: Karmarkar's original proposal, now called the *projective* method, dual and primal-dual *path-following* methods, and *potential reduction* methods. Collectively they are called *interior-point algorithms*. There are now several excellent monographs devoted to this subject. Among these are [82], [98], [101].

On the heels of interior-point algorithms for linear programming came natural extensions of these methods for quadratic programming and certain classes of LCPs. Since the casting of optimality criteria for a linear

programming—or a convex quadratic—programming problem as an LCP (q,M) leads to the formation of a positive semidefinite coefficient matrix M, it is not surprising that there should be interior-point algorithms for such problems. These are commonly called monotone LCPs inasmuch as $x^{\mathrm{T}}Mx \geq 0$ for all x implies that the mapping $F(x) = q + Mx$ is monotone in the sense that

$$\langle F(x) - F(y), x - y \rangle \geq 0 \quad \text{for all} \quad x, y. \tag{11}$$

In 1988, Y. Ye [99] introduced the use of interior-point methods for the (monotone) LCP. In that same year, studied **P**-matrix LCPs. He and others such as M.J. Todd, S. Mizuno, N. Megiddo, and M. Kojima (in various combinations) actively explored this line of research, and the results poured out. This effort provided the background for the publication of the slender but powerful volume [67] in which the matrix class $\mathbf{P}_*$ was introduced. But, as noted above, it would be a while before the full connection of this class with the sufficient matrices would be revealed.

Around this time, Ye and Pardalos [102] described a "condition" number $\gamma(qM)$ for the LCP (q,M) that characterizes the difficulty of using a potential reduction algorithm to solve that instance of the LCP (q,M) it belongs to a class they call **G**. Notice that **G** is a class of linear complementarity problems, not a class of matrices. Nonetheless, they note that $(q,M) \in \mathbf{G}$ whenever $M \in \mathbf{PSD}$. (In this case, the choice of q does not matter.) They also point out that (i) $(q,M) \in \mathbf{G}$ whenever $M \in \mathbf{CP}$ (copositive) and $q \geq 0$, and (ii) $(q,M) \in \mathbf{G}$ whenever M^{-1} exists and is copositive and $M^{-1}q \leq 0$. Problems with $q \geq 0$ or with $M^{-1}q \leq 0$ are trivial ($x = 0$ is a solution of the first and $x = q$ is a solution of the second), but Ye and Pardalos remark that the potential reduction algorithm often finds alternate solutions to such problems. Ye's paper [100] gives a polynomial potential reduction algorithm for the **P**-matrix LCP.

In 1997, B. Jansen, C. Roos, and T. Terlaky [59] published a family of polynomial algorithms for the monotone LCP and soon thereafter [87], F.A. Potra and R. Sheng introduced a large-step infeasible-interior-point algorithm that can be used on the $\mathbf{P}_*$-matrix LCP. Actually, they assume that the matrix of the LCP belongs to $\mathbf{P}_*(\kappa)$ for some κ, as it must if it belongs to $\mathbf{P}_*$. The authors show that if a given LCP with such a matrix is solvable (which in this case is equivalent to saying it is feasible) then the algorithm converges from arbitrary positive starting points. The number of iterations depends on the quality of the starting point.

While the interior-point methods and the contest to improve their worst-case behavior bounds were occupying peoples' attention, linear complementarily problems of many kinds were also being solved by the

more traditional pivoting methods, such as Lemke's algorithm [70], the principal pivoting method [23], and a procedure called the criss-cross method [91]. The combination of these methods with sufficient matrices deserves a few words. In the paper [1] it is shown that Lemke's algorithm will process (i.e., either solve or reveal the infeasibility of) any LCP (q, M) in which $M \in \mathbf{P}_0 \cap \mathbf{Q}_0$. Since row sufficient matrices belong to $\mathbf{P}_0 \cap \mathbf{Q}_0$, it follows that Lemke's algorithm will process such LCPs. It was shown in [12] that the principal pivoting method processes LCPs with row sufficient matrices. Moreover, the least-index degeneracy resolution rule that was known [7] for LCPs with PSD- and P-matrices was extended to **SU** [14]. In this case, both row and column sufficiency are needed. Finally, in [50] it is shown that a matrix M is sufficient if and only if the criss-cross method processes all instances of the LCPs (q, M) and (q, M^{T}).

Software for solving complementarity problems (including of course the LCP) is available. For example, the web-based NEOS server enables one to use T.F. Rutherford's MILES [39], and the celebrated product PATH by S.P. Dirkse and M.C. Ferris [26]. In 1997, the Mathematical Programming Society awarded Dirkse and Ferris with the Beale-Orchard-Hays Prize for excellence in computational mathematical programming. MILES requires GAMS input, whereas for PATH the input can be in Fortran, AMPL, or GAMS.

4. APPLICATIONS

From the outset, the linear complementarily problem was viewed as a model of the KKT conditions of an inequality-constrained linear or quadratic programming problem, but other applications were noticed as well. These included the computation of Nash equilibrium points for bimatrix games via the Lemke-Howson algorithm [72], the solution of problems in contact mechanics, market equilibrium problems, and optimal stopping problems. Still another problem to which the linear complementarity problem applies is the computation of convex hulls of finite sets in the plane.

All the applications mentioned above are briefly discussed in [18], but most of what is described there was already available by 1978. A much more comprehensive and up-to-date survey of applications of engineering and economic applications of (primarily nonlinear) complementarity problems is available in the (1997) paper [36] by Ferris and Pang. Both of these authors have written extensively on a broad range of applications of complementarity problems and variational inequalities. More on applications of complementarity will be found in [55] and [57]. Still more recent (2003) is the monumental two-volume work of Facchinei and Pang [33] which

opens with an impressive collection of source problems. In the latter presentation, the emphasis tends to be on variational inequalities, but there are plenty of complementarity problems in it as well. One of these is on the pricing of American options. In 1990, this finance problem was treated by Jaillet, Lamberton and Lapeyre [58] who used a combination of variational inequalities and a finite-dimensional discretization to an LCP. According to Facchinei and Pang [33, p. 119] "although not explicitly using the LCP framework, Brennan and Schwartz" [6] "are arguably the earliest authors who used an iterative LCP algorithm for solving the American option pricing problem". The subject is developed in [53].

Under favorable conditions, the solution of nonlinear complementarity problems becomes another application of the LCP. This, of course, is through linearization methods and *sequential linear complementarity problem* (SLCP) solution. The idea is to linearize the (differentiable) mapping F of a nonlinear complementarity problem (over the nonnegative orthant) at a current iterate, say x^k and then solve the LCP (q, M) where

$$q = F(x^k) - \nabla F(x^k) x^k \quad \text{and} \quad M = \nabla F(x^k) \equiv \left[\frac{\partial F_i(x^k)}{\partial x_j} \right] \tag{12}$$

This approach has been used extensively by L. Mathiesen in [76], [77] and (jointly with C.D. Kolstad) in [68]. The application in each of these publications is an economic equilibrium problem. The LCPs are typically solved using Lemke's algorithm. The convergence arguments given in the Kolstad-Mathiesen paper are based on theorems of Pang and Chan [83].

The recent global interest in the restructuring and design of electricity markets has given rise to numerous opportunities to build equilibrium and complementarity models, some of which are of the linear type. Such models aim to provide both quantities and prices for electric power generation, transmission and distribution systems. See [78], [52], [51], and [86].

5. GENERALIZATIONS

The casting of complementarity problems in abstract spaces constitutes one sort of generalization that dates from the previously mentioned work of Karamardian (some of which, incidentally, benefitted from the celebrated Hartman-Stampacchia theorem [49]). But this is not the sort of generalization we have in mind here. Instead, the title of this section refers to other related forms of the problem. The first of these to be called a generalized linear complementarity problem was introduced in 1970 by

Cottle and Dantzig [15]. It is now called the *vertical linear complementarity problem* (VLCP). Briefly, the problem is to satisfy the system

$$x \geq 0, \quad y = q + Nx \geq 0, \quad x_j \prod_{k=1}^{m_j} y_j^k = 0 \tag{13}$$

As may be inferred from the complementarity conditions above, the matrix N has $\sum_{j=1}^{n} m_j$ rows and n columns. The rows are grouped into a "stack" of submatrices N_j. With this formulation, some of the definitions ordinarily used for square matrices can be applied to "representative submatrices," that is, $n \times n$ matrices whose jth row comes from the jth block, N_j. Thus, for instance a generalized **P**-matrix is a vertical block matrix for which each representative submatrix belongs to **P**.

Interest in the VLCP seems to have been nonexistent until 1989 when B.P. Szanc's Ph.D. thesis [90] was completed. Two years later another Ph.D. thesis [28], that of A.A. Ebiefung, came along. He and M. Kostreva produced a series of papers [30], [31], and [32] involving this model and its applications. See also [29] and [47].

Although it seems to have been named after the VLCP, the *horizontal linear complementarity problem* (HLCP) can be said to date back to a seminal paper of Samelson, Thrall, and Wesler [88] published in 1958. (In this paper it was first proved that **P** is the class of matrices M for which the (standard) LCP (q, M) has one and only one solution, regardless of which q is used.) The underlying mathematical problem Samelson, Thrall, and Wesler studied was that of solving the system

$$Ax + By = c, \quad x, y \geq 0, \quad x^{\mathrm{T}} y = 0, \tag{14}$$

They sought conditions on the $n \times n$ matrices A and B such that this system would have a unique solution for every $c \in R^n$. In so doing, they were able to manipulate the problem into the one we think of as the standard LCP.

It seems that the recent interest in the HLCP arose in the wake of the research on interior point methods for linear and quadratic programming. This literature includes [69], [93], and [42]. The last two of these are concerned with the question of reducing the horizontal LCP to a standard LCP. An infeasible interior-point algorithm for the HLCP can be found in [103].

Yet another generalization of the LCP is the so-called *extended linear complementarity problem* (XLCP) introduced by Mangasarian and Pang

[75]. This problem can be expressed as that of finding a solution of the system

$$Ax + By \in K, \quad x, y \geq 0, \quad x^{\mathrm{T}} y = 0, \tag{15}$$

where K is a polyhedral convex set. In this problem, the matrices A and B are of the same order, but not necessarily square. In [43], Gowda studied the XLCP, he introduced and characterized the column-sufficiency, row-sufficiency, and **P**-properties. He then specialized these properties to the HLCP and VLCP.

Another style of generalized linear complementarity problem is given in [24]. In this case, the authors seek *all* nonnegative solutions of the system

$$Mx = 0 \quad \text{and} \quad \sum_{i=1}^{\ell} \prod_{k \in \mathcal{B}i} x_k = 0. \tag{16}$$

This form of the problems appears to be of interest in electrical circuit theory.

And finally, there is another *extended linear complementarity problem* (ELCP) devised by B. De Schutter and B. De Moor [25]. The problem is to find a solution of the system

$$\begin{aligned} &\sum_{j=1}^{m} \prod_{i \in \phi_j} (Ax - c)_i = 0 \\ &Ax - c \geq 0 \\ &Bx = d \end{aligned} \tag{17}$$

or show that no such vector exists. The data for this model have the following specifications: $A \in R^{p \times n}$, $B \in R^{q \times n}$, $c \in R^p$, $d \in R^q$, and $\phi_j \subseteq \{1, \ldots p\}$ for $j = 1, \ldots, m$. The authors demonstrate the formulation of all the above generalizations (and more) in terms of the ELCP.

They also note that the general ELCP is **NP**-hard since it includes the standard LCP which, as shown by S.J. Chung [8], is **NP**-complete.

REFERENCES

[1] M. Aganagic and R.W. Cottle. A constructive characterization of $\mathbf{Q}_0$ with nonnegative principal minors. *Mathematical Programming* 37 (1987) 223-252.

[2] D. Baraff. Issues in Computing Contact Forces for Non-penetrating rigid bodies. *Algorithmica* 10 (1993) 292-352.

[3] A. Berman and R.J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences.* New York: Academic Press, 1979.

[4] J.R. Birge and A. Gana. Computational complexity of Van der Heyden's variable dimension algorithm and Dantzig-Cottle's principal pivoting method for solving LCPs. *Mathematical Programming* 26 (1983) 205-214.

[5] K.C. Border. *Fixed Point Theorems with Applications to Economics and Game Theory.* Cambridge: Cambridge University Press, 1985.

[6] Brennan and Schwartz. Valuation of American put options. *Journal of Finance.* 32 (1977) 449-462.

[7] Y-Y. Chang. Least-index resolution of degeneracy in linear complementarity problems. Technical Report 79-14, Department of Operations Research, Stanford University, 1979.

[8] S.J. Chung. NP-completeness of the linear complementarity problem. *Journal of Optimization Theory and Applications* 60 (1989) 393-399.

[9] R.W. Cottle. Some recent developments in linear complementarity theory. In [16, pp. 97-104]

[10] R.W. Cottle. Completely-**Q** matrices. *Mathematical Programming* 19 (1980) 347-351.

[11] R.W. Cottle. Observations on a class of nasty linear complementarity problems. *Discrete Applied Mathematics* 2 (1980) 89-111.

[12] R.W. Cottle. The principal pivoting method revisited. *Mathematical Programming* 48 (1990) 369-385.

[13] R.W. COTTLE. Are $\mathbf{P}_*$-matrices just sufficient? Talk at 36th Joint National ORSA/TIMS Meeting, Phoenix, Arizona, November 1, 1993.

[14] R.W. Cottle and Y-Y. Chang. Least-index resolution of degeneracy in linear complemen-tarity problems with sufficient matrices. SIAM *Journal on Matrix Analysis and Applications.* 13 (1992) 1131-1141.

[15] R.W. Cottle and G.B. Dantzig. A generalization of the linear complementarity problem. *Journal of Combinatorial Theory* 8 (1970) 79-90.

[16] R.W. Cottle, F. Giannessi and, J.-L. Lions. *Variational Inequalities and Complementarity Problems.* Chichester: John Wiley & Sons, 1980.

[17] R.W. Cottle and S-M. Guu. Two characterizations of sufficient matrices. *Linear Algebra and its Applications* 17 (1992) 65-74.

[18] R.W. Cottle J-S. Pang, and R.E. Stone. *The Linear Complementarity Problem.* Boston: Academic Press, 1992.

[19] R.W. Cottle, J-S. Pang, and V. Venkateswaran. Sufficient matrices and the linear complementarity problem. *Linear Algebra and its Applications* 114/115, (1989) 231-249.

[20] M. C. Coutinho. *Dynamic Simulations of Multi-Body Systems.* New York: Springer, 1991.

[21] G.E. Coxson. The P matrix problem is co-NP complete. *Mathematical Programming* 64 (1994) 173-178.

[22] J. Crank. *Free and Moving Boundary Problems.* Oxford: Oxford University Press, 1984.

[23] G.B. Dantzig and R.W. Cottle. Positive (semi-)definite programming, in (J. Abadie, ed.) *Nonlinear Programming. Amsterdam*: North-Holland, 1967, pp. 55-73.

[24] B. De Moor, L. Vandenberghe, and J. Vandewalle. The generalized linear complementarity problem and an algorithm to find all its solutions. *Mathematical Programming* 57 (1992) 415-426.

[25] B. De Schutter and B. De Moor. The extended linear complementarity problem. *Mathematical Programming* 71 (1995) 289-325.

[26] S.P. Dirkse and M.C. Ferris. The PATH solver: A non-monotone stabilization scheme for mixed complementarity problems. *Optimization Methods and Software* 5 (1995) 123-156.

[27] B.C. Eaves. The linear complementarity problem, *Management Science* 17 (1971) 612-634.

[28] A.A. Ebiefung. *The Generalized Linear Complementarity Problem and its Applications.* Ph.D. thesis. Clemson University, Clemson, S.C., 1991.

[29] A.A. Ebiefung. Existence theory and Q-matrix characterization for the generalized linear complementarity problem. *Linear Algebra and its Applications* 223/224 (1995) 155-169.

[30] A.A. Ebiefung and M.M. Kostreva. Global solvability of generalized linear complementarity problems and a related class of polynomial complementarity problems, in (C. Floudas P. Pardalos, eds.) *Recent Advances in Global Optimization.* Princeton, N.J.: Princeton University Press, 1992, pp. 102-124.

[31] A.A. Ebiefung and M.M. Kostreva. Generalized P_0- and Z-matrices, *Linear Algebra and its Applications* 195 (1993) 165-179.

[32] A. A. Ebiefung and M. M. Kostreva. The generalized Leontief input-output model and its application to the choice of new technology, *Annals of Operations Research* 44 (1993) 161-172.

[33] F. Facchinei and J-S. Pang. Finite-Dimensional *Variational Inequalities and Complementarity Problems* (*Volumes* I and II. New York: Springer-Verlag, 2003.

[34] Y. Fathi. Computational complexity of LCPs associated with positive definite matrices. Mathematical Programming 17 (1979) 335-344.

[35] M.C. Ferris and C. Kanzow. Complementarity and related problems, in (P. Pardalos and M. Resende, eds.) *Handbook of Applied Optimization.* Oxford: Oxford University Press, 2002.

[36] M.C. Ferris and J-S. Pang. Engineering and economic applications of complementarity problems. SIAM Review 39 (1997) 669-713.

[37] M.C. Ferris, O.L. Mangasarian, and J-S. Pang, eds. *Complementarity: Applications, Algorithms and Extensions.* Dordrecht: Kluwer Academic Publishers, 2001.

[38] M.C. Ferris and J-S. Pang, eds. *Complementarity and Variational Inequalities.* Philadelphia: Society for Industrial and Applied Mathematics, 1997.

[39] www.gams/com/solvers/mpsge/syntax.htm.

[40] A. Gana *Studies in the Complementarity Problem.* Ph.D. thesis. University of Michigan, Ann Arbor, 1982.

[41] F. Giannessi and A. Maugeri. *Variational Inequalities and Network Equilibrium Problems.* New York: Plenum Press, 1995.

[42] M.S. Gowda. On reducing a monotone horizontal LCP to an LCP. *Applied Mathematics Letters.* 8 (1995) 97-100.

[43] M.S. Gowda. On the extended linear complementarity problem. *Mathematical Programming* 72 (1995) 33-50.

[44] O. Güler. Generalized linear complementarity problems. *Mathematics of Operations Research* 20 (1955) 441-448.

[45] S-M. Guu and R.W. Cottle. On a subclass of P_0. *Linear Algebra and its Applications* 223/224 (1995) 325-335.

[46] G.J. Habetler and A.J. Price. Existence theory for generalized nonlinear complementarily problems. *Joumal of Optimization Theory and Applicatins* 7 (1971) 223-239.

[47] G.J. Habetler and B.P. Szanc. Existence and uniqueness of solutions for the generalized linear complementarity problem. *Joumal of Optimization Theory and Applications* 84 (1995) 103-116.

[48] P.T. Harker and J-S. Pang. Finite-dimensional variational inequalities and nonlinear complementarity problems. *Mathematical Programming, Series B* 48 (1990) 161-220.

[49] P. Hartman and G. Stampacchia. On some nonlinear elliptic differential functional equations. *Acta Mathematica* 115 (1966) 153-188.

[50] D. Den Hertog, C. Roos, and T. Terlaky. The linear complementarity problem, sufficient matrices, and the criss-cross method. *Linear Algebra and its Applications* 187 (1993) 1-14.

[51] B.F Hobbs Linear complementarity models of Nash-Cournot competition in bilateral and POOLCO power markets. *IEEE Transactions on Power Systems* 16 (2001) 194-202.

[52] B.F. Hobbs, C.B. Metzler, and J-S. Pang. Strategie gaming analysis for electric power networks: An MPEC approach. IEEE Transactions on Power Systems 15 (2000) 638-645.

[53] J. Huang and J-S. Pang. Option pricing and linear complementarity. *The Journal of Computational Finance.* 2 (1998) 31-60.

[54] A. Ingleton. A problem in linear inequalities. Proceedings of the *London Mathematical Society* 16 (1966) 519-536.

[55] G. Isac. Complementarity Problems [Lecture Notes in Mathematics 1528]. Berlin: Springer Verlag, 1992.

[56] G. Isac. *Topological Methods in Complementarity Theory.* Dordrecht: Kluwer Academic Publishers, 2000.

[57] G. Isac, V.A. Bulavsky, and V.V. Kalashnikov. *Complementarity, Equilibrium, Efficiency, and Economics.* Dordrecht: Kluwer Academic Publishers, 2002.

[58] P. Jaillet, D. Lamberton, and B. Lapeyre. *Acta Applicandae Mathematicae* 21 (1990) 263-289.

[59] B.Jansen, C. Roos, and T. Terlaky. A family of polynomial affine scaling algorithms for positive semidefinite linear complementarity problems. *SIAM Joumal on Optimization* 7 (1997) 126-140.

[60] S. Karamardian. The nonlinear complementarity problem with applications, part 1. *Joumal of Optimization Theory and Applications* 4 (1969) 87-98.

[61] S. Karamardian. Generalized complementarity problems. *Joumal of Optimization Theory and Applications* 8 (1971) 161-168.

[62] S. Karamardian. The complementarity problem. *Mathematical Programming* 2 (1972) 107-129.

[63] N.K. Karmarkar. A new polynomial-time algorithm in linear programming. *Combinatoria* 4 (1984) 373-395.

[64] L. Khachiyan. A polynomial algorithm in linear programming (in Russian). *Doklady Akademia Nauk* SSSR 224 (1979) 1093-1096. [English translation: Soviet Mathematica Doklady 20 (1979) 191-194.]

[65] L. Khachiyan. Polynomial algorithms in linear programming (in Russian). *Zhumal Vychisitel'noi Matematiki i Matematicheskoi Fiziki* 20 (1980) 51-68. [English translation: U.S.S.R. *Computational Mathematics and Mathematical Physics* 20, (1980) 53-72.]

[66] V. Klee and G.J. Minty. How good is the simplex algorithm? in (O. Shisha, ed.) Inequalities III. New York: Academic Press, 1972.

[67] M. Kojima, N. Megiddo, T. Noma, and A. Yoshise. *A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems.* [Lecture Notes in Computer Science, v. 538.] Berlin: Springer-Verlag, 1091.

[68] C.D. Kolstad and L. Mathiesen. Computing Cournot-Nash equilibria. Operations Research 39 (1991) 739-748.

[69] D. Kuhn and R. Lówen. Piecewise affine bijections of R^n, and the equation $Sx^+ - Tx^- = y$. *Linear Algebra and its Applications* 96 (1987) 109-129.

[70] C.E. Lemke. Bimatrix equilibrium points and mathematical programming. *Management Science* 11 (1965) 681-689.

[71] E. Lemke. A survey of complementarity theory. In [16, pp. 213-239].

[72] E. Lemke and J.T. Howson, JR. Equilibrium points of bimatrix games. *SIAM Journal on Applied Mathematics* 12 (1964) 413-423.

[73] Q. Luo and J-S. Pang . Error bounds for analytic systems and their applications. *Mathematical Programming* 67 (1994) 1-28.

[74] Q. Luo, J-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints.* Cambridge: Cambridge University Press, 1996.

[75] O.L. Mangasarian and J-S. Pang. The extended linear complementarity problem. *SIAM Journal on Matrix Analysis and Applications* 16 (1995) 359-368.

[76] L. Mathiesen. Computational experience in solving equilibrium models by a sequence of linear complementarity problems. *Operations Research* 33 (1985) 1225-1250.

[77] L. Mathiesen. An algorithm based on a sequence of linear complementarity problems applied to a Walrasian equilibrium model: An example: *Mathematical Programming* 37 (1987) 1-18.

[78] C.B. Metzler. *Complementarity Modeis of Competitive Oligopolistic Electric Power Markets.* Ph.D. thesis. The Johns Hopkins University, Baltimore, Md., 2000.

[79] K.G. Murty. On the number of solutions to the linear complementarity problem and spanning properties of complementary cones. *Linear Algebra and its Applications* 5 (1972) 65-108.

[80] K.G. Murty. Computational complexity of complementary pivot methods. *Mathematical Programming Study* 7 (1978) 61-73.

[81] K.G. Murty *Linear Complementarity, Linear and Nonlinear Programming.* Berlin: Heldermann Verlag, 1988.

[82] Y. Nesterov and A. Nemirovski. *Interior-Point Algorithms in Convex Programming.* Philadelphia: SIAM, 1994.

[83] J-S. Pang and D. Chan. Iterative methods for variational and complementarity problems. Mathematical Programming 24 (1982) 284-313.

[84] T.D. Parsons. Applications of principal pivoting, in (H.W. Kuhn, ed.) *Proceedings of the Princeton Symposium on Mathematical Programming.* Princeton, N.J.: Princeton University Press, 1970.

[85] F. Pfeiffer And C. Glocker. *Multibody Dynamics with Unilateral Contacts.* New York: John Wiley & Sons, 1996.

[86] H. Pieper. *Algorithms for Mathematical Programs with Equilibrium Constraints with Applications to Deregulated Electricity Markets.* Ph.D. thesis. Stanford University, Stanford, Calif., 2001.

[87] F.A. Potra and R. Sheng. A large-step infeasible-interior-point method for the P_*-matrix LCP. *SIAM Journal on Optimization* 7 (1997) 318-335.

[88] H. Samelson, R.M. THRALL, AND O. *WESLER.* A partition theorem for Euclidean n-space. *Proceedings of the American Mathematical Society* 9 (1958) 805-807.

[89] R.E. Stone. *Geometric Aspects of the Linear Complementarity Problem.* Ph.D. thesis. Stanford University, Stanford, Calif., 1981.

[90] B.P. Szanc. *The Generalized Complementarity Problem.* Ph.D. thesis. Rensselaer Polytechnic Institute, Troy, N.Y., 1989.

[91] T. Terlaky. A convergent criss-cross method. *Mathematische Operationsforschung und Statistik, ser. Optimization* 16 (1985)683-690.

[92] M.J. Todd. The many facets of linear programming. Mathematical Programming 91 (2002) 401-416.

[93] R.H. Tütüncu and M.J. Todd. Reducing horizontal linear complementarity problems. *Linear Algebra and its Applications* 223/224 (1993) 717-729.

[94] H. Väliaho. Criteria for sufficient matrices. *Linear Algebra and its Applications* 233 (1996) 109-129.

[95] H. Väliaho. $\mathbf{P}_*$-matrices are just sufficient. *Linear Algebra and its Applications* 239 (1996) 103-108.

[96] H. Väliaho. Determining the handicap of a sufficient matrix. *Linear Algebra and its Applications* 253 (1997) 279-298.

[97] S.J. Wr.Ight. An infeasible-interior-point algorithm for linear complementarity problems. *Mathematical Programming* 67 (1994) 29-51.

[98] S.J. Wrjght. *Primal-Dual Interior-Point Methods.* Philadelphia: SIAM, 1997.

[99] Y. Ye *Interior Algorithms for Linear, Quadratic, and Linearly Constrained Convex Programs.* Ph.D. thesis. Stanford University, Stanford, Calif., 1988.

[100] Y. Ye. A further result on the potential reduction algorithm for the P-matrix linear complementarity problem in (P.M. Pardalos, ed.) *Advances in Optimization and Parallel Computing.* Amsterdam: North-Holland, 1992, pp. 310-316.

[101] Y. Ye. *Interior Point Methods.* New York: John Wiley & Sons, 1997.

[102] Y. Ye And P. Pardalos. A class of linear complementarity problems solvable in polynomial time. *Linear Algebra and its Applications* 152 (1991) 3-17.

[103] Y. Zhang. On the convergence of a class of infeasible interior-point algorithms for the horizontal linear complementarity problem. *SIAM Journal on Optimization* 4 (1994) 208-227.

VARIATIONAL INEQUALITIES IN VECTOR OPTIMIZATION

G.P. Crespi,[1] I. Ginchev[2] and M. Rocca[3]
University of Valle d'Aosta, Faculty of Economics, Aosta, Italy;[1] Technical University of Varna, Dept. of Mathematics, Varna, Bulgaria;[2] University of Insubria, Dept. of Economics, Varese, Italy[3]

Abstract: In this paper we investigate the links among generalized scalar variational inequalities of differential type, vector variational inequalities and vector optimization problems. The considered scalar variational inequalities are obtained through a nonlinear scalarization by means of the so called "oriented distance" function [14,15].

In the case of Stampacchia-type variational inequalities, the solutions of the proposed ones coincide with the solutions of the vector variational inequalities introduced by Giannessi [8]. For Minty-type variational inequalities, analogous coincidence happens under convexity hypotheses. Furthermore, the considered variational inequalities reveal useful in filling a gap between scalar and vector variational inequalities. Namely, in the scalar case Minty variational inequalities of differential type represent a sufficient optimality condition without additional assumptions, while in the vector case the convexity hypothesis is needed. Moreover it is shown that vector functions admitting a solution of the proposed Minty variational inequality enjoy some well-posedness properties, analogously to the scalar case [4].

1. INTRODUCTION

Given a map F from $\mathbb{R}^n$ to $\mathbb{R}^n$ and a nonempty set $K \subseteq \mathbb{R}^n$, we say that a point $x^* \in K$ is a solution of a Stampacchia variational inequality when [13]:

$$VI(F,K) \qquad \langle F(x^*), y - x^* \rangle \geq 0, \qquad \forall y \in K .$$

Analogously we say that $x^* \in K$ is a solution of a Minty variational inequality when [17]:

$$MVI(F,K) \quad \langle F(y), x^* - y \rangle \leq 0, \qquad \forall y \in K .$$

In particular, when the variational inequality admits a primitive minimization problem (that is the function f to minimize is such that $F = f'$) and K is a convex set, $VI(f',K)$ and $MVI(f',K)$ have strong links with this problem. Roughly speaking, $VI(f',K)$ is a necessary condition for the minimization of the function f over the set K, which becomes also sufficient when f is convex. On the contrary, $MVI(f',K)$ is a sufficient condition for the minimization of f over the set K, which becomes necessary if f is convex. Recently it has been observed also [4] that the existence of a solution of $MVI(f',K)$ has some implications on the well-posedness of the related optimization problem.

Variational inequalities in the sense of Minty and Stampacchia have been extended to the case where F is a point-to-set map from $\mathbb{R}^n$ to $2^{\mathbb{R}^n}$ (see for instance [10]). In this case a point $x^* \in K$ is a solution of a Stampacchia variational inequality, when there exists $\xi^* \in F(x^*)$, such that $\langle \xi^*, y - x^* \rangle \geq 0$, $\forall y \in K$. Analogously $x^* \in K$ is said a solution of a Minty variational inequality when it holds $\langle v, x^* - y \rangle \leq 0$, $\forall y \in K$ and $\forall v \in F(y)$.

Furthermore a vector extension of Minty and Stampacchia variational inequalities has been introduced by F. Giannessi [8,9], who has also given some links between the solutions of vector variational inequalities and the solutions of a vector optimization problem. Roughly speaking, it has been proved that Stampacchia vector variational inequalities represent a necessary condition for optimality (that becomes sufficient under convexity assumptions). Analogously to the scalar case it is proved that Minty vector variational inequality is a necessary and sufficient optimality condition under convexity assumptions. But a gap with the scalar case arises, namely that convexity is needed also to prove that Minty vector variational inequality is a sufficient optimality condition.

In this paper we introduce a generalization of scalar variational inequalities (of differential type) and we investigate their links with vector variational inequalities and vector optimization problems. The considered variational inequalities are obtained through a nonlinear scalarization, which makes use of the so called "oriented distance" function [14,15]. We show that the solutions of the proposed variational inequalities coincide with the solutions of some variational inequalities for point to set maps. In the case of

Stampacchia-type variational inequalities the links of the proposed ones with vector optimization coincide with those holding for vector valued ones. For Minty-type variational inequalities analogous coincidence holds under convexity assumptions. We show that if the convexity hypothesis is dropped, the proposed Minty variational inequalities provide a stronger solution concept with respect to Minty vector variational inequalities and are useful in filling the previously mentioned gap.

Moreover it is shown that vector functions admitting a solution of the proposed Minty variational inequality enjoy well-posedness properties analogously to the scalar case [4].

The paper is structured as follows. In section 2 we recall some known results about Minty variational inequalities and scalar optimization. Section 3 presents the concept of "oriented distance function" and its application in the scalarization of vector optimality concepts. Section 4 deals with variational inequalities and vector optimization.

2. SCALAR VARIATIONAL INEQUALITIES

We are concerned with the following optimization problem:

$$P(\phi,K) \qquad \min \phi(x), \qquad x \in K \subseteq \mathbb{R}^n,$$

where $\phi : \mathbb{R}^n \to \mathbb{R}$. A point $x^* \in K$ is a solution of $P(\phi,K)$ when $\phi(x) - \phi(x^*) \geq 0, \forall x \in K$. The solution is strong when $\phi(x) - \phi(x^*) > 0, \ \forall x \in K \setminus \{x^*\}$.

In this section we assume that ϕ is a function defined and directionally differentiable on an open set containing K. We recall that the directional derivative of ϕ at a point x in the direction $d \in \mathbb{R}^n$ is defined as:

$$\phi'(x;d) = \lim_{t \to 0^+} \frac{\phi(x+td) - \phi(x)}{t},$$

when this limit exists and is finite. We deal with the following variational problems:

$VI(\phi',K)$ Find a point $x^* \in K$ such that $\phi'(x^*; y - x^*) \geq 0$, $\forall y \in K$.

$MVI(\phi',K)$ Find a point $x^* \in K$ such that $\phi'(y; x^* - y) \leq 0$, $\forall y \in K$.

Observe that the previous problems reduce to the classical Stampacchia and Minty variational inequalities when ϕ is differentiable, which induces us to use the classical abbreviations *VI* and *MVI*.

Definition 1.

i) Let K be a nonempty subset of $\mathbb{R}^n$. The set $\ker K$ consisting of all $x \in K$ such that $(y \in K, t \in [0,1]) \Rightarrow x + t(y - x) \in K$ is called the kernel of K.

ii) A nonempty set K is star-shaped if $\ker K \neq \varnothing$.

In the following we use the abbreviation st-sh for star-shaped. It is known (see e.g. [18]) that the set $\ker K$ is convex for an arbitrary st-sh set K.

Definition 2. *A function ϕ defined on $\mathbb{R}^n$ is called increasing along rays at a point x^* (for short, $f \in IAR(x^*)$) if the restriction of this function on the ray $\mathbb{R}_{x^*,x} = \{x^* + \alpha x \mid \alpha \geq 0\}$ is increasing for each $x \in \mathbb{R}^n$. (A function g of one real variable is called increasing if $t_2 \geq t_1$ implies $g(t_2) \geq g(t_1)$.)*

Definition 3. *Let $K \subseteq \mathbb{R}^n$ be a st-sh set and $x^* \in \ker K$. A function ϕ defined on K is called increasing along rays at x^* (for short, $\phi \in IAR(K, x^*)$), if the restriction of this function on the intersection $\mathbb{R}_{x^*,x} \cap K$ is increasing, for each $x \in K$.*

Proposition 1. [4]

i) If $\phi \in IAR(K, x^*)$, then x^* is a solution of $P(\phi, K)$.

ii) $\phi \in IAR(K, x^*)$ if and only if $x^* \in \ker lev_{\leq c}\phi$ for every $c \geq \phi(x^*)$ (here $lev_{\leq c}\phi := \{x \in K \mid \phi(x) \leq c\}$).

The following result can be deduced from Theorem 2 in [4].

Proposition 2.

i) Let x^* be a solution of $MVI(\phi', K)$ and $x^* \in \ker K$. Then $\phi \in IAR(K, x^*)$.

ii) Let $\phi \in IAR(K, x^*)$. Then x^* is a solution of $MVI(\phi', K)$.

Remark 1. If x^* is a strong solution of $MVI(\phi', K)$ (i.e. $\phi'(y; x^* - y) < 0, \forall y \in K \setminus \{x^*\}$), in the previous Proposition we can easily

conclude with the same proof that ϕ is strictly increasing along rays starting at x^*.

The following result has an immediate proof and we omit it.

Proposition 3.

i) Let $x^* \in \ker K$. If $x^* \in K$ is a solution of $P(\phi, K)$, then x^* solves $VI(\phi', K)$.

ii) Let K be a convex set. If ϕ is convex and $x^* \in K$ solves $VI(\phi', K)$ then x^* is a solution of $P(\phi, K)$.

Proposition 4.

i) Let $x^* \in \ker K$. If $x^* \in K$ is a (strong) solution of $MVI(\phi', K)$ then x^* is a (strong) solution of $P(\phi, K)$.

ii) Let K be a convex set. If $x^* \in K$ solves $P(\phi, K)$ and ϕ is convex, then x^* solves $MVI(\phi, K)$.

Proof:

i) Since x^* is a solution of $MVI(\phi', K)$, then $\phi \in IAR(K, x^*)$ and hence x^* solves $P(\phi, K)$. Analogously when x^* is a strong solution of $MVI(\phi', K)$.

ii) If ϕ is convex and $x^* \in K$ solves $P(\phi, K)$, then $\phi \in IAR(K, x^*)$ and so x^* solves $MVI(\phi', K)$.

□

Problems $VI(\phi', K)$ and $MVI(\phi', K)$ can be linked by the following result, analogous to the classical Minty's Lemma.

Proposition 5.

i) Let $x^* \in \ker K$. If $x^* \in K$ solves $MVI(\phi', K)$ and $\phi'(\cdot; d)$ is upper semicontinuous (u.s.c.) along rays starting at x^* for every $d \in \mathbb{R}^n$, then x^* is a solution of $VI(\phi', K)$.

ii) Let K be a convex set. If $x^* \in K$ solves $VI(\phi', K)$ and ϕ is convex, then x^* solves $MVI(\phi', K)$.

Proof:

i) We begin proving that under the assumptions, if $x^* \in K$ solves $MVI(\phi', K)$, then x^* is such that $\phi'(y; y - x^*) \geq 0, \forall y \in K$. Since x^* solves $MVI(\phi', K)$, we know that $\phi \in IAR(K, x^*)$ and since

$x^* \in \ker K$, the set $\{\mathbb{R}_{x^*,y} \cap K\}$ is convex and hence has a nonempty relative interior $\mathrm{ri}\{\mathbb{R}_{x^*,y} \cap K\}$. If $y \in \mathrm{ri}\{\mathbb{R}_{x^*,y} \cap K\}$, for $t > 0$ "small enough" we have $y + t(y - x^*) = x^* + (1+t)(y - x^*) \in \mathbb{R}_{x^*,y} \cap K$ and hence $\phi(y + t(y - x^*)) \geq \phi(y)$, from which it follows easily $\phi'(y; y - x^*) \geq 0$. Let now $y \in \{\mathbb{R}_{x^*,y} \cap K\} \setminus \mathrm{ri}\{\mathbb{R}_{x^*,y} \cap K\}$. Hence we have $y = \lim y_k$, for some sequence $y_k \in \mathrm{ri}\{\mathbb{R}_{x^*,y} \cap K\}$, that is $y_k = x^* + t_k(y - x^*)$. It holds:

$$0 \leq \phi'(y_k; y_k - x^*) = \phi'(x^* + t_k(y - x^*); t_k(y - x^*))$$

and hence:

$$0 \leq \limsup_{k \to +\infty} \phi'(x^* + t_k(y - x^*); y - x^*) \leq \phi'(y; y - x^*),$$

where the last inequality follows since $\phi'(\cdot, d)$ is u.s.c. along rays starting at x^*.

Let now $z \in K$ and consider the point $z(t) := x^* + t(z - x^*)$, $t \in (0,1]$. We have $0 \leq \phi'(z(t); z(t) - x^*) = \phi'(z(t); t(z - x^*))$ and hence $\phi'(z(t); (z - x^*)) \geq 0$. Passing to the limit as $t \to 0^+$ and taking into account the fact that $\phi'(\cdot; y - x^*)$ is u.s.c. along rays starting at x^*, we get $\phi'(x^*; y - x^*) \geq 0$.

ii) If $x^* \in K$ solves $VI(\phi', K)$, then x^* solves $P(\phi, K)$ and $\phi \in IAR(K, x^*)$. Hence x^* solves $MVI(\phi', K)$.

□

Now we recall the notion of Tykhonov well-posedness for problem $P(\phi, K)$.

Definition 4. *A sequence $x^k \in K$ is a minimizing sequence for $P(\phi, K)$, when $\phi(x^*) \to \inf_K \phi(x)$.*

Definition 5. *Problem $P(\phi, K)$ is Tykhonov well-posed when it admits a unique solution x^* and every minimizing sequence for $P(\phi, K)$ converges to x^*.*

For $\varepsilon > 0$ we set:

$$L^{\phi}(\varepsilon) = \{x \in K : \phi(x) \leq \inf_{K} \phi + \varepsilon\}.$$

Theorem 1. [7]

i) If $P(\phi, K)$ is Tykhonov well-posed, then $\operatorname{diam} L^{\phi}(\varepsilon) \to 0$ as $\varepsilon \to 0^{+}$, or equivalently $\inf_{\varepsilon > 0} \operatorname{diam} L^{\phi}(\varepsilon) = 0$ (here $\operatorname{diam} A$ denotes the diameter of the set A).

ii) Let ϕ be lower semicontinuous and bounded from below. If $\inf_{\varepsilon > 0} \operatorname{diam} L^{\phi}(\varepsilon) = 0$, then $P(f, K)$ is Tykhonov well-posed.

Theorem 2. *[4] Let K be a closed subset of $\mathbb{R}^n$, $x^* \in \ker K$ and $f \in IAR(K, x^*)$. If $P(\phi, K)$ admits a unique solution, then it is Tykhonov well-posed.*

3. SCALAR CHARACTERIZATIONS OF VECTOR OPTIMALITY CONCEPTS

Let C be a closed, convex, pointed cone with nonempty interior. Let M be any of the cones C^c, $C \setminus \{0\}$, C and $\operatorname{int} C$. The vector optimization problem (see e.g. [19]) corresponding to M, where $f : \mathbb{R}^n \to \mathbb{R}^l$, is written as:

$$VP(f, K) \qquad v\text{-}\min_{M} f(x), \qquad x \in K.$$

This amounts to find a point $x^* \in K$ (called the optimal solution), such that there is no $y \in K \setminus \{x^*\}$ with $f(y) \in f(x^*) - M$. The optimal solutions of the vector problem corresponding to $-C^c$ (respectively, $C \setminus \{0\}$, C and $\operatorname{int} C$) are called ideal solutions (respectively, efficient solutions, strongly efficient solutions and weakly efficient solutions). We will denote the efficient solutions as e-solutions and the weakly efficient solutions as w-solutions.

Let us now recall the notion of "oriented distance" function, introduced by Hiriart-Hurruty [14,15].

Definition 6. *For a set $A \subseteq \mathbb{R}^l$ let the oriented distance function $\Delta_A : \mathbb{R}^l \to \mathbb{R} \cup \{\pm\infty\}$ be defined as:*

$$\Delta_A(y) = d_A(y) - d_{\mathbb{R}^l \setminus A}(y),$$

where $d_A(y) := \inf_{a \in A} \|y - a\|$ is the distance from the point y to the set A.

Function Δ_A has been recently used in [20] to characterize several notions of efficient point of a given set $D \subseteq \mathbb{R}^l$. In [12] it has been proved that when A is a closed, convex, pointed cone, then we have:

$$\Delta_{-A}(y) = \max_{\xi \in A' \cap S} \langle \xi, y \rangle,$$

where $A' := \{x \in \mathbb{R}^l \mid \langle x, a \rangle \geq 0, \forall a \in A\}$ is the positive polar of the set A and S the unit sphere in $\mathbb{R}^l$.

In this section we use function Δ_{-A} in order to give scalar characterizations of several notions of efficiency for problem $VP(f,K)$. Furthermore, some results characterize pointwise well-posedness of problem $VP(f,K)$ [6] through function Δ_{-A}.

Given a point $\hat{x} \in K$, consider the function:

$$\phi_{\hat{x}}(x) = \max_{\xi \in C' \cap S} \langle \xi, f(x) - f(\hat{x}) \rangle.$$

Clearly $\phi_{\hat{x}}(x) = \Delta_{-C}(f(x) - f(\hat{x}))$. We consider the problem:

$$P(\phi_{\hat{x}}, K) \qquad \min \phi_{\hat{x}}(x), \qquad x \in K.$$

The following Theorem can be found in [11].

Theorem 3.

i) The point $x^* \in K$ is a strong e-solution of $VP(f,K)$ if and only if x^* is a strong solution of $P(\phi_{x^*}, K)$.

ii) The point $x^* \in K$ is a w-solution of $VP(f,K)$ if and only if x^* is a solution of $P(\phi_{x^*}, K)$.

The next result slightly extends Theorem 3.

Theorem 4.

i) The point $x^* \in K$ is a strong e-solution of $VP(f,K)$ if and only if there exists a point $\hat{x} \in K$, such that x^* is a strong solution of $P(\phi_{\hat{x}}, K)$.

ii) The point $x^* \in K$ is a w-solution of $VP(f,K)$ if and only if there exists a point $\hat{x} \in K$, such that x^* is a solution of $P(\phi_{\hat{x}}, K)$.

Proof. We prove only i), since the proof of ii) is analogous. Let x^* be a strong e-solution of $VP(f,K)$. Then from Theorem 3 we know that x^* is a strong solution of $P(\phi_{x^*},K)$ and necessity is proved.

Now, assume that for some $\hat{x} \in K$, x^* is a strong solution of $P(\phi_{\hat{x}},K)$, i.e. $\phi_{\hat{x}}(x^*) < \phi_{\hat{x}}(x), \forall x \in K \setminus \{x^*\}$, or equivalently:

$$\max_{\xi \in C' \cap S} \langle \xi, f(x^*) - f(\hat{x}) \rangle < \max_{\xi \in C' \cap S} \langle \xi, f(x) - f(\hat{x}) \rangle =$$
$$\max_{\xi \in C' \cap S} \langle \xi, f(x) - f(x^*) + f(x^*) - f(\hat{x}) \rangle \leq$$
$$\max_{\xi \in C' \cap S} \langle \xi, f(x) - f(x^*) \rangle + \max_{\xi \in C' \cap S} \langle \xi, f(x^*) - f(\hat{x}) \rangle, \forall x \in K \setminus \{x^*\}.$$

Hence $\max_{\xi \in C' \cap S} \langle \xi, f(x) - f(x^*) \rangle > 0, \forall x \in K \setminus \{x^*\}$, i.e. x^* is a strong solution of $P(\phi_{x^*}(x),K)$. From the previous Theorem we obtain that x^* is a strong e-solution of $VP(f,K)$.

□

Now we recall the notion of pointwise well-posedness for problem $VP(f,K)$ [6]. Let $k \in C$, $\alpha > 0$, $v \in K$ and set:

$$L(v,k,\alpha) = \{x \in K \mid f(x) \in f(v) + \alpha k - C\}.$$

Definition 7. *Problem $VP(f,K)$ is said to be pointwise well-posed at the e-solution x^* when:*

$$\inf_{\alpha > 0} \operatorname{diam} L(x^*,k,\alpha) = 0, \quad \text{for each } k \in C.$$

Theorem 5. *Let f be a continuous function and let $x^* \in K$ be an e-solution of $VP(f,K)$. Problem $VP(f,K)$ is pointwise well-posed at x^* if and only if problem $P(\phi_{x^*},K)$ is Tykhonov well-posed.*

Proof. Since x^* is an e-solution of $VP(f,K)$, then x^* is also a w-solution of $VP(f,K)$ and hence (Theorem 3) a solution of $P(\phi_{x^*},K)$, with $\phi_{x^*}(x^*) = 0$. Let $P(\phi_{x^*},K)$ be Tykhonov well-posed. If for some $k \in C$ and $\alpha > 0$, $x \in L(x^*,k,\alpha)$, then, for some $c \in C$, we have $f(x) - f(x^*) = -c + \alpha k$ and so:

$$\phi_{x^*}(x) = \max_{\xi \in C' \cap S} \langle \xi, f(x) - f(x^*) \rangle = \max_{\xi \in C' \cap S} \langle \xi, -c + \alpha k \rangle \leq$$

$$\leq \max_{\xi \in C' \cap S} \langle \xi, -c \rangle + \alpha \max_{\xi \in C' \cap S} \langle \xi, k \rangle \leq \alpha \max_{\xi \in C' \cap S} \langle \xi, k \rangle.$$

(the last inequality follows since for every $\xi \in C' \cap S$, we have $\langle \xi, -c \rangle \leq 0$). Hence we have $x \in L^{\phi_{x^*}}\left(\alpha \max_{\xi \in C' \cap S} \langle \xi, k \rangle\right)$. It follows that $\forall \alpha > 0$ and $\forall k \in C$, we have:

$$L(x^*, k, \alpha) \subseteq L^{\phi_{x^*}}\left(\alpha \max_{\xi \in C' \cap S} \langle \xi, k \rangle\right)$$

and so, $\forall k \in C$:

$$\inf_{\alpha > 0} \operatorname{diam} L(x^*, k, \alpha) \leq \inf_{\alpha > 0} \operatorname{diam} L^{\phi_{x^*}}\left(\alpha \max_{\xi \in C' \cap S} \langle \xi, k \rangle\right).$$

Since $P(\phi_{x^*}, K)$ is Tykhonov well-posed, we have:

$$\inf_{\alpha > 0} \operatorname{diam} L^{\phi_{x^*}}\left(\alpha \max_{\xi \in C' \cap S} \langle \xi, k \rangle\right) = 0$$

and hence $\inf_{\alpha > 0} \operatorname{diam} L(x^*, k, \alpha) = 0$, that is $VP(f, K)$ is pointwise well-posed at x^*.

Assume now that $VP(f, K)$ is pointwise well-posed at x^*. We prove that there exists a point $\bar{k} \in \operatorname{int} C$ such that for every $\alpha > 0$ it holds:

$$L^{\phi_{x^*}}(\alpha) \subseteq L(x^*, \bar{k}, \alpha).$$

For every $k \in \operatorname{int} C$ and $\xi \in C' \cap S$ we have $\langle \xi, k \rangle > 0$. Choose a vector $\bar{k} \in \operatorname{int} C$ with $\min_{\xi \in C' \cap S} \langle \xi, \bar{k} \rangle > 1$. If, ab absurdo, for some $\alpha > 0$ there exists a point $x \in L^{\phi_{x^*}}(\alpha) \setminus L(x^*, \bar{k}, \alpha)$, then we have $f(x) - f(x^*) \notin -C + \alpha \bar{k}$. It follows the existence of a point $\bar{\xi} \in C' \cap S$ such that $\langle \bar{\xi}, f(x) - f(x^*) - \alpha \bar{k} \rangle > 0$ and so:

$$\langle \bar{\xi}, f(x) - f(x^*) \rangle > \alpha \langle \bar{\xi}, \bar{k} \rangle,$$

from which:

$$\max_{\xi\in C'\cap S}\left\langle \xi, f(x)-f(x^*)\right\rangle > \alpha\left\langle \bar{\xi},\bar{k}\right\rangle \geq \alpha \min_{\xi\in C'\cap S}\left\langle \xi,\bar{k}\right\rangle > \alpha,$$

that is $\phi_{x^*}(x)-\phi_{x^*}(x^*)>\alpha$ and hence the absurdo $x\notin L^{\phi_{x^*}}(\alpha)$. So we have:

$$L^{\phi_{x^*}}(\alpha)\subseteq L(x^*,\bar{k},\alpha),\forall\alpha>0.$$

Since $VP(f,K)$ is pointwise well-posed at x^*, we have $\inf_{\alpha>0}\operatorname{diam}L(x^*,\bar{k},\alpha)=0$ and so also $\inf_{\alpha>0}\operatorname{diam}L^{\phi_{x^*}}(\alpha)=0$, which, recalling Theorem 1, completes the proof. □

In the scalar case it is known that if ϕ is a convex function with a unique (strong) minimizer over K, then problem $P(\phi,K)$ is Tykhonov well-posed. Now we extend this property to the vector case.

Definition 8. *The function $f:K\subseteq\mathbb{R}^n\to\mathbb{R}^l$ is said to be C–convex when:*

$$f(\lambda x+(1-\lambda)y)-[\lambda f(x)+(1-\lambda)f(y)]\in -C \quad \forall x,y\in K,\quad \forall\lambda\in[0,1].$$

The following result has an almost immediate proof and we omit it.

Proposition 6. *Let $f:\mathbb{R}^n\to\mathbb{R}^l$ be a C-convex function. Then $\forall\hat{x}\in K$, the function $\phi_{\hat{x}}(x)$ is convex.*

Theorem 6. *If $f:\mathbb{R}^n\to\mathbb{R}^l$ is C-convex, then f is pointwise well-posed at any strong e-solution of $VP(f,k)$.*

Proof: Assume that f is C-convex and let x^* be a strong e-solution of $VP(f,K)$. Then, from Theorem 3 x^* is the unique minimizer of the convex function $\phi_{x^*}(x)$ over K and a classical result (see [7]) states that problem $P(\phi_{x^*},K)$ is Tykhonov well-posed. The thesis then follows from Theorem 5

□

Remark 2. If we consider $C=\mathbb{R}^l_+$ and define $\tilde{\phi}_{\hat{x}}(x)=\max\{f_i(x)-f_i(\hat{x}),i=1,...,l\}$, then it can be proved [4] that in the results presented in this paper, function $\phi_{\hat{x}}(x)$ can be replaced by $\tilde{\phi}_{\hat{x}}(x)$.

4. VARIATIONAL INEQUALITIES AND VECTOR OPTIMIZATION

Vector variational inequalities (of Stampacchia type) have been first introduced in [8]. Later a vector formualtion of Minty variational inequality has been proposed as well (see e.g. [9]). Both the inequalities involve a matrix valued function $F:\mathbb{R}^n \to \mathbb{R}^{l\times n}$ and a feasible region $K \subseteq \mathbb{R}^n$. We consider the following sets:

$$\Omega(x) := \{u \in \mathbb{R}^l \mid u = F(x)(y-x), y \in K\},$$
$$\Theta(x) := \{w \in \mathbb{R}^l \mid w = F(y)(y-x), y \in K\}.$$

Definition 9.

i) A vector $x^* \in K$ is a solution of a strong vector variational inequality of Stampacchia type when:

$$VVI^s(F,K) \qquad \Omega(x^*) \cap (-C) = \{0\}.$$

ii) A vector $x^* \in K$ is a solution of a weak vector variational inequality of Stampacchia type when:

$$VVI(F,K) \qquad \Omega(x^*) \cap (-\mathrm{int}C) = \varnothing.$$

Definition 10.

i) A vector $x^* \in K$ is a solution of a strong vector variational inequality of Minty type when:

$$MVVI^s(F,K) \qquad \Theta(x^*) \cap (-C) = \{0\}.$$

ii) A vector $x^* \in K$ is a solution of a weak vector variational inequality of Minty type when:

$$MVVI(F,K) \qquad \Theta(x^*) \cap (-\mathrm{int}C) = \varnothing.$$

In the sequel we will deal mainly with weak vector variational inequalities of Stampacchia and Minty type (for short VVI and MVVI, respectively).

The following result (see [9]) extends the classical Minty's Lemma to the vector case.

Lemma 1. *Let K be a convex set and let F be hemicontinuous and C-monotone. Then x^* is a solution of $MVVI(F,K)$ if and only if it solves $VVI(F,K)$.*

Similarly to the scalar case, we consider a function $f:\mathbb{R}^n \to \mathbb{R}^l$, that we assume to be differentiable on an open set containing K. We denote by f' the Jacobian of f.

The following results (see [8,9]) link $VVI(f',K)$ and $MVVI(f',K)$ to vector optimization.

Proposition 7. *Let $f:\mathbb{R}^n \to \mathbb{R}^l$ be differentiable on an open set containing K.*

i) If $x^* \in \ker K$ is a a w-solution of $VP(f,K)$, then it solves also $VVI(f',K)$.

ii) If K is a convex set, f is C–convex and x^* is a solution of $VVI(f',K)$, then it is a w-solution of $VP(f,K)$.

Some refinements of the relations between VVI and efficiency have been given in [2].

Proposition 8. *Let $C=\mathbb{R}^l_+$ and K be a convex set. If f is C–convex and differentiable on an open set containing K, then $x^* \in K$ is a w-solution of $VP(f,K)$ if and only if it is a solution of $MVVI(f',K)$.*

Remark 3. The previous result has been extended to an arbitrary ordering cone C (closed, convex, pointed and with nonempty interior) in [3], under the hypothesis that f' is hemicontinuous at x^*, i.e. that the restriction of f' on any ray starting at x^* is continuous. This assumption is not really additional with respect to Proposition 8, since the Jacobian of every $\mathbb{R}^l_+$-convex and differentiable function is hemicontinuous.

In particular, Proposition 8 gives an extension to the vector case of Proposition 4 (for differentiable functions). Anyway, in Proposition 8, convexity is needed also for proving that $MVVI(f',K)$ is a sufficient condition for optimality, while in the scalar case, convexity is needed only in the proof of the necessary part.

The next example shows that the convexity assumption in Proposition 8 cannot be dropped.

Example 1. *Let* $C=\mathbb{R}^2_+$, $K=[-\frac{2}{\pi},0]$ *and consider a function* $f:\mathbb{R}\to\mathbb{R}^2$, $f(x)=\begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix}$, *defined as follows. We set:*

$$f_1(x)=\begin{cases} x^2\sin\dfrac{1}{x}-x^2, & x\neq 0 \\ 0, & x=0 \end{cases}$$

and observe that $-2x^2\le f_1(x)\le 0$, $\forall x\in K$ *and* f_1 *is differentiable on* K. *Function* f_1 *has a countable number of local minimizers and of local maximizers over* K. *The local maximizers of* f_1 *are the points* $y_k=-\frac{1}{\frac{\pi}{2}+2k\pi}$, $k=0,1,...$ *and* $f_1(y_k)=0$. *If we denote by* x_k, $k=0,1,...$ *the local minimizers of* f *over* K, *we have* $y_k<x_k<y_{k+1}$, $\forall k=0,1,...$.

Function f_2 *is defined on* K *as:*

$$f_2(x)=\begin{cases} -\dfrac{f_1(x_k)}{2}\left[\cos\left(\dfrac{\pi x}{x_k-y_k}+\dfrac{\pi(x_k-2y_k)}{x_k-y_k}\right)-1\right], & x\in[y_k,x_k) \\ -\dfrac{f_1(x_{k+1})}{2}\left[\cos\left(\dfrac{\pi x}{y_{k+1}-x_k}+\dfrac{\pi(2y_{k+1}-3x_k)}{y_{k+1}-x_k}\right)-1\right], & x\in[x_k,y_{k+1}) \\ 0, & x=0 \end{cases}$$

for $k=0,1,...$. *It is easily seen that also* f_2 *is differentiable on* K. *The graphs of* f_1 *and* f_2 *are plotted in figure 1.*

The points $x\in[-\frac{2}{\pi},x_0]$ *are w-solutions, while the other points in* K *are not w-solutions. In particular,* $x^*=0$ *is an ideal maximal point (i.e.* $f(x)-f(x^*)\in\mathbb{R}^2_-$, $\forall x\in K$*). Anyway, it is easy to see that any* point *of* K *is a solution of* $MVVI(f',K)$.

In order to fill the gap between Proposition 8 and the analogous scalar result, we consider function $\phi_{\bar{x}}$ introduced in the previous section. From now on we assume that f is a function of class C^1 on an open set containing K. The following Theorem resumes some classical properties of function $\phi_{\bar{x}}$.

Theorem 7. *[5]*

(i) $\phi_{\bar{x}}$ *is directionally differentiable and*

$$\phi'_{\hat{x}}(x;d) = \max_{\xi \in R_{\hat{x}}(x)} \xi^{\top} f'(x)d,$$

where $R_{\hat{x}}(x) = \{\xi \in C' \cap S : \phi_{\hat{x}}(x) = \langle \xi, f(x) - f(\hat{x}) \rangle\}$.

ii) $\phi'_{\hat{x}}(x;\cdot)$ *is sublinear and can be expressed as*:

$$\phi'_{\hat{x}}(x;d) = \max_{v \in \partial \phi_{\hat{x}}(x)} \langle v, d \rangle,$$

where $\partial \phi_{\hat{x}}(x) = \operatorname{conv}\{\xi^{\top} f'(x), \xi \in R_{\hat{x}}(x)\}$ *(here* conv A *denotes the convex hull of the set A).*

Now we consider the following problems:

$VI(\phi'_{\hat{x}}, K)$ For a given $\hat{x} \in K$, find a point $x^* \in K$ such that $\phi'_{\hat{x}}(x^*; y - x^*) \geq 0, \forall y \in K$.

$MVI(\phi'_{\hat{x}}, K)$ For a given $\hat{x} \in K$, find a point $x^* \in K$ such that $\phi'_{\hat{x}}(y; x^* - y) \leq 0, \forall y \in K$

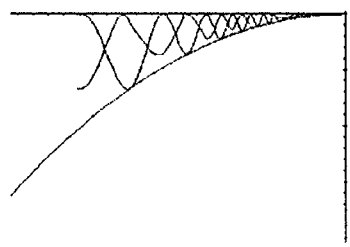


Figure 1. $f_1(x)$ and $f_2(x)$.

Remark 4. Clearly, Proposition 5 provides some links between these two problems. Since, under the made assumptions, $\phi'_{\hat{x}}(.;d)$ is u.s.c. [5], then any solution of Problem MVI $(\phi'_{\hat{x}}, K)$ is a solution of VI $(\phi'_{\hat{x}}, K)$. Conversely, if f is C-convex, then $\phi_{\hat{x}}$ is convex (see Proposition 6) and hence Proposition 5 states that every solution of Problem VI $(\phi'_{\hat{x}}, K)$ is also a solution of MVI $(\phi'_{\hat{x}}, K)$.

The next results state the equivalence betweeen the previous problems and generalized variational inequalities for point to set maps [10].

Proposition 9. *Let K be a convex set. Problem $VI(\phi'_{\hat{x}}, K)$ is equivalent to the following generalized variational inequality of Stampacchia type :*

$VI(\partial\phi_{\hat{x}}, K)$ For some given $\hat{x} \in K$, find a point $x^* \in K$, such that $\exists v \in \partial\phi_{\hat{x}}(x^*)$ for which $\langle v, x^* - y \rangle \leq 0$.

Proof. $VI(\partial\phi_{\hat{x}}, K) \Rightarrow VI(\phi'_{\hat{x}}, K)$ is obvious. Instead, assume that x^* solves $VI(\phi'_{\hat{x}}, K)$, i.e. $\phi'_{\hat{x}}(x^*; y - x^*) \geq 0, \forall y \in K$. This means:

$$\max_{v \in \partial\phi_{\hat{x}}(x^*)} \langle v, y - x^* \rangle \geq 0, \forall y \in K$$

and the result follows from Lemma 1 in [1]. □

Similarly we get the following result which we state without the obvious proof.

Proposition 10. *Let K be a convex set. Problem $MVI(\phi'_{\hat{x}}, K)$ is equivalent to the following generalized variational inequality of Minty type:*

$MVI(\partial\phi_{\hat{x}}, K)$ For a given $\hat{x} \in K$, find a point $x^* \in K$ such that for every $v \in \partial\phi_{\hat{x}}(y)$ and for every $y \in K$.

Now we prove that the solutions of problem $VI(\phi'_{\hat{x}}, K)$ coincide with the solutions of $VVI(f', K)$.

Proposition 11. *Let K be a convex set. If $x^* \in K$ solves problem $VI(\phi'_{\hat{x}}, K)$ for some $\hat{x} \in K$, then x^* is a solution of $VVI(f', K)$. Conversely, if $x^* \in K$ solves $VVI(f', K)$, then x^* solves problem $VI(\phi'_{x^*}, K)$.*

Proof. Assume first that x^* solves problem $VI(\phi'_{\hat{x}}, K)$ for some $\hat{x} \in K$. Then from Proposition 9 we know that x^* solves $VI(\partial\phi_{\hat{x}}, K)$, i.e. there exists $v^* \in \partial\phi_{\hat{x}}(x^*)$, such that $\langle v^*, y - x^* \rangle \geq 0, \forall y \in K$. By Caratheodory Theorem $v^* = \sum_{i=1}^{r} \lambda_i \xi_i^{\top} f'(x^*)$, with $0 < r \leq n+1$, $\lambda_i \geq 0$, $\sum_{i=1}^{r} \lambda_i = 1$, $\xi_i \in R_{\hat{x}}(x^*)$. This means $\sum_{i=1}^{r} \lambda_i \xi_i^{\top} f'(x^*)(y - x^*) \geq 0$, $\forall y \in K$. Ab absurdo assume that for some $\bar{y} \in K$ it holds $f'(x^*)(\bar{y} - x^*) \in -\operatorname{int} C$. Hence, for

every $\xi \in C' \cap S$, we must have $\xi^\top f'(x^*)(\overline{y} - x^*) < 0$ and this conradicts the previous inequality.

Assume now that $x^* \in K$ solves $VVI(f', K)$ and observe that since K is convex, also $\Omega(x^*)$ is a convex set. Since $\Omega(x^*) \cap -\text{int}\, C = \varnothing$, then from the well known Separation Theorem, we have the existence of a vector $\xi \in C' \cap S$ such that $\xi^\top f'(x^*)(y - x^*) \geq 0$. Now, observe that we have $\phi'_{x^*}(x^*; y - x^*) = \max_{\xi \in R_{x^*}(x^*)} \xi^\top f'(x^*)(y - x^*)$ and $R_{x^*}(x^*) = C' \cap S$. So the previous inequality implies $\phi'_{x^*}(x^*; y - x^*) \geq 0$.

Remark 5. In [16] it has been proved that, under the hypotheses of the previous result, the set of the solutions of $VVI(f', K)$ coincide also with the set of the solutions of the scalar variational inequalities $VI(\xi^\top f', K)$, $\xi \in C'$.

Now we turn our attention to problem $MVI(\phi'_{\hat{x}}, K)$.

Theorem 8. *Let $x^* \in K$ solve $MVI(\phi'_{\hat{x}}, K)$. Then x^* solves $MVVI(f', K)$.*

Proof: Let x^* solve $MVI(\phi'_{\hat{x}}, K)$ and ab absurdo assume that x^* does not solve $MVVI(f', K)$. Hence, for some $\overline{y} \in K$ we have $f'(\overline{y})(\overline{y} - x^*) \in -\text{int}\, C$ and so $\xi^\top f'(\overline{y})(\overline{y} - x^*) < 0$, $\forall \xi \in C' \cap S$. This contradicts the fact that x^* solves $MVI(\phi'_{\hat{x}}, K)$, i.e. that $\max_{\xi \in R_{\hat{x}}(y)} f'(y)(y - x^*) \geq 0, \forall y \in K$. □

The converse of the previous result holds under convexity assumptions.

Theorem 9. *Let K be a convex set and f be a C-convex function. If $x^* \in K$ solves $MVVI(f', K)$, then x^* solves problem $MVI(\phi'_{x^*}, K)$.*

Proof. We know that, if f is C-convex and x^* solves $MVVI(f', K)$, then x^* is a w-solution of $VP(f, K)$ (Proposition 8 and Remark 3) and hence x^* is a solution of $P(\phi_{x^*}, K)$ (Theorem 3). Since f is C-convex, from Proposition 6 we know that $\phi_{x^*}(x)$ is convex and then $\phi_{x^*} \in IAR(K, x^*)$. It follows that $\phi_{x^*}(y; x^* - y) \leq 0$ (recall Proposition 2) and the proof is complete. □

The convexity assumption in the previous result cannot be dropped as the following example shows. Hence, when convexity assumptions do not hold $MVI(\phi_{\hat{x}}, K)$ defines a stronger solution concept then $MVVI(f', K)$.

Example 2. *Consider the function of Example 1 that clearly is not* $\mathbb{R}^2_+$ *convex. The point* $x^* = 0$ *is a solution of* $MVI(f', K)$*, but there is no* $\hat{x} \in [-\frac{2}{\pi}, 0]$ *such that* x^* *solves* $MVI(\phi'_{\hat{x}}, K)$*, with* $\phi_{\hat{x}}(x) = \max\{f_1(x) - f_1(\hat{x}); f_2(x) - f_2(\hat{x})\}$ *(recall Remark 2).*

The next result states that $MVI(\phi'_{\hat{x}}, K)$ is a sufficient optimality condition.

Theorem 10. *Let* $x^* \in \ker K$ *be a solution of* $MVI(\phi'_{\hat{x}}, K)$ *for some* $\hat{x} \in K$*. Then* x^* *is a w-solution of* $VP(f, K)$.

Proof. Since x^* solves $MVI(\phi'_{\hat{x}}, K)$, then x^* is a solution of $P(\phi_{\hat{x}}, K)$ and hence a w-solution of $VP(f, K)$ (recall Proposition 4 and Theorem 4). □

Theorem 11. *Let* $x^* \in \ker K$*. If* x^* *is a strong solution of* $MVI(\phi'_{\hat{x}}, K)$ *for some* $\hat{x} \in K$*, then* x^* *is a strong* e*-solution of* $VP(f, K)$*. Furthermore, if* x^* *is a strong solution of* $MVI(\phi'_{x^*}, K)$*, then* $VP(f, K)$ *is pointwise well-posed at* x^*.

Proof. If x^* is a strong solution of problem $MVI(\phi'_{\hat{x}}, K)$ for some $\hat{x}$, then x^* is a strong e-solution of $VP(f, K)$ (apply Proposition 4 and Theorem 4). Assume in particular that x^* is a strong solution of problem $MVI(\phi'_{x^*}, K)$, i.e.:

$$\phi'_{x^*}(y; x^* - y) < 0, \forall y \in K \setminus \{x^*\}.$$

Then, combining Propositions 2 and 4 and Theorems 2 and 5, the proof is complete. □

Example 3. *Consider the function* $f : \mathbb{R} \to \mathbb{R}^2$ *defined as* $f(x) = (x, \log|x-1|) = (f_1(x), f_2(x))$*, let* $C = \mathbb{R}^2_+$ *and* $K = [-1/2, 1/2]$*. It is easy to check that* $x^* = 0$ *solves* $MVVI(f', K)$ *and* x^* *is an e-solution of* $VP(f, K)$ *(and hence also a w-solution). Anyway, Proposition 8 would not have allowed such a conclusion, since* f *is not* C*-convex. Instead, considering function* $\phi_{x^*}(x) = \max\{f_1(x), f_2(x)\}$ *one gets that* x^* *is a strong*

solution of $MVI(\phi_{x^*},K)$ *and hence* x^* *is a strong e-solution of* $VP(f,K)$. *Furthermore* $VP(f,K)$ *is pointwise well-posed at* x^*.

REFERENCES

[1] Blum E., Oettli W.: From optimization and variational inequalities to equilibrium problems. Math. Student 63, no. 1-4, 123–145, 1994.

[2] Crespi G.P.: Proper efficiency and vector variational inequalities, Journal of Information and Optimization Sciences, Vol. 23, No. 1, pp. 49-62, 2002.

[3] Crespi G.P., Guerraggio A., Rocca, M.: Minty variational inequality and optimization: scalar and vector case, to appear in the Proceedings of the VII Symposium on Generalized Convexity-Monotonicity, Hanoi, Vietnam, August 2002.

[4] Crespi G.P., Ginchev I., Rocca M. : Existence of solutions and star-shapedness in Minty variational inequality. J. Global Optim., to appear.

[5] Demyanov V.F., Rubinov A.M.: Constructive Nonsmooth Analysis, Peter Lang, Frankfurt am Main, 1995.

[6] Dentcheva D., Helbig, S.: On variational principles, level sets, well-posedness, and ε - solutions in vector optimization. J. Optim. Theory Appl. 89, no. 2, 325–349, 1996.

[7] Dontchev A.L., Zolezzi T.: Well-posed optimization problems,Springer, Berlin, 1993.

[8] Giannessi, F.: Theorems of the alternative, quadratic programs and complementarity problems, in Variational Inequalities and Complementarity Problems. Theory and applications (R.W. Cottle, F. Giannessi, J.L. Lions eds.), Wiley, New York, pp. 151-186, 1980.

[9] Giannessi F.: On Minty variational principle, in New Trends in Mathematical Programming, Kluwer, Dordrecht pp. 93-99, 1997.

[10] Giannessi F.: On a connection among separation, penalization and regularization for variational inequalities with point to set operators. Rend. Circ. Mat. Palermo (2) Suppl. No. 48, 137–145, 1997.

[11] Ginchev I., Guerraggio A., Rocca M.: From scalar to vector optimization. Appl. Math., to appear.

[12] Ginchev I, Hoffman A.: Approximation of set-valued functioons by single-valued ones. Discuss. Math. Differ. Incl. Control Optim. 22, no. 1, 33–66, 2002.

[13] Kinderlehrer, D., Stampacchia, G. An introduction to variational inequalities and their applications, Academic Press, New York, 1980.

[14] Hiriart-Urruty J.-B.: New concepts in nondifferentiable programming. Analyse non convexe, Bull. Soc. Math. France 60, 57–85, 1979.

[15] Hiriart-Urruty J.-B.: Tangent cones, generalized gradients and mathematical programming in Banach spaces. Math. Oper. Res. 4, 79–97, 1979.

[16] Lee G.M., Kim D.S., Lee B.S., Yen N.D.: Vector variational inequalities as a tool for studying vector optimization problems. Nonlinear Anal. 34, 745-765, 1998.

[17] Minty, G.J.: On the generalization of a direct method of the calculus of variations, Bulletin of American Mathematical Society, Vol. 73, pp. 314-321, 1967.

[18] Rubinov, A.M.: Abstract convexity and global optimization, Kluwer, Dordrecht, 2000.

[19] Sawaragi, Y., Nakayama, H., Tanino, T.: Theory of multiobjective optimization. Mathematics in Science and Engineering, 176. Academic Press, Inc., Orlando, FL, 1985.

[20] Zaffaroni A.: Degrees of efficiency and degrees of minimality. SIAM J. Optimization, to appear.

VARIATIONAL INEQUALITIES FOR GENERAL EVOLUTIONARY FINANCIAL EQUILIBRIUM

P. Daniele
Dept. of Mathematics, University of Catania, Catania, Italy

1. INTRODUCTION

In a previous paper (see [6]), we studied an evolutionary financial equilibrium problem in the case of quadratic utility function

$$V_i(x_i(t), y_i(t)) = \int_0^T \left\{ \begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix}^T Q^i(t) \begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix} - r(t) \times [x_i(t) - y_i(t)] \right\} dt,$$

where $Q^i(t) = \begin{bmatrix} Q^i_{11}(t) & Q^i_{12}(t) \\ Q^i_{21}(t) & Q^i_{22}(t) \end{bmatrix}$ is a $2n \times 2n$ variance–covariance matrix. Now we intend to extend this particular model to a general case in which the utility function is given by

$$U_i(t, x_i(t), y_i(t), r(t)) = u_i(t, x_i(t), y_i(t)) + r(t)(x_i(t) - y_i(t)),$$

where $u_i(t, x_i(t), y_i(t))$ is a concave and differentiable function. The assumption of concavity on $u_i(t, x_i(t), y_i(t))$ is essential in order to obtain a characterization of the evolutionary financial equilibrium and the existence

of the financial equilibrium. This fact perfectly agrees with many other situations in which the concavity (or convexity) plays an essential role. Moreover, like the previous paper, also this one is devoted to the evolutionary case. Although equilibrium excludes time, time is, nevertheless, central in both the physical–technological world as well as in the socio–economic world. Then we cannot neglect it in our investigations. In the same context take place the papers [2] and [3] - [4] which discuss other time–dependent applications using the same approach applied to financial equilibrium problems.

This paper is organized as follows. In Section 2, we develop the model, provide the equilibrium conditions, and give the variational inequality formulation. We also identify the underlying network structure of the problem both out of and in the equilibrium state. In Section 3, we provide some theoretical results, whereas in Section 4 we give the proof of the variational inequality formulation and establish an existence result. In Section 5, we summarize the results of this paper and provide suggestions for future research.

2. THE EVOLUTIONARY FINANCIAL MODEL

In this section, we present the evolutionary financial model and give the variational inequality formulation of the equilibrium conditions. The functional setting in which we study this evolutionary model is the Lebesgue space $L^2([0,T],R^p)$, which appears to be the appropriate setting since it allows us to obtain equilibrium conditions equivalent to a variational inequality involving the L^2–scalar product in $[0,T]$. The time dependence of the model in the $L^2([0,T])$ sense allows the model to follow the financial behavior, even in the presence of possibly very irregular evolution, whereas the equilibrium conditions are required to hold almost everywhere (see [2], [3], [4] for analogous problems). In this setting, the variance–covariance matrices associated with the sectors' risk perceptions will be required to have $L^\infty([0,T])$–entries.

Analytically, consider a financial economy consisting of m sectors, with a typical sector denoted by i, and with n instruments, with a typical financial instrument denoted by j, in the period $\mathcal{T}=[0,T]$. Examples of sectors include: households, domestic businesses, banks, and other financial institutions, as well as state and local governments. Examples of financial instruments, in turn, are: mortgages, mutual funds, savings deposits, money market funds, etc.

Let $s_i(t)$ denote the total financial volume held by sector i at the time t, which is considered to depend on the time $t\in[0,T]$. At time t, denote the

amount of instrument j held as an asset in sector i's portfolio by $x_{ij}(t)$ and the amount of instrument j held as a liability in sector i's portfolio by $y_{ij}(t)$. The assets in sector i's portfolio are grouped into the column vector $x_i(t) = [x_{i1}(t),\ x_{i2}(t),\ ...,\ x_{ij}(t),\ ...,\ x_{in}(t)]^T$ and the liabilities in sector i's portfolio are grouped into the column vector $y_i(t) = [y_{i1}(t), y_{i2}(t), ..., y_{ij}(t), ..., y_{in}(t)]^T$. Moreover, group the sector asset vectors into the matrix

$$x(t) = \begin{bmatrix} x_1(t) \\ \cdots \\ x_i(t) \\ \cdots \\ x_n(t) \end{bmatrix} = \begin{bmatrix} x_{11}(t) & \cdots & x_{1j}(t) & \cdots & x_{1n}(t) \\ \cdots & & & & \\ x_{i1}(t) & \cdots & x_{ij}(t) & \cdots & x_{in}(t) \\ \cdots & & & & \\ x_{m1}(t) & \cdots & x_{mj}(t) & \cdots & x_{mn}(t) \end{bmatrix}$$

and group the sector liability vectors into the matrix

$$y(t) = \begin{bmatrix} y_1(t) \\ \cdots \\ y_i(t) \\ \cdots \\ y_n(t) \end{bmatrix} = \begin{bmatrix} y_{11}(t) & \cdots & y_{1j}(t) & \cdots & y_{1n}(t) \\ \cdots & & & & \\ y_{i1}(t) & \cdots & y_{ij}(t) & \cdots & y_{in}(t) \\ \cdots & & & & \\ y_{m1}(t) & \cdots & y_{mj}(t) & \cdots & y_{mn}(t) \end{bmatrix}.$$

We generalize the quadratic financial model and we assume that each sector seeks to maximize its utility, where the utility function $U_i(t, x_i(t),\ y_i(t),\ r(t))$ is given by:

$$U_i(t, x_i(t), y_i(t), r(t)) = u_i(t, x_i(t), y_i(t)) + r(t)(x_i(t) - y_i(t)).$$

The quadratic financial model is a particular case of this general model which can be obtained again setting

$$-u_i(t, x_i(t), y_i(t)) = \begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix}^T Q^i(t) \begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix}.$$

Hence $-u_i(t, x_i(t), y_i(t))$ represents a general form of the aversion to the risk.

We suppose that the sector's utility function $u_i(t, x_i(t), y_i(t))$ is defined on $[0,T]\times R^n \times R^n$, measurable in t and continuous with respect to x_i and y_i. Moreover we assume that $\dfrac{\partial u_i}{\partial x_{ij}}$ and $\dfrac{\partial u_i}{\partial y_{ij}}$ exist and that they are measurable in t and continuous with respect to x_i and y_i. Further we require that the following growth conditions hold:

$$|u_i(t,x,y)| \le \alpha_i(t)\|x\|\|y\|, \ \forall x, y \in R^n, \text{ a.e. in } [0,T], \ \forall i = 1,...,n, \tag{1}$$

and

$$\left|\frac{\partial u_i(t,x,y)}{\partial x_{ij}}\right| \le \beta_{ij}(t)\|y\|, \quad \left|\frac{\partial u_i(t,x,y)}{\partial y_{ij}}\right| \le \gamma_{ij}(t)\|x\|, \tag{2}$$

where α_i, β_{ij}, γ_{ij} are non negative functions of $L^\infty([0,T])$. Finally, we suppose that the function $u_i(t, x_i(t), y_i(t))$ is concave.

Assuming as the functional setting the Lebesgue space $L^2([0,T], R^p)$, the set of feasible assets and liabilities becomes:

$$P_i = \{(x_i(t), y_i(t)) \in L^2([0,T], R^{2n}):$$
$$\sum_{j=1}^{n} x_{ij}(t) = s_i(t), \quad \sum_{j=1}^{n} y_{ij}(t) = s_i(t) \text{ a.e. in } [0,T],$$
$$x_{ij}(t) \ge 0, \quad y_{ij}(t) \ge 0, \text{ a.e. in } [0,T]\}.$$

In Figure 1, we depict the network structure associated with the above feasible set and the financial economy out of equilibrium. The set of feasible assets and liabilities associated with each sector corresponds to budget constraints.

We now can give the following definition of an equilibrium of the financial model.

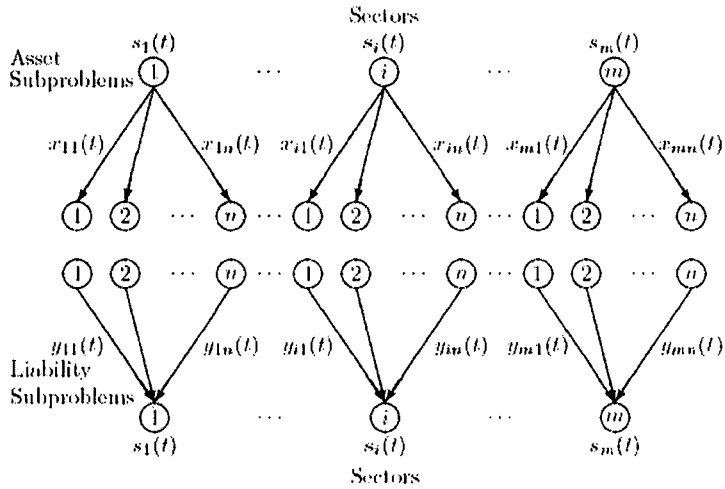


Fig. 1. Network structure of the sectors' optimization problems a.e. in $[0,T]$

Definition 1. *A vector of sector assets, liabilities, and instrument prices* $(x^*(t),\ y^*(t),\ r^*(t)) \in \prod_{i=1}^{m} P_i \times L^2([0,T], R_+^n)$ *is an equilibrium of the evolutionary financial model if and only if it satisfies the system of inequalities*

$$-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial x_{ij}} - r_j^*(t) - \mu_i^{(1)}(t) \geq 0, \tag{3}$$

and

$$-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial y_{ij}} + r_j^*(t) - \mu_i^{(2)}(t) \geq 0, \tag{4}$$

and equalities

$$x_{ij}^*(t)\left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial x_{ij}} - r_j^*(t) - \mu_i^{(1)}(t)\right] = 0, \tag{5}$$

$$y_{ij}^*(t)\left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial y_{ij}} + r_j^*(t) - \mu_i^{(2)}(t)\right] = 0, \tag{6}$$

where $\mu_i^1(t),\ \mu_i^2(t) \in L^2([0,T])$ *are Lagrangean functions, for all sectors* $i:\ i=1,2,...,m,$ *and for all instruments* $j:\ j=1,2,...,n,$ *and the condition*

$$\begin{cases} \sum_{i=1}^{m}\left(x_{ij}^*(t)-y_{ij}^*(t)\right)\geq 0, & \text{a.e. in } [0,T] \\ \sum_{i=1}^{m}\left(x_{ij}^*(t)-y_{ij}^*(t)\right)r_j^*(t)=0, \quad r^*(t)\in L^2([0,T],R_+^n), \end{cases} \tag{7}$$

simultaneously, where a.e. means that the condition holds almost everywhere.

The meaning of this definition is the following: to each financial volume $s_i(t)$ invested by the sector i, we associate the functions $\mu_i^{(1)}(t)$ and $\mu_i^{(2)}(t)$ related, respectively, to the assets and to the liabilities and which represent the "equilibrium utilities" per unit of the sector i. The financial volume invested in the instrument j as assets $x_{ij}^*(t)$ is greater than or equal to zero if the j-th component

$$-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial x_{ij}}-r_j^*(t)$$

of the utility is equal to $\mu_i^{(1)}(t)$, whereas if

$$-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial x_{ij}}-r_j^*(t)>\mu_i^{(1)}(t),$$

then $x_{ij}^*(t)=0$. The same occurs for the liabilities. It is remarkable that the equilibrium definition is, in a sense, the same as that given by the Wardrop (1952) principle which states that in the case of user-optimization on congested transportation networks (see [1]) the user (which is a traveller in that case) rejects the less convenient (or more costly) choice (which, in the context of a transportation network, is a path or route).

The functions $\mu_i^{(1)}(t)$ and $\mu_i^{(2)}(t)$ are Lagrangean functions associated with the constraints $\sum_{j=1}^{n} x_{ij}(t)-s_i(t)=0$ and $\sum_{j=1}^{n} y_{ij}(t)-s_i(t)=0$, respectively. The fact that they are unknown a priori has no influence because, as we shall see by means of Theorem 1, Definition 1 is equivalent

to a variational inequality in which $\mu_i^{(1)}$ and $\mu_i^{(2)}$ do not appear. Nevertheless, by the use of Theorem 3, they can be obtained.

Conditions (7), which represent the equilibrium condition for the prices, express the equilibration of the total assets and the total liabilities of each instrument; namely, if the price of instrument j is positive, then the amount of the assets is equal to amount of liabilities; if there is an excess supply of an instrument in the economy, then its price must be zero.

Moreover, if we consider the group of conditions (3)–(6) for a fixed $r(t)$, then we realize (see Section 3) that they are necessary and sufficient conditions to ensure that (x_i^*, y_i^*) is the maximum of the problem:

$$\max_{P_i} \int_0^T \{u_i(t, x_i(t), y_i(t)) + r(t) \times [x_i(t) - y_i(t)]\}\, dt, \quad \forall (x_i(t), y_i(t)) \in P_i. \tag{8}$$

Problem (8) means that each sector maximizes his utility. The functional $u_i(t, x(t), y(t))$ is concave and, using assumption 1, it also belongs to $L^1([0, T])$, as well as $r(t) \times [x_i(t) - y_i(t)]$. Moreover, since it is continuous in virtue of (1) (see [8]), it is also upper semicontinuos in the set P_i which is weakly compact, then such a maximum exists (see [9], Lemma 2.11, pag. 15).

We now state the variational inequality formulation of the governing equilibrium conditions, the proof of which is given in Section 4.

Theorem 1. *A vector* $(x^*(t), y^*(t), r^*(t)) \in \prod_{i=1}^{m} P_i \times L^2([0,T], R_+^n)$ *is an evolutionary financial equilibrium if and only if it satisfies the following variational inequality: Find* $(x^*(t), y^*(t), r^*(t)) \in \prod_{i=1}^{m} P_i \times L^2([0,T], R_+^n)$ *:*

$$\sum_{i=1}^{m}\int_0^T\{\sum_{j=1}^{n}\left[-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial x_{ij}}-r_j^*(t)\right]\times\left[x_{ij}(t)-x_{ij}^*(t)\right]$$
$$+\sum_{j=1}^{n}\left[-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial y_{ij}}+r_j^*(t)\right]\times\left[y_{ij}(t)-y_{ij}^*(t)\right] \tag{9}$$
$$+\sum_{j=1}^{n}\left(x_i^*(t)-y_i^*(t)\right)\times\left[r(t)-r_j^*(t)\right]\}\,dt\geq 0,$$
$$\forall(x(t),y(t),r(t))\in\prod_{i=1}^{m}P_i\times L^2([0,T],R_+^n).$$

Such an integral exists as a consequence of condition (2).

In the subsequent section we will prove the equivalence between problem (8) and conditions (3)–(6). In addition, we will establish the equivalence between condition (7) and a suitable variational inequality. Observe, that due to conditions (7), in equilibrium, we have that for each financial instrument, its price times the total amount of the instrument as an asset minus the total amount as a liability is exactly equal to zero.

3. SOME THEORETICAL RESULTS

The proof of the equivalence between Definition 1 and the variational inequality formulation is obtained by showing that conditions (3)–(6) are equivalent to problem (8), which, in turn, is equivalent to a first variational inequality and that conditions (7) are equivalent to a second variational inequality. From these two variational inequalities, we then derive variational inequality (9).

We start by establishing the equivalence between problem (8) and a variational inequality. This proof is standard (see [17] for a similar argument), but we recall it for the reader's convenience.

Theorem 2. $\left(x_i^*(t),y_i^*(t)\right)$ *is a solution to (8) if and only if* $\left(x_i^*(t),y_i^*(t)\right)$ *is a solution to the variational inequality*

$$\int_0^T -\sum_{j=1}^{n}\left[\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial x_{ij}} + r^*(t)\right]\times[x_{ij}(t)-x_{ij}^*(t)]\,dt$$

$$+\int_0^T -\sum_{j=1}^{n}\left[\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial y_{ij}} - r^*(t)\right]\times[y_{ij}(t)-y_{ij}^*(t)]\,dt \geq 0, \qquad (10)$$

$$\forall\left(x_i(t),y_i(t)\right)\in P_i,$$

where $r_j^*(t)$ *denotes the price for instrument* j *at the time* $t\in[0,T]$.

Proof: Let us assume that $\left(x_i^*(t),y_i^*(t)\right)$ is a solution to problem (8). Then for all $\left(x_i(t),y_i(t)\right)\in P_i$ the function

$$F(\lambda)=\int_0^T\{u_i(t,\lambda x_i^*(t)+(1-\lambda)x_i(t),\lambda y_i^*(t)+(1-\lambda)y_i(t))$$

$$+r(t)\times\left[\lambda x_i^*(t)+(1-\lambda)x_i(t)-\lambda y_i^*(t)-(1-\lambda)y_i(t)\right]\}\,dt,\quad \lambda\in[0,1]$$

admits the maximum solution when $\lambda=1$ and $F'(1)\geq 0$. Hence, we can consider the derivative of $F(\lambda)$ with respect to λ and we obtain:

$$\frac{\partial}{\partial\lambda}\int_0^T\{u_i(t,\lambda x_i^*(t)+(1-\lambda)x_i(t),\lambda y_i^*(t)+(1-\lambda)y_i(t))$$

$$+r(t)\times\left[\lambda x_i^*(t)+(1-\lambda)x_i(t)-\lambda y_i^*(t)-(1-\lambda)y_i(t)\right]\}\,dt=$$

$$\int_0^T\{\sum_{j=1}^{n}\frac{\partial u_i(t,\lambda x_i^*(t)+(1-\lambda)x_i(t),\lambda y_i^*(t)+(1-\lambda)y_i(t))}{\partial x_{ij}}(x_{ij}^*(t)-x_{ij}(t))$$

$$+\sum_{j=1}^{n}\frac{\partial u_i(t,\lambda x_i^*(t)+(1-\lambda)x_i(t),\lambda y_i^*(t)+(1-\lambda)y_i(t))}{\partial y_{ij}}(y_{ij}^*(t)-y_{ij}(t))\}\,dt$$

$$+\int_0^T\sum_{i=1}^{m}r(t)\times\left[x_{ij}^*(t)-x_{ij}(t)-y_{ij}^*(t)+y_{ij}(t)\right]dt.$$

So we obtain:

$$F'(1)=\int_0^T\{\sum_{j=1}^n\frac{\partial u_i(t,x_i^*(t),y_i^*(t))}{\partial x_{ij}}\cdot(x_{ij}^*(t)-x_{ij}(t))$$

$$+\sum_{j=1}^n\frac{\partial u_i(t,x_i^*(t),y_i^*(t))}{\partial y_{ij}}\cdot(y_{ij}^*(t)-y_{ij}(t))\}\,dt$$

$$+\int_0^T\sum_{i=1}^m r(t)\times\left[x_{ij}^*(t)-x_{ij}(t)-y_{ij}^*(t)+y_{ij}(t)\right]dt\geq 0,\quad\forall\left(x_i(t),y_i(t)\right)\in P_i,$$

namely, the variational inequality (10).

Vice versa, let us assume that $\left(x_i^*(t),y_i^*(t)\right)$ is solution to problem (10). Since the function $\mathcal{U}_i\left(x_i(t),y_i(t)\right)=\int_0^T U_i\left(x_i(t),y_i(t)\right)dt$ is concave, then for all $\left(x_i^*(t),y_i^*(t)\right)\in P_i$ the following estimate holds:

$$-\mathcal{U}_i\left(\lambda x_i(t)+(1-\lambda)x_i^*(t),\lambda y_i(t)+(1-\lambda)y_i^*(t)\right)$$
$$\leq-\lambda\mathcal{U}_i\left(x_i(t),y_i(t)\right)-(1-\lambda)\mathcal{U}_i\left(x_i^*(t),y_i^*(t)\right),$$

namely, $\forall\lambda\in(0,1]$:

$$-\frac{\mathcal{U}_i\left(x_i^*(t)+\lambda(x_i(t)-x_i^*(t)),y_i^*(t)+\lambda(y_i(t)-y_i^*(t))\right)-\mathcal{U}_i\left(x_i^*(t),y_i^*(t)\right)}{\lambda}$$
$$\leq-\mathcal{U}_i\left(x_i(t),y_i(t)\right)+\mathcal{U}_i\left(x_i^*(t),y_i^*(t)\right).\qquad(11)$$

When $\lambda\to 0$, the left-hand side of (11) converges to:

$$-\left[\frac{d}{d\lambda}\mathcal{U}_i\left(x_i^*(t)+\lambda(x_i(t)-x_i^*(t)),y_i^*(t)+\lambda(y_i(t)-y_i^*(t))\right)\right]_{\lambda=0}$$
$$=-\int_0^T\sum_{j=1}^n\{\frac{\partial u_i(t,x_i^*(t),y_i^*(t))}{\partial x_{ij}}\cdot(x_{ij}(t)-x_{ij}^*(t))$$
$$+\frac{\partial u_i(t,x_i^*(t),y_i^*(t))}{\partial y_{ij}}\cdot(y_{ij}(t)-y_{ij}^*(t))$$
$$+r_j^*(t)(x_{ij}(t)-x_{ij}^*(t)-y_{ij}(t)+y_{ij^*}(t))\}\,dt,$$

that is the left-hand side of the variational inequality (10), which is ≥ 0 $\forall \left(x_i(t), y_i(t)\right) \in P_i$ and, hence,

$$0 \leq -\mathcal{U}_i\left(x_i(t), y_i(t)\right) + \mathcal{U}_i\left(x_i^*(t), y_i^*(t)\right), \quad \forall \left(x_i(t), y_i(t)\right) \in P_i.$$

It means that $\left(x_i^*(t), y_i^*(t)\right)$ is solution to the problem (8). □

Let $\left(x_i^*(t), y_i^*(t)\right)$ be the solution to problem (8) for a given $r^*(t)$ and, hence, to variational inequality (10). Now we can prove the following characterization of the solution.

Theorem 3. $\left(x_i^*(t), y_i^*(t)\right)$ *is a solution to (8) or to (10) if and only if a.e. in* $[0,T]$ *it satisfies the conditions:*

$$-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial x_{ij}} - r_j^*(t) - \mu_i^{(1)}(t) \geq 0, \tag{3}$$

$$-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial y_{ij}} + r_j^*(t) - \mu_i^{(2)}(t) \geq 0, \tag{4}$$

$$x_{ij}^*(t)\left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial x_{ij}} - r_j^*(t) - \mu_i^{(1)}(t)\right] = 0, \tag{5}$$

$$y_{ij}^*(t)\left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial y_{ij}} + r_j^*(t) - \mu_i^{(2)}(t)\right] = 0, \tag{6}$$

where $\mu_i^1(t),\ \mu_i^2(t) \in L^2([0,T])$ *are Lagrangean functions.*

Proof: The proof of this equivalence is based on the infinite-dimensional Lagrangean theory, which has proven to be a powerful tool in determining essential properties of optimization problems (see [9], [5]). This theory proceeds in the following way.

Let us consider the function

$$L\left(x_i(t), y_i(t), \lambda_i^{(1)}(t), \lambda_i^{(2)}(t), \mu_i^{(1)}(t), \mu_i^{(2)}(t)\right)$$

$$= \Psi(x_i(t), y_i(t)) - \int_0^T \sum_{j=1}^n \lambda_{ij}^{(1)}(t) x_{ij}(t)\, dt - \int_0^T \sum_{j=1}^n \lambda_{ij}^{(2)}(t) y_{ij}(t)\, dt$$

$$- \int_0^T \mu_i^{(1)}(t)\left(\sum_{j=1}^n x_{ij}(t) - s_i(t)\right) dt - \int_0^T \mu_i^{(2)}(t)\left(\sum_{j=1}^n y_{ij}(t) - s_i(t)\right) dt,$$

where

$$\Psi(x_i(t), y_i(t))$$
$$= \int_0^T \sum_{j=1}^{n} \left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial x_{ij}} - r_j^*(t) \right] \times [x_{ij}(t) - x_{ij}^*(t)]\, dt$$
$$+ \int_0^T \sum_{j=1}^{n} \left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial y_{ij}} + r_j^*(t) \right] \times [y_{ij}(t) - y_{ij}^*(t)]\, dt,$$
$$\left(x_i(t), y_i(t)\right) \in L^2([0,T], R^{2n}) \text{ and } \left(\lambda_i^1(t), \lambda_i^2(t), \mu_i^1(t), \mu_i^2(t)\right) \in C$$
$$= \{\lambda_i^{(1)}(t),\ \lambda_i^{(2)}(t) \in L^2([0,T], R^n),\quad \lambda_i^{(1)}(t),\ \lambda_i^{(2)}(t) \geq 0,$$
$$\mu_i^{(1)}(t),\ \mu_i^{(2)}(t) \in L^2([0,T]);\ i = 1, 2, \ldots, m\}.$$

By means of Lagrangean Theory (cf. [5]), it is possible to prove that there exist $\lambda_i^{(1)}(t)$, $\lambda_i^{(2)}(t)$, $\mu_i^{(1)}(t)$, $\mu_i^{(2)}(t)$, such that $\lambda_i^{(1)}(t) \geq 0$, $\lambda_i^{(2)}(t) \geq 0$ and

$$\int_0^T \sum_{j=1}^{n} \lambda_{ij}^{(1)}(t) x_{ij}^*(t)\, dt = 0 \Rightarrow \lambda_{ij}^{(1)}(t) x_{ij}^*(t) = 0 \text{ a.e. in } [0,T]$$
$$\int_0^T \sum_{j=1}^{n} \lambda_{ij}^{(2)}(t) y_{ij}^*(t)\, dt = 0 \Rightarrow \lambda_{ij}^{(2)}(t) y_{ij}^*(t) = 0 \text{ a.e. in } [0,T].$$

Moreover, using the characterization of the solution by means of a saddle point (see [5]), we obtain:

$$\mathcal{L}\left(x_i(t), y_i(t), \lambda_i^{(1)}(t), \lambda_i^{(2)}(t), \mu_i^{(1)}(t), \mu_i^{(2)}(t)\right)$$
$$= \int_0^T \sum_{j=1}^{n} \left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial x_{ij}} - r_j^*(t) - \lambda_{ij}^{(1)}(t) - \mu_i^{(1)}(t) \right]$$
$$\times [x_{ij}(t) - x_{ij}^*(t)]\, dt \qquad (12)$$
$$+ \int_0^T \sum_{j=1}^{n} \left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial y_{ij}} + r_j^*(t) - \lambda_{ij}^{(2)}(t) - \mu_i^{(2)}(t) \right]$$
$$\times [y_{ij}(t) - y_{ij}^*(t)]\, dt \geq 0, \quad \forall \left(x_i(t), y_i(t)\right) \in L^2([0,T], R^{2n}).$$

If we set

$$\varepsilon_1(t) = -\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial x_{ij}} - r_j^*(t) - \lambda_{ij}^{(1)}(t) - \mu_i^{(1)}(t)$$

and

$$\varepsilon_2(t) = -\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial y_{ij}} - r_j^*(t) - \lambda_{ij}^{(1)}(t) - \mu_i^{(1)}(t),$$

by choosing

$$x_i(t) = x_i^*(t) + \varepsilon_1(t),$$

and

$$y_i(t) = y_i^*(t) + \varepsilon_2(t),$$

we get:

$$\int_0^T \sum_{j=1}^n \left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial x_{ij}} - r_j^*(t) - \lambda_{ij}^{(1)}(t) - \mu_i^{(1)}(t) \right]^2 dt$$

$$+ \int_0^T \sum_{j=1}^n \left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial y_{ij}} + r_j^*(t) - \lambda_{ij}^{(2)}(t) - \mu_i^{(2)}(t) \right]^2 dt \geq 0.$$

By choosing

$$x_i(t) = x_i^*(t) - \varepsilon_1(t),$$

and

$$y_i(t) = y_i^*(t) - \varepsilon_2(t),$$

we obtain:

$$-\int_0^T \sum_{j=1}^n \left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial x_{ij}} - r_j^*(t) - \lambda_{ij}^{(1)}(t) - \mu_i^{(1)}(t) \right]^2 dt$$

$$-\int_0^T \sum_{j=1}^{n}\left[-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial y_{ij}}+r_j^*(t)-\lambda_{ij}^{(2)}(t)-\mu_i^{(2)}(t)\right]^2 dt \geq 0.$$

Hence,

$$-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial x_{ij}}-r_j^*(t)-\mu_i^{(1)}(t)=\lambda_{ij}^{(1)}(t)\geq 0$$

and

$$-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial y_{ij}}+r_j^*(t)-\mu_i^{(2)}(t)=\lambda_{ij}^{(2)}(t)\geq 0.$$

Moreover, taking into account that $\lambda_i^{(1)}(t)x_i^*(t)=0$, $\lambda_i^{(2)}(t)y_i^*(t)=0$, we get:

$$x_{ij}^*(t)\left[-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial x_{ij}}-r_j^*(t)-\mu_i^{(1)}(t)\right]=0$$

and

$$y_{ij}^*(t)\left[-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial y_{ij}}+r_j^*(t)-\mu_i^{(2)}(t)\right]=0.$$

Conversely, if estimates (3)–(6) hold, then we show that the variational inequality (10) holds. From (3), we obtain:

$$\sum_{j=1}^{n}\left[-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial x_{ij}}-r_j^*(t)-\mu_i^{(1)}(t)\right]\times[x_{ij}(t)-x_{ij}^*(t)]\geq 0,$$

and taking into account that $\sum_{j=1}^{n} x_{ij}(t)=s_i(t)$ and $\sum_{j=1}^{n} x_{ij}^*(t)=s_i(t)$ a.e. in $[0,T]$, we get:

$$\sum_{j=1}^{n}[-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial x_{ij}}-r_j^*(t)]\times[x_{ij}(t)-x_{ij}^*(t)]\geq 0,$$

and then

$$\int_0^T \sum_{j=1}^{n} [-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial x_{ij}} - r_j^*(t)] \times [x_{ij}(t) - x_{ij}^*(t)] dt \geq 0. \tag{13}$$

Similarly, one obtains:

$$\sum_{j=1}^{n} [-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial y_{ij}} + r_j^*(t)] \times [y_{ij}(t) - y_{ij}^*(t)] \geq 0$$

and, subsequently,

$$\int_0^T \sum_{j=1}^{n} [-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial y_{ij}} + r_j^*(t)] \times [y_{ij}(t) - y_{ij}^*(t)] dt \geq 0. \tag{14}$$

Summing now inequalities (13) and (14) for all i, we conclude that for $(x^*(t), y^*(t)) \in \prod_{i=1}^{n} P_i$:

$$\sum_{i=1}^{m} \int_0^T \sum_{j=1}^{n} [-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial x_{ij}} - r_j^*(t)] \times [x_{ij}(t) - x_{ij}^*(t)] dt$$
$$+ \int_0^T \sum_{j=1}^{n} [-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial y_{ij}} + r_j^*(t)] \times [y_{ij}(t) - y_{ij}^*(t)] dt \geq 0,$$
$$\forall (x_i(t), y_i(t)) \in P_i,$$

and the proof is complete. □

We now describe the variational inequality associated with the instrument prices $r_j(t)$.

The equilibrium condition related to the prices of the instruments is

$$\begin{cases} \sum_{i=1}^{m} (x_{ij}^*(t) - y_{ij}^*(t)) \geq 0 & \text{a.e. in } [0,T] \\ \sum_{i=1}^{m} (x_{ij}^*(t) - y_{ij}^*(t)) r_j^*(t) = 0, \quad r^*(t) \in L^2([0,T], R_+^n). \end{cases} \tag{7}$$

Following the same proof of [6], we get the following theorem.

Theorem 4. *Condition (7) is equivalent to*

$$\begin{cases} \text{Find } r^*(t) \in L^2([0,T], R_+^n) \text{ such that} \\ \int_0^T \sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] \times \left[r_j(t) - r_j^*(t)\right] dt \geq 0, \\ \qquad \forall r(t) \in L^2([0,T], R_+^n). \end{cases} \tag{15}$$

4. VARIATIONAL INEQUALITY FORMULATION PROOF AND EXISTENCE THEOREM

Following the proof of Theorem 3, we can now prove Theorem 1.

Proof of Theorem 1: From the results of the preceding section, it immediately follows that if $(x^*(t), y^*(t), r^*(t)) \in \prod_{i=1}^m P_i \times L^2([0,T], R_+^n)$ is a financial equilibrium, then it satisfies the variational inequalities (10) and (15), hence, the variational inequality (9) and vice versa. □

We now establish the following existence theorem.

Theorem 5 (existence). *If* $(x^*(t), y^*(t), r^*(t)) \in \prod_{i=1}^m P_i \times L^2([0,T], R_+^n)$ *is an equilibrium, then the equilibrium asset and liability vector* $(x^*(t), y^*(t))$ *is a solution to the variational inequality:*

$$\sum_{i=1}^m \int_0^T \{\sum_{j=1}^n \left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial x_{ij}}\right] \times \left[x_{ij}(t) - x_{ij}^*(t)\right]$$

$$+\sum_{j=1}^n \left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial y_{ij}}\right] \times \left[y_{ij}(t) - y_{ij}^*(t)\right]\} dt \geq 0, \tag{16}$$

$$\forall (x(t), y(t)) \in S,$$

where

$$S \equiv \left\{ (x(t), y(t)) \in \prod_{i=1}^{m} P_i; \quad \sum_{i=1}^{m} (x_{ij}(t) - y_{ij}(t)) \geq 0, j = 1, 2, ..., n \right\}.$$

Conversely, if $(x^*(t), y^*(t))$ *is a solution to (17), then there exists an* $r^*(t) \in L^2([0,T], R_+^n)$ *such that* $(x^*(t), y^*(t), r^*(t))$ *is an equilibrium.*

Proof: Assume that $(x^*(t), y^*(t), r^*(t)) \in \prod_{i=1}^{m} P_i \times L^2([0,T], R_+^n)$ is an equilibrium. Then $(x^*(t), y^*(t), r^*(t))$ satisfies (9). In (9) let us set:

$$x_i(t) = x_i^*(t), \quad y_i(t) = y_i^*(t), \quad r(t) = 0, \quad \text{a.e. in } [0,T] \text{ and } \forall i = 1, ..., m,$$

then we get:

$$-\sum_{i=1}^{m} \int_0^T \sum_{j=1}^{n} \left[x_{ij}^*(t) - y_{ij}^*(t) \right] \times r_j^*(t)\, dt \geq 0. \tag{17}$$

Let us now set in (9) $r(t) = r^*(t)$ and we obtain:

$$\begin{aligned} &\sum_{i=1}^{m} \int_0^T \left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial x_{ij}} \right] \times \left[x_{ij}(t) - x_{ij}^*(t) \right] \\ &+ \sum_{j=1}^{n} \left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial y_{ij}} \right] \times \left[y_{ij}(t) - y_{ij}^*(t) \right] \} dt \\ &\geq \int_0^T \sum_{j=1}^{n} r_j^*(t) \left[\sum_{i=1}^{m} (x_{ij}(t) - y_{ij}(t)) - \sum_{i=1}^{m} (x_{ij}^*(t) - y_{ij}^*(t)) \right] dt. \end{aligned} \tag{18}$$

But the right-hand side of inequality (19) is ≥ 0, because of (17) and the constraint set S. Thus, we have established that $(x^*(t), y^*(t))$ satisfies (17).

Observe that $u_i(t, x_i(t), y_i(t))$ is concave, then $-u_i(t, x_i(t), y_i(t))$ is convex and its gradient is monotone. From assumption (2), it is also hemicontinuous along line segments and, hence, it is lower semicontinuous

with respect to the strong and weak topology. In virtue of Corollary 5.1 in [4] the variational inequality (17) admits a solution.

Now, in order to prove the existence of $r^*(t) \in L^2([0,T], R_+^n)$ such that $(x^*(t), y^*(t), r^*(t))$ is an equilibrium, let us apply the Lagrange Multiplier Theorem (see [5]) to the function:

$$L\left(x(t), y(t), \lambda^{(1)}(t), \lambda^{(2)}(t), \mu^{(1)}(t), \mu^{(2)}(t), r(t)\right)$$
$$= \Phi(x(t), y(t)) - \sum_{i=1}^{m} \int_0^T \sum_{j=1}^{n} \lambda_{ij}^{(1)}(t) x_{ij}(t)\, dt - \sum_{i=1}^{m} \int_0^T \sum_{j=1}^{n} \lambda_{ij}^{(2)}(t) y_{ij}(t)\, dt$$
$$- \sum_{i=1}^{m} \int_0^T \mu_i^{(1)}(t) \left[\sum_{j=1}^{n} x_{ij}(t) - s_i(t) \right] dt - \sum_{i=1}^{m} \int_0^T \mu_i^{(2)}(t) \left[\sum_{j=1}^{n} y_{ij}(t) - s_i(t) \right] dt$$
$$- \int_0^T \sum_{j=1}^{n} r_j(t) \left[\sum_{i=1}^{m} (x_{ij}(t) - y_{ij}(t)) \right] dt$$

where

$$\Phi(x(t), y(t)) = \sum_{i=1}^{m} \int_0^T \sum_{j=1}^{n} \left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial x_{ij}} \right] \times [x_{ij}(t) - x_{ij}^*(t)]\, dt$$
$$+ \int_0^T \sum_{j=1}^{n} \left[-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial y_{ij}} \right] \times [y_{ij}(t) - y_{ij}^*(t)]\, dt,$$

and

$$\left(\lambda_i^1(t), \lambda_i^2(t), \mu_i^1(t), \mu_i^2(t), r(t)\right) \in \overline{C}$$
$$= \{\lambda^{(1)}(t),\ \lambda^{(2)}(t) \in L^2([0,T], R^{nm}),\quad \lambda_i^{(1)}(t),\ \lambda_i^{(2)}(t) \geq 0,$$
$$\mu_i^{(1)}(t),\ \mu_i^{(2)}(t) \in L^2([0,T], R^m);\ i = 1, 2, \ldots, m;$$
$$r(t) \in L^2([0,T], R^n),\quad r(t) \geq 0 \text{ a.e. in } [0,T]\}.$$

We get that, besides $\lambda_i^{(1)}(t)$, $\lambda_i^{(2)}(t)$, $\mu_i^{(1)}(t)$ and $\mu_i^{(2)}(t)$, there exists an $r^*(t) \in L^2([0,T], R_+^n)$ corresponding to the constraints defining S. For such a pattern $(x^*(t), y^*(t), r^*(t))$ we have:

$$-\frac{\partial u_i(t, x_i(t), y_i(t))}{\partial x_{ij}} - r_j^*(t) - \mu_i^{(1)}(t) = \lambda_{ij}^{(1)} \geq 0, \tag{3}$$

$$-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial y_{ij}}+r_j^*(t)-\mu_i^{(2)}(t)=\lambda_{ij}^{(2)}\geq 0, \tag{4}$$

$$x_{ij}^*(t)\left[-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial x_{ij}}-r_j^*(t)-\mu_i^{(1)}(t)\right]=0, \tag{5}$$

$$y_{ij}^*(t)\left[-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial y_{ij}}+r_j^*(t)-\mu_i^{(2)}(t)\right]=0, \tag{6}$$

and

$$\begin{cases} \sum_{i=1}^{m}\left(x_{ij}^*(t)-y_{ij}^*(t)\right)\geq 0 & \text{a.e. in } [0,T] \\ \sum_{i=1}^{m}\left(x_{ij}^*(t)-y_{ij}^*(t)\right)r_j^*(t)=0, \quad r^*(t)\in L^2([0,T],R_+^n). \end{cases} \tag{7}$$

From Conditions (3)–(7), a.e. in $[0,T]$, we obtain:

$$\sum_{i=1}^{m}\sum_{j=1}^{n}[-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial x_{ij}}-r_j^*(t)]\times\left[x_{ij}(t)-x_{ij}^*(t)\right]$$
$$+\sum_{i=1}^{m}\sum_{j=1}^{n}[-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial y_{ij}}+r_j^*(t)]\times\left[y_{ij}(t)-y_{ij}^*(t)\right]\geq 0 \tag{19}$$

and

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left[x_{ij}^*(t)-y_{ij}^*(t)\right]\times\left[r_j(t)-r_j^*(t)\right]\geq 0. \tag{20}$$

Summing now (19) and (20) and integrating the result, we obtain:

$$\sum_{i=1}^{m}\int_0^T\{\sum_{j=1}^{n}[-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial x_{ij}}-r_j^*(t)]\times\left[x_{ij}(t)-x_{ij}^*(t)\right]$$

$$+\sum_{j=1}^{n}[-\frac{\partial u_i(t,x_i(t),y_i(t))}{\partial y_{ij}}+r_j^*(t)]\times\left[y_{ij}(t)-y_{ij}^*(t)\right]+$$

$$+\sum_{j=1}^{n}\left(x_{ij}^*(t)-y_{ij}^*(t)\right)\times\left[r_j(t)-r_j^*(t)\right]\}dt\geq 0,$$

and the proof is complete.

5. SUMMARY AND CONCLUSIONS

In this paper, we have proposed a new framework for the modeling, analysis, and computation of financial equilibrium problems through a novel evolutionary model. In contrast to earlier multi-sector, multi-instrument financial equilibrium models, the new model allows for variance–covariance matrices to be time-dependent, as well as the sector financial volumes. We described the behavior of the financial sectors, derived the equilibrium conditions, and then established the equivalent infinite-dimensional variational inequality formulation. We provided the network structure of the problem, both out of and in equilibrium. We also proved the existence of an equilibrium pattern.

Future research will include exploring a variety of modeling extensions, including, but not limited to, such policy interventions as price bounds and taxes. In addition, it would be interesting to apply the results herein to financial networks with intermediation (cf. [15]).

REFERENCES

[1] Dafermos, S. C., and Sparrow, F. T. (1969), "The Traffic Assignment Problem for a General Network", *Journal of Research of the National Bureau of Standards* **B**, 91-118.

[2] Daniele, P., and Maugeri, A. (2001), "On Dynamical Equilibrium Problems and Variational Inequalities", in *Equilibrium Problems: Nonsmooth Optimization and Variational Inequality Models*, pp. 59-59, F. Giannessi, A. Maugeri, and P. M. Pardalos, editors, Kluwer Academic Publishers, Dordrecht, The Netherlands.

[3] Daniele, P., Maugeri, A., and Oettli, W. (1998), "Variational Inequalities and Time-Dependent Traffic Equilibria", *Comptes Rendus de l' Académie des Sciences Paris*, t. 326, serie I, 1059–1062.

[4] Daniele, P., Maugeri, A., and Oettli, W. (1999), "Time-Dependent Traffic Equilibria", *Journal of Optimization Theory and its Applications*, 543-555.

[5] Daniele, P. (1999), "Lagrangean Function for Dynamic Variational Inequalities", *Rendiconti del Circolo Matematico di Palermo*, Serie II, Suppl. 58, 101-119.

[6] Daniele, P. (2003), "Variational Inequalities for Evolutionary Financial Equilibrium", *Innovations in Financial and Economic Networks*, A. Nagurney Ed. (2003), pp. 84-108.

[7] Dong, J., Zhang, D., and Nagurney, A. (1996), "A Projected Dynamical Systems Model of General Financial Equilibrium with Stability Analysis", *Mathematical and Computer Modelling*, 35-44.

[8] Fucik, S., Kufner, A. (1980) "Nonlinear Differential Equations", Elsevier Sci. Publ. Co., New York.

[9] Jahn, J. (1996), *Introduction to the Theory of Nonlinear Optimization*, Springer-Verlag, Berlin, Germany.

[10] Markowitz, H. M. (1959), *Portfolio Selection: Efficient Diversification of Investments*, Wiley & Sons, New York.

[11] Nagurney, A. (1994), "Variational Inequalities in the Analysis and Computation of Multi-Sector, Multi-Instrument Financial Equilibria," *Journal of Economic Dynamics and Control*, 161-184.

[12] Nagurney, A. (1999), *Network Economics - A Variational Inequality Approach*, second and revised edition, Kluwer Academic Publishers, Dordrecht, The Netherlands.

[13] Nagurney, A. (2001), "Finance and Variational Inequalities", *Quantitative Finance*, 309-317.

[14] Nagurney, A., Dong, J., and Hughes, M. (1992), "Formulation and Computation of General Financial Equilibrium", *Optimization*, 339-354.

[15] Nagurney, A., and Ke, K. (2001), "Financial Networks with Intermediation", *Quantitative Finance*, 441-451.

[16] Nagurney, A., and Zhang, D. (1996), *Projected Dynamical Systems and Variational Inequalities with Applications*, Kluwer Academic Publishers, Boston, Massachusetts.

[17] Stampacchia, G. (1969), "Variational Inequalities, Theory and Applications of Monotone Operators", *Proceedings of a NATO Advanced Study Institute (Venice, 1968)*, Oderisi, Gubbio, 101–192.

[18] Wardrop, J. G. (1952), "Some Theoretical Aspects of Road Traffic Research", *Proceedings of the Institute of Civil Engineers*, Part II, pp. 325-378.

VARIATIONAL CONTROL PROBLEMS WITH CONSTRAINTS VIA EXACT PENALIZATION*

V.F. Demyanov,[1] F. Giannessi[2] and G.Sh. Tamasyan[1]
Applied Mathematics Dept., St. Petersburg State University, Staryi Peterhof, St. Petersburg, Russia;[1] Dept. of Mathematics, University of Pisa, Pisa, Italy[2]

Abstract: The Exact Penalization approach to solving constrained problems of Calculus of Variations described in [17] is extended to the case of variational problems where the functional and the constraints contain a control function. The constraints are of both the equality- and inequality-type constraints. The initial constrained problem is reduced to an unconstrained one. The related unconstrained problem is essentially nonsmooth. Necessary optimality conditions are derived.

Key words: Calculus of Variations, Equality- and Inequality-type Constraints, Necessary optimality conditions, Penalty Function, Exact Penalty Function, Nonsmooth Analysis.

AMS subject classifications: 90C30, 52A20.

1. INTRODUCTION

In the paper a special equality- and inequality-type constrained problem of Calculus of Variations where control functions are present [1], [17] is treated by making use of the penalty function approach. The paper is a continuation and generalization of the paper [8] where classical variational

* The work was supported by the Russian Foundation for Fundamental Studies (RFFI) under Grant No 03-01-00668.

problems were studied by means of the exact penalization technique. The idea is to reduce the initial constrained problem to an unconstrained one using exact penalty functions. Such an approach was first proposed by [13,14] and later by [21] for solving convex programming problems (see [2,15,16,18] and bibliography therein). Some new conditions for application of exact penalties were stated in [12,6,9]. These conditions turns out to be useful for solving problems of Optimal Control and Calculus of Variations (see [10]).

In the present paper we consider a special problem of Calculus of Variations with equality- and inequality-type constraints. The functional is of the form

$$I(x) = I(y,u) = \int_0^T F(y(t), y'(t), u(t), u(y(t)), t)\, dt,$$

where

$$y \in C^1[0,T], u \in C^1[t_0,t_1], y(0) = y_0, y(T) = y_1, \int_{t_0}^{t_1} u(t)\, dt = 1,$$

$$l(y(t), y'(t), u(t), u(y(t)), t) = 0, y'(t) > 0 \ \forall t \in [0,T],\ u(t) > 0 \ \forall t \in [t_0,t_1],$$

t_0 and t_1 are specified below.

The paper is organized as follows. In section 2 the problem is formulated. In Section 3 an equivalent statement of the problem is given. Local minima and penalty functions are discussed in Section 4. Properties of the function describing the constraints are studied in Section 5. An Exact penalty function is introduced and studied in Section 6. Necessary optimality conditions are established in Section 7.

2. STATEMENT OF THE PROBLEM

Let $T > 0$ be fixed. By $C^1[0,T]$ we denote the class of continuously differentiable functions defined on $[0,T]$.

Let a function $l(y, y_1, u, u_1, t)$ be continuous with $\frac{\partial l}{\partial y}$, $\frac{\partial l}{\partial y_1}$, $\frac{\partial l}{\partial u}$ and $\frac{\partial l}{\partial u_1}$ in all its arguments on $R^4 \times [0,T]$.

Fix $y_0 \in R$ and $y_1 \in R$, such that $y_0 < y_1$. Put $t_0 = \min\{0, y_0\}$, $t_1 = \max\{T, y_1\}$ and introduce the set of functions

$$\Omega = \{[y,u] \mid y \in C^1[0,T], u \in C^1[t_0,t_1], y(0) = y_0, y(T) = y_1,$$
$$\int_{t_0}^{t_1} u(t)\,dt = 1,\ l(y(t), y'(t), u(t), u(y(t)), t) = 0,\ y'(t) > 0\ \forall t \in [0,T],$$
$$u(t) > 0\ \forall t \in [t_0,t_1]\}. \tag{2.1}$$

Assume that

$$\Omega \neq \varnothing.$$

Put $x = [y,u]$ and consider the functional

$$I(x) = I(y,u) = \int_0^T F(y(t), y'(t), u(t), u(y(t)), t)\,dt, \tag{2.2}$$

where the function $F(y, y_1, u, u_1, t)$ is continuous together with respect to $\frac{\partial F}{\partial y}$, y_1, $\frac{\partial F}{\partial u}$ and u_1 in all its arguments on $R^4 \times [0,T]$. An arc $x(t) = [y(t), u(t)] \in \Omega$ is called *admissible.*

An admissible arc $x^* = [y^*, u^*] \in \Omega$ is called a *strong extremal* of the functional $I(x)$ defined by (2), if there exists $\varepsilon > 0$, such that

$$I(x^*) \le I(x) \quad \forall x \in \Omega \cap B_\varepsilon(x^*), \tag{2.3}$$

where

$$B_\varepsilon(x^*) =$$
$$= \{x \in C^1[0,T] \times C^1[t_0,t_1] \mid \max_{0 \le t \le T} |y(t) - y^*(t)| + \max_{t_0 \le t \le t_1} |u(t) - u^*(t)| < \varepsilon\}. \tag{2.4}$$

If

$$I(x^*) \le I(x) \quad \forall x \in \Omega \cap \tilde{B}_\varepsilon(x^*), \tag{2.5}$$

where

$$\tilde{B}_\varepsilon(x^*) = \{x \in B_\varepsilon(x^*) \mid \max_{0 \le t \le T} |y'(t) - y^*(t)| + \max_{t_0 \le t \le t_1} |u'(t) - u^*(t)| < \varepsilon\}, \tag{2.6}$$

then the arc $x^*(t) \in \Omega$ is called a *weak extremal* of the functional (2.2).

Every strong extremal is a weak one as well (the converse is, generally speaking, not true).

3. AN EQUIVALENT STATEMENT OF THE PROBLEM

Let us refomulate the above stated problem. Let

$$y(t) = y_0 + \int_0^t e^{z_1(\tau)}\, d\tau, \quad u(t) = e^{u_0 + \int_0^t z_2(\tau) d\tau},$$

where

$$z_1(t) \in C[0,T], \quad z_2(t) \in C[t_0,t_1], \quad u_0 \in R.$$

It follows from (2.1) that

$$y(0) = y_0, \; y'(t) = e^{z_1(t)} > 0 \;\; \forall t \in [0,T]; \qquad u(t) > 0 \; \forall t \in [t_0,t_1].$$

Put

$$Z = \{[z_1, z_2, u_0] \,|\, z_1 \in C[0,T], z_2 \in C[t_0,t_1], u_0 \in R,$$
$$y_0 + \int_0^T e^{z_1(t)}\, dt = y_1, \int_{t_0}^{t_1} e^{u_0 + \int_0^t z_2(\tau) d\tau}\, dt = 1,$$
$$l(y_0 + \int_0^t e^{z_1(\tau)}\, d\tau, e^{z_1(t)}, e^{u_0 + \int_0^t z_2(\tau) d\tau}, e^{u_0 + \int_0^{y_0 + \int_0^t e^{z_1(\tau)} d\tau} z_2(\tau) d\tau}, t) = 0$$
$$\forall t \in [0,T]\}. \tag{3.1}$$

where $C[t_0,t_1]$ – is the family of bounded and continuous on $[t_0,t_1]$ functions.

Introduce the functional $f(z) =$

$$= \int_0^T F(y_0 + \int_0^t e^{z_1(\tau)}\, d\tau, e^{z_1(t)}, e^{u_0 + \int_0^t z_2(\tau) d\tau}, e^{u_0 + \int_0^{y_0 + \int_0^t e^{z_1(\tau)} d\tau} z_2(\tau) d\tau}, t) dt. \tag{3.2}$$

Set

$$h_0(T,z) := y_0 - y_1 + \int_0^T e^{z_1(\tau)}\, d\tau, \tag{3.3}$$

$$h_1(t_0,t_1,z) := \int_{t_0}^{t_1} e^{u_0+\int_0^t z_2(\tau)d\tau}\, dt - 1, \tag{3.4}$$

$$h_2(t,z) := l(y_0 + \int_0^t e^{z_1(\tau)}d\tau, e^{z_1(t)}, e^{u_0+\int_0^t z_2(\tau)d\tau}, e^{u_0+\int_{t_0}^{y_0+\int_0^t e^{z_1(\tau)}d\tau} z_2(\tau)d\tau}, t). \tag{3.5}$$

The set Z can be represented in the form

$$Z = \{ z = [z_1, z_2, u_0] \in C[0,T] \times C[t_0,t_1] \times R \mid h_0(T,z) = 0,$$

$$h_1(t_0,t_1,z) = 0,\ h_2(t,z) = 0\ \forall t \in [0,T]\}, \tag{3.6}$$

As in [8] it is easy to show that the following holds.

Lemma 3.1 *The problem*

$$I(x) \longrightarrow \min_{x\in\Omega} \tag{3.7}$$

is equivalent to the problem

$$f(z) \longrightarrow \min_{z\in Z} \tag{3.8}$$

in the following sense: if $x^* = [y^*, u^*] \in \Omega$ *– is a solution of the problem (3.7), then the function* $z^* = [z_1^*(t), z_2^*(t), u_0^*] = [\ln(y^*(t)), \frac{u^*(t)}{u^*(t)}, \ln(u^*(t_0))]$ *is a solution of the problem (3.8); and vice versa, if* $z^* \in Z$ *is a minimizer of the function f on the set Z, then the function*

$$x^* = [y_0 + \int_0^t e^{z_1^*(\tau)}\, d\tau, e^{u_0^*+\int_0^t z_2^*(\tau)d\tau}]$$

is a minimizer of the problem (3.7).

Proof. Let $x^* \in \Omega$ and

$$I(x^*) \le I(x) \quad \forall x \in \Omega.$$

Put $z^* = [\ln(y^*(t)), \dfrac{u^*(t)}{u^*(t)}, \ln(u^*(t_0))]$. Take any $z \in Z$. The function

$$x = [y,u] = [y_0 + \int_0^t e^{z_1(\tau)}\, d\tau, e^{u_0 + \int_0^t z_2(\tau) d\tau}]$$

belongs to the set $C^1[0,T] \times C^1[t_0,t_1]$ and satisfies the conditions

$$y(0) = y_0,\ y(T) = y_1,\ \int_{t_0}^{t_1} u(t)\, dt = 1,$$

$$l(y(t), y'(t), u(t), u(y(t)), t) = 0,\quad y'(t) > 0\ \ \forall t \in [0,T],\quad u(t) > 0\ \ \forall t \in [t_0,t_1],$$

i.e. $x \in \Omega$. Therefore

$$f(z^*) = I(x^*) \le I(x) = f(x).$$

Since $z \in Z$ is arbitrary, then $f(z^*) = \min\limits_{z \in Z} f(z)$, hence, z^* is a solution of the problem (3.8).

Now let $z^* \in Z$ and

$$f(z^*) \le f(z) \quad \forall z \in Z.$$

Put

$$x^* = [y^*, u^*] = [y_0 + \int_0^t e^{z_1^*(\tau)}\, d\tau, e^{u_0^* + \int_0^t z_2^*(\tau) d\tau}].$$

Take any $x \in \Omega$ and put $z = [\ln(y'(t)), \dfrac{u'(t)}{u(t)}, \ln(u(t_0))]$. It follows from (2.1), (3.1) and (3.6) that

$$h_0(T,z) = 0,\ h_1(t_0,t_1,z) = 0,\ \ h_2(t,z) = 0,\ \forall t \in [0,T]$$

i.e. $z \in Z$, therefore

$$I(z^*) = f(x^*) \le f(x) = I(x).$$

Due to the arbitrariness of $x \in \Omega$

$$I(x^*) = \min_{x \in \Omega} I(x),$$

i.e. x^* is a solution of the problem (3.7).

Thus, if $z^* = [z_1^*, z_2^*, u_0^*] \in Z$ is a global minimizer of the functional f on the set Z, then the point $x^* = [y^*, u^*]$, where

$$y^* = y_0 + \int_0^{\cdot} e^{z_1^*(\tau)}\, d\tau, \; u^* = e^{u_0^* + \int_0^{\cdot} z_2^*(\tau) d\tau} \in \Omega$$

and is a global minimizer of the functional $I(x)$ on Ω, and, conversely, if $x^* = [y^*, u^*] \in \Omega$ is a global minimizer of the functional $I(x)$ on Ω, then

$$z^*(t) = [\ln(y^*(t)), \frac{u^*(t)}{u^*(t)}, \ln(u^*(t_0))] \in Z$$

is a global minimizer of f on Z.

4. LOCAL MINIMA

Consider the problem (3.8). On the set $U = C[0,T] \times C[t_0, t_1] \times R$ let us introduce a metric $\rho(z, \overline{z})$, where $z = [z_1(t), z_2(t), u_0] \in U$, $\overline{z} = [\overline{z}_1(t), \overline{z}_2(t), \overline{u}_0] \in U$. As ρ it is possible to choose, for example, one of the following functions:

$$\rho_1(z, \overline{z}) =$$

$$= \max_{t \in [0,T]} \left| \int_0^{\cdot} (z_1(\tau) - \overline{z}_1(\tau)) d\tau \right| + \max_{t \in [t_0, t_1]} \left| \int_{t_0}^{\cdot} (z_2(\tau) - \overline{z}_2(\tau)) d\tau \right| + |u_0 - \overline{u}_0|, \quad (4.1)$$

$$\rho_2(z, \overline{z}) = \int_0^T |z_1(t) - \overline{z}_1(t)|\, dt + \int_{t_0}^{t_1} |z_2(t) - \overline{z}_2(t)|\, dt + |u_0 - \overline{u}_0|, \quad (4.2)$$

$$\rho_3(z,\overline{z}) = \max_{t\in[0,T]} |z_1(t)-\overline{z}_1(t)| + \max_{t\in[t_0,t_1]} |z_2(t)-\overline{z}_2(t)| + \frac{1}{t_1-t_0}|u_0-\overline{u}_0|, \quad (4.3)$$

$$\rho_4(z,\overline{z}) = [\int_0^T (z_1(t)-\overline{z}_1(t))^2 dt + \int_{t_0}^{t_1} (z_2(t)-\overline{z}_2(t))^2 dt + (u_0-\overline{u}_0)^2]^{\frac{1}{2}}. \quad (4.4)$$

The following relations exist between the metrics ρ_i $(i \in 1:4)$:

$$\rho_1(z,\overline{z}) \le \rho_2(z,\overline{z}) \le (t_1-t_0)\rho_3(z,\overline{z}),$$
$$\rho_4(z,\overline{z}) \le \sqrt{t_1-t_0}\,\rho_3(z,\overline{z}). \quad (4.5)$$

Thus, the metric ρ_3 majorizes the metrics ρ_2, ρ_1 and ρ_4, while the metric ρ_2 majorizes the metric ρ_1.

The following inclusions hold:

$$\{z \mid \rho_3(z,\overline{z}) < \frac{\varepsilon}{t_1-t_0}\} \subset \{z \mid \rho_2(z,\overline{z}) < \varepsilon\} \subset \{z \mid \rho_1(z,\overline{z}) < \varepsilon\},$$
$$\{z \mid \rho_3(z,\overline{z}) < \frac{\varepsilon}{\sqrt{t_1-t_0}}\} \subset \{z \mid \rho_4(z,\overline{z}) < \varepsilon\}.$$

Therefore one concludes that a local minimizer of the function f on the set Z in the metric ρ_1 is a local minimizer of f in both the metrics ρ_2 and ρ_3, while a local minimizer of f in the metric ρ_2 is a local minimizer of f in the metric ρ_3 and a local minimizer of f in the metric ρ_4 is a local minimizer of f in the metric ρ_3.

It is not difficult to see that if $z^* \in Z$ is a local minimizer of the functional f on the set Z with the metric ρ_1, then the function

$$x^* = [y^*, u^*] = [y_0 + \int_0^{\cdot} e^{z_1^*(\tau)}\, d\tau, e^{u_0^* + \int_0^{\cdot} z_2^*(\tau) d\tau}]$$

belongs to the set Ω and is a strong extremal of the functional I on the set Ω.

If $z^* \in Z$ is a local minimizer of the functional f on the set Z with the metric ρ_4, ρ_3 or ρ_2, then the function $x^* = [y^*, u^*]$ is a weak extremal of the functional I on Ω (but not necessarily a strong extremal).

Conversely, if $x^* \in \Omega$ is a weak extremal of the problem (3.7), then the function $z^*(t) = [\ln(y^*(t)), \dfrac{u^*(t)}{u^*(t)}, \ln(u^*(t_0))]$ is a local minimizer of the function f on the set Z in the metric ρ_3.

If $x^*(t) \in \Omega$ is a strong extremal of the problem (3.7), then the function $z^*(t) = [\ln(y^*(t)), \dfrac{u^*(t)}{u^*(t)}, \ln(u^*(t_0))]$ is a local minimizer of the function f on Z in the metrics ρ_1, ρ_2 and ρ_3 (but not necessarily in the metric ρ_4).

Remark 4.1. *It has already been observed that the metric ρ_3 majorizes the metrics ρ_4, ρ_2 and ρ_1, while the metric ρ_2 majorizes the metric ρ_1. It is not difficult to show that neither pair of metrics (ρ_1, ρ_2), (ρ_1, ρ_3), (ρ_2, ρ_3) and (ρ_4, ρ_3) is a pair of equivalent metrics.*

Remark 4.2. *The set Z defined by (3.6) can be represented in the equivalent form*

$$Z = \{ z = [z_1, z_2, u_0] \in U \mid \varphi(z) = 0 \}, \tag{4.6}$$

where

$$\varphi(z) = [\int_0^T h_2^2(t, z)\, dt + h_1^2(t_0, t_1, z) + h_0^2(T, z)]^{\frac{1}{2}}, \tag{4.7}$$

and h_0 is given by (3.3), h_1 – by (3.4), h_2 – by (3.5).

We have

$$\varphi(z) \geq 0 \quad \forall z \in C[0,T] \times C[t_0, t_1] \times R.$$

The relation (4.5) implies that if

$$\varphi(z) \geq a \rho_3(z, Z),$$

where $a > 0$, then

$$\varphi(z) \geq \frac{a}{t_1 - t_0} \rho_2(z, Z) \geq \frac{a}{t_1 - t_0} \rho_1(z, Z).$$

If $\varphi(z) \geq a\rho_2(z,Z)$, then $\varphi(z) \geq a\rho_1(z,Z)$.

5. PROPERTIES OF THE FUNCTION φ

5.1 The classical variation of z.

Thus, we consider the problem of minimizing the functional $f(z) =$

$$= \int_0^T F(y_0 + \int_0 e^{z_1(\tau)}\, d\tau, e^{z_1(t)}, e^{u_0 + \int_0 z_2(\tau)d\tau}, e^{u_0 + \int_0^{y_0 + \int_0 e^{z_1(\tau)}d\tau} z_2(\tau)d\tau}, t)\, dt, \quad (5.1)$$

on the set $Z \subset C[0,T] \times C[t_0,t_1] \times R = U$, defined by the relation (4.6).

Let us study the properties of the function φ. Fix $z \in U$. Let us choose and fix an arbitrary $v = [v_1, v_2, v_0] \in U$ and for some $\varepsilon \geq 0$ put

$$z_\varepsilon = z(t) + \varepsilon v(t) = [z_{1\varepsilon}, z_{2\varepsilon}, u_{0\varepsilon}], \quad (5.2)$$

where

$$z_{1\varepsilon} = z_1(t) + \varepsilon v_1(t), \; z_{2\varepsilon} = z_2(t) + \varepsilon v_2(t), \; u_{0\varepsilon} = u_0 + \varepsilon v_0.$$

Then

$$\begin{aligned} y_\varepsilon(t) &= y_0 + \int_0 e^{z_{1\varepsilon}(\tau)}\, d\tau = y_0 + \int_0 e^{z_1(\tau) + \varepsilon v_1(\tau)}\, d\tau = \\ &= y_0 + \int_0 e^{z_1(\tau)} e^{\varepsilon v_1(\tau)}\, d\tau = \\ &= y_0 + \int_0 e^{z_1(\tau)} [1 + \varepsilon v_1(\tau) + o(\varepsilon)] d\tau = \\ &= y_0 + \int_0 e^{z_1(\tau)}\, d\tau + \varepsilon \int_0 e^{z_1(\tau)} v_1(\tau)\, d\tau + o(\varepsilon) = \\ &= y(t) + \varepsilon \int_0 e^{z_1(\tau)} v_1(\tau)\, d\tau + o(\varepsilon), \end{aligned} \quad (5.3)$$

$$u_\varepsilon(t) = e^{u_{0\varepsilon} + \int_0 z_{2\varepsilon}(\tau)d\tau} = e^{u_0 + \varepsilon v_0 + \int_0 (z_2(\tau) + \varepsilon v_2(\tau))d\tau} =$$

$$= e^{u_0 + \int_0 z_2(\tau)d\tau + \varepsilon(v_0 + \int_0 v_2(\tau)d\tau)} =$$

$$= u(t) + \varepsilon u(t)(v_0 + \int_{t_0}^{t} v_2(\tau)\,d\tau) + o(\varepsilon). \tag{5.4}$$

It follows from (5.3) that

$$u(y_\varepsilon(t)) = e^{u_0 + \int_0^{y_\varepsilon(t)} z_2(\tau)d\tau} = e^{u_0 + \int_0^{y(t)+\varepsilon\int_0^t e^{z_1(\tau)}v_1(\tau)d\tau + o(\varepsilon)} z_2(\tau)d\tau} =$$

$$= e^{u_0 + \int_0^{y(t)} z_2(\tau)d\tau + \int_{y(t)}^{y(t)+\varepsilon\int_0^t e^{z_1(\tau)}v_1(\tau)d\tau + o(\varepsilon)} z_2(\tau)d\tau} =$$

$$= e^{u_0 + \int_0^{y(t)} z_2(\tau)d\tau + \varepsilon z_2(y(t))\int_0^t e^{z_1(\tau)}v_1(\tau)d\tau + o(\varepsilon)} =$$

$$= u(y(t)) + \varepsilon u(y(t)) z_2(y(t)) \int_0^t e^{z_1(\tau)} v_1(\tau)\,d\tau + o(\varepsilon). \tag{5.5}$$

Using (5.3), (5.4) and (5.5), one gets

$$u_\varepsilon(y_\varepsilon(t)) = u(y_\varepsilon(t))[1 + \varepsilon(v_0 + \int_{t_0}^{y_\varepsilon(t)} v_2(\tau)\,d\tau)] + o(\varepsilon) =$$

$$= u(y_\varepsilon(t))[1 + \varepsilon(v_0 + \int_{t_0}^{y(t)+\varepsilon\int_0^t e^{z_1(\tau)}v_1(\tau)d\tau} v_2(\tau)\,d\tau)] + o(\varepsilon) =$$

$$= u(y_\varepsilon(t))[1 + \varepsilon(v_0 + \int_{t_0}^{y(t)} v_2(\tau)d\tau + \int_{y(t)}^{y(t)+\varepsilon\int_0^t e^{z_1(\tau)}v_1(\tau)d\tau} v_2(\tau)d\tau)] + o(\varepsilon) =$$

$$= u(y_\varepsilon(t))[1 + \varepsilon(v_0 + \int_{t_0}^{y(t)} v_2(\tau)d\tau + \varepsilon v_2(y(t))\int_0^t e^{z_1(\tau)}v_1(\tau)d\tau + o(\varepsilon))] + o(\varepsilon) =$$

$$= [u(y(t)) + \varepsilon u(y(t)) z_2(y(t)) \int_0^t e^{z_1(\tau)} v_1(\tau)\,d\tau + o(\varepsilon)] \times$$

$$\times [1 + \varepsilon(v_0 + \int_{t_0}^{y(t)} v_2(\tau)\,d\tau) + o(\varepsilon)] + o(\varepsilon) =$$

$$= u(y(t)) + \varepsilon u(y(t))[v_0 + \int_{t_0}^{y(t)} v_2(\tau)\,d\tau + z_2(y(t)) \int_0^t e^{z_1(\tau)} v_1(\tau)\,d\tau] +$$

$$+ o(\varepsilon). \tag{5.6}$$

The relations (3.5), (5.2), (5.3), (5.4) and (5.6) imply

$$h_2(t, z_\varepsilon) = l(y(t) + \varepsilon \int_0^t e^{z_1(\tau)} v_1(\tau)\,d\tau + o(\varepsilon),$$

$$e^{z_1(t)} + \varepsilon e^{z_1(t)} v_1(t) + o(\varepsilon), u(t) + \varepsilon u(t)(v_0 + \int_{t_0}^{t} v_2(\tau)\,d\tau) + o(\varepsilon),$$

$$u(y(t))+\varepsilon u(y(t))[v_0+\int_{t_0}^{y(t)} v_2(\tau)\,d\tau+z_2(y(t))\int_0^t e^{z_1(\tau)}v_1(\tau)\,d\tau]+\mathrm{o}(\varepsilon),t)=$$
$$=h_2(t,z)+\varepsilon H_2(t,z,v)+\mathrm{o}(\varepsilon,t), \tag{5.7}$$

where

$$H_2(t,z,v)=$$
$$=\frac{\partial l(t)}{\partial y}\int_0^t e^{z_1(\tau)}v_1(\tau)\,d\tau+\frac{\partial l(t)}{\partial y'}e^{z_1(t)}v_1(t)+\frac{\partial l(t)}{\partial u}u(t)(v_0+\int_{t_0}^t v_2(\tau)\,d\tau)+$$
$$+\frac{\partial l(t)}{\partial u(y)}u(y(t))[v_0+\int_{t_0}^{y(t)} v_2(\tau)\,d\tau+z_2(y(t))\int_0^t e^{z_1(\tau)}v_1(\tau)\,d\tau], \tag{5.8}$$
$$\frac{\mathrm{o}(\varepsilon,t)}{\varepsilon}\underset{\varepsilon\downarrow 0}{\to}0.$$

(3.3) and (5.2) yeild

$$h_0(T,z_\varepsilon)=y_0-y_1+\int_0^T e^{z_1(t)+\varepsilon v_1(t)}\,dt=y_0-y_1+\int_0^T e^{z_1(t)}e^{\varepsilon v_1(t)}\,dt=$$
$$=y_0-y_1+\int_0^T e^{z_1(t)}[1+\varepsilon v_1(t)+\mathrm{o}(\varepsilon)]\,dt=$$
$$=y_0-y_1+\int_0^T e^{z_1(t)}\,dt+\varepsilon\int_0^T e^{z_1(t)}v_1(t)\,dt+\mathrm{o}(\varepsilon)=$$
$$=h_0(T,z)+\varepsilon\int_0^T e^{z_1(t)}v_1(t)\,dt. \tag{5.9}$$

We conclude from (3.4) and (5.2) that

$$h_1(t_0,t_1,z_\varepsilon)=$$
$$=\int_{t_0}^{t_1} e^{u_{0\varepsilon}+\int_0^t z_{2\varepsilon}(\tau)d\tau}\,dt-1=\int_{t_0}^{t_1} e^{u_0+\varepsilon v_0+\int_0^t z_2(\tau)d\tau+\varepsilon\int_0^t v_2(\tau)d\tau}\,dt-1=$$
$$=\int_{t_0}^{t_1} e^{u_0+\int_0^t z_2(\tau)d\tau}\,e^{\varepsilon(v_0+\int_0^t v_2(\tau)d\tau)}\,dt-1=$$
$$=\int_{t_0}^{t_1} e^{u_0+\int_0^t z_2(\tau)d\tau}[1+\varepsilon(v_0+\int_{t_0}^t v_2(\tau)\,d\tau)+\mathrm{o}(\varepsilon)]\,dt-1=$$
$$=h_1(t_0,t_1,z)+\varepsilon\int_{t_0}^{t_1}(v_0+\int_{t_0}^t v_2(\tau)\,d\tau)u(t)\,dt+\mathrm{o}(\varepsilon). \tag{5.10}$$

5.2 The case $z \notin Z$.

First consider the case $\varphi(z) > 0$, i.e. $z \notin Z$.

Theorem 5.1 *If $z \notin Z$, then the function $\varphi(z)$ is Gâteaux differentiable at the point z.*

Proof We have (see (4.7), (5.7), (5.9) and (5.10))

$$\varphi(z_\varepsilon) = [\int_0^T h_2^2(t, z_\varepsilon)\,dt + h_1^2(t_0, t_1, z_\varepsilon) + h_0^2(T, z_\varepsilon)]^{\frac{1}{2}} =$$

$$= [\int_0^T (h_2(t,z) + \varepsilon H_2(t,z,v) + o(\varepsilon,t))^2\,dt +$$

$$+ (h_1(t_0,t_1,z) + \varepsilon \int_{t_0}^{t_1} (v_0 + \int_{t_0}^{t} v_2(\tau)\,d\tau)u(t)\,dt + o(\varepsilon))^2 +$$

$$+ (h_0(T,z) + \varepsilon \int_0^T e^{z_1(t)} v_1(t)\,dt)^2]^{\frac{1}{2}} =$$

$$= \{\int_0^T [h_2^2(t,z) + 2\varepsilon h_2(t,z) H_2(t,z,v) + o(\varepsilon,t)]\,dt +$$

$$+ [h_1^2(t_0,t_1,z) + 2\varepsilon h_1(t_0,t_1,z) \int_{t_0}^{t_1} (v_0 + \int_{t_0}^{t} v_2(\tau)\,d\tau)u(t)\,dt + o(\varepsilon)] +$$

$$+ [h_0^2(T,z) + 2\varepsilon h_0(T,z) \int_0^T e^{z_1(t)} v_1(t)\,dt + o(\varepsilon)]\}^{\frac{1}{2}} =$$

$$= \varphi(z) + \varepsilon\{\int_0^T \frac{h_2(t,z)}{\varphi(z)} H_2(t,z,v)\,dt +$$

$$+ \frac{h_1(t_0,t_1,z)}{\varphi(z)} \int_{t_0}^{t_1} (v_0 + \int_{t_0}^{t} v_2(\tau)\,d\tau)u(t)\,dt +$$

$$+ \frac{h_0(T,z)}{\varphi(z)} \int_0^T e^{z_1(t)} v_1(t)\,dt\} + o(\varepsilon) =$$

$$= \varphi(z) + \varepsilon\varphi'(z,v) + o(\varepsilon), \tag{5.11}$$

where

$$\frac{o(\varepsilon)}{\varepsilon} \underset{\varepsilon \downarrow 0}{\to} 0,$$

$$\varphi'(z,v) = \lim_{\varepsilon \downarrow 0} \frac{\varphi(z_\varepsilon) - \varphi(z)}{\varepsilon} =$$

$$= \int_0^T \frac{h_2(t,z)}{\varphi(z)} H_2(t,z,v)\,dt + \frac{h_1(t_0,t_1,z)}{\varphi(z)} \int_{t_0}^{t_1} (v_0 + \int_{t_0}^{t} v_2(\tau)\,d\tau)u(t)\,dt +$$

$$+\frac{h_0(T,z)}{\varphi(z)}\int_0^T e^{z_1(t)}v_1(t)\,dt.$$

It follows from (5.8) that

$$\begin{aligned}
&\varphi'(z,v)=\\
&=\int_0^T\frac{h_2(t,z)}{\varphi(z)}H_2(t,z,v)\,dt+\frac{h_1(t_0,t_1,z)}{\varphi(z)}\int_{t_0}^{t_1}(v_0+\int_{t_0}^{t}v_2(\tau)\,d\tau)u(t)\,dt+\\
&+\frac{h_0(T,z)}{\varphi(z)}\int_0^T e^{z_1(t)}v_1(t)\,dt=\\
&=\int_0^T\frac{h_2(t,z)}{\varphi(z)}\{\frac{\partial l(t)}{\partial y}\int_0^t e^{z_1(\tau)}v_1(\tau)\,d\tau+\frac{\partial l(t)}{\partial y'}e^{z_1(t)}v_1(t)+\\
&+\frac{\partial l(t)}{\partial u}u(t)(v_0+\int_{t_0}^{t}v_2(\tau)\,d\tau)+\\
&+\frac{\partial l(t)}{\partial u(y)}u(y(t))[v_0+\int_{t_0}^{y(t)}v_2(\tau)\,d\tau+z_2(y(t))\int_0^t e^{z_1(\tau)}v_1(\tau)\,d\tau]\}\,dt+\\
&+\frac{h_1(t_0,t_1,z)}{\varphi(z)}\int_{t_0}^{t_1}(v_0+\int_{t_0}^{t}v_2(\tau)\,d\tau)u(t)\,dt+\frac{h_0(T,z)}{\varphi(z)}\int_0^T e^{z_1(t)}v_1(t)\,dt=\\
&=I_1+I_2+I_3,
\end{aligned}\tag{5.12}$$

where

$$\begin{aligned}
I_1=\{&\frac{h_1(t_0,t_1,z)}{\varphi(z)}\int_{t_0}^{t_1}u(t)\,dt+\int_0^T\frac{h_2(t,z)}{\varphi(z)}[\frac{\partial l(t)}{\partial u}u(t)+\\
&+\frac{\partial l(t)}{\partial u(y)}u(y(t))]\,dt\}v_0,
\end{aligned}\tag{5.13}$$

$$\begin{aligned}
I_2=\int_0^T\{&\frac{h_2(t,z)}{\varphi(z)}\frac{\partial l(t)}{\partial y}\int_0^t e^{z_1(\tau)}v_1(\tau)\,d\tau+\frac{h_2(t,z)}{\varphi(z)}\frac{\partial l(t)}{\partial y'}e^{z_1(t)}v_1(t)+\\
&+\frac{h_2(t,z)}{\varphi(z)}\frac{\partial l(t)}{\partial u(y)}u(y(t))z_2(y(t))\int_0^t e^{z_1(\tau)}v_1(\tau)\,d\tau+\\
&+\frac{h_0(T,z)}{\varphi(z)}e^{z_1(t)}v_1(t)\}\,dt,
\end{aligned}\tag{5.14}$$

$$I_3 = \int_0^T \{\frac{h_2(t,z)}{\varphi(z)} \frac{\partial l(t)}{\partial u} u(t) \int_{t_0}^{t} v_2(\tau)\, d\tau +$$
$$+ \frac{h_2(t,z)}{\varphi(z)} \frac{\partial l(t)}{\partial u(y)} u(y(t)) \int_{t_0}^{y(t)} v_2(\tau)\, d\tau\}\, dt +$$
$$+ \frac{h_1(t_0,t_1,z)}{\varphi(z)} \int_{t_0}^{t_1} u(t)(\int_{t_0}^{t} v_2(\tau)\, d\tau)\, dt. \tag{5.15}$$

Thus, at the point $z \notin Z$ the function φ is Dini directionally differentiable. Let us transform the formula (5.12). Put

$$\frac{h_0(T,z)}{\varphi(z)} := w_0(z) \in R, \tag{5.16}$$

$$\frac{h_1(t_0,t_1,z)}{\varphi(z)} := w_1(z) \in R, \tag{5.17}$$

$$\frac{h_2(t,z)}{\varphi(z)} := w_2(t,z) \in C[0,T]. \tag{5.18}$$

Clearly,

$$\|w(z)\| = [\int_0^T w_2^2(t,z)\, dt + w_1^2(z) + w_0^2(z)]^{\frac{1}{2}} = 1. \tag{5.19}$$

Integrating by parts in (5.14) and using the relation

$$\int_0^T a(t)(\int_0^t b(\tau) d\tau)\, dt = \int_0^T (\int_t^T a(\tau) d\tau) b(t)\, dt,$$

we get

$$I_1 = \{w_1(z) \int_{t_0}^{t_1} u(t)\, dt + \int_0^T w_2(t,z)[\frac{\partial l(t)}{\partial u} u(t) + \frac{\partial l(t)}{\partial u(y)} u(y(t))]\, dt\} v_0, \tag{5.20}$$

$$I_2 = \int_0^T \{\int_t^T w_2(\tau,z) \frac{\partial l(\tau)}{\partial y} d\tau + w_2(t,z) \frac{\partial l(t)}{\partial y'} +$$

$$+\int_t^T w_2(\tau,z)\frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))z_2(y(\tau))\,d\tau + w_0(z)\}e^{z_1(t)}v_1(t)\,dt =$$

$$= \int_0^T \{w_0(z) + w_2(t,z)\frac{\partial l(t)}{\partial y'} +$$

$$+\int_t^T w_2(\tau,z)[\frac{\partial l(\tau)}{\partial y} + \frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))z_2(y(\tau))]\,d\tau\}e^{z_1(t)}v_1(t)\,dt. \quad (5.21)$$

It is not difficult to show that

$$\int_0^T a(t)(\int_{t_0} b(\tau)d\tau)\,dt = \int_0^T (\int_t^T a(\tau)d\tau)b(t)\,dt + \int_{t_0}^0 (\int_0^T a(\tau)d\tau)b(t)\,dt, \quad (5.22)$$

$$\int_{t_0}^{t_1} a(t)(\int_{t_0} b(\tau)d\tau)\,dt = \int_{t_0}^{t_1} (\int_t^{t_1} a(\tau)d\tau)b(t)\,dt. \quad (5.23)$$

Now using in the first summand of (5.15) the relation (5.22), and in the third summand the relation (5.23), we have

$$I_3 = \int_0^T (\int_t^T w_2(\tau,z)\frac{\partial l(\tau)}{\partial u}u(\tau)\,d\tau)v_2(t)\,dt +$$

$$+\int_{t_0}^0 (\int_0^T w_2(\tau,z)\frac{\partial l(\tau)}{\partial u}u(\tau)\,d\tau)v_2(t)\,dt +$$

$$+\int_0^T w_2(t,z)\frac{\partial l(t)}{\partial u(y)}u(y(t))(\int_{t_0}^{y(t)} v_2(\tau)\,d\tau)\,dt +$$

$$+w_1(z)\int_{t_0}^{t_1} (\int_t^{t_1} u(\tau)\,d\tau)v_2(t)\,dt, \quad (5.24)$$

To transform the third summand in (5.24) we need the following property:

$$\int_0^T a(t)(\int_{t_0}^{y(t)} b(\tau)d\tau)\,dt =$$

$$= \int_{y_0}^{y(T)} (\int_{g(t)}^T a(\tau)d\tau)b(t)\,dt + \int_{t_0}^{y_0} (\int_0^T a(\tau)d\tau)b(t)\,dt, \quad (5.25)$$

where the function $g(t)$ is inverse to the function $y(t)$ (i.e. $y(g(t)) = t$).

Let us prove (5.25). Since the function $y(t)$ is continuously differentiable and monotonically increasing ($y'(t) > 0\ \forall t \in [0,T]$ (see (2.1)), then there exists the single-valued inverse function $g(t)$

($g'(t) > 0 \; \forall t \in [t_0, t_1]$), which is also monotonically increasing and continuously differentiable. The relation $y(0) = y_0$ yields $g(y_0) = 0$.

We have

$$S := \int_0^T a(t)\left(\int_{t_0}^{y(t)} b(\tau)d\tau\right)dt = \int_0^T \left(\int_{t_0}^{y(t)} b(\tau)d\tau\right) d\int_0^t a(\tau)d\tau =$$

$$= \left(\int_{t_0}^{y(t)} b(\tau)d\tau\right)\left(\int_0^t a(\tau)d\tau\right)\Big|_0^T - \int_0^T \left(\int_0^t a(\tau)d\tau\right) d\int_{t_0}^{y(t)} b(\tau)d\tau =$$

$$= \left(\int_{t_0}^{y(T)} b(\tau)d\tau\right)\left(\int_0^T a(\tau)d\tau\right) - \int_0^T \left(\int_0^t a(\tau)d\tau\right) y'(t)b((y(t))\,dt =$$

$$= \int_{t_0}^{y(T)} \left(\int_0^T a(t)dt\right)b(t)\,dt - \int_0^T \left(\int_0^t a(\tau)d\tau\right)b((y(t))\,dy(t). \qquad (5.26)$$

In the second integral of (5.26) let us change the variables: $y(t) = \gamma$. Then

$t = g(\gamma)\; dy(t) = d\gamma,$

putting $t = 0$ *one gets* $\gamma = y(0) = y_0,$

putting $t = T$ *one gets* $\gamma = y(T).$

We have

$$S = \int_{t_0}^{y(T)} \left(\int_0^T a(t)dt\right)b(t)\,dt - \int_{y_0}^{y(T)} \left(\int_0^{g(\gamma)} a(\tau)d\tau\right)b(\gamma)\,d\gamma =$$

$$= \int_{y_0}^{y(T)} \left(\int_{g(t)}^T a(t)dt\right)b(t)\,dt + \int_{t_0}^{y_0} \left(\int_0^T a(\tau)d\tau\right)b(t)\,dt.$$

The relations (5.24) and (5.25) yield

$$I_3 = \int_0^T \left(\int_t^T w_2(\tau, z)\frac{\partial l(\tau)}{\partial u}u(\tau)\,d\tau\right)v_2(t)\,dt +$$

$$+ \int_{t_0}^0 \left(\int_0^T w_2(\tau, z)\frac{\partial l(\tau)}{\partial u}u(\tau)\,d\tau\right)v_2(t)\,dt +$$

$$+ \int_0^T w_2(t, z)\frac{\partial l(t)}{\partial u(y)}u(y(t))\left(\int_{t_0}^{y(t)} v_2(\tau)\,d\tau\right)dt +$$

$$+ w_1(z)\int_{t_0}^{t_1}\left(\int_t^{t_1} u(\tau)\,d\tau\right)v_2(t)\,dt =$$

$$= \int_0^T \left(\int_t^T w_2(\tau, z)\frac{\partial l(\tau)}{\partial u}u(\tau)\,d\tau\right)v_2(t)\,dt +$$

$$+\int_{t_0}^{0}\Big(\int_0^T w_2(\tau,z)\frac{\partial l(\tau)}{\partial u}u(\tau)\,d\tau\Big)v_2(t)\,dt+$$
$$+\int_{y_0}^{y(T)}\Big(\int_{g(t)}^T w_2(\tau,z)\frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))\,d\tau\Big)v_2(t)\,dt+$$
$$+\int_{t_0}^{y_0}\Big(\int_0^T w_2(\tau,z)\frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))\,d\tau\Big)v_2(t)\,dt+$$
$$+w_1(z)\int_{t_0}^{t_1}\Big(\int_t^{t_1}u(\tau)\,d\tau\Big)v_2(t)\,dt.$$

Put

$$G_{21}(t,z)=\int_t^T w_2(\tau,z)\frac{\partial l(\tau)}{\partial u}u(\tau)\,d\tau, \tag{5.27}$$

$$G_{22}(t,z)=\int_{g(t)}^T w_2(\tau,z)\frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))\,d\tau. \tag{5.28}$$

Let us introduce the functions

$$\sigma_1(t):=\begin{cases}1, & if\ t\in[0,T],\\ 0, & if\ t\notin[0,T],\end{cases}\qquad \sigma_2(t):=\begin{cases}1, & if\ t\in[t_0,0)\ and\ t_0\neq 0,\\ 0, & if\ t\notin[t_0,0)\ or\ t_0=0,\end{cases} \tag{5.29}$$

$$\sigma_3(t):=\begin{cases}1, & if\ t\in[y_0,y(T)],\\ 0, & if\ t\notin[y_0,y(T)],\end{cases}\ \sigma_4(t):=\begin{cases}1, & if\ t\in[t_0,y_0)\ and\ t_0\neq y_0,\\ 0, & if\ t\notin[t_0,y_0)\ or\ t_0=y_0,\end{cases} \tag{5.30}$$

$$\sigma_5(t):=\begin{cases}1, & if\ t\in[t_0,t_1],\\ 0, & if\ t\notin[t_0,t_1].\end{cases} \tag{5.31}$$

Put

$$T_1=\max\{t_1,y(T)\}, \tag{5.32}$$

where $t_1=\max\{T,y_1\}$.

Using (5.27)-(5.31) one gets the following relation for I_3:

$$I_3 = \int_{t_0}^{T_1} [\sigma_1(t)G_{21}(t,z) + \sigma_2(t)G_{21}(0,z) +$$
$$+\sigma_3(t)G_{22}(t,z) + \sigma_4(t)G_{22}(y_0,z) + \tag{5.33}$$
$$+\sigma_5(t)w_1(z)\int_t^{T_1} u(\tau)\,d\tau]v_2(t)\,dt = \int_{t_0}^{T_1} G_2(t,z)v_2(t)\,dt.$$

Thus, the relations (5.12), (5.20), (5.21) and (5.33) show that if $z \notin Z$, then the function $\varphi(z)$ is Dini directionally differentiable at the point z and

$$\varphi'(z,v) = (G(z),v) = \int_0^T G_1(t,z)v_1(t)\,dt + \int_{t_0}^{T_1} G_2(t,z)v_2(t)\,dt + G_0(z)v_0, \tag{5.34}$$

where

$$G(z) = [G_1(t,z), G_2(t,z), G_0(z)] \in C[0,T] \times C[t_0,T_1] \times R, \tag{5.35}$$

$$G_0(z) = w_1(z)\int_{t_0}^{T_1} u(t)\,dt + \int_0^T w_2(t,z)[\frac{\partial l(t)}{\partial u}u(t) + \frac{\partial l(t)}{\partial u(y)}u(y(t))]\,dt, \tag{5.36}$$

$$G_1(t,z) = [w_0(z) + w_2(t,z)\frac{\partial l(t)}{\partial y'} +$$
$$+\int_t^T w_2(\tau,z)[\frac{\partial l(\tau)}{\partial y} + \frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))z_2(y(\tau))]d\tau]e^{z_1(t)}, \tag{5.37}$$

$$G_2(t,z) = \sigma_1(t)G_{21}(t,z) + \sigma_2(t)G_{21}(0,z) + \sigma_3(t)G_{22}(t,z) +$$
$$+\sigma_4(t)G_{22}(y_0,z) + \sigma_5(t)w_1(z)\int_t^{T_1} u(\tau)\,d\tau =$$
$$= \sigma_1(t)\int_t^T w_2(\tau,z)\frac{\partial l(\tau)}{\partial u}u(\tau)\,d\tau + \sigma_2(t)\int_0^T w_2(\tau,z)\frac{\partial l(\tau)}{\partial u}u(\tau)\,d\tau +$$

$$+\sigma_3(t)\int_{g(t)}^{T} w_2(\tau,z)\frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))\,d\tau +$$

$$+\sigma_4(t)\int_{0}^{T} w_2(\tau,z)\frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))\,d\tau + \tag{5.38}$$

$$+\sigma_5(t)w_1(z)\int_{t}^{t_1} u(\tau)\,d\tau,$$

the functions $\sigma_i(t)$ are defined in (5.29) – (5.31).

Here $w(z)=[w_2(t,z),w_1(z),w_0(z)]$ are described by (5.16)–(5.18) and satisfy (5.19). The relation (5.34) means that the function φ is Gâteaux differentiable at the point $z\in Z$, and the point $G(z)\in U$ can be viewed as the "gradient" of $\varphi(z)$ at the point z.

5.3 The case $z\in Z$.

Now let us consider the case $\varphi(z)=0$ (i.e. $z\in Z$).

Theorem 5.2 *If $z\in Z$, then the function φ is Dini directionally differentiable at the point z.*

Proof. It follows from (4.7), (5.7), (5.9) and (5.10) that

$$\varphi(z_\varepsilon)=[\int_0^T h_2^2(t,z_\varepsilon)\,dt+h_1^2(t_0,t_1,z_\varepsilon)+h_0^2(T,z_\varepsilon)]^{\frac{1}{2}}=$$

$$=[\int_0^T (h_2(t,z)+\varepsilon H_2(t,z,v)+o(\varepsilon,t))^2\,dt+$$

$$+(h_1(t_0,t_1,z)+\varepsilon\int_{t_0}^{t_1}(v_0+\int_{t_0}^{t} v_2(\tau)\,d\tau)u(t)\,dt+o(\varepsilon))^2+$$

$$+(h_0(T,z)+\varepsilon\int_0^T e^{z_1(t)}v_1(t)\,dt)^2]^{\frac{1}{2}}. \tag{5.39}$$

Since $\varphi(z)=0$, then

$$h_0(T,z)=0,\ h_1(t_0,t_1,z)=0,\ h_2(t,z)=0\ \ \forall t\in[0,T]. \tag{5.40}$$

Hence

$$\varphi(z_\varepsilon)=\varepsilon\|H_2(z,v);v\|,$$

where

$$\|H_2(z,v);v\| =$$

$$= [\int_0^T H_2^2(t,z,v)dt + (\int_{t_0}^{t_1}[v_0 + \int_{t_0}^{t} v_2(\tau)\,d\tau]u(t)dt)^2 + (\int_0^T e^{z_1(t)}v_1(t)dt)^2]^{\frac{1}{2}}.$$

$$\varphi'(z,v) = \lim_{\varepsilon\downarrow 0}\frac{\varphi(z_\varepsilon)-\varphi(z)}{\varepsilon} = \|H_2(z,v);v\| =$$

$$= \max_{\|w\|\le 1}[\int_0^T H_2(t,z,v)w_2(t)\,dt + w_1\int_{t_0}^{t_1}[v_0 + \int_{t_0}^{t} v_2(\tau)\,d\tau]u(t)\,dt +$$

$$+ w_0\int_0^T e^{z_1(t)}v_1(t)\,dt]. \tag{5.41}$$

Here

$$w = [w_2, w_1, w_0],\ w_2(t)\in C[0,T],\ \ w_1\in R,\ \ w_0\in R,$$

$$\|w\| = [\int_0^T w_2^2(t)\,dt + w_1^2 + w_0^2]^{\frac{1}{2}} = 1. \tag{5.42}$$

Substituting (5.8) in (5.41) one gets

$$\varphi'(z,v) =$$

$$= \max_{w\in W}[\int_0^T w_2(t)\{\frac{\partial l(t)}{\partial y}\int_0^t e^{z_1(\tau)}v_1(\tau)\,d\tau + \frac{\partial l(t)}{\partial y'}e^{z_1(t)}v_1(t) +$$

$$+\frac{\partial l(t)}{\partial u}u(t)(v_0 + \int_{t_0}^{t} v_2(\tau)\,d\tau) +$$

$$+\frac{\partial l(t)}{\partial u(y)}u(y(t))[v_0 + \int_{t_0}^{y(t)} v_2(\tau)\,d\tau + z_2(y(t))\int_0^t e^{z_1(\tau)}v_1(\tau)\,d\tau]\}\,dt +$$

$$+w_1\int_{t_0}^{t_1}(v_0 + \int_{t_0}^{t} v_2(\tau)\,d\tau)u(t)\,dt + w_0\int_0^T e^{z_1(t)}v_1(t)\,dt] =$$

$$= J_1 + J_2 + J_3, \tag{5.43}$$

where

$$J_1 = \{w_1\int_{t_0}^{t_1} u(t)\,dt + \int_0^T w_2(t)[\frac{\partial l(t)}{\partial u}u(t) + \frac{\partial l(t)}{\partial u(y)}u(y(t))]\,dt\}v_0, \tag{5.44}$$

$$J_2 = \int_0^T \{w_2(t)\frac{\partial l(t)}{\partial y}\int_0^{} e^{z_1(\tau)}v_1(\tau)\,d\tau + w_2(t)\frac{\partial l(t)}{\partial y'}e^{z_1(\tau)}v_1(t) +$$

$$+w_2(t)\frac{\partial l(t)}{\partial u(y)}u(y(t))z_2(y(t))\int_0^{} e^{z_1(\tau)}v_1(\tau)\,d\tau + w_0 e^{z_1(t)}v_1(t)\}\,dt, \quad (5.45)$$

$$J_3 = \int_0^T \{w_2(t)\frac{\partial l(t)}{\partial u}u(t)\int_{t_0}^{} v_2(\tau)\,d\tau +$$

$$+w_2(t)\frac{\partial l(t)}{\partial u(y)}u(y(t))\int_{t_0}^{y(t)} v_2(\tau)\,d\tau\}\,dt + \quad (5.46)$$

$$+w_1\int_{t_0}^{t_1} u(t)(\int_{t_0}^{} v_2(\tau)\,d\tau)\,dt,$$

$$W = \{w = [w_2, w_1, w_0] \mid w_0 \in R,\ w_1 \in R,\ w_2 \in C[0,T],$$

$$\|w\| = [\int_0^T w_2^2(t)\,dt + w_1^2 + w_0^2]^{\frac{1}{2}} \le 1\}. \quad (5.47)$$

As in the case $z \notin Z$, integrating by parts in (5.44), (5.45), (5.46) and making use of the condition

$$\int_{t_0}^{t_1} u(t)\,dt = 1, \quad T_1 = t_1,$$

we have

$$J_1 = \{w_1 + \int_0^T w_2(t)[\frac{\partial l(t)}{\partial u}u(t) + \frac{\partial l(t)}{\partial u(y)}u(y(t))]\,dt\}v_0, \quad (5.48)$$

$$J_2 = \int_0^T \{w_0 + w_2(t)\frac{\partial l(t)}{\partial y'} +$$

$$+\int_t^T w_2(\tau)[\frac{\partial l(\tau)}{\partial y} + \frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))z_2(y(\tau))]\,d\tau\}e^{z_1(t)}v_1(t)\,dt, \quad (5.49)$$

$$J_3 = \int_0^T (\int_t^T w_2(\tau)\frac{\partial l(\tau)}{\partial u}u(\tau)\,d\tau)v_2(t)\,dt +$$

$$+\int_{t_0}^0 (\int_0^T w_2(\tau)\frac{\partial l(\tau)}{\partial u}u(\tau)\,d\tau)v_2(t)\,dt +$$

$$+\int_{y_0}^{y(T)}\Big(\int_{g(t)}^{T} w_2(\tau)\frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))\,d\tau\Big)v_2(t)\,dt+$$

$$+\int_{t_0}^{y_0}\Big(\int_{0}^{T} w_2(\tau)\frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))\,d\tau\Big)v_2(t)\,dt+$$

$$+w_1\int_{t_0}^{t_1}\Big(\int_{t}^{t_1} u(\tau)\,d\tau\Big)v_2(t)\,dt=$$

$$=\int_{t_0}^{t_1}[\sigma_1(t)A_{21}(t)+\sigma_2(t)A_{21}(0)+\sigma_3(t)A_{22}(t)+$$

$$+\sigma_4(t)A_{22}(y_0)+w_1\int_{t}^{t_1}u(\tau)\,d\tau]v_2(t)\,dt=$$

$$=\int_{t_0}^{t_1}A_2(t)v_2(t)\,dt, \tag{5.50}$$

where $\sigma_i(t)$ are defined in (5.29)-(5.30), and

$$A_{21}(t)=\int_{t}^{T} w_2(\tau)\frac{\partial l(\tau)}{\partial u}u(\tau)\,d\tau, \tag{5.51}$$

$$A_{22}(t)=\int_{g(t)}^{T} w_2(\tau)\frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))\,d\tau. \tag{5.52}$$

From (5.43), (5.48), (5.49) and (5.50) one can conclude that at the point $z\in Z$ the function $\varphi(z)$ is Dini directionally differentiable, moreover, it is even subdifferentiable, i.e.

$$\varphi'(z,v)=\max_{A\in\partial\varphi(z)}(A,v), \tag{5.53}$$

where

$$(A,v)=\int_{t_0}^{t_1}A_2(t)v_2(t)\,dt+\int_{0}^{T}A_1(t)v_1(t)\,dt+A_0v_0,$$

$$\partial\varphi(z)=\{A=[A_1(t),A_2(t),A_0]\in U\mid$$

$$A_1(t)=$$

$$=[w_0+w_2(t)\frac{\partial l(t)}{\partial y'}+\int_{t}^{T}w_2(\tau)[\frac{\partial l(\tau)}{\partial y}+\frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))z_2(y(\tau))]d\tau]e^{z_1(t)},$$

$$A_2(t)=\sigma_1(t)A_{21}(t)+\sigma_2(t)A_{21}(0)+\sigma_3(t)A_{22}(t)+$$

$$+\sigma_4(t)A_{22}(y_0)+w_1\int_t^{t_1}u(\tau)\,d\tau=$$

$$=\sigma_1(t)\int_t^T w_2(\tau)\frac{\partial l(\tau)}{\partial u}u(\tau)\,d\tau+\sigma_2(t)\int_0^T w_2(\tau)\frac{\partial l(\tau)}{\partial u}u(\tau)\,d\tau+$$

$$+\sigma_3(t)\int_{g(t)}^T w_2(\tau)\frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))\,d\tau+$$

$$+\sigma_4(t)\int_0^T w_2(\tau)\frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))\,d\tau+w_1\int_t^{t_1}u(\tau)\,d\tau,$$

$$A_0=w_1+\int_0^T w_2(t)[\frac{\partial l(t)}{\partial u}u(t)+\frac{\partial l(t)}{\partial u(y)}u(y(t))]\,dt, \qquad (5.54)$$

$$w=[w_2,w_1,w_0]\in W\},$$

and the set W is defined in (5.47).

Remark 5.1. *Arguing as in the prof of Theorem 5.2, one gets the following representation for the functional* $f(z)$:

$$f(z_\varepsilon)=f(z)+\varepsilon\int_0^T[\frac{\partial F(t)}{\partial y}\int_0^t e^{z_1(\tau)}v_1(\tau)\,d\tau+$$

$$+\frac{\partial F(t)}{\partial y'}e^{z_1(t)}v_1(t)+\frac{\partial F(t)}{\partial u}u(t)(v_0+\int_{t_0}^t v_2(\tau)\,d\tau)+$$

$$+\frac{\partial F(t)}{\partial u(y)}u(y(t))(v_0+\int_{t_0}^{y(t)}v_2(\tau)\,d\tau+z_2(y(t))\int_0^t e^{z_1(\tau)}v_1(\tau)\,d\tau)]\,dt+o(\varepsilon)=$$

$$=f(z)+\varepsilon(B(z),v)+o(\varepsilon), \qquad (5.55)$$

where

$$(B(z),v)=\int_{t_0}^{t_1}B_2(t,z)v_2(t)\,dt+\int_0^T B_1(t,z)v_1(t)\,dt+B_0(z)v_0,$$

$$B_1(t,z)=\{\frac{\partial F(t)}{\partial y'}+\int_t^T[\frac{\partial F(\tau)}{\partial y}+\frac{\partial F(\tau)}{\partial u(y)}u(y(\tau))z_2(y(\tau))]\,d\tau\}e^{z_1(t)}, \qquad (5.56)$$

$$B_2(t,z)=\sigma_1(t)B_{21}(t,z)+\sigma_2(t)B_{21}(0,z)+\sigma_3(t)B_{22}(t,z)+$$

$$+\sigma_4(t)B_{22}(y_0,z)=\sigma_1(t)\int_t^T\frac{\partial F(\tau)}{\partial u}u(\tau)\,d\tau+$$

$$+\sigma_2(t)\int_0^T\frac{\partial F(\tau)}{\partial u}u(\tau)\,d\tau+$$

$$+\sigma_3(t)\int_{g(t)}^T\frac{\partial F(\tau)}{\partial u(y)}u(y(\tau))\,d\tau+\sigma_4(t)\int_0^T\frac{\partial F(\tau)}{\partial u(y)}u(y(\tau))\,d\tau, \qquad (5.57)$$

$$B_0(z)=\int_0^T[\frac{\partial F(t)}{\partial u}u(t)+\frac{\partial F(t)}{\partial u(y)}u(y(t))]\,dt, \qquad (5.58)$$

$$\frac{o(\varepsilon)}{\varepsilon}\underset{\varepsilon\downarrow 0}{\to}0,$$

and $\sigma_i(t)$ *is defined in (5.29) – (5.30).*

6. AN EXACT PENALTY FUNCTION

6.1 Properties of the function G.

Let us consider again the case $z\notin Z$. It was shown in Subsection 5.1 that the function φ is Gâteaux differentiable at a point $z\notin Z$, and the corresponding "gradient" $G(t,z)$ of the function φ $G(t,z)$ (in the space $U=C[0,T]\times C[t_0,t_1]\times R$) is given by the relations (5.35)–(5.38).

Since it is known that $y_0<y_1$ and $T>0$, the following 63 cases of allocation of these points (i.e. the points $y_0,0,T,y(T),y_1$) are possible, namely

$1\,case:\ y_0<0<T<y(T)<y_1$; $2\,case:\ y_0<0<T<y(T)=y_1$;

$3\,case:\ y_0<0<T<y_1<y(T)$; $4\,case:\ y_0<0<T=y_1<y(T)$;

$5\,case:\ y_0<0<y_1<T<y(T)$; $6\,case:\ y_0<0=y_1<T<y(T)$;

$7\,case:\ y_0<y_1<0<T<y(T)$; $8\,case:\ y_0<0<T=y(T)<y_1$;

$9\,case:\ y_0<0<T=y(T)=y_1$; $10\,case:\ y_0<0<y_1<T=y(T)$;

$11\,case:\ y_0<0=y_1<T=y(T)$; $12\,case:\ y_0<y_1<0<T=y(T)$;

$13\,case:\ y_0=0<T<y(T)<y_1$; $14\,case:\ y_0=0<T<y(T)=y_1$;

$15\,case:\ y_0 = 0 < T < y_1 < y(T);$ $16\,case:\ y_0 = 0 < T = y_1 < y(T);$
$17\,case:\ y_0 = 0 < y_1 < T < y(T);$ $18\,case:\ y_0 = 0 < y(T) < T < y_1;$
$19\,case:\ y_0 = 0 < y(T) < T = y_1;$ $20\,case:\ y_0 = 0 < y(T) < y_1 < T;$
$21\,case:\ y_0 = 0 < y(T) = y_1 < T;$ $22\,case:\ y_0 = 0 < y_1 < y(T) < T;$
$23\,case:\ y_0 < 0 < y(T) < T < y_1;$ $24\,case:\ y_0 < 0 < y(T) < T = y_1;$
$25\,case:\ y_0 < 0 < y(T) < y_1 < T;$ $26\,case:\ y_0 < 0 < y(T) = y_1 < T;$
$27\,case:\ y_0 < 0 < y_1 < y(T) < T;$ $28\,case:\ y_0 < 0 = y_1 < y(T) < T;$
$29\,case:\ y_0 < y_1 < 0 < y(T) < T;$ $30\,case:\ y_0 < y(T) < 0 < T < y_1;$
$31\,case:\ y_0 < y(T) < 0 < T = y_1;$ $32\,case:\ y_0 < y(T) < 0 < y_1 < T;$
$33\,case:\ y_0 < y(T) < 0 = y_1 < T;$ $34\,case:\ y_0 < y(T) < y_1 < 0 < T;$
$35\,case:\ y_0 < y(T) = y_1 < 0 < T;$ $36\,case:\ y_0 < y_1 < y(T) < 0 < T;$
$37\,case:\ 0 < y_0 < y(T) < T < y_1;$ $38\,case:\ 0 < y_0 < y(T) < T = y_1;$
$39\,case:\ 0 < y_0 < y(T) < y_1 < T;$ $40\,case:\ 0 < y_0 < y(T) = y_1 < T;$
$41\,case:\ 0 < y_0 < y_1 < y(T) < T;$ $42\,case:\ 0 < y_0 < y(T) = T < y_1;$
$43\,case:\ 0 < y_0 < y(T) = T = y_1;$ $44\,case:\ 0 < y_0 < y_1 < y(T) = T;$
$45\,case:\ 0 < y_0 < T < y(T) < y_1;$ $46\,case:\ 0 < y_0 < T < y(T) = y_1;$
$47\,case:\ 0 < y_0 < T < y_1 < y(T);$ $48\,case:\ 0 < y_0 < T = y_1 < y(T);$
$49\,case:\ 0 < y_0 < y_1 < T < y(T);$ $50\,case:\ 0 < T < y_0 < y(T) < y_1;$
$51\,case:\ 0 < T < y_0 < y(T) = y_1;$ $52\,case:\ 0 < T < y_0 < y_1 < y(T);$
$53\,case:\ y_0 < 0 = y(T) < T < y_1;$ $54\,case:\ y_0 < 0 = y(T) < T = y_1;$
$55\,case:\ y_0 < 0 = y(T) < y_1 < T;$ $56\,case:\ y_0 < 0 = y(T) = y_1 < T;$
$57\,case:\ y_0 < y_1 < 0 = y(T) < T;$ $58\,case:\ 0 < y_0 = T < y(T) < y_1;$
$59\,case:\ 0 < y_0 = T < y(T) = y_1;$ $60\,case:\ 0 < y_0 = T < y_1 < y(T);$
$61\,case:\ y_0 = 0 < T = y(T) < y_1;$ $62\,case:\ y_0 = 0 < T = y(T) = y_1;$

$63\,case:\ y_0 = 0 < y_1 < T = y(T).$

For the cases 1–53, 58 and 60 the following proposition holds.

Theorem 6.1 *Let $S_i \subset U$ be a bounded set (in the metric ρ_i). If the conditions*

$$|\frac{\partial l(t)}{\partial y}|\geq b_0>0,\ |\frac{\partial l(t)}{\partial y'}|\geq b_1>0,\ |\frac{\partial l(t)}{\partial u}|\geq b_2>0,$$
$$|\frac{\partial l(t)}{\partial u(y)}|\geq b_3>0\ \ \forall t\in[0,T],\forall z\in S_i\setminus Z \tag{6.1}$$

hold then there exist $a_1>0$ *and* $a_2>0$, *such that*

$$\|G(z)\|=[\int_{t_0}^{T_1}G_2^2(t,z)dt+\int_0^T G_1^2(t,z)dt+G_0^2(z)]^{\frac{1}{2}}\geq a_1\ \ \forall z\in S_i\setminus Z, \tag{6.2}$$

$$b(z)=\sup_{t\in[t_0,T_1]}|G_2(t,z)|+\sup_{t\in[0,T]}|G_1(t,z)|+|G_0(z)|\geq a_2\ \ \forall z\in S_i\setminus Z. \tag{6.3}$$

Proof. Let $z\in S_i\setminus Z$. First let us show that

$$\|G(z)\|\neq 0. \tag{6.4}$$

Here **0** is the zero element of the space $U=C[0,T]\times C[t_0,T_1]\times R$, i.e.

$$0=[0_{C[0,T]},0_{C[t_0,T_1]},0].$$

Consider the case *2*, i.e. $y_0<0<T<y(T)=y_1$. In this case $T_1=t_1$ $\sigma_4(t)=0,\ \sigma_5(t)=1;\forall t\in[t_0,t_1]$ (see (5.29)–(5.31)), since $t_0=\min\{0,y_0\}=y_0$. Assume that (6.4) is not valid, that is

$$G_1(t,z)=0\ \ \forall t\in[0,T],\quad G_2(t,z)=0\ \ \forall t\in[t_0,t_1],\quad G_0(z)=0. \tag{6.5}$$

The relations (5.35)–(5.38) yield

$$G_2(t,z)=\sigma_1(t)\int_t^T w_2(\tau,z)\frac{\partial l(\tau)}{\partial u}u(\tau)d\tau+$$
$$+\sigma_2(t)\int_0^T w_2(\tau,z)\frac{\partial l(\tau)}{\partial u}u(\tau)d\tau+ \tag{6.6}$$
$$+\sigma_3(t)\int_{g(t)}^T w_2(\tau,z)\frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))d\tau+w_1(z)\int_t^{t_1}u(\tau)d\tau=0,$$

where $w=[w_2, w_1, w_0]$ satisfies (5.16)-(5.19).

Differentiating (6.6) w.r.t. t for $t \in (T, y(T)]$ and taking into account that in this interval $\sigma_3(t)=1$, $\sigma_1(t)=0$ and $\sigma_2(t)=0$, one gets

$$w_2(g(t),z)\frac{\partial l(g(t))}{\partial u(y)}u(y(g(t)))g'(t)+w_1(z)u(t)=0 \quad \forall t \in (T, y(T)].$$

Since

$$u(t)>0 \ \forall t \in [t_0, t_1], \quad u(y(g(t)))=u(t), \tag{6.7}$$

then dividing the both parts by $u(t)$, we have

$$w_2(g(t),z)\frac{\partial l(g(t))}{\partial u(y)}g'(t)+w_1(z)=0 \quad \forall t \in (T, y(T)]. \tag{6.8}$$

Again differentiating (6.6) w.r.t. t for $t \in [0, T)$ and taking into account that in this interval $\sigma_1(t)=1$, $\sigma_3(t)=1$ and $\sigma_2(t)=0$, one gets

$$w_2(t,z)\frac{\partial l(t)}{\partial u}u(t)+w_2(g(t),z)\frac{\partial l(g(t))}{\partial u(y)}u(y(g(t)))g'(t)+$$
$$+w_1(z)u(t)=0, \quad \forall t \in [0, T).$$

Dividing the both parts of this equality by $u(t)$ (see (6.7)), we have

$$w_2(t,z)\frac{\partial l(t)}{\partial u}+w_2(g(t),z)\frac{\partial l(g(t))}{\partial u(y)}g'(t)+w_1(z)=0 \quad \forall t \in [0, T). \tag{6.9}$$

Since the functions in the left-hand parts of the relations (6.8) and (6.9) are continuous, then tending in (6.8) t to T from the right, and in (6.9) t – to T from the left, one gets

$$w_2(T,z)\frac{\partial l(T)}{\partial u}=0.$$

However, since (see (6.1))

$$|\frac{\partial l(t,z)}{\partial u}| \geq b_2 > 0 \quad \forall t \in [t_0, t_1],$$

then

$$w_2(T,z) = 0. \tag{6.10}$$

It follows from the expression for $G_1(t,z)$ (see (5.37)) at $t = T$ that

$$w_0(z) = -w_2(T,z)\frac{\partial l(T)}{\partial y'}.$$

It follows from this relation and (6.10) that

$$w_0(z) = 0. \tag{6.11}$$

The relations (6.5) and (5.37) imply

$$G_1(t,z) = w_0(z) + w_2(t,z)\frac{\partial l(t)}{\partial y'} +$$
$$+\int_t^T w_2(\tau,z)[\frac{\partial l(\tau)}{\partial y} + \frac{\partial l(\tau)}{\partial u(y)}u(y(\tau))z_2(y(\tau))]d\tau = 0 \quad \forall t \in [0,T]. \tag{6.12}$$

Since $w_0(z) = 0$ and $w_2(T,z) = 0$, then there exists a unique (namely, the zero one) solution of the homogeneous integral Volterra equation (6.12):

$$w_2(t,z) = 0 \ \forall t \in [0,T]. \tag{6.13}$$

The relation (6.5) and (5.36) imply

$$G_0(z) = w_1(z)\int_{t_0}^{t_1} u(t)\,dt + \int_0^T w_2(t,z)[\frac{\partial l(t)}{\partial u}u(t) + \frac{\partial l(t)}{\partial u(y)}u(y(t))]\,dt = 0. \tag{6.14}$$

Since

$$u(t) > 0 \ \forall t \in [t_0, t_1]$$

then $\int_0^T u(t)\,dt > 0$ and (6.14) yields

$$w_1(z) = 0. \tag{6.15}$$

The relations (6.11), (6.13) and (6.15) imply

$$\|w(z)\| = [\int_0^T w_2^2(t,z)\,dt + w_1^2(z) + w_0^2(z)]^{\frac{1}{2}} = 0,$$

which contradicts (5.19). Thus, (6.4) is proved.

Now let us assume that (6.2) does not hold. Then there exists a sequence

$$\{z_k\} = \{z_{1k}, z_{2k}, u_{0k}\},$$

such that

$$z_k \in S_i \setminus Z \quad \forall k, \quad \|G(z_k)\| \underset{k\to\infty}{\to} 0. \tag{6.16}$$

Since

$$\begin{aligned}
&\|G(z_k)\| = \int_{t_0}^{t_1} G_2^2(t,z_k)\,dt + \int_0^T G_1^2(t,z_k)\,dt + G_0^2(z_k) = \\
&= \int_{t_0}^{t_1} [\sigma_1(t)G_{21}(t,z_k) + \sigma_2(t)G_{21}(0,z_k) + \sigma_3(t)G_{22}(t,z_k) + \\
&+ w_1(z_k)\int_t^{t_1} u_k(\tau)\,d\tau]^2\,dt + \int_0^T G_1^2(t,z_k)\,dt + G_0^2(z_k),
\end{aligned} \tag{6.17}$$

where (see (5.35) – (5.38))

$$\begin{aligned}
G_0(z_k) = w_1(z_k)\int_{t_0}^{t_1} u_k(t)\,dt + \int_0^T w_2(t,z_k)[\frac{\partial l_k(t)}{\partial u}u_k(t) + \\
+ \frac{\partial l_k(t)}{\partial u(y)}u_k(y_k(t))]\,dt,
\end{aligned} \tag{6.18}$$

$$\begin{aligned}
&G_1(t,z_k) = \{w_0(z_k) + w_2(t,z_k)\frac{\partial l_k(t)}{\partial y'} + \\
&+ \int_t^T w_2(\tau,z_k)\{\frac{\partial l_k(\tau)}{\partial y} + \frac{\partial l_k(\tau)}{\partial u(y)}u_k(y_k(\tau))z_{2k}(y_k(\tau))\}\,d\tau\}e^{z_{1k}(t)},
\end{aligned} \tag{6.19}$$

$$G_{21}(t,z_k)=\int_t^T w_2(\tau,z_k)\frac{\partial l_k(\tau)}{\partial u}u_k(\tau)\,d\tau, \tag{6.20}$$

$$G_{22}(t,z_k)=\int_{g_k(t)}^T w_2(\tau,z_k)\frac{\partial l_k(\tau)}{\partial u(y)}u_k(y_k(\tau))\,d\tau, \tag{6.21}$$

$$y_k(t)=y_0+\int_0^t e^{z_{1k}(\tau)}\,d\tau,\quad u_k(t)=e^{u_{0k}+\int_0^t z_{2k}(\tau)d\tau},$$
$$\frac{\partial l_k(t)}{\partial y}=\frac{\partial l(y_k(t),y_k(t),u_k(t),u_k(y_k(t)),t)}{\partial y},$$

then (6.16) and (6.17) for $t\in(T,y(T)]$ imply

$$G_{22}(t,z_k)+w_1(z_k)\int_t^{\vartheta_1}u_k(\tau)\,d\tau\underset{k\to\infty}{\to}0\quad\forall t\in(T,y(T)]. \tag{6.22}$$

Here

$$G_{22}(t,z_k)=\int_{g_k(t)}^T w_2(\tau,z_k)\frac{\partial l_k(\tau)}{\partial u(y)}u_k(y_k(\tau))\,d\tau\quad\forall t\in(T,y(T)].$$

The relation (6.22) is valid for all $t\in(T,y(T)]$, therefore integrating (6.22) w.r.t. t and dividing by $u_k(t)$, one gets

$$w_2(g_k(t),z_k)\frac{\partial l_k(g_k(t))}{\partial u(y)}g_k(t)+w_1(z_k)\underset{k\to\infty}{\to}0\quad\forall t\in(T,y(T)]. \tag{6.23}$$

It follows from (6.16) and (6.17) for $t\in[0,T]$ that

$$G_{21}(t,z_k)+G_{22}(t,z_k)+w_1(z_k)\int_t^{\vartheta_1}u_k(\tau)\,d\tau\underset{k\to\infty}{\to}0\quad\forall t\in(T,y(T)], \tag{6.24}$$

where

$$G_{21}(t,z_k)=\int_t^T w_2(\tau,z_k)\frac{\partial l_k(\tau)}{\partial u}u_k(\tau)\,d\tau\quad\forall t\in[0,T].$$

The relation (6.24) holds for all $t \in [0,T]$, therefore integrating (6.24) w.r.t. t and dividing by $u_k(t)$, one gets

$$w_2(t,z_k)\frac{\partial l_k(t)}{\partial u} + w_2(g_k(t),z_k)\frac{\partial l_k(g_k(t))}{\partial u(y)} g_k(t) + w_1(z_k) \underset{k\to\infty}{\to} 0 \quad \forall t \in [0,T]. \tag{6.25}$$

Since the functions in the left-hand sides of (6.24) ane (6.25) are continuous, then passing t in (6.24) to T from the right, and in (6.25) to T from the left, one gets

$$w_2(T,z_k)\frac{\partial l_k(T)}{\partial u} \underset{k\to\infty}{\to} 0. \tag{6.26}$$

However (see (6.1))

$$\left|\frac{\partial l_k(t,z)}{\partial u}\right| \geq b_2 > 0 \quad \forall t \in [t_0,t_1],$$

therefore (6.26) yields

$$w_2(T,z_k) \underset{k\to\infty}{\to} 0. \tag{6.27}$$

Taking into account the expression for $G_1(t,z_k)$ (see (6.19)) and putting $t = T$, we get

$$w_0(z_k) = -w_2(T,z_k)\frac{\partial l_k(T)}{\partial y'},$$

but for $t = T$

$$w_2(T,z_k) \underset{k\to\infty}{\to} 0,$$

therefore

$$w_0(z_k) \underset{k\to\infty}{\to} 0. \tag{6.28}$$

The relations (6.19), (6.1) and (6.28) yield the following integral Volterra equation:

$$e^{-z_{1k}(t)}G_1(t,z_k) = w_2(t,z_k)\frac{\partial l_k(t)}{\partial y'} +$$
$$+\int_t^T w_2(\tau,z_k)\{\frac{\partial l_k(\tau)}{\partial y}+\frac{\partial l_k(\tau)}{\partial u(y)}u_k(y_k(\tau))z_{2k}(y_k(\tau))\}\,d\tau, \qquad (6.29)$$

whose kernel

$$[\frac{\partial l_k(t)}{\partial y'}]^{-1}\{\frac{\partial l_k(\tau)}{\partial y}+\frac{\partial l_k(\tau)}{\partial u(y)}u_k(y_k(\tau))z_{2k}(y_k(\tau))\}$$

is bounded on S_i (remind that the set S_i is bounded), therefore the properties of the solutions of integral equations show that

$$w_2(t,z_k) = f(t,z_k)+\int_t^T R_k(t,\tau)f(\tau,z_k)\,d\tau, \qquad (6.30)$$

where $R_k(t,\tau)$ is the resolvent of the equation (6.29), and its norm is bounded by the same constant for all k' s,

$$f(t,z_k) = G_1(t,z_k)e^{-z_{1k}(t)}[\frac{\partial l_k(t)}{\partial y'}]^{-1}.$$

Due to the continuous dependence of the solutions of integral equations on the right-hand sides, one concludes from the relations (6.16), (6.17) and (6.30) that

$$w_2(t,z_k) \underset{k\to\infty}{\to} 0 \quad \forall t\in[0,T]. \qquad (6.31)$$

Since

$$u_k(t) > 0\ \forall t\in[t_0,t_1]$$

then $\int_0^T u_k(t)\,dt > 0$, therefore (see (6.18) and (6.31))

$$w_1(z_k) \underset{k\to\infty}{\to} 0. \qquad (6.32)$$

It follows from (6.28), (6.31) and (6.32) that

$$\|w(z_k)\| = [\int_0^T w_2^2(t,z_k)\,dt + w_1^2(z_k) + w_0^2(z_k)]^{\frac{1}{2}} \underset{k\to\infty}{\to} 0,$$

which contradicts the relation $\|w(z_k)\| = 1 \ \forall k$. Thus (6.2) is proved. The relation (6.3) is proved analogously. In a similar way one can show the relations (6.2)-(6.3) in all remaining cases.

For the cases 54–57, 59, 61–63 the following proposition holds.

Theorem 6.2 *Assume that the relation (6.1) holds and*
for the cases 54–57 and 59 at least one of the conditions 1) or 3) below takes place;
for the case 61 at least one of the conditions 1) or 2) is valid;
for the cases 62 and 63 the condition 1) is satisfied:
1) there exists $\overline{t} \in [0,T]$, *such that*

$$l(y(\overline{t}), y'(\overline{t}), u(\overline{t}), u(y(\overline{t})), \overline{t}) = 0 \quad \forall z \in S_i \setminus Z, \tag{6.33}$$

2)

$$\frac{\partial l(0,z)}{\partial u} + \frac{\partial l(0,z)}{\partial u(y)} g'(0) \neq 0 \ , \frac{\partial l(T,z)}{\partial u} + \frac{\partial l(T,z)}{\partial u(y)} g'(T) \neq 0, \quad \forall z \in S_i \setminus Z, \tag{6.34}$$

3) there exists $\overline{t} \in [0,T]$, *such that*

$$\frac{\partial l(\overline{t},z)}{\partial u} y'(\overline{t}) - \frac{\partial l(\overline{t},z)}{\partial u(y)} \neq 0 \quad \forall z \in S_i \setminus Z. \tag{6.35}$$

Then there exist $a_1 > 0$ *and* $a_2 > 0$, *such that*

$$\|G(z)\| = [\int_{t_0}^{t_1} G_2^2(t,z)\,dt + \int_0^T G_1^2(t,z)\,dt + G_0^2(z)]^{\frac{1}{2}} \geq a_1 \ \forall z \in S_i \setminus Z, \tag{6.36}$$

$$b(z) = \sup_{t\in[t_0,t_1]} |G_2(t,z)| + \sup_{t\in[0,T]} |G_1(t,z)| + |G_0(z)| \geq a_2 \quad \forall z \in S_i \setminus Z. \tag{6.37}$$

Proof is analogous to that of Theorem 6.1.

Corollary 6.1 *If $z^* \in Z$ is a local minimizer of the functional f on the set Z (in the metric ρ_i), then there exist $a > 0$ and $\delta > 0$ such that in some neighbourhood $B_{i\delta}(z^*) = \{z \in U \mid \rho_i(z, z^*) < \delta\}$ of the point z^* the function φ satisfies the condition*

$$\varphi_i^{\downarrow}(z) = \inf_{\|v\|=1} \{\int_{t_0}^{T_1} G_2(t,z)v_2(t)\,dt + \int_0^T G_1(t,z)v_1(t)\,dt + G_0(z)v_0\} =$$
$$= \inf_{\|v\|=1} (G(z), v) = -\|G(z)\| \leq -a < 0 \quad \forall z \in B_{i\delta}(z^*) \setminus Z. \tag{6.38}$$

Theorem 6.3 *Let a function f be Lipschitz on the set $Z_\varepsilon = \{z \in U \mid \varphi(z) < \varepsilon\}$ (in the metric ρ_i). If $z^* \in Z$ is a local minimizer of f on the set Z (in the metric ρ_i), then a $\lambda^* < \infty$ exists such that for $\lambda > \lambda^*$ the point z^* is a local minimizer of the functional $\Phi_\lambda = f(z) + \lambda\varphi(z)$ on U (in the same metric ρ_i).*

Proof follows from Theorem 4.1 of [7] and Corollary 6.1.

7. NECESSARY CONDITIONS FOR AN EXTREMUM

Let $z^* \in Z$ be a local minimizer of the functional f on the set Z in the metric ρ_i. Assume that in some neighbourhood $B_{i\delta}(z^*)$ of the point z^* the function φ satifies the condition (6.38) and that f is Lipschitz on $B_{i\delta}(z^*)$ in the metric ρ_i. Then for $\lambda > \lambda^*$ the point z^* is a local minimizer of the function $\Phi_\lambda(z)$ in the metric ρ_i. Fix $v = [v_1, v_2, v_0] \in U$ and for $\varepsilon \geq 0$ put

$$z_\varepsilon(t) = z^* + \varepsilon v(t). \tag{7.1}$$

Note that

$$\rho_i(z_\varepsilon, z^*) \underset{\varepsilon\downarrow 0}{\to} 0 \quad \forall i \in 1:4. \tag{7.2}$$

It follows from (5.53) and (5.55) that

$$\Phi_\lambda(z_\varepsilon) = \Phi_\lambda(z^*) + \varepsilon\{(B(z^*), v) + \lambda \max_{A \in \partial\varphi(z^*)} (A, v)\} + o(\varepsilon, v), \qquad (7.3)$$

where

$$\frac{o(\varepsilon, v)}{\varepsilon} \underset{\varepsilon \downarrow 0}{\to} 0 \quad \forall v \in U,$$

$$(B(z^*), v) = \int_{t_0}^{t_1} B_2(t, z^*) v_2(t)\, dt + \int_0^T B_1(t, z^*) v_1(t)\, dt + B_0(z^*) v_0,$$

$$B_1(t, z^*) = \{\frac{\partial F(t, z^*)}{\partial y'} +$$

$$+ \int_t^T [\frac{\partial F(\tau, z^*)}{\partial y} + \frac{\partial F(\tau, z^*)}{\partial u(y)} u^*(y^*(\tau)) z_2^*(y^*(\tau))]\, d\tau\} e^{z_1^*(t)}, \qquad (7.4)$$

$$B_2(t, z^*) =$$

$$= \sigma_1(t) B_{21}(t, z^*) + \sigma_2(t) B_{21}(0, z^*) + \sigma_3(t) B_{22}(t, z^*) + \sigma_4(t) B_{22}(y_0, z^*) =$$

$$= \sigma_1(t) \int_t^T \frac{\partial F(\tau, z^*)}{\partial u} u^*(\tau)\, d\tau + \sigma_2(t) \int_0^T \frac{\partial F(\tau, z^*)}{\partial u} u^*(\tau)\, d\tau +$$

$$+ \sigma_3(t) \int_{g^*(t)}^T \frac{\partial F(\tau, z^*)}{\partial u(y)} u^*(y^*(\tau))\, d\tau + \sigma_4(t) \int_0^T \frac{\partial F(\tau, z^*)}{\partial u(y)} u^*(y^*(\tau))\, d\tau, \qquad (7.5)$$

$$B_0(z^*) = \int_0^T [\frac{\partial F(t, z^*)}{\partial u} u^*(t) + \frac{\partial F(t, z^*)}{\partial u(y)} u^*(y^*(t))]\, dt, \qquad (7.6)$$

$$(A, v) = \int_{t_0}^{t_1} A_2(t) v_2(t)\, dt + \int_0^T A_1(t) v_1(t)\, dt + A_0 v_0,$$

$$\partial\varphi(z^*) = \{A = [A_1(t), A_2(t), A_0] \in U \mid$$

$$A_1(t) = [w_0 + w_2(t) \frac{\partial l(t, z^*)}{\partial y'} +$$

$$+ \int_t^T w_2(\tau) [\frac{\partial l(\tau, z^*)}{\partial y} + \frac{\partial l(\tau, z^*)}{\partial u(y)} u^*(y^*(\tau)) z_2^*(y^*(\tau))]\, d\tau] e^{z_1^*(t)},$$

$$A_2(t) = \sigma_1(t)A_{21}^*(t) + \sigma_2(t)A_{21}^*(0) + \sigma_3(t)A_{22}^*(t) +$$

$$+\sigma_4(t)A_{22}^*(y_0) + w_1\int_t^{\vartheta_1} u^*(\tau)\,d\tau =$$

$$= \sigma_1(t)\int_t^T w_2(\tau)\frac{\partial l(\tau, z^*)}{\partial u}u^*(\tau)\,d\tau + \sigma_2(t)\int_0^T w_2(\tau)\frac{\partial l(\tau, z^*)}{\partial u}u^*(\tau)\,d\tau +$$

$$+\sigma_3(t)\int_{g^*(t)}^T w_2(\tau)\frac{\partial l(\tau, z^*)}{\partial u(y)}u^*(y^*(\tau))\,d\tau +$$

$$+\sigma_4(t)\int_0^T w_2(\tau)\frac{\partial l(\tau, z^*)}{\partial u(y)}u^*(y^*(\tau))\,d\tau + \; + w_1\int_t^{\vartheta_1} u^*(\tau)\,d\tau,$$

$$A_0 = w_1 + \int_0^T w_2(t)[\frac{\partial l(t, z^*)}{\partial u}u^*(t) + \frac{\partial l(t, z^*)}{\partial u(y)}u^*(y^*(t))]\,dt,$$

$$w = [w_2, w_1, w_0] \in W\}, \tag{7.7}$$

$$W = \{w = [w_2, w_1, w_0] \mid w_0 \in R,\ w_1 \in R,\ w_2 \in C[0,T],$$

$$\|w\| = [\int_0^T w_2^2(t)\,dt + w_1^2 + w_0^2]^{\frac{1}{2}} \le 1\}. \tag{7.8}$$

Then (7.3) implies

$$\Phi_\lambda(z_\varepsilon) = \Phi_\lambda(z^*) + \varepsilon\Phi_\lambda(z^*, v) + o(\varepsilon, v), \tag{7.9}$$

$$\Phi_\lambda(z^*, v) = \max_{C \in \partial\Phi_\lambda(z^*)}(C, v) =$$

$$= \max_{C \in \partial\Phi_\lambda(z^*)}[\int_{\vartheta_0}^{\vartheta_1} C_2(t)v_2(t)\,dt + \int_0^T C_1(t)v_1(t)\,dt + C_0v_0], \tag{7.10}$$

where

$$\partial\Phi_\lambda(z^*) = \{C = [C_1(t), C_2(t), C_0] \in U \mid$$

$$C_1(t) = [\frac{\partial F(t, z^*)}{\partial y'} + \int_t^T [\frac{\partial F(\tau, z^*)}{\partial y} +$$

$$\frac{\partial F(\tau, z^*)}{\partial u(y)}u^*(y^*(\tau))z_2^*(y^*(\tau))]\,d\tau + \lambda w_0 + \lambda w_2(t)\frac{\partial l(t, z^*)}{\partial y'} +$$

$$+\int_t^T \lambda w_2(\tau)[\frac{\partial l(\tau, z^*)}{\partial y} + \frac{\partial l(\tau, z^*)}{\partial u(y)}u^*(y^*(\tau))z_2^*(y^*(\tau))]\,d\tau]e^{z_1^*(t)},$$

$$C_2(t) = \sigma_1(t)[B_{21}^*(t) + \lambda A_{21}^*(t)] + \sigma_2(t)[B_{21}^*(0) + \lambda A_{21}^*(0)] +$$
$$+\sigma_3(t)[B_{22}^*(t) + \lambda A_{22}^*(t)] + \sigma_4(t)[B_{22}^*(y_0) + \lambda A_{22}^*(y_0)] +$$
$$+w_1 \int_t^1 u^*(\tau)\,d\tau =$$
$$= \sigma_1(t)[\int_t^T \frac{\partial F(\tau, z^*)}{\partial u} u^*(\tau)\,d\tau + \int_t^T \lambda w_2(\tau) \frac{\partial l(\tau, z^*)}{\partial u} u^*(\tau)\,d\tau] +$$
$$+\sigma_2(t)[\int_0^T \frac{\partial F(\tau, z^*)}{\partial u} u^*(\tau)\,d\tau + \int_0^T \lambda w_2(\tau) \frac{\partial l(\tau, z^*)}{\partial u} u^*(\tau)\,d\tau] +$$
$$+\sigma_3(t)[\int_{g^*(t)}^T \frac{\partial F(\tau, z^*)}{\partial u(y)} u^*(y^*(\tau))\,d\tau +$$
$$+\int_{g^*(t)}^T \lambda w_2(\tau) \frac{\partial l(\tau, z^*)}{\partial u(y)} u^*(y^*(\tau))\,d\tau] +$$
$$+\sigma_4(t)[\int_0^T \frac{\partial F(\tau, z^*)}{\partial u(y)} u^*(y^*(\tau))\,d\tau +$$
$$+\int_0^T \lambda w_2(\tau) \frac{\partial l(\tau, z^*)}{\partial u(y)} u^*(y^*(\tau))\,d\tau] + \lambda w_1 \int_t^1 u^*(\tau)\,d\tau,$$
$$C_0 = \int_0^T [\frac{\partial F(t, z^*)}{\partial u} u^*(t) + \frac{\partial F(t, z^*)}{\partial u(y)} u^*(y^*(t))]\,dt +$$
$$+\lambda w_1 + \int_0^T \lambda w_2(t)[\frac{\partial l(t, z^*)}{\partial u} u^*(t) + \frac{\partial l(t, z^*)}{\partial u(y)} u^*(y^*(t))]\,dt,$$
$$w = [w_2, w_1, w_0] \in W\}. \tag{7.11}$$

Since the point z^* is a local minimizer of the functional $\Phi_\lambda(z)$ in the metric ρ_i and since (see (7.2))

$$\rho_i(z_\varepsilon, z^*) \underset{\varepsilon \downarrow 0}{\to} 0 \quad \forall i \in 1:4,$$

then (see [11,20]) the following relation should hold:

$$\Phi_\lambda(z^*, v) \geq 0, \quad \forall v \in U. \tag{7.12}$$

It was shown in [3] that the condition (7.12) is equivalent to the relation

$$0 \in \partial \Phi_\lambda(z^*), \tag{7.13}$$

where $0 \in U$ is an element of the space U. Like in [17], it is possible to prove that the conditions (7.11) and (7.13) imply that there exists an element $w = [w_2, w_1, w_0] \in W$ such that

$$\frac{\partial F(t,z^*)}{\partial y'} + \int_t^T \{\frac{\partial F(\tau,z^*)}{\partial y} + \frac{\partial F(\tau,z^*)}{\partial u(y)} u^*(y^*(\tau)) z_2^*(y^*(\tau))\}\, d\tau +$$
$$+\lambda w_0 + \lambda w_2(t) \frac{\partial l(t,z^*)}{\partial y'} +$$
$$+ \int_t^T \lambda w_2(\tau) \{\frac{\partial l(\tau,z^*)}{\partial y} + \frac{\partial l(\tau,z^*)}{\partial u(y)} u^*(y^*(\tau)) z_2^*(y^*(\tau))\}\, d\tau = 0 \qquad (7.14)$$
$$\forall t \in [0,T],$$

$$\sigma_1(t)[\int_t^T \frac{\partial F(\tau,z^*)}{\partial u} u^*(\tau)\, d\tau + \int_t^T \lambda w_2(\tau) \frac{\partial l(\tau,z^*)}{\partial u} u^*(\tau)\, d\tau] +$$
$$+\sigma_2(t)[\int_0^T \frac{\partial F(\tau,z^*)}{\partial u} u^*(\tau)\, d\tau + \int_0^T \lambda w_2(\tau) \frac{\partial l(\tau,z^*)}{\partial u} u^*(\tau)\, d\tau] +$$
$$+\sigma_3(t)[\int_{g^*(t)}^T \frac{\partial F(\tau,z^*)}{\partial u(y)} u^*(y^*(\tau))\, d\tau +$$
$$\int_{g^*(t)}^T \lambda w_2(\tau) \frac{\partial l(\tau,z^*)}{\partial u(y)} u^*(y^*(\tau))\, d\tau] +$$

$$+\sigma_4(t)[\int_0^T \frac{\partial F(\tau,z^*)}{\partial u(y)} u^*(y^*(\tau))\, d\tau +$$

$$+ \int_0^T \lambda w_2(\tau) \frac{\partial l(\tau,z^*)}{\partial u(y)} u^*(y^*(\tau))\, d\tau] + \lambda w_1 \int_t^{t_1} u^*(\tau)\, d\tau = 0$$
$$\forall t \in [t_0, t_1], \qquad (7.15)$$

$$\int_0^T [\frac{\partial F(t,z^*)}{\partial u} u^*(t) + \frac{\partial F(t,z^*)}{\partial u(y)} u^*(y^*(t))]\, dt +$$

$$+\lambda w_1 + \int_0^T \lambda w_2(t)[\frac{\partial l(t,z^*)}{\partial u}u^*(t)+\frac{\partial l(t,z^*)}{\partial u(y)}u^*(y^*(t))]dt = 0. \quad (7.16)$$

Put

$$\Psi_2(t) = \lambda\varphi_2(t), \quad \Psi_1 = \lambda\varphi_1, \quad \Psi_0 = \lambda\varphi_0. \quad (7.17)$$

The relations (7.8), (7.14)-(7.16) and (7.17) yield the following

Theorem 7.1 *Let* $z^* = [z_1^*, z_2^*, u_0^*] \in Z, x^* = [y^*, u^*] \in \Omega$. *Assume that in some neighbourhood* $B_{i\delta}(z^*) = \{z \mid \rho_i(z,z^*) < \delta\}$ *of the point* z^* *the function* φ *satisfies the relation (6.38) and the function* f *is Lipschitz in* $B_{i\delta}(z^*)$ *in the metric* ρ_i. *For a point* $z^* \in Z$ *to be a global or local minimizer of the functional (3.2) on the set* Z *in the metric* ρ_i, *it is necessary that there exist some constants* $\Psi_0, \Psi_1 \in R$ *and a function* $\Psi_2(t) \in C[0,T]$, *satisfying the conditions*

$$\frac{\partial F(t,z^*)}{\partial y'}+\int_t^T\{\frac{\partial F(\tau,z^*)}{\partial y}+\frac{\partial F(\tau,z^*)}{\partial u(y)}u^*(y^*(\tau))z_2^*(y^*(\tau))\}d\tau +$$
$$+\Psi_0+\Psi_2(t)\frac{\partial l(t,z^*)}{\partial y'}+$$
$$+\int_t^T \Psi_2(\tau)\{\frac{\partial l(\tau,z^*)}{\partial y}+\frac{\partial l(\tau,z^*)}{\partial u(y)}u^*(y^*(\tau))z_2^*(y^*(\tau))\}d\tau = 0$$
$$\forall t \in [0,T], \quad (7.18)$$

$$\sigma_1(t)[\int_t^T \frac{\partial F(\tau,z^*)}{\partial u}u(\tau)d\tau+\int_t^T \Psi_2(\tau)\frac{\partial l(\tau,z^*)}{\partial u}u^*(\tau)d\tau]+$$
$$+\sigma_2(t)[\int_0^T \frac{\partial F(\tau,z^*)}{\partial u}u(\tau)d\tau+\int_0^T \Psi_2(\tau)\frac{\partial l(\tau,z^*)}{\partial u}u^*(\tau)d\tau]+$$
$$+\sigma_3(t)[\int_{g^*(t)}^T \frac{\partial F(\tau,z^*)}{\partial u(y)}u^*(y^*(\tau))d\tau$$
$$+\int_{g^*(t)}^T \Psi_2(\tau)\frac{\partial l(\tau,z^*)}{\partial u(y)}u^*(y^*(\tau))d\tau]+$$

$$+\sigma_4(t)[\int_0^T \frac{\partial F(\tau,z^*)}{\partial u(y)} u^*(y^*(\tau))\,d\tau + \int_0^T \Psi_2(\tau)\frac{\partial l(\tau,z^*)}{\partial u(y)} u^*(y^*(\tau))\,d\tau] +$$

$$+\Psi_1 \int_t^{t_1} u^*(\tau)\,d\tau = 0 \quad \forall t \in [t_0,t_1], \tag{7.19}$$

$$\int_0^T [\frac{\partial F(t,z^*)}{\partial u} u^*(t) + \frac{\partial F(t,z^*)}{\partial u(y)} u^*(y^*(t))]\,dt +$$

$$+\Psi_1 + \int_0^T \Psi_2(t)[\frac{\partial l(t,z^*)}{\partial u} u^*(t) + \frac{\partial l(t,z^*)}{\partial u(y)} u^*(y^*(t))]\,dt = 0. \tag{7.20}$$

ACKNOWLEDGMENT

The Authors express their gratitude to Dr. A. Boyarsky of Concordia University of Canada for having posed the problem.

REFERENCES

[1] Bliss G.A. (1946), *Lectures on the Calculus of Variations*. Chicago, Univ. of Chicago Press.

[2] Boukary D., Fiacco A.V. (1995), Survey of penalty, exact-penalty and multiplier methods from 1968 to 1993. *Optimization*, Vol. 32, No. 4, pp. 301–334.

[3] Demyanov V.F., Vasiliev L.V. (1985) *Nondifferentiable optimization*. New-York, Springer-Optimization Software.

[4] Demyanov V.F. (1992), Nonsmooth problems in Calculus of Variations. In *Lecture Notes in Economics and Math. Systems. v. 382. Advances in Optimization*. Eds.W.Oettli, D.Pallaschke, pp. 227–238. Berlin, Springer Verlag.

[5] Demyanov V.F. (1992a), Calculus of Variations in nonsmooth presentation. In *Nonsmooth Optimization: Methods and Applications*. Ed. F.Giannessi, pp. 76–91. Singapore, Gordon and Breach.

[6] Demyanov V.F. (1994), Exact penalty functions in nonsmooth optimization problems. *Vestnik of St. Petersburg University*, ser. 1, issue 4 (No. 22), pp. 21-27.

[7] Demyanov V.F. (2000), *Conditions for an extremum and variational problems*. St. Petersburg, St. Petersburg University Press.

[8] Demyanov V.F. (2003) Constrained problems of Calculus of Variations via Penalization Technique. In: *Equilibrium Problems and Variational Models*. A.Maugeri, F.Giannessi (Eds.), pp. 79-108. Kluwer Academic Publishers.

[9] Demyanov V.F., Di Pillo G., Facchinei F. (1998), Exact penalization via Dini and Hadamard conditional derivatives. *Optimization Methods and Software*, Vol. 9, pp. 19–36.

[10] Demyanov V.F., Giannessi F., and Karelin V.V. (1998), Optimal Control Problems via Exact Penalty Functions. *Journal of Global Optimization*, Vol. 12, No. 3, pp.215–223.

[11] Demyanov V.F., Rubinov A.M. (1995), *Constructive Nonsmooth Analysis*. Frankfurt a/M., Peter Lang Verlag.

[12] Di Pillo G., Facchinei F. (1989), Exact penalty functions for nondifferentiable programming problems. In *Nonsmooth Optimization and Related Topics*. Eds. F.H. Clarke, V.F. Demyanov and F. Giannessi, pp. 89-107. New York, Plenum.

[13] Eremin I.I. (1966), On the penalty method in convex programming. In *Abstracts of ICM–66*.Section 14., Moscow, 1966.

[14] Eremin I.I. (1967). A method of "penalties" in Convex Programming. *Soviet Mathematics Doklady* (4), 748–751.

[15] Fletcher R. (1983), Penalty functions. In *Mathematical programming: the state of the art*. (Eds. A. Bachen, M. Grötschel, B. Korte), Springer–Verlag, Berlin, pp. 87–114.

[16] Han S., Mangasarian O. (1979), Exact penalty functions in nonlinear programming. *Mathematical Programming*, Vol.17, pp. 251–269.

[17] Hestenes M.R. (1966), *Calculus of Variations and Optimal Control Theory*. New York, John Wiley & Sons.

[18] Kaplan A.A., Tichatschke R. (1994), *Stable methods for ill-posed variational problems*. Berlin, Akademie Verlag.

[19] Pontryagin L.S., Boltyanskii V.G., Gamkrelidze R.V., and Mishchenko E.F. (1962), *The Mathematical theory of Optimal Processes*. New York, Interscience Publishers.

[20] Rockafellar R.T. (1970) *Convex Analysis*. Princeton. N.J., Princeton University Press.

[21] Zangwill W.L. (1967), Nonlinear programming via penalty functions. *Management Science*, Vol. 13, pp. 344–358.

CONTINUOUS SETS AND NON-ATTAINING FUCTIONALS IN REFLEXIVE BANACH SPACES

Emil Ernst[1] and Michel Théra[2*]
Laboratory of Modelisation in Mechanics and Thermodynamics, Faculty of Science and Techniques of Saint Jérôme, Saint Jérôme, France ;[1] Laco, University of Limoges, Limoges Cedex, France[2]

Abstract: In this paper we prove, in the framework of reflexive Banach spaces, that a linear and continuous functional f achieves its supremum on every small ε - uniform perturbation of a closed convex set C containing no lines, if and only if f belongs to the norm-interior of the barrier cone of C. This result is applied to prove that every closed convex subset C of a reflexive Banach space X which contains no lines is continuous if and only if every small ε -uniform perturbation of C does not allow non-attaining linear and continuous functionals. Finally, we define a new class of non-coercive variational inequalities and state a corresponding open problem.

Key words and phrases: Continuous closed convex set, non-attaining functional, well-positioned set, non-coercive variational inequalities.

1. INTRODUCTION AND NOTATIONS

Throughout the paper, we suppose that X is a reflexive Banach space with continuous dual X^*. The norms in X and X^* will be denoted by $\|\cdot\|$ and $\|\cdot\|_*$, and the primal and dual closed unit balls of X and X^* by $\mathbb{B}_X$

* The research of Michel Théra has been supported by NATO Collaborative Linkage Grant 978488.

and $\mathbb{B}_{X^*}$, respectively. Given a closed convex subset C of X and $\varepsilon > 0$, we call *ε-uniform perturbation* of C every closed convex set C_ε which satisfies:

$$C_\varepsilon \subseteq C + \varepsilon \mathbb{B}_X \text{ and } C \subseteq C_\varepsilon + \varepsilon \mathbb{B}_X. \tag{1}$$

The main purpose of this note is to determine, when a closed convex subset C of X is given, the class of all the linear continuous functionals on X which attain their supremum on every ε-uniform perturbation C_ε of C for a small ε.

The main result of this note (Theorem 1, Section 2) claims that a linear continuous functional f achieves its supremum on every ε-uniform perturbation C_ε of C (ε sufficiently small) if and only if f belongs to the norm-interior of the barrier cone of C, that is the cone of all the linear continuous functionals bounded from above on C. Using a recent characterization of the interior of the barrier cone ([1]) we deduce (Corollary 1) that f reaches its supremum on every ε-uniform perturbation C_ε of C (ε sufficiently small) if and only if C is well-positioned and f belongs to the norm-interior of the negative polar cone of the recession cone of C.

Theorem 1 gives thus a necessary and sufficient condition for a linear continuous functional on X to achieve its supremum not only on a given closed convex set C but also on any ε-uniform perturbation C_ε of C (ε sufficiently small).

Throughout the paper, as customary, given a closed convex set C, by a non-attaining functional we mean a linear continuous functional bounded from above on C which does not reach its supremum on C. In Theorem 1 we characterize the class of all closed convex subsets C of a reflexive Banach space X such that any sufficiently small ε-uniform perturbation C_ε of C disallows non-attaining functionals. Under theses assumptions, we show (Proposition 1, Section 3) that all the sufficiently small ε-uniform perturbations of a closed convex set C disallow non-attaining functionals if and only if C is a continuous set, in the sense of Gale and Klee ([6]) for sets in finite dimensional spaces. The reader is referred to [4]) for a definition and several properties of infinite-dimensional continuous sets.

Finally, by virtue of a remark of Del Piero [5], we use Proposition 1 to define a class of non-coercive variational inequalities for which a natural necessary condition for the existence of solutions is also sufficient, and this for every small ε-uniform perturbation of the data involved in the problem. This result should be applicable to variational problems as they often arise in finance or in engineering problems in which data are known only with a certain precision and it is desired that further refinement of the data should not cause substantial changes in the existence of a solution.

We end up this note with an open question: can we characterize the class of all semi-coercive variational inequalities for which the above mentioned necessary condition is also sufficient?

For the convenience of the reader, we now introduce some additional notations. As usual, $j: X^* \to X$ is the *duality mapping* given by $\langle f, j(f)\rangle = \|f\|_*^2$ and $\|j(f)\| = \|f\|_*$, (see for example [8]),

$$S^\circ = \{f \in X^* : \langle f, w\rangle \leq 0 \ \forall w \in S\}$$

is the negative polar cone of the set S of X, and S° reduces to the orthogonal

$$S^\perp = \{f \in X^* : \langle f, w\rangle = 0 \ \forall w \in S\}$$

when S is a linear subspace of X. The linear subspace of X parallel to the largest linear manifold contained in C will be denoted by $l(C)$:

$$l(C) = C^\infty \cap (-C^\infty).$$

We will use the notations Int$S$ and BdS to denote respectively the norm-interior and the norm-boundary of a set S in X or in X^*. We recall that the *recession cone* (see [7]) to the closed convex set S is the closed convex cone S^∞ defined by

$$S^\infty = \{v \in X : \ \forall \lambda > 0, \forall x_0 \in S, x_0 + \lambda v \in S\},$$

and that a set S is called *linearly bounded* whenever $S^\infty = \{0\}$.

If $\Phi: X \to \mathbb{R} \cup \{+\infty\}$ is an extended-real-valued function, Dom Φ is the set of all $x \in X$ for which $\Phi(x)$ is finite, and we say that Φ is *proper* if Dom $\Phi \neq \varnothing$. When Φ is a proper lower semi-continuous convex function, the *recession* function Φ^∞ of Φ is the proper lower semi-continuous convex function whose epigraph is the recession cone to the epigraph of Φ, i.e., $\text{epi}\Phi^\infty = (\text{epi}\Phi)^\infty$. Equivalently

$$\Phi^\infty(x) = \lim_{t \to +\infty} \frac{\Phi(x_0 + tx)}{t},$$

where x_0 is any element such that $\Phi(x_0)$ is finite. Given a closed convex subset S of X, the domain of the *support function* given by

$$\sigma_S(f) := \sup_{x \in S} \langle f, x \rangle$$

is the *barrier cone* of S:

$$\mathcal{B}(S) = \{f \in X^* : \sigma_S(f) < +\infty\} = \mathrm{Dom}\sigma_S.$$

Finally, we use the symbols "$\to$" and "$\rightharpoonup$" to denote the strong convergence and the weak convergence on X, respectively.

2. THE MAIN RESULT

Let us consider a closed and convex set C which contains no lines and a continuous linear functional f on X. The following two lemmata collect conditions on C and on f allowing us to construct, for every $\varepsilon > 0$, an ε-uniform perturbation C_ε of C on which f does not achieve its supremum.

Lemma 1. *Let C be a non-void closed convex subset of X which contains no lines, and $f \in \mathcal{B}(C)$, $\|f\|_* = 1$ for which there is $w \in C^\infty$, $\|w\| = 1$, such that $\langle f, w \rangle = 0$. Then, for every $\varepsilon > 0$, there is an ε-uniform perturbation C_ε of C such that f does not reach its supremum on C_ε.*

Proof of Lemma 1: If f does not reach its supremum on C, then take $C_\varepsilon = C$ and Lemma 1 follows. Suppose now that f attains its supremum on C at $\overline{x}$, i.e.,

$$\langle f, \overline{x} \rangle \geq \langle f, x \rangle \; \forall x \in C.$$

Set

$$D = \left\{ \overline{x} + \nu w + \mu j(f) : 0 \leq \nu, 0 \leq \mu \leq \frac{\nu}{1+\nu} \varepsilon \right\},$$

and take C_ε as the closed convex hull of C and D (denoted by $\overline{\mathrm{co}}(C, D)$) $C_\varepsilon = \overline{\mathrm{co}}(C, D)$. Remark that $D \subset A + \varepsilon \mathbb{B}_X$, where $A = \overline{x} + \mathbb{R}_+ w$ is a half-line in C. Accordingly, $C_\varepsilon \subset C + \varepsilon \mathbb{B}_X$. As obviously $C \subset C_\varepsilon \subset C_\varepsilon + \varepsilon \mathbb{B}_X$, it follows that the closed convex set C_ε is an ε-uniform perturbation of C.

On D, the supremum of f is $\langle f,\overline{x}\rangle+\varepsilon$, while on C the supremum of f is $\langle f,\overline{x}\rangle$. Hence, the supremum of f on C_ε is $\langle f,\overline{x}\rangle+\varepsilon$. Let us show that f does not reach this value on C_ε.

Suppose by contradiction that there is $\tilde{x}\in C_\varepsilon$ such that

$$\langle f,\tilde{x}\rangle=\langle f,\overline{x}\rangle+\varepsilon.$$

As $\tilde{x}\in\overline{\text{co}}(C,D)$, select a sequence $(a_n)_{n\in\mathbb{N}^*}\subset\text{co}(C,D)$ norm-converging to $\tilde{x}$. As obviously D is closed and convex, for every $a_n\in\text{co}(C,D)$ we may pick $d_n\in D$ (that is $d_n=\overline{x}+\nu_n w+\mu_n j(f)$ for some $0\le\nu_n$ and $0\le\mu_n\le\frac{\nu_n}{1-\nu_n}\varepsilon$), $c_n\in C$ and $\lambda_n\in[0,1]$ such that

$$a_n=\lambda_n c_n+(1-\lambda_n)d_n.$$

As $\langle f,w\rangle=0$ we deduce that

$$\langle f,a_n\rangle-\langle f,\overline{x}\rangle=\lambda_n\left(\langle f,c_n\rangle-\langle f,\overline{x}\rangle\right)+(1-\lambda_n)\mu_n.$$

Since $\lim_{n\to+\infty}a_n=\tilde{x}$, then

$$\lim_{n\to+\infty}\left(\mu_n+\lambda_n(\langle f,c_n\rangle-\langle f,\overline{x}\rangle)-\mu_n\right)=\varepsilon.$$

As $\mu_n\le\varepsilon$ and $\lambda_n(\langle f,c_n\rangle-\langle f,\overline{x}\rangle-\mu_n)\le 0$, the previous relation implies that

$$\lim_{n\to+\infty}\left(\lambda_n(\langle f,c_n\rangle-\langle f,\overline{x}\rangle-\mu_n)\right)=0 \text{ and } \lim_{n\to+\infty}\mu_n=\varepsilon. \tag{2}$$

Using the fact that $\langle f,c_n\rangle-\langle f,\overline{x}\rangle-\mu_n\le-\mu_n$, we derive

$$\limsup_{n\to\infty}\left(\langle f,c_n\rangle-\langle f,\overline{x}\rangle-\mu_n\right)\le-\varepsilon. \tag{3}$$

Combining relations (3) and (2) it follows that $(\lambda_n)\to 0$, while since $(\mu_n)\to\varepsilon$ we deduce that $(\nu_n)\to\infty$. Let us observe that

$$\lambda_n c_n=-(1-\lambda_n)\nu_n w+a_n-(1-\lambda_n)(\overline{x}+\mu_n j(f)) \tag{4}$$

and that $(1-\lambda_n)\nu_n > 0$, for n large enough ($1-\lambda_n \to 1$ and $\nu_n \to \infty$). Hence, dividing by $(1-\lambda_n)\nu_n$ we obtain

$$\frac{\lambda_n}{(1-\lambda_n)\nu_n} c_n = -w + \frac{a_n - (1-\lambda_n)(\overline{x} + \mu_n j(f))}{(1-\lambda_n)\nu_n}. \tag{5}$$

Being convergent, the sequence $(a_n)_{n\in\mathbb{N}^*}$ is bounded, so the previous relation implies that

$$\frac{\lambda_n}{(1-\lambda_n)\nu_n} c_n \to -w.$$

As $t_n := \frac{\lambda_n}{(1-\lambda_n)\nu_n} \to 0$, $c_n \in C$ and $t_n c_n \to -w$ as $n \to +\infty$, it follows that $-w \in C^\infty$, that is $0 \neq w \in l(C)$, contradicting the fact that the closed convex set C contains no lines. As a result, f does not achieve its supremum on the ε-uniform perturbation C_ε of C. □

In order to state the second condition ensuring the existence of at least one ε-uniform perturbation C_ε of C on which f does not attain its supremum, let us recall the concept of *well-positioned convex sets*, introduced by Adly *et al.* in a recent paper ([1]).

Definition 1. *A nonempty subset C of the normed vector space X is well-positioned if there exist $x_0 \in X$ and $g \in X^*$ such that:*

$$\langle g, x - x_0 \rangle \geq \|x - x_0\|, \quad \forall x \in C.$$

It follows directly from the definition that when C is well-positioned, the sets $x + \lambda C$ and B are well-positioned for every $x \in X$, $\lambda \in \mathbb{R}$ and $\varnothing \neq B \subset C$.

Lemma 2. *Let C be a nonempty closed convex subset of X containing no lines and which is not well-positioned, and $f \in \mathcal{B}(C)$, $\|f\|_* = 1$. Then, for every $\varepsilon > 0$, there is an ε-uniform perturbation C_ε of C on which f does not attain its supremum.*

Proof of Lemma 2: Notice first that when f does not reach its supremum on C it is sufficient to set $C_\varepsilon = C$, and when there is $w \in C^\infty$, $\|w\| = 1$, such

that $\langle f, w\rangle = 0$, we may apply Lemma 1. Consider now the remaining case, that is when

$$\langle f, w\rangle < 0 \quad \forall w \in C^{\infty}, \ \|w\| = 1, \tag{6}$$

and f attains its supremum on C at $\overline{x} \in C$. In order to achieve the proof of Lemma 2 we shall construct an ε-uniform perturbation C_{ε} of C on which f does not attain its supremum.

Take $\overline{y} = \overline{x} + \varepsilon j(f)$ and consider

$$B = \left\{x \in \overline{\mathrm{co}}(\overline{y}, C) : \langle f, x\rangle = \langle f, \overline{x}\rangle\right\}. \tag{7}$$

Obviously, B is a closed convex set. As $\overline{\mathrm{co}}(\overline{y}, C)^{\infty} = C^{\infty}$, it follows that $B^{\infty} \subset C^{\infty}$; taking into account relation (6) we deduce that

$$\langle f, w\rangle < 0 \quad \forall w \in B^{\infty}, \ \|w\| = 1. \tag{8}$$

On the other hand, relation (7) implies that

$$\langle f, w\rangle = 0 \quad \forall w \in B^{\infty}. \tag{9}$$

Combining relations (8) and (9), it follows that $B^{\infty} = \{0\}$. Accordingly, B is a linearly bounded closed convex set.

Let us now prove that B is unbounded. Indeed, by contradiction, let us suppose that $B \subseteq \rho \mathbb{B}_X$ for some $\rho > 0$. Let $x \in C$; observe that the convex combination

$$z = \frac{\langle f, \overline{x}\rangle - \langle f, x\rangle}{\varepsilon + \langle f, \overline{x}\rangle - \langle f, x\rangle}\overline{y} + \frac{\varepsilon}{\varepsilon + \langle f, \overline{x}\rangle - \langle f, x\rangle}x \tag{10}$$

of $\overline{y}$ and x belongs to C_{ε}, and, as $\langle f, z\rangle = \langle f, \overline{x}\rangle$, we deduce that $z \in B$. Accordingly, $\|z\| \le \rho$; in addition, as $\overline{x} \in B$, we have $\|\overline{x}\| \le \rho$, and therefore $\|z - \overline{x}\| \le 2\rho$.

Standard calculations yield

$$z - \overline{x} = \frac{\varepsilon\left((x - \overline{y}) + \langle f, \overline{y} - x\rangle j(f)\right)}{\langle f, \overline{y} - x\rangle},$$

from which we obtain

$$\frac{\left\|(x-\overline{y})+\langle f,\overline{y}-x\rangle j(f)\right\|}{\langle f,\overline{y}-x\rangle}\leq\frac{2\rho}{\varepsilon}.$$

Hence

$$\|x-\overline{y}\|\leq\left\langle-\left(1+\frac{2\rho}{\varepsilon}\right)f,x-\overline{y}\right\rangle\quad\forall x\in C. \tag{11}$$

As relation (11) contradicts the fact that the set C is not well-positioned we obtain the unboundedness of B. Accordingly, the set $B-\overline{x}$ is an unbounded linearly bounded closed convex set, and thus (see [2]), there is a linear continuous functional g such that

$$\inf_{y\in B}\langle g,y-\overline{x}\rangle<\langle g,x-\overline{x}\rangle\leq 1\quad\forall x\in B. \tag{12}$$

Now, take

$$C_\varepsilon=\Big\{x\in\overline{\text{co}}(\overline{y},C):\tag{13}$$
$$\langle g,x+\langle f,\overline{x}-x\rangle j(f)-\overline{x}\rangle+\frac{3}{\varepsilon}\langle f,x\rangle\leq 2+\frac{3}{\varepsilon}\langle f,\overline{x}\rangle\Big\}.$$

Obviously, C_ε is a closed convex set is included in $\overline{\text{co}}(\overline{y},C)$, whence is included in $C_\varepsilon\subseteq C+\varepsilon\mathbb{B}_X$. Moreover, as (see relation (10))

$$x+\langle f,\overline{x}-x\rangle j(f)-\overline{x}=\left(1+\frac{\langle f,\overline{x}-x\rangle}{\varepsilon}\right)(z-\overline{x}),$$

we deduce from relation (12) that

$$\langle g,x+\langle f,\overline{x}-x\rangle j(f)-\overline{x}\rangle\leq 1+\frac{\langle f,\overline{x}-x\rangle}{\varepsilon}. \tag{14}$$

Taking into account that $\langle f,x\rangle\leq\langle f,\overline{x}\rangle$ for every $x\in C$ we deduce from (14) that

$$\langle g,x+\langle f,\overline{x}-x\rangle j(f)-\overline{x}\rangle+\frac{3}{\varepsilon}\langle f,x\rangle\leq 1+\frac{3}{\varepsilon}\langle f,\overline{x}\rangle\quad\forall x\in C,$$

and therefore by virtue of Definition (13) of C_ε we have $C \subset C_\varepsilon$. Accordingly, C_ε is an ε-uniform perturbation of C in the sense of (1).

Fix $z \in B$; for $\lambda \in [0,1]$ put $z(\lambda) = \lambda z + (1-\lambda)\overline{y}$ and define

$$h(\lambda) = \langle g, z(\lambda) + \langle f, \overline{x} - z(\lambda)\rangle j(f) - \overline{x}\rangle + \frac{3}{\varepsilon}\langle f, z(\lambda)\rangle.$$

We have

$$h(0) = 3 + \frac{3}{\varepsilon}\langle f, \overline{x}\rangle > 2 + \frac{3}{\varepsilon}\langle f, \overline{x}\rangle,$$

while

$$h(1) = \langle g, z - \overline{x}\rangle + \frac{3}{\varepsilon}\langle f, \overline{x}\rangle \le 1 + \frac{3}{\varepsilon}\langle f, \overline{x}\rangle < 2 + \frac{3}{\varepsilon}\langle f, \overline{x}\rangle.$$

As the map $h : [0,1] \to \mathbb{R}$ is continuous, there is $\overline{\lambda} \in (0,1)$ such that

$$h(\overline{\lambda}) = 2 + \frac{3}{\varepsilon}\langle f, \overline{x}\rangle. \tag{15}$$

Obviously $z(\overline{\lambda}) \in \overline{\mathrm{co}}(\overline{y}, C)$ and thus $z(\overline{\lambda}) \in C_\varepsilon$. Relation (15) yields

$$\langle f, z(\overline{\lambda})\rangle = \langle f, \overline{x}\rangle + \varepsilon\left(1 - \frac{3}{3 - \langle g, z - \overline{x}\rangle}\right),$$

and thus

$$\sup_{x \in C_\varepsilon}\langle f, x\rangle \ge \langle f, \overline{x}\rangle + \varepsilon\left(1 - \frac{3}{3 - \inf\limits_{z \in B}\langle g, z - \overline{x}\rangle}\right). \tag{16}$$

On the other hand, for every $x \in C_\varepsilon$

$$\langle f, x\rangle \le \langle f, \overline{x}\rangle + \varepsilon\left(1 - \frac{3}{3 - \langle g, z - \overline{x}\rangle}\right). \tag{17}$$

Hence, relations (12), (15) and (17) infer that for every $x \in C_\varepsilon$ we have

$$\langle f,x\rangle < \langle f,\bar{x}\rangle + \varepsilon\left(1 - \frac{3}{3 - \inf_{z\in B}\langle g, z-\bar{x}\rangle}\right) = \sup_{y\in C_\varepsilon}\langle f,y\rangle.$$

Accordingly, the linear continuous functional f does not reach its supremum on the ε-uniform perturbation C_ε of C, the proof of Lemma 2 is thus complete. □

The main result of this note characterizes all the linear continuous functionals which achieve their supremum on every sufficiently small ε-uniform perturbation of a given closed and convex set.

Theorem 1. *Let C be a non-void closed convex subset of X and f a non-null linear continuous functional. Then, there is $\varepsilon > 0$ such that f reaches its supremum on every closed convex subset C_ε of X fulfilling relation (1) if and only if f belongs to the norm-interior of the barrier cone of C.*

Proof of Theorem 1: Consider $f \in \operatorname{Int} \mathcal{B}(C)$. From Corollary 2.1 of [1] select R_f and $\gamma_f \in \mathbb{R}$ such that

$$\langle f,x\rangle \le R_f - \gamma_f \|x\| \quad \forall x \in C. \tag{18}$$

Using relations (1) and (18) we deduce that

$$\langle f,x\rangle \le \varepsilon\|f\|_* + R_f - \gamma_f\|x\| \quad \forall x \in C_\varepsilon. \tag{19}$$

Remark that $f \in \mathcal{B}(C_\varepsilon)$ and consider a maximizing sequence $(x_n)_{n\in\mathbb{N}^*} \subset C_\varepsilon$ of f, i.e., a sequence satisfying

$$\langle f,x_n\rangle \to \sup_{y\in C_\varepsilon}\langle f,y\rangle. \tag{20}$$

Accordingly, for n large enough we have

$$\langle f,x_n\rangle \ge \sup_{y\in C_\varepsilon}\langle f,y\rangle - 1,$$

and from (19) it follows that

$$\|x_n\| \le \frac{1}{\gamma_n}\left(1 + \varepsilon\|f\|_* + R_f - \sup_{y\in C_\varepsilon}\langle f,y\rangle\right).$$

The sequence $(x_n)_{n\in\mathbb{N}^*} \subset C_\varepsilon$ is therefore bounded, and, as X is reflexive and C is a closed and convex (thus weakly closed) set, the sequence $(x_n)_{n\in\mathbb{N}^*}$ has a weak cluster point $w \in C_\varepsilon$. From relation (20) we derive that

$$\langle f, w\rangle = \lim_{n\to\infty}\langle f, x_n\rangle = \sup_{y\in C_\varepsilon}\langle f, y\rangle,$$

which means that f attains its supremum on C_ε.

In order to prove that a continuous linear functional f which achieves its supremum on every ε-uniform perturbation C_ε of a closed convex set C belongs to the norm-interior of the barrier cone of C, let us first remark that every such functional must be bounded from above on C.

If we suppose that C is a not well-positioned, from Lemma 2 we deduce that, for every $\varepsilon > 0$, there is an ε-uniform perturbation C_ε of C (in the sense of (1)) on which f does not attain its supremum, a contradiction. If we suppose that there is $w \in C^\infty$ such that $\langle f, w\rangle = 0$, Lemma 1 proves that there is an ε-uniform perturbation C_ε of C on which f does not attain its supremum, once again a contradiction.

Accordingly, if the continuous linear functional f achieves its supremum on every ε-uniform perturbation C_ε of a closed convex set C, then C is necessarily a well-positioned closed and convex set, and the following relation holds

$$\langle f, w\rangle < 0 \quad \forall w \in C^\infty, w \neq 0. \tag{21}$$

In order to achieve the proof, let us prove that f belongs to the norm-interior of the barrier cone of C. By contradiction we suppose that $f \in \mathrm{Bd}\, C$. Since C is well-positioned, the norm-interior of the convex set $\mathcal{B}(C)$ is nonempty. Hence, there exists some $w \in X^{**}$ of norm 1 such that

$$\langle f, w\rangle \geq \langle h, w\rangle \quad \forall h \in \mathcal{B}(C).$$

Because X is reflexive we (may) consider that $w \in X$. The set $\mathcal{B}(C)$ is a cone, thus

$$\langle f, w\rangle \geq 0 \geq \langle h, w\rangle \quad \forall h \in \mathcal{B}(C). \tag{22}$$

Accordingly,

$$w \in [\mathcal{B}(C \cap L)]^\circ = (C \cap L)^\infty = C^\infty \cap V(L),$$

and from relation (22) it follows that $\langle f, w\rangle = 0$. Lemma 1 implies that there is at least one ε-uniform perturbation of C on which f does not achieve its supremum, contradiction which completely achieves the proof of Theorem 1. □

By virtue of Proposition 2.1 in [1], we deduce the following consequence of Theorem 1.

Corollary 1. *Let C be a nonempty closed convex subset of X and f a non-null continuous linear functional. The following two statements are equivalent:*

(a) f achieves its supremum on every ε-uniform perturbation of C;

(b) C is well-positioned and f belongs to the norm-interior of the negative polar cone of the recession cone of C,
$\langle f, w\rangle < 0 \quad \forall w \in C^{\infty}, w \neq 0.$

3. CONTINUOUS CLOSED CONVEX SETS AND NECESSARY CONDITIONS FOR NON-COERCIVE VARIATIONAL INEQUALITIES

The last section of this note is concerned with a new characterization of continuous closed and convex sets, as defined by Gale and Klee [6] (see also [4])):

Definition 2. *The closed convex set C of X is called continuous if its support functional $\sigma_C : X^* \to \mathbb{R}$ is continuous on $X^* \setminus \{0\}$.*

Observe that in a Banach space X, every lower semicontinuous convex function $h : X \to \mathbb{R} \cup \{+\infty\}$, is norm-continuous at $x \in X$ if and only if x belongs to the set $(X \setminus \mathrm{Dom}\, h) \cup \mathrm{Int}(\mathrm{Dom}\, h)$. Applying this remark to the support function $\sigma_C : X^* \to \mathbb{R} \cup \{+\infty\}$ of a closed convex subset C of X we deduce that C is continuous if and only if

$$\mathcal{B}(C) = \{0\} \cup \mathrm{Int}\mathcal{B}(C).$$

The previous remark and Theorem 1 lead to the following result.

Proposition 1. *Let C be a nonempty closed convex subset of X. Then every linear continuous functional bounded from above on C achieves its supremum on every ε-uniform perturbation C_ε of C if and only if C is continuous.*

We recall that an operator is called *semi-coercive* if there exist some positive constant $\kappa > 0$ and some closed subspace U of X such that if $\mathrm{dist}_U(x)$ denotes the distance from x to U, we have

$$\langle Av - Au, v - u \rangle \geq \kappa \left(\mathrm{dist}_U(v-u)\right)^2 \quad \forall u, v \in X$$
$$A(x+u) = A(x) \quad \forall x \in X \text{ and } u \in U, \text{ and } A(X) \subseteq U^{\perp}.$$

The class of semi-coercive operators contains for instance the projection operator onto a closed subspace of a Hilbert space.

Let K be a closed convex subset of X, f be an element in X^*, A be a semi-coercive operator from X to X^*, $\Phi : X \to \mathbb{R} \cup \{+\infty\}$ be a lower semi-continuous convex function that we assume to be bounded from below, and suppose that $K \cap \mathrm{Dom}\Phi \neq \varnothing$. We call semi-coercive variational inequality the problem of

$$\begin{aligned} &\text{finding } u \in K \cap \mathrm{Dom}\Phi \text{ such that} \\ &\langle Au - f, v - u \rangle + \Phi(v) - \Phi(u) \geq 0, \ \forall v \in K. \end{aligned} \tag{23}$$

Proposition 1 has a direct application in the theory of semi-coercive variational inequalities. Indeed, when the variational inequality is governed by an operator which is bounded, semi-coercive and pseudo-monotone (in the sense of Brézis [3], page 132), it has been noticed that if a solution to (23) exists, then the energy functional

$$\mathcal{F}(x) = \kappa \left(\mathrm{dist}_U(x)\right)^2 + I_K(x) + \Phi(x) - \langle f, x \rangle \quad \forall x \in X,$$

where I_K denotes the indicator functional of K is bounded from below on X and that if the energy functional is coercive on X then (23) has a solution (see for instance the proof of Proposition 3.1 in [1]).

Remark that $\mathcal{F}$ is bounded from below if and only if the linear continuous functional

$$(f, -1) : (X \times \mathbb{R})^* \to \mathbb{R}, \langle (f, -1), (x, a) \rangle = \langle f, x \rangle - a, \quad \forall (x, a) \in X \times \mathbb{R}$$

is bounded from above in $X \times \mathbb{R}$ on the epigraph of Ψ defined by

$$\Psi(x) = \kappa \left(\mathrm{dist}_U(x)\right)^2 + I_K(x) + \Phi(x)$$

and that $\mathcal{F}$ is coercive if and only if $(f,-1)$ belongs to the norm-interior of the barrier cone of the same epigraph.

Thus, Proposition 1 and Proposition 3.1 from [1] imply that, if the epigraph of Ψ is a continuous subset of $X \times \mathbb{R}$ then the boundedness from below of the energy functional $\mathcal{F}$ is a necessary and sufficient condition for the existence of a solution to the variational inequality (23) for every ε - uniform perturbation of the data involved in the problem. As every continuous set is well-positioned, we use Theorem 4.1 from [1] to deduce that, whenever the epigraph of Ψ is a continuous subset of $X \times \mathbb{R}$, then the energy functional $\mathcal{F}$ is bounded from below if and only if

$$\langle f,u\rangle < \Phi^{\infty}(u), \quad \forall u \in K^{\infty} \cap U,\ u \neq 0.$$

The following result summarizes the previous reasoning.

Proposition 2. *If the epigraph of* Ψ *is a continuous subset of* $X \times \mathbb{R}$, *then relation*

$$\langle f,u\rangle < \Phi^{\infty}(u), \quad \forall u \in K^{\infty} \cap U,\ u \neq 0$$

(equivalent to the boundedness from below of the energy functional $\mathcal{F}$*) is a necessary and sufficient condition for the existence of a solution to the variational inequality (23). Moreover, the existence of a solution is achieved also for every instance involving a bounded and semi-coercive operator* A_{ε}, *a linear functional* f_{ε}, *a proper lower semi-continuous convex function* Φ_{ε} *that is bounded from below, and a closed convex set* K_{ε} *such that* $K_{\varepsilon} \cap \mathrm{Dom}\,\Phi_{\varepsilon} \neq \varnothing$, *and*

$$\begin{aligned}
&\|A(x) - A_{\varepsilon}(x)\|_{*} < \varepsilon, \quad \forall x \in X \\
&\qquad \|f - f_{\varepsilon}\|_{*} < \varepsilon, \\
&K \subset K_{\varepsilon} + \varepsilon B_X \text{ and } K_{\varepsilon} \subset K + \varepsilon B_X, \\
&\Phi(x) - \varepsilon \leq \Phi_{\varepsilon}(x) \leq \Phi(x) + \varepsilon, \quad \forall x \in X.
\end{aligned}$$

Finally, let us remark that Proposition 2 does not provide a complete characterization of semi-coercive variational inequalities for which the necessary condition involving the boundedness from below of the energy functional $\mathcal{F}$ is also sufficient for the existence of solutions, characterization which at our knowledge, remains an open problem.

REFERENCES

[1] S. Adly, E. Ernst and M. Théra, *Stability of non-coercive variational inequalities*, Commun. Contem. Math., 4 (2002), 145–160.

[2] S. Adly, E. Ernst And M. Théra, *On the closedness of the algebraic difference of closed convex sets*, J. Math. Pures Appl., 82(2003), 1219–1249.

[3] H. Brezis, *Equations et inéquations non-linéaires dans les espaces vectoriels en dualité*, Ann. Inst. Fourier, Grenoble, 18(1968), 115 – 175.

[4] P. Coutat, M. Volle and J. E. Martinez-Legaz, *Convex functions with continuous epigraph or continuous level sets*, J. Optim. Theory Appl., 88 (1996), 365–379.

[5] G. Del Piero, *A Condition for Statical Admissibility in Unilateral Structural Analysis*, Theoretical and Numerical Non-smooth Mechanics, International Colloquium in honor of the 80^{th} birthday of Jean Jacques Moreau, 17-19 November 2003, Montpellier, France

[6] D. Gale and V. Klee, *Continuous convex sets*, Math. Scand. 7 (1959), 370 –391.

[7] R. T. Rockafellar, *Convex Analysis*, Princeton Mathematical Series 28, Princeton University Press, 1970.

[8] E. Zeidler, *Nonlinear Functional Analysis and its Applications II*, Springer-Verlag, 1990.

EXISTENCE AND MULTIPLICITY RESULTS FOR A NON LINEAR HAMMERSTEIN INTEGRAL EQUATION

F. Faraci
Dept. of Mathematics, University of Catania, Catania, Italy

Abstract: In this paper we study the solvability of a nonlinear Hammerstein integral equation by using a variational principle of B. Ricceri and methods of critical point theory. In particular we do not require any positivity assumption on the kernel of the equation. Our results can be applied to higher order elliptic boundary value problem with changing sign kernel.

1. INTRODUCTION

In the present paper we deal with the following nonlinear Hammerstein integral equation

$$x(s) = \int_{\Omega} k(s,t) f(t,x(t)) dt, \tag{1}$$

where $\Omega \subset \mathbb{R}^N$ is a bounded domain, $k : \Omega \times \Omega \to \mathbb{R}$ is a measurable and symmetric kernel and $f : \Omega \times \mathbb{R} \to \mathbb{R}$ is a Carathéodory function.

Under suitable hypotheses on the kernel k, and growth assumptions on the nonlinearity f, we prove existence and multiplicity results for equation (1).

There is a wide literature dealing with the problem of solving equation (1), see for instance [7,1,3,]. In these papers the authors apply a suitable

splitting theorem for linear integral operator and study an equivalent equation to (1), setting the new problem in the space of the 2–nd power summable functions. The solutions of the original problem belong to $\mathbb{L}_p$ for some $p > 2$.

Following an idea contained in [10], we work in a more abstract context, that is however the most natural one for equation (1). We prove that the solutions belong to a suitable energy space, compactly embedded into $\mathbb{L}_p$ for some $p > 2$. The idea of introducing these spaces goes back to the theory of Hilbert scales generated by linear operators (see [8]).

We notice that no positivity assumptions are made on the kernel. In a previous result (see [3]), we assumed $k \geq 0$ in order to prove the existence of at least two solutions to the equation

$$x(s) = \lambda \int_{\Omega} k(s,t) f(t, x(t)) dt,$$

where λ is a positive parameter. Actually, in many concrete examples, the kernel k arising as a Green function of a differential operator, is positive. In these cases it is possible to apply methods of positive operators to this equation in order to obtain information on the location of the solution.

Our approach is variational: since the kernel k is symmetric, it is possible to associate to (1) an energy functional whose critical points are the solutions of the Hammerstein equation. We apply a recent variational principle by Ricceri ([14]), a powerful tool for the localization of minima of integral functionals. We are able then to consider those cases where the kernel may change sign (as in higher order elliptic equations).

The scheme of the paper is the following. In section 2 we define the energy space as well as the energy functional representing our problem; in section 3 we prove our main existence-localization theorem. This result is then applied in section 4 to some concrete examples of nonlinearities. We conclude the paper with an application of the previous results to a polyharmonic boundary value problem.

2. PRELIMINARIES

Let us introduce the variational setting of our problem. In this section we define the energy space $\mathbb{E}$ and the energy functional J for the Hammerstein equation (1) on $\mathbb{E}$ whose critical points are precisely the solutions to (1).

2.1 The energy space

The problem of constructing a suitable function space where to set equation (1) was previously considered by Moroz and Zabreiko in [10].

Equation (1) can be written in the operator form

$$x = \mathbf{K}\mathbf{f}x,$$

where $\mathbf{f}x = f(\cdot, x)$ is the nonlinear superposition operator generated by the function f and $\mathbf{K}$ is the linear integral operator

$$\mathbf{K}x(s) = \int_{\Omega} k(s,t)x(t)dt.$$

The symmetry assumption on the kernel k implies the self–adjointness of $\mathbf{K}$ in $\mathbb{L}_2$ (see [11] for regular operators and [15] for general case). For our purposes we will need some additional information about the properties of the space $\mathbb{E}$ determined by specific properties of the linear operator $\mathbf{K}$.

Throughout the sequel it is assumed that $\mathbf{K}$ is a bounded compact operator from $\mathbb{L}_{p'}$ into $\mathbb{L}_p$ where $\frac{1}{p'} + \frac{1}{p} = 1$ and $1 < p' < 2 < p$.

$\mathbf{K}$ satisfies the latter condition if for example (see [7])

$$\int_{\Omega\times\Omega} |k(s,t)|^p \, ds \, dt < \infty.$$

It is further assumed that $\mathbf{K}$ is positive–definite on $\mathbb{L}_2$, that is $(\mathbf{K}x, x) > 0$ for all $x \in \mathbb{L}_2 \setminus \{0\}$ where we are denoting by $(\cdot,\cdot)$ the scalar product in $\mathbb{L}_2$.

Let $\mathbf{L}$ be the left inverse to $\mathbf{K}$ defined on the domain

$$D(\mathbf{L}) = \{y \in \mathbb{L}_2 : \mathbf{K}x = y \text{ has as solution } x \in \mathbb{L}_2\}$$

by the formula

$$\mathbf{L}y = x$$

where x is the unique solution of $\mathbf{K}x = y$.

The operator $\mathbf{L}$ is unbounded, positive definite and selfadjoint in $\mathbb{L}_2$ and it satisfies

$$\mathbf{L}\mathbf{K}x = x \quad \text{for all} \quad x \in \mathbb{L}_2.$$

Let us define a bilinear form on $D(\mathbf{L})$ by the formula

$$\langle x, y\rangle = (x, \mathbf{L}y).$$

Clearly $\langle\cdot,\cdot\rangle$ is an inner product on $D(\mathbf{L})$. In particular we have

$$\langle x, x\rangle \geq \|\mathbf{K}\|^{-1}\|x\|_2^2 \quad \text{for all } x \in D(\mathbf{L}), \tag{2}$$

where $\|\mathbf{K}\|$ stands for the norm of $\mathbf{K}$ in $\mathbb{L}_2$.

It is well known that if $\mathbf{L}$ is symmetric, densely defined on $\mathbb{L}_2$ and it satisfies (2), then the form $\langle\cdot,\cdot\rangle$ is closable in $\mathbb{L}_2$, i.e. it possesses a closed extension in $\mathbb{L}_2$. If we denote by $\mathbb{E}$ the domain of the closure of $\langle\cdot,\cdot\rangle$ in $\mathbb{L}_2$, then $\mathbb{E}$ is a Hilbert space with respect to the norm $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$, densely and continuously embedded in $\mathbb{L}_2$. $\mathbb{E}$ is called *the energy space* for the operator $\mathbf{K}$.

Lemma 1 *The space* $\mathbb{E}$ *has the following properties:*

1. $\langle x, x\rangle \geq \|\mathbf{K}\|^{-1}\|x\|_2^2$ *for all* $x \in \mathbb{E}$.
2. *The embedding* $\mathbb{E} \subset \mathbb{L}_p$ *is compact.*

Proof. *1.* If $x \in \mathbb{E}$, by the definition of the closure of a form, there exists a sequence $\{x_n\} \subset D(\mathbf{L})$ tending to x in $\mathbb{L}_2$ and satisfying $\langle x_n - x_m, x_n - x_m\rangle \to 0$ as $n, m \to +\infty$. Since $\|x\|^2 = \lim_{n\to\infty}\langle x_n, x_n\rangle$, passing to the limit in (2), we get our claim.

2. The proof follows from the classical Krasnosel'skii Krein factorization theorem which states that under our assumptions on $\mathbf{K}$, the square root of $\mathbf{K}$, $\mathbf{K}^{\frac{1}{2}}$ acting from $\mathbb{L}_2$ into $\mathbb{L}_p$ is compact. The thesis follows noticing that $\mathbb{E} = \mathbf{K}^{\frac{1}{2}}(\mathbb{L}_2)$.

2.2 The energy functional

Let us introduce now the energy functional on $\mathbb{E}$ for the Hammerstein equation (1) defined by

$$J(x) = \frac{1}{2}\|x\|^2 + \Phi(x),$$

where

$$\Phi(x) = -\int_\Omega F(s,x(s))ds \quad \text{and} \quad F(s,x) = \int_0^x f(s,t)dt.$$

Our main assumption is:
(f) there exist $r \in (1,2)$, $q \in (2,p)$ and $a \in \mathbb{L}_{\frac{p}{p-r}}$, $b \in \mathbb{L}_{\frac{p}{p-q}}$, $c \in \mathbb{L}_{p'}$ such that

$$|f(s,x)| \le a(s)|x|^{r-1} + b(s)|x|^{q-1} + c(s).$$

The previous assumption allows us to deduce some important properties of J. It is easy to prove the following lemma.

Lemma 2 *Let assume condition (f). Then the functional Φ is well–defined and sequentially weakly continuous on $\mathbb{E}$. Moreover it is continuously differentiable on $\mathbb{E}$. In particular J is continuously differentiable on $\mathbb{E}$ with the derivative given by*

$$J'(x)(h) = (\mathbf{L}x,h) - \int_\Omega f(s,x(s))h(s)ds \quad \text{for all } h \in \mathbb{E}.$$

Remark 3 Any critical point of J is a solution to the Hammerstein equation (1).

We conclude this section with the statement of our main tool, a recent variational principle of B. Ricceri which provides a powerful instrument for the existence and the localization of local minima of the energy functional.

Theorem 4 *([14, Theorem 2.5]) Let $\mathbb{E}$ be a Hilbert space and $\Phi, \Psi : \mathbb{E} \to \mathbb{R}$ two sequentially weakly lower semicontinuous functionals. Assume that Ψ is strongly continuous and coercive on $\mathbb{E}$, that is $\lim_{\|x\| \to +\infty} \Psi(x) = +\infty$. For each $\rho > \inf_{\mathbb{E}} \Psi$ set*

$$\varphi(\rho) := \inf_{\Psi^\rho} \frac{\Phi(x) - \inf_{cl_w \Psi^\rho} \Phi}{\rho - \Psi(x)}, \tag{3}$$

where $\Psi^\rho := \{x \in \mathbb{E} : \Psi(x) < \rho\}$ and $cl_w \Psi^\rho$ is the closure of Ψ^ρ in the weak topology of $\mathbb{E}$. Then, for each $\rho > \inf_{\mathbb{E}} \Psi$ and each $\mu > \varphi(\rho)$, the restriction of the functional $\Phi + \mu\Psi$ to Ψ^ρ has a global minimum point in Ψ^ρ.

3. EXISTENCE–LOCALIZATION THEOREM

This section is devoted to our main existence–localization theorem where we apply the variational principle of B. Ricceri to the energy functional J which can be expressed as the sum of two sequentially weakly lower semicontinuous and continuously differentiable terms.

Before stating our result we need to introduce the best constant of the embedding of $\mathbb{E}$ into $\mathbb{L}_p$, that is

$$S := \sup\{\|x\|_p : x \in \mathbb{E}, \|x\| \le 1\}.$$

Theorem 5 *Let assume condition* (f).

If there exists $\rho_* > 0$ *such that*

$$\Gamma(\rho_*) := S^r \|a\|_{\frac{p}{p-r}} \rho_*^{r-1} + S^q \|b\|_{\frac{p}{p-q}} \rho_*^{q-1} + S\|c\|_{p'} < \rho_*, \tag{4}$$

then, J has a local minimum in $\mathbb{E}$ whose norm is less than ρ_*.

Proof. Let us define on $\mathbb{E}$ the functionals

$$\Psi(x) := \|x\|^2 \quad \text{and}$$
$$\Phi(x) = -\int_\Omega F(s, x(s))ds.$$

Ψ and Φ are sequentially weakly lower semicontinuous on $\mathbb{E}$; Ψ is clearly strongly continuous and coercive. We claim that there exists a positive ρ such that the inequality

$$\varphi(\rho^2) = \inf_{\|x\| < \rho} \frac{\Phi(x) - \inf_{\|y\| \le \rho} \Phi(y)}{\rho^2 - \|x\|^2} < \frac{1}{2} \tag{5}$$

holds, in order to apply Theorem 4 with $\mu = \frac{1}{2}$. Then, we will have proved the existence of a local minimum x for J lying in Ψ^{ρ^2}, i.e. satisfying $\|x\| < \rho$.

Following an idea contained in [2], we introduce the function

$$\alpha(\rho) := \sup_{\|x\| \le \rho} \int_\Omega F(s, x(s))ds$$

for $\rho > 0$. It is easily seen that α is well defined and non decreasing in $]0,+\infty[$.

Let us prove now that

$$\limsup_{\tau \to 0} \frac{\alpha(\rho+\tau)-\alpha(\rho)}{\tau} < \rho \tag{6}$$

for some $\rho > 0$.

By direct computations, if $\rho > 0$ and $0 < |\tau| < \rho$, one has

$$\frac{\alpha(\rho+\tau)-\alpha(\rho)}{\tau} \le \frac{1}{|\tau|} \sup_{\|x\|_p \le 1} \int_\Omega \left| \int_{\rho x(s)}^{(\rho+\tau)x(s)} |f(s,t)|\,dt \right| ds \le$$

$$\le \frac{1}{|\tau|} \sup_{\|x\|_p \le 1} \left\{ \|a\|_{\frac{p}{p-r}} \left| \frac{(\rho+\tau)^r-\rho^r}{r} \right| \|x\|_p^r + \|b\|_{\frac{p}{p-q}} \left| \frac{(\rho+\tau)^q-\rho^q}{q} \right| \|x\|_p^q + \|c\|_{p'} \|x\|_p \right\} \le$$

$$\le \frac{S^r}{r} \|a\|_{\frac{p}{p-r}} \left| \frac{(\rho+\tau)^r-\rho^r}{\tau} \right| + \frac{S^q}{q} \|b\|_{\frac{p}{p-q}} \left| \frac{(\rho+\tau)^q-\rho^q}{\tau} \right| + S\|c\|_{p'}.$$

Passing to the maximum limit we obtain that

$$\limsup_{\tau \to 0} \frac{\alpha(\rho+\tau)-\alpha(\rho)}{\tau} \le \Gamma(\rho).$$

Assuming the existence of a positive ρ_* such that $\Gamma(\rho_*) < \rho_*$, (6) holds.

Now, condition (6) implies the inequality

$$\inf_{\rho>0} \inf_{\sigma<\rho} \frac{\alpha(\rho)-\alpha(\sigma)}{\rho^2-\sigma^2} < \frac{1}{2},$$

which is equivalent to our claim (5). □

4. APPLICATIONS

In this section we consider specific examples of nonlinearities in order to handle condition (4). Combining our existence–localization theorem with the Mountain Pass Theorem, we obtain a multiplicity result for the Hammerstein equation (1).

4.1 A superlinear case

In this first application, we deal with a superlinear nonhomogeneous nonlinearity where a positive parameter appears:

$$f_\lambda(s,x) = b(s)g(x) + \lambda c(s), \qquad (f_1)$$

Here $0 \neq c \in \mathbb{L}_{p'}$, $0 < b \in \mathbb{L}_{\frac{p}{p-q}}$ for some $q \in (2,p)$ and g is a continuous real function such that $g(0) \equiv 0$ and satisfying the assumptions:

$$|g(x)| \leq k|x|^{q-1} \text{ for some } k \in \mathbb{R},\ k > 0 \text{ and for every } x \in \mathbb{R}; \qquad (g_1)$$

$$\text{there exist } \tau > 2,\ R_\tau > 0 \text{ such that if } G(x) := \int_0^x g(t)dt \qquad (g_2)$$

$$0 < \tau G(x) < g(x)x \quad \text{for } |x| \geq R_\tau.$$

Our result reads as follows.

Theorem 6 *Let assume conditions* (f_1), (g_1) *and* (g_2). *Then, there exists a positive* $\lambda_* > 0$ *such that for each* $\lambda \in (0,\lambda_*)$ *equation (1) has at least two solutions.*

Proof. We denote by J_λ the energy functional associated to $f_\lambda(s,x)$. We are going to apply Theorem 5 to our nonlinearity that clearly satisfies (f). We can estimate the function $\Gamma(\rho)$ given by the formula

$$\Gamma(\rho) = S^q k \|b\|_{\frac{p}{p-q}} \rho^{q-1} + S\lambda \|c\|_{p'}.$$

It is easily seen that there exists a positive λ_* such that for every $\lambda \in (0,\lambda_*)$,

$$\rho_*(\lambda) = 2\lambda S \|\gamma\|_{p'}$$

satisfies condition

$$\Gamma(\rho_*(\lambda)) < \rho_*(\lambda).$$

Thus, Theorem 5 ensures the existence of a local minimum x_λ, whose norm is less than $\rho_*(\lambda)$.

In the next step we prove that J_λ is unbounded below. From assumption (g_2), it follows that

$$G(x) \geq c_1 |x|^{\tau} - c_2$$

for some $c_1, c_2 > 0$ and for every $x \in \mathbb{R}$.

Let us choose then a function $0 \neq x \in \mathbb{E}$ and $\sigma > 0$.

One has

$$\int_\Omega F_\lambda(s, \sigma x(s))ds \geq c_1 \sigma^{\tau} \int_\Omega b(s)|x(s)|^{\tau} ds - c_2 \int_\Omega b(s)ds + \lambda\sigma \int_\Omega c(s)x(s)ds.$$

Passing to the limit, for $\sigma \to +\infty$ we have that

$$J_\lambda(\sigma x) = \frac{\sigma^2}{2}\|x\|^2 - \int_\Omega F_\lambda(s, \sigma x(s))ds \to -\infty$$

that is our claim.

In a standard way it is possible to prove that J_λ satisfies the Palais–Smale condition.

Hence, all the assumptions of the Mountain Pass Theorem are satisfied (see, e.g. [5,13]), so we deduce that J_λ has a second critical point y_λ, different from x_λ. □

4.2 A nonlinearity with sublinear and superlinear terms

In this second application we deal with a nonlinearity where a combination of sublinear and superlinear terms appears:

$$f_\lambda(s,x) = b(s)g(x) + \lambda a(s)x|x|^{r-2}. \qquad (f_2)$$

Here, $0 < a \in \mathbb{L}_{\frac{p}{p-r}}$ for some $r \in (1,2)$, $b \in \mathbb{L}_{\frac{p}{p-q}}$ for some $q \in (2,p)$ and g is as in the previous application.

Remark 7 We notice that $f_\lambda(s,0) = 0$ and $x = 0$ is a trivial solution of (1).

For proving the next result we need to recall the following definition. A critical point y of J_λ is said a *mountain pass type critical point* of J_λ if there exists arbitrary small $\rho > 0$ such that the set

$$\{J_\lambda^c \cap B_\rho(y)\} \setminus \{y\}$$

is nonempty and not path connected, where $c = J_\lambda(y)$, $J_\lambda^c = \{x \in \mathbb{E} : J_\lambda(x) \leq c\}$, and $B_\rho(y)$ is the open ball of radius ρ centered at y.

The following two lemmas will be useful in the sequel.

Lemma 8 *Let assume conditions* (f_2), (g_1) *and* (g_2). *Put* $M_\rho := \{x \in B_\rho(0) : J_\lambda(x) \geq 0\}$. *Then, there exists* $\overline{\rho} > 0$ *such that*

$$\frac{d}{d\sigma} J_\lambda(\sigma x)|_{\sigma=1} > 0 \tag{7}$$

for any $x \in M_{\overline{\rho}}$.

Proof. See [9].

Lemma 9 *Let assume conditions* (f_2), (g_1) *and* (g_2). *If* $\overline{\rho}$ *is as in Lemma 8, then the set* $\{J_\lambda^0 \cap B_{\overline{\rho}} - (0)\} \setminus \{0\}$ *is pathwise connected.*

Proof. We notice that the set $J_\lambda^0 \cap B_{\overline{\rho}}(0)$ is starshaped with respect to the origin. Let us suppose that $J_\lambda(\sigma_0 x_0) > 0$ for some $x_0 \in J_\lambda^0 \cap B_{\overline{\rho}}(0)$ and $\sigma_0 \in (0,1)$. (7) implies that

$$\frac{d}{d\sigma} J_\lambda(\sigma\sigma_0 x_0)|_{\sigma=1} > 0.$$

Then, $J_\lambda(\sigma x_0) > 0$ for all $\sigma \in [\sigma_0, 1]$. In particular, we obtain $J_\lambda(x_0) > 0$ in contrast with the definition of x_0.

Let us prove now that $\{J_\lambda^0 \cap B_{\overline{\rho}}(0)\} \setminus \{0\}$ is a retract of the set $B_{\overline{\rho}}(0) \setminus \{0\}$. Let x be in $M_{\overline{\rho}}$. Lemma 8 implies the existence of a unique solution $\sigma(x) \in (0,1]$ solution of $J_\lambda(\sigma x) = 0$. The uniqueness follows from the starshapedness of $J_\lambda^0 \cap B_{\overline{\rho}}$.

By (7) we have

$$\frac{d}{d\sigma} J_\lambda(\sigma\sigma(x)x)_{\sigma=1} > 0.$$

The continuity of the function $\sigma(x)$ in a neighborhood of x in $M_{\overline{\rho}}$ follows from the Implicit Function Theorem. In particular one has that $\sigma : M_{\overline{\rho}} \to (0,1]$ is continuous. We define then $r : B_{\overline{\rho}}(0) \to \{J_\lambda^0 \cap B_{\overline{\rho}}(0)\} \setminus \{0\}$ by the formula

$$r(x) = \begin{cases} \sigma(x)x, & x \in M_{\overline{\rho}}, \\ x, & x \in \{J_\lambda^0 \cap B_{\overline{\rho}}(0)\} \setminus \{0\}. \end{cases}$$

It is possible to prove that r is a retraction of $\{B_{\bar{\rho}}\} \setminus \{0\}$ into $\{J_\lambda^0 \cap B_{\bar{\rho}}\} \setminus \{0\}$. In particular r is continuous as it follows from the continuity of σ. Moreover the restriction of r to $\{J_\lambda^0 \cap B_{\bar{\rho}}(0)\} \setminus \{0\}$ is the identity map. Now, $\{B_{\bar{\rho}}\} \setminus \{0\}$ is contractible in itself. By [12] the retract of a contractible in itself set is also contractible in itself. Therefore $\{J_\lambda^0 \cap B_{\bar{\rho}}\} \setminus \{0\}$ is contractible in itself. In particular, $\{J_\lambda^0 \cap B_{\bar{\rho}}\} \setminus \{0\}$ is pathwise connected as we claimed. □

Our result reads as follows.

Theorem 10 *Let assume conditions* (f_2), (g_1) *and* (g_2). *Then, there exists a positive* λ_* *such that for each* $\lambda \in (0, \lambda_*)$ *equation (1) has at least two nontrivial solutions.*

Proof. The scheme of the proof follows the one of the previous theorem: the first solution comes from an application of Theorem 5 to J_λ; by using the Mountain Pass Theorem we are able to prove the existence of a second critical point to J_λ.

The nonlinearity f_λ clearly satisfies (f). Let us consider the function $\Gamma(\rho)$ given by the formula

$$\Gamma(\rho) = S^q k \|b\|_{\frac{p}{p-q}} \rho^{q-1} + \lambda S^r \|a\|_{\frac{p}{p-r}} \rho^{r-1}.$$

It is easily seen that there exists a positive λ_* such that for every $\lambda \in (0, \lambda_*)$,

$$\rho_*(\lambda) = \frac{1}{S}\left(\frac{\lambda \|a\|_{\frac{p}{p-r}}}{k\|b\|_{\frac{p}{p-q}}} \frac{(2-r)}{(q-2)}\right)^{\frac{1}{q-r}}$$

satisfies

$$\Gamma(\rho_*(\lambda)) < \rho_*(\lambda).$$

Theorem 5 ensures the existence of a local minimum x_λ whose norm is less than $\rho_*(\lambda)$. We need to prove that x_λ is different to zero.

Fix $\lambda \in (0, \lambda_*)$. J_λ satisfies the following inequality:

$$J_\lambda(\sigma x) \leq \frac{1}{2}\sigma^2 \|x\|^2 + \frac{k}{q}|\sigma|^q \int_\Omega b(s)|x(s)|^q \, ds - \frac{\lambda}{r}|\sigma|^r \int_\Omega a(s)|x(s)|^r \, ds < 0$$

as σ tends to zero.

Since x_λ is the global minimum of the restriction of J_λ to a suitable open ball centered at zero, x_λ is different to zero.

J_λ is unbounded from below and satisfies the Palais–Smale condition.

Thus from a suitable version of the Mountain Pass Theorem (see, e.g., [5]), we have the following alternative: either J_λ has a mountain pass type critical point y_λ different to x_λ or its set of critical points is infinite. We know from Lemma 9 that $x=0$ is not a mountain pass type critical point. Hence, J_λ has at least two nontrivial critical points as we claimed. □

5. AN EXAMPLE

In this section we give an application of our results. Let us consider the following semilinear polyharmonic problem

$$\begin{cases} (-\Delta)^m x = f(s,x) & \text{in } \Omega, \\ \mathcal{D}_m x = 0 & \text{on } \partial\Omega. \end{cases} \tag{8}$$

where $m \in \mathbb{N}$ is an integer, $(-\Delta)^m$ is the m–harmonic Laplace operator, $\mathcal{D}_m x := (\mathcal{D}^k x)_{k\in\mathbb{N}^n}$ $(0 \le |k| \le m-1)$ is the boundary operator, $\Omega \subset \mathbb{R}^N$ is a bounded domain with the boundary $\partial\Omega$ of the class C^{2m+1} and $f : \Omega \times \mathbb{R} \to \mathbb{R}$ is a Carathéodory function.

We can solve the above problem by introducing an equivalent Hammerstein equation having as kernel a suitable Green function. Such Green function $G_{m,N}(s,t)$ exists, symmetric and satisfies the estimate (see [6])

$$|G_{m,N}(s,t)| \le \begin{cases} c\,|s-t|^{2m-N} & \text{if } m < \frac{N}{2}, \\ |\log|s-t|| + c & \text{if } m = \frac{N}{2}, \\ c & \text{if } m > \frac{N}{2}. \end{cases}$$

It is possible to prove that the integral operator $\mathbf{K}$ satisfies all the assumptions in Section 1 with p satisfying the following estimates

$$p < \begin{cases} \dfrac{2N}{N-2m} & \text{if } m < \frac{N}{2}, \\ \infty & \text{if } m \ge \frac{N}{2}. \end{cases}$$

Hence, we can apply the results in the previous sections.

Remark 11 We point out that the Green function $G_{m,N}(s,t)$ changes sign on many model domains (see, e.g. [4]). So, the classical positivity methods can not be applied to (8) while the results of the present paper can be used to study (8).

ACKNOWLEDGEMENTS

Parts of this paper were completed during the visit of the first author to the University of Bristol. It is a pleasure to thank the University for support and hospitality. A special thank goes to Vitaly Moroz for his clear suggestions.

REFERENCES

[1] A. Ambrosetti, P.H. Rabinowitz, *Dual variational methods in critical point theory and applications* J. Functional Analysis 14 (1973) 349–381.

[2] G. Anello, G. Cordaro, *An existence and localization theorem for the solutions of a Dirichlet problem*, Ann. Polon. Math., to appear.

[3] F. Faraci, *Bifurcation theorems for Hammerstein nonlinear integral equations* Glasgow Math. J. 44 (2002) 471–481.

[4] H. Grunau, G. Sweers, *Positivity for equations involving polyharmonic operators with Dirichlet boundary conditions*, Math. Ann. 307 (1997) 589–626.

[5] H. Hofer, *A geometric description of the neighbourhood of a critical point given by the mountain-pass theorem*, J. London Math. Soc. (2) 31 (1985), 566–570.

[6] V. Kozlov, V. Maz'ya, J. Rossmann, Elliptic boundary value problems in domain with point singularities. Mathematical Surveys and Monographs, Vol. 52, A.M.S., 1997.

[7] M.A. Krasnoselskiĭ, Topological methods in the theory of non-linear integral equations. Macmillan, New York, 1964.

[8] S.G. Krein, Ju.I. Petunin , *Scales of Banach spaces* (Russian).Uspehi Mat. Nauk (1966) 2 89–168.

[9] V. Moroz, *On the Morse critical groups for indefinite sublinear elliptic problems*, Nonlinear Anal. Ser. A: Theory Methods. 52 (2003), 1441–1453.

[10] V. Moroz, P. Zabreiko, *On the Hammerstein equations with natural growth conditions*. Z. Anal. Anwendungen 18 (1999), 625–638.

[11] A. Povolotskii, P. Zabreiko, *On the theory of Hammerstein equations* (Russian). Ukrain. Mat. Zh. 22 (1970), 150–162.

[12] E.H. Spanier, Algebraic topology, McGraw-Hill Book Co., New–York, 1966.

[13] P.H. Rabinowitz, *Minimax methods in critical point theory with applications to differential equations*, CBMS Regional Conf. Ser. in Math, 65 A.M.S., R.I., 1986.

[14] B. Ricceri, *A general variational principle and some of its applications*, J. Comput. Appl. Math. 113 (2000), 401–410.

[15] P. Zabreiko, *On the theory of Integral Operators* (Russian), Thesis, Voronej State University, 1968.

DIFFERENTIABILITY OF WEAK SOLUTIONS OF NONLINEAR SECOND ORDER PARABOLIC SYSTEMS WITH QUADRATIC GROWTH AND NON LINEARITY $q \geq 2$

L. Fattorusso
D.I.M.E.T. Faculty of Engineering, University of Reggio Calabria, Reggio Calabria, Italy

Abstract: Let Ω be a bounded open subset of $\mathbb{R}^n$, let $X=(x,t)$ be a point of $\mathbb{R}^n \times \mathbb{R}^N$. In the cylinder $Q=\Omega\times(-T,0)$, $T>0$, we deduce the local differentiability result

$$u \in L^2(-a,0,H^2(B(\sigma),\mathbb{R}^N)) \cap H^1(-a,0,L^2(B(\sigma),\mathbb{R}^N))$$

for the solutions u of the class $L^q(-T,0,H^{1,q}(\Omega,\mathbb{R}^N)) \cap C^{0,\lambda}(\bar{Q},\mathbb{R}^N)$ ($0<\lambda<1$, N integer ≥ 1) of the non linear parabolic system

$$-\sum_{i=1}^{n} D_i a^i(X,u,Du) + \partial u \partial t = B^0(X,u,Du)$$

with quadratic growth and non linearity $q \geq 2$. This result had been obtained making use of the interpolation theory and an imbedding theorem of Gagliardo-Nirenberg type for functions u belonging to $W^{1,q} \cap C^{0,\lambda}$.

1. INTRODUCTION

Let Ω be an open bounded subset of $\mathbb{R}^n$ ($n>2$) of generic point $x=(x_1,x_2,...,x_n)$, Q the cylinder $\Omega\times(-T,0)$ ($0<T<+\infty$); here N is an integer >1, $(|)_k$ and $\|\cdot\|_k$ are the scalar product and the norm in $\mathbb{R}^k$,

respectively. We will drop the subscript k when there is no fear of confusion.

We define

$$B(x^0,\sigma)=\left\{x\in\mathbb{R}^n : |\, x_i-x_i^0\,|<\sigma, i=1,...,n\right\}$$

If $u:Q\to\mathbb{R}^N$, we set $Du=(D_1u,...,D_nu)$ where, as usual, $D_i=\frac{\partial}{\partial x_i}$. Clearly $Du\in\mathbb{R}^{nN}$ and we denote by $p=(p^1,...,p^n)$, $p^i\in\mathbb{R}^N$, a typical vector of $\mathbb{R}^{nN}$ and let $V(p)=\left(1+\|p\|^2\right)^{\frac{1}{2}}$.

Let $u\in L^q(-T,0,H^{1,q}(\Omega,\mathbb{R}^N))\cap C^{0,\lambda}(\overline{Q},\mathbb{R}^N)$ $(0<\lambda<1)$ [1] be a solution in Q to the second order nonlinear parabolic system of variational type

$$-\sum_{i=1}^{n}D_ia^i(X,u,Du)+\frac{\partial u}{\partial t}=B^0(X,u,Du) \tag{1.0}$$

[1] By $H^{m,p}\left(\Omega,\mathbb{R}^N\right), m=0,1,2...,\ 1<p<\infty,$ we will denote the usual Sobolev space

$$H^{0,p}\left(\Omega,\mathbb{R}^N\right)=L^p\left(\Omega,\mathbb{R}^N\right) \text{ e } |u|_{0,p,\Omega}\,\|u\|_{0,p,r}=\left\{\int_\Omega\|u\|^p\,dx\right\}^{1/p},\ 1<p<\infty.$$

If $1\le p<\infty$ and m, j are integers ≥ 0, we denote

$$|u|_{j,p,\Omega}=\left[\int_\Omega\left(\sum_{|\alpha|=j}\|D^\alpha u\|^2\right)^{p/2}dx\right]^{1/p},\quad \|u\|_{m,p,\Omega}=\left\{\sum_{j=0}^{m}|u|_{j,p,\Omega}^p\right\}^{1/p}.$$

if $p=2$, we shall use thenotation $H^s,|\cdot|_{s,\Omega},\|\cdot\|_{s,\Omega}$

By $H^{\theta,r}(\Omega,\mathbb{R}^N),0<\theta<1,1<r<\infty$, we will denote the Slobodeckij space of those vectors $u\in L^r(\Omega,\mathbb{R}^N)$ such that

$$|\,u\,|_{\theta,r,\Omega}^r=\int_\Omega dx\int_\Omega\frac{\|u(x)-u(y)\|^r}{\|x-y\|^{n+\theta r}}dy<+\infty$$

By $H^{m+\theta,r}(\Omega,\mathbb{R}^N)$, $m=1,2,...,0<\theta<1,1<r<\infty$ we will denote the space of those vectors $u\in H^{m,r}(\Omega,\mathbb{R}^N)$ such that $D^\alpha u\in H^{\theta,r}(\Omega,\mathbb{R}^N)$, $\forall\,|\,\alpha\,|=m$.

If $r=2$ we shall use the notation $H^{m+\theta},m=0,1,2,...,0\le\theta<1$ instead of $H^{m+\theta,2}$.

By $C^{0,\lambda}(\Omega,\mathbb{R}^N)$, $0<\lambda<1$, we shall denote the space of those vectors $u\in C^0(\overline{\Omega},\mathbb{R}^N)$ for which $[u]_{\lambda,\overline{\Omega}}=\sup_{x,y\in\overline{\Omega},x\neq y}\dfrac{\|u(x)-u(y)\|}{\|x-y\|^\lambda}<+\infty$

In Q the Hölder continuity is considered with respect to the parabolic metric

$$d(X,Y)=\max\left\{\|x-y\|,|\,t-\tau\,|^{1/2}\right\},\quad X=(x,t),\quad Y=(y,\tau)$$

in the sense that

$$\int_Q \left\{ \sum_{i=1}^{n} \left(a^i(X,u,Du) \mid D_i\varphi \right) - \left(u \mid \frac{\partial \varphi}{\partial t} \right) \right\} dX$$

$$= \int_Q \left(B^0(X,u,Du) \mid \varphi \right) dX, \quad \forall \varphi \in C_0^\infty(Q, \mathbb{R}^N), \tag{1.1}$$

where $X = (x,t)$ and $a^i(X,u,p)$, $i = 1,...,n$, and $B^0(X,u,p)$ are vectors of $\mathbb{R}^N$ defined on $\Lambda = Q \times \mathbb{R}^N \times \mathbb{R}^{nN}$, satisfying the following conditions:

the vector $B^0(X,u,p)$ is measurable in X, continuous in (u,p), and, for each $(X,u,p) \in \Lambda$, with $\|u\| \le k$, $p \in \mathbb{R}^{nN}$ (1.2)

$$\|B^0(X,u,p)\| + \sum_{s=1}^{n} \left\| \frac{\partial B^0}{\partial x_s} \right\| + \sum_{k=1}^{N} \left\| \frac{\partial B^0}{\partial u_k} \right\| \le M(k) V^q(p)$$

$$\sum_{k=1}^{N} \sum_{j=1}^{n} \left\| \frac{\partial B^0}{\partial p_k^j} \right\| \le c(k) V^{q-1}(p)$$

the vectors $a^i(X,u,p)$, $i = 1,2,...,n$, are of class C^1 in $\bar{Q} \times \mathbb{R}^N \times \mathbb{R}^{nN}$ and, for each $(X,u,p) \in \Lambda$ with $\|u\| \le k$ (1.3)

$$\|a^i\| + \sum_{s=1}^{n} \left\| \frac{\partial a^i}{\partial x_s} \right\| + \sum_{k=1}^{N} \left\| \frac{\partial a^i}{\partial u_k} \right\| \le c(k) V^{q-1}(p), \quad i = 1,2,...,n$$

$$\sum_{k=1}^{N} \sum_{j=1}^{n} \left\| \frac{\partial a^i}{\partial p_k^j} \right\| \le M(k) V^{q-2}(p), \quad i = 1,2,...,n$$

there exists $\nu(k) > 0$ such that (1.4)

$$\sum_{i,j=1}^{n}\sum_{h,k=1}^{N}\frac{\partial a_k^i(X,u,p)}{\partial p_k^j}\xi_h^i\xi_k^j \geq \nu(k)\|\xi\|^2$$

for each $\xi=(\xi^1|\xi^2|...|\xi^n)\in\mathbb{R}^{nN}$ and for each $(X,u,p)\in\Lambda$ with $\|u\|\leq k$.

In the work [4] it had been examined the local differentiability with respect to the spatial derivatives of the solutions

$$u\in L^q(-T,0,H^{1,q}(\Omega,\mathbb{R}^N))\cap C^{0,\lambda}(Q,\mathbb{R}^N),\quad q\geq 2,\quad 0<\lambda<1 \tag{1.5}$$

to the system (1.1), proving that, under the assumptions of monotony and non linearity $q>2$, for each cube $B(\sigma)=B(x^0,\sigma)\subset\subset\Omega$ and $\forall a\in(0,T)$ it results

$$u\in L^q(-a,0,H^{1+\theta,q}(B(\sigma),\mathbb{R}^N)),\quad \forall\theta\in\left(0,\frac{2}{q}\right)$$

and this result is analogous to that which I had obtained in [3] under the assumptions of non linearity $q=2$, under the boundedness conditions for the derivatives $\frac{\partial a^j}{\partial p_k^i}$ and of strong ellipticity.

In the paper [5] I had considered again the problem of differentiability, under assumptions of monotony and non linearity $1<q<2$, always achieving results of the same type.

The aim of this paper is to obtain for the solutions (1.5) of the system (1.1), under the assumptions (1.2), (1.3), (1.4) and of non linearity $q>2$, the result of

$$u\in L^2(-a,0,H^2(B(\sigma),\mathbb{R}^N))\cap H^1(-a,0,L^2(B(\sigma),\mathbb{R}^N))$$

for each cube $B(\sigma)=B(x^0,\sigma)\subset\subset\Omega$ and $\forall a\in(0,T)$, making use of the interpolation theory and an imbedding theorem of Gagliardo-Nirenberg type for functions u belonging to $W^{1,q}\cap C^{0,\lambda}$.

This paper extends the result which had been obtained by Marino-Maugeri in [6] in the case of nonlinearity $q=2$ and it is analogous to the regularity result which had been obtained by Campanato in [2] for elliptic systems with nonlinearity $q>2$.

2. SOME NOTATIONS AND PRELIMINARY RESULTS

In this section we list a few lemmas that will be needed in the sequel of the work and which are already well known in the mathematical literature.

Let $B(\sigma)=B(x^0,\sigma)$, $(x^0\in\mathbb{R}^n,\sigma>0)$ a cube of $\mathbb{R}^n$ defined by

$$B(\sigma)=\{x\in\mathbb{R}^n:|x_i-x_i^0|<\sigma,\quad i=1,...,2\}.$$

If $u:B(\sigma)\times(-T,0)\to\mathbb{R}^N$, $(T>0)$ and $X=(x,t)\in B(\tau\sigma)\times(-T,0)$, $\tau\in(0,1)$, $|h|<(1-\tau)\sigma$, then we define

$$\tau_{i,h}u(X)=u(x+he^i,t)-u(X),\quad i=1,2,...,n$$

where $\{e^s\}_{s=1,..,n}$ is the standard base of $\mathbb{R}^n$.

Lemma 2.1. *If* $u\in L^q(-b,-\rho,H^{1,q}(B(\sigma),\mathbb{R}^N))$, $q>1$, $0\le\rho<b$, *then* $\forall\tau\in(0,1)$ *and* $\forall|h|<(1-\tau)\sigma$

$$\int_{-b}^{-\rho}dt\int_{B(\tau,\sigma)}\left\|\tau_{i,h}u\right\|^q dx\le|h|^q\int_{-b}^{-\rho}dt\int_{B(\sigma)}\left\|D_iu\right\|^q dx,\quad i=1,2,\ldots,n.$$

See for instance [1], Cap. I, Lemma 3.VI.

Lemma 2.2. *If* $v\in L^p(-a,0,L^p(B(2\sigma),\mathbb{R}^N))$, $a,\sigma>0$, $1<p<+\infty$, *and there exists* $M>0$ *such that*

$$\int_{-a}^{0}dt\int_{B(\sigma)}\left\|\tau_{i,h}v\right\|^p dx\le|h|^p M,\quad\forall|h|<\sigma,i=1,2,\ldots,n.$$

then $v\in L^p(-a,0,H^{1,p}(B(\sigma),R^N))$ *and*

$$\int_{-a}^{0}dt\int_{B(\sigma)}\left\|D_iv\right\|^p dx\le M,\quad i=1,2,\ldots,n.$$

The proof is the same of Theorem 3.X in [1].

Lemma 2.3. *Let* N *be positive integer and* Ω *a cube of* $\mathbb{R}^n$. *If*

$$u\in H^{1+\theta,q}(\Omega,\mathbb{R}^N)\cap C^{0,\lambda}(\Omega,\mathbb{R}^N)$$

with $1<r<\infty$, $0<\theta<1$ *and* $0<\lambda<1$, *then* $u\in W^{1,p}(\Omega,\mathbb{R}^N)$ *and there exists a constant* c *(depending on* $\Omega,\theta,\lambda,n,a,q$*) such that:*

$$\|u\|_{1,p,\Omega}\le c\|u\|^{a}_{1+\theta,q,\Omega}\|u\|^{1-a}_{C^{0,\lambda}(\Omega,\mathbb{R}^N)},$$

where

$$\frac{1}{p}=\frac{1}{n}+a\left(\frac{1}{q}-\frac{1+\theta}{n}\right)-(1-a)\frac{\lambda}{n},\quad \forall a\in\left]\frac{1-\lambda}{1+\theta-\lambda},1\right[.$$

In particular, if $1-\lambda<\theta<1$, *for* $a=\frac{1}{2}$ *we get*

$$u\in W^{1,p}(\Omega,\mathbb{R}^N)$$

and there exists a constant c *(depending on* $\Omega,\theta,\lambda,n,a,q$*) such that*

$$\|u\|_{1,p,\Omega}\le c\|u\|^{\frac{1}{2}}_{1+\theta,q,\Omega}\|u\|^{\frac{1}{2}}_{C^{0,\lambda}(\Omega,\mathbb{R}^N)}$$

where $p=2q+\dfrac{2q^2(\theta+\lambda-1)}{n-q(\theta+\lambda-1)}(>2q)$

See [6] Theorem 2.2 for $m=1$, $r=q$, $s=0$, $j=1$.

3. DIFFERENTIABILITY OF THE SOLUTIONS TO THE SYSTEM (1.1)

Let $u\in L^q(-T,0,H^{1,q}(\Omega,\mathbb{R}^N))\cap C^{0,\lambda}(\overline{Q},\mathbb{R}^N)$, $0<\lambda<1$, $q\ge 2$, be a solution to the system (1.1) and let us suppose that the assumptions (1.2), (1.3) and (1.4) are fulfilled; in what follows we shall set

$$k=\sup_{Q}\|u\|,\quad U=[u]_{\lambda,Q}=\sup_{X,Y\in Q,X\ne Y}\frac{\|u(X)-u(Y)\|}{d^{\lambda}(X,Y)}$$

where $d(X,Y)$ is the parabolic metric

$$d(X,Y)=\max\left\{\|x-y\|,|t-\tau|^{\frac{1}{2}}\right\},\quad X=(x,t),Y=(y,\tau).$$

Now we show the following

Theorem 3.1. *If* $u \in L^q(-T,0,H^{1,q}(\Omega,\mathbb{R}^N)) \cap C^{0,\lambda}(\bar{Q},\mathbb{R}^N)$, $0<\lambda<1$, $q \geq 2$, *is a solution to the system (1.1), if the assumptions (1.2), (1.3) and (1.4) hold, then,* $\forall B(3\sigma) = B(x^0,3\sigma) \subset\subset \Omega$, $\forall a,b \in (0,T)$, $a<b$, *it results:*

$$u \in L^2(-a,0,H^2(B(\sigma),\mathbb{R}^N)) \cap H^1(-a,0,L^2(B(\sigma),\mathbb{R}^N)) \tag{3.1}$$

and the following estimate holds:

$$\begin{aligned} &\int_{-a}^{0} \left(|u|^2_{2,B(\sigma)} + \left|\frac{\partial u}{\partial t}\right|^2 \right) dt \leq \\ &\leq c(\nu,k,U,\lambda,\sigma,q,a,b,n)\left\{1 + \int_{-b}^{0} |u|^q_{1,q,B(3\sigma)}\, dt\right\} \end{aligned} \tag{3.2}$$

Proof. Fixed $B(3\sigma) = B(x^0,3\sigma) \subset\subset \Omega$, $a,b \in (0,T)$, with $a<b$, let $\psi(x) \in C_0^\infty(\mathbb{R}^n)$ be a real function which has the following properties:

$$0 \leq \psi \leq 1, \quad \psi = 1 \text{ in } B(\sigma), \quad \psi = 0 \text{ in } \mathbb{R}^n \setminus B(2\sigma), \quad \|D\psi\| \leq \frac{c}{\sigma}. \tag{3.3}$$

Let $\rho_m(t)$, with m integer $> 2/a$, be a function defined on $\mathbb{R}$ by this way

$$\rho_m(t) = \begin{cases} 1 & \text{if } -a \leq t \leq \dfrac{-2}{m} \\ 0 & \text{if } t \geq \dfrac{-1}{m} \text{ or } t \leq b \\ \dfrac{t+b}{b-a} & \text{if } -b < t < -a \\ -(mt+1) & \text{if } \dfrac{-2}{m} < t < \dfrac{-1}{m} \end{cases} \tag{3.4}$$

Finally let $\{g_s(t)\}$ be a sequence of symmetric mollifying functions

$$\begin{cases} g_s(t) \in C_o^\infty(\mathbb{R}), \quad g_s(t) \geq 0, \quad g_s(t) = g_s(-t) \\ \text{supp } g_s \subset \left[-\frac{1}{s}, \frac{1}{s}\right] \\ \int g_s(t)dt = 1 \end{cases} \tag{3.5}$$

Having fixed i integer, $1 \leq i \leq n$, and h such that $|h| < \min\left\{1, \frac{\sigma}{2}\right\}$, if we set $b^* = \frac{a+b}{2}$, let us assume in (1.1), for each $m > \frac{2}{a}$ and for each $s > \max\left\{m, \frac{1}{T-b}\right\}$,

$$\varphi = \tau_{i,-h}\{\psi^2 \rho_m[(\rho_m \tau_{i,h} u) * g_s]\}.$$

Then we get

$$\begin{aligned} &\int_Q \sum_{j=1}^n \left(\tau_{i,h} a^j(X,u,Du) | D_j\{\psi^2 \rho_m[(\rho_m \tau_{i,h} u) * g_s]\}\right) dX = \\ &= \int_Q \left(\tau_{i,h} u | \psi^2 \{\rho_m[(\rho_m \tau_{i,h} u) * g_s]\}'\right) dX + \\ &+ \int_Q \left(B^0(X,u,Du) | \tau_{i,-h}\{\psi^2 \rho_m[(\rho_m \tau_{i,h} u) * g_s]\}\right) dX. \end{aligned} \tag{3.6}$$

Furthermore

$$\begin{aligned} &\tau_{i,h} a^j(X, u(X), Du(X)) = \\ &= \int_0^1 \frac{\partial}{\partial \eta} a^j(x + \eta h e^i, t, u(X) + \eta \tau_{i,h} u(X), Du(X) + \eta \tau_{i,h} Du(X)) d\eta = \\ &= h \frac{\partial \tilde{a}^j}{\partial x_i} + \sum_{k=1}^N (\tau_{i,h} u_k(X)) \frac{\partial \tilde{a}^j}{\partial u_k} + \sum_{r=1}^n \sum_{k=1}^N (\tau_{i,h} D_r u_k(X)) \frac{\partial \tilde{a}^j}{\partial p_k^r} \end{aligned} \tag{3.7}$$

where, if $b = b(X,u,p)$ is a vector of $\mathbb{R}^N$, for the sake of simplicity let us set

$$\tilde{b}(x) = \int_0^1 b(x + h\eta e^i, t, u(X) + \eta\tau_{i,h}u(X), Du(X) + \eta\tau_{i,h}Du(X))d\eta. \quad (3.8)$$

Therefore, from (3.6) we obtain that:

$$\begin{aligned}
&\int_Q \psi^2\rho_m \sum_{j,r=1}^{n}\sum_{k=1}^{N}\left((\tau_{i,h}D_r u_k(X))\frac{\partial\tilde{a}^j}{\partial p_k^r}|(\rho_m\tau_{i,h}D_j u) * g_s\right)dX = \\
&= -2\int_Q \psi\rho_m \sum_{j,r=1}^{n}\sum_{k=1}^{N}\left((\tau_{i,h}D_r u_k(X))\frac{\partial\tilde{a}^j}{\partial p_k^r}|D_j\psi[(\rho_m\tau_{i,h}u) * g_s]\right)dX - \\
&-\int_Q \sum_{j,r=1}^{n}\sum_{k=1}^{N}\left((\tau_{i,h}u_k(X))\frac{\partial\tilde{a}^j}{\partial u_k}|D_j\left\{\psi^2\rho_m[(\rho_m\tau_{i,h}u) * g_s]\right\}\right)dX - \quad (3.9)\\
&-h\int_Q \sum_{j=1}^{n}\left(\frac{\partial\tilde{a}^j}{\partial x_i}|D_j\left\{\psi^2\rho_m[(\rho_m\tau_{i,h}u) * g_s]\right\}\right)dX + \\
&+\int_Q \psi^2\rho_m(\tau_{i,h}u|(\rho_m\tau_{i,h}u) * g_s)dX + \\
&+\int_Q \left(B^0(X,u,Du)|\tau_{i,-h}\{\psi^2\rho_m[(\rho_m\tau_{i,h}u) * g_s]\}\right)dX
\end{aligned}$$

taking into account that

$$D_j\{\psi^2\rho_m[(\rho_m\tau_{i,h}u) * g_s]\} = \psi^2\rho_m[(\rho_m\tau_{i,h}D_j u) * g_s] + 2\psi\rho_m D_j\psi[(\rho_m\tau_{i,h}u) * g_s]$$

and that, by symmetry of the $g_s(t)$

$$\int_Q \left(\tau_{i,h}u|\psi^2\rho_m[(\rho_m\tau_{i,h}u) * g_s]'\right)dX = 0.$$

And so, from (3.9), taking the limit for $s \to +\infty$, we obtain that:

$$\begin{aligned}
&A = \int_Q \psi^2\rho_m^2 \sum_{j,r=1}^{n}\sum_{k=1}^{N}\left((\tau_{i,h}D_r u_k(X))\frac{\partial\tilde{a}^j}{\partial p_k^r}|\tau_{i,h}D_j u\right)dX = \\
&= -2\int_Q \psi\rho_m^2 \sum_{j,r=1}^{n}\sum_{k=1}^{N}\left((\tau_{i,h}D_r u_k(X))\frac{\partial\tilde{a}^j}{\partial p_k^r}|D_j\psi\,\tau_{i,h}u\right)dX - \\
&-\int_Q \sum_{j,r=1}^{n}\sum_{k=1}^{N}\left((\tau_{i,h}u_k)\frac{\partial\tilde{a}^j}{\partial u_k}|D_j(\psi^2\rho_m^2\tau_{i,h}u)\right)dX - \quad (3.10)
\end{aligned}$$

$$-h\int_Q \sum_{j=1}^{n}\left(\frac{\partial \tilde{a}^j}{\partial x_i}|D_j(\psi^2\rho_m^2\tau_{i,h}u)\right)dX +$$
$$+\int_Q \psi^2\rho_m\rho_m\left\|\tau_{i,h}u\right\|^2 dX +$$
$$+\int_Q \left(B^0(X,u,Du)|\tau_{i,-h}(\psi^2\rho_m^2\tau_{i,h}u)\}\right)dX =$$
$$= B + C + D + E + F.$$

By the assumption (1.4) and from Lemma 2.VI of [2], the integral in the left-hand side can be estimated in the following way

$$A \geq \nu(k)\int_{-b^*}^{-1/m} dt\int_{B(2\sigma)} \psi^2\rho_m^2\left\|\tau_{i,h}Du\right\|^2(1+\|Du\|^2)^{\frac{q-2}{2}}dx \geq$$
$$\geq \nu C(K,q)\int_{-b^*}^{-1/m} dt\int_{B(2\sigma)} \psi^2\rho_m^2\left\|\tau_{i,h}Du\right\|^2(1+\|Du\|+\left\|\tau_{i,h}\,Du\right\|)^{q-2}dx \tag{3.11}$$

On the other hand from (3.8) and by the assumption (1.3) it follows that

$$\sum_{k=1}^{N}\sum_{r=1}^{n}\left\|\frac{\partial \tilde{a}^j}{\partial p_k^r}\right\| \leq M(k)(1+\|Du\|+\left\|\tau_{i,h}\,Du\right\|)^{q-2} \tag{3.12}$$

$$\left\|\frac{\partial \tilde{a}^j}{\partial x_i}\right\| + \sum_{k=1}^{N}\left\|\frac{\partial \tilde{a}^j}{\partial u_k}\right\| \leq M(k)(1+\|Du\|+\left\|\tau_{i,h}\,Du\right\|)^{q-1}. \tag{3.13}$$

Then, we obtain that

$$|B| \leq c(k,q,\sigma,n)\int_{-b^*}^{-1/m} dt\int_{B(2\sigma)} \psi\rho_m^2(1+\|Du\|+\left\|\tau_{i,h}\,Du\right\|)^{q-2}\left\|\tau_{i,h}u\right\|\left\|\tau_{i,h}Du\right\|dX$$
$$\leq c(k,q,\sigma,n)\left(\int_{-b^*}^{-1/m} dt\int_{B(2\sigma)} \psi^2\rho_m^2(1+\|Du\|+\left\|\tau_{i,h}\,Du\right\|)^{q-2}\left\|\tau_{i,h}Du\right\|^2 dx\right)^{\frac{1}{2}}\cdot$$
$$\cdot\left(\int_{-b^*}^{-1/m} dt\int_{B(2\sigma)} \psi^2\rho_m^2(1+\|Du\|+\left\|\tau_{i,h}\,Du\right\|)^{q-2}\left\|\tau_{i,h}u\right\|^2 dx\right)^{\frac{1}{2}}$$

and from this inequality it follows, $\forall\varepsilon > 0$, that

$$| B | \leq \frac{\varepsilon}{3} \int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} \psi^2 \rho_m^2 (1 + \|Du\| + \|\tau_{i,h} Du\|)^{q-2} \|\tau_{i,h} Du\|^2 dx + \tag{3.14}$$

$$+ c(k,q,\sigma,n,\varepsilon) \int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} (1 + \|Du\| + \|\tau_{i,h} Du\|)^q \|\tau_{i,h} u\|^2 dx$$

Analogously, we have

$$| C | \leq c(k,q,\sigma,n) \int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} (1 + \|Du\| + \|\tau_{i,h} Du\|)^{q-1} \|\tau_{i,h} u\| \cdot$$

$$\cdot (\psi^2 \rho_m^2 \|\tau_{i,h} Du\| + c(\sigma) \psi \rho_m^2 \|\tau_{i,h} u\|) dx \leq$$

$$\leq c(k,q,\sigma,n) \int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} (1 + \|Du\| + \|\tau_{i,h} Du\|)^{q-1} \psi^2 \rho_m^2 \|\tau_{i,h} u\| \|\tau_{i,h} Du\|^2 dx +$$

$$+ c(k,q,\sigma,n) \int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} (1 + \|Du\| + \|\tau_{i,h} Du\|)^{q-1} \psi \rho_m^2 \|\tau_{i,h} u\|^2 dx \leq$$

$$\leq c(k,q,\sigma,n) \left(\int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} \psi^2 \rho_m^2 (1 + \|Du\| + \|\tau_{i,h} Du\|)^{q-2} \|\tau_{i,h} Du\|^2 dx \right)^{\frac{1}{2}} \cdot$$

$$\cdot \left(\int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} (1 + \|Du\| + \|\tau_{i,h} Du\|) \|\tau_{i,h} u\|^2 dx \right)^{\frac{1}{2}} +$$

$$+ c(k,q,\sigma,n) \int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} \psi^2 \rho_m^2 (1 + \|Du\| + \|\tau_{i,h} Du\|)^q \|\tau_{i,h} u\|^2 dx.$$

Then, $\forall \varepsilon > 0$ it follows:

$$| C | \leq \frac{\varepsilon}{3} \int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} \psi^2 \rho_m^2 (1 + \|Du\| + \|\tau_{i,h} Du\|)^{q-2} \|\tau_{i,h} Du\|^2 dx + \tag{3.15}$$

$$+ c(k,q,\sigma,n,\varepsilon) \int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} (1 + \|Du\| + \|\tau_{i,h} Du\|)^q \|\tau_{i,h} u\|^2 dx.$$

Moreover, by the assumption (1.3) and from lemma 2.I we obtain that

$$
\begin{aligned}
|D| \le c(k,n,q)\,|h| \int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} (1+\|Du\|+\|\tau_{i,h}\,Du\|)^{q-1} \cdot \\
\cdot (\psi^2\rho_m^2\,\|\tau_{i,h}Du\| + c(\sigma)\psi\rho_m^2\,\|\tau_{i,h}u\|)dx \le \\
\le c(k,n,q)\,|h| \left(\int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} \psi^2\rho_m^2(1+\|Du\|+\|\tau_{i,h}\,Du\|)^{q-2}\,\|\tau_{i,h}Du\|^2\,dx\right)^{\frac{1}{2}} \cdot \\
\cdot \left(\int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} \psi^2\rho_m^2(1+\|Du\|+\|\tau_{i,h}\,Du\|)^{q}\,dx\right)^{\frac{1}{2}} + \\
+ c(k,\sigma,n,q)\,|h| \left(\int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} \psi\rho_m^2\,\|\tau_{i,h}\,u\|^q\,dx\right)^{\frac{1}{q}} \cdot \\
\cdot \left(\int_{-b^*}^{-1/m} dt \int_{B(2\sigma)}^{2} (1+\|Du\|+\|\tau_{i,h}\,Du\|)^{q}\,dx\right)^{\frac{q-1}{q}}.
\end{aligned}
$$

Then, $\forall \varepsilon > 0$ it follows that

$$
\begin{aligned}
|D| \le \varepsilon \int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} \psi^2\rho_m^2(1+\|Du\|+\|\tau_{i,h}\,Du\|)^{q-2}\,\|\tau_{i,h}Du\|^2\,dx + \\
+ c(k,q,n,\varepsilon)\,|h|^2 \int_{-b^*}^{-1/m} dt \int_{B(3\sigma)} (1+\|Du\|)^q\,dx + \\
+ c(k,q,n,\sigma)\,|h|^2 \left(\int_{-b^*}^{-1/m} dt \int_{B(3\sigma)} (1+\|Du\|)^q\,dx\right)^{\frac{1}{q}} \cdot \qquad (3.16) \\
\cdot \left(c(q)\int_{-b^*}^{-1/m} dt \int_{B(3\sigma)} (1+\|Du\|)^q\,dx\right)^{\frac{q-1}{q}} \le \\
\le \varepsilon \int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} \psi^2\rho_m^2(1+\|Du\|+\|\tau_{i,h}\,Du\|)^{q-2}\,\|\tau_{i,h}\,Du\|^2\,dx + \\
+ c(k,\sigma,n,q,\varepsilon)\,|h|^2 \int_{-b^*}^{-1/m} dt \int_{B(3\sigma)} (1+\|Du\|)^q\,dx.
\end{aligned}
$$

Moreover we have

$$
\begin{aligned}
|E| &= \int_Q \psi\rho_m\rho_m\,\|\tau_{i,h}u\|^2\,dx \le \\
&\le |h|^2 \int_{-b^*}^{-a} dt \int_{B(2\sigma)} \psi^2\rho_m\rho_m\,\|Du\|^2\,dx \le \qquad (3.17) \\
&\le \frac{|h|^2}{b-a} \int_{-b^*}^{-a} dt \int_{B(2\sigma)} (1+\|Du\|)^q\,dx
\end{aligned}
$$

taking into account that

$$\rho_m f_m \begin{cases} \le 0 & \text{if} \quad \dfrac{-2}{m} \le t \le -\dfrac{1}{m} \\ = 0 & \text{if } t \le -b \text{ or } t \ge -\dfrac{1}{m} \text{ or } -a \le t \le -\dfrac{2}{m} \\ \le \dfrac{1}{b-a} & \text{if} \quad -b \le t \le -a \end{cases}$$

By the assumption (1.2), moreover we have

$$\begin{aligned} &| F | \le \int_Q \left\| B^0(X,u,Du) \right\| \left\| \tau_{i,-h}(\psi^2 \rho_m^2 \tau_{i,h} u) \right\| dX \le \\ &\le c(k,q) \int_{-b^*}^{-1/m} dt \int_{B(\frac{5}{2}\sigma)} (1 + \|Du\|^2)^{\frac{q}{2}} \left\| \tau_{i,-h}(\psi^2 \tau_{i,h} u) \right\| dx. \end{aligned} \tag{3.18}$$

From (3.10)-(3.18), with $\varepsilon = \nu 6$ in (3.14), (3.15), (3.16), it follows, for each integer i, $1 \le i \le n$, and for each $|h| < \min\left\{1, \dfrac{\sigma}{2}\right\}$

$$\begin{aligned} &\frac{\nu}{2} c(k,q) \int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} \psi^2 \rho_m^2 \left\| \tau_{i,h} Du \right\|^2 (1 + \|Du\| + \left\| \tau_{i,h} Du \right\|)^{q-2} dx \le \\ &\le c(k,\sigma,q,a,b,n,\nu) \, | h |^2 \int_{-b^*}^{-1/m} dt \int_{B(3\sigma)} (1 + \|Du\|)^q dx + \\ &+ c(k,\sigma,q,n,\nu) \int_{-b^*}^{-1/m} dt \int_{B(2\sigma)} \left\| \tau_{i,h} u \right\|^2 (1 + \|Du\| + \left\| \tau_{i,h} Du \right\|)^q dx + \\ &+ \int_{-b^*}^{-1/m} dt \int_{B(\frac{5}{2}\sigma)} (1 + \|Du\|^2)^{\frac{q}{2}} \tau_{i,-h}(\psi^2 \tau_{i,h} u) dx. \end{aligned} \tag{3.19}$$

Let us consider now the last integral that appears at the right hand side of (3.19). From Theorem 3.III of [4] (with $\sigma_0 = 3\sigma, a = b^*$) we deduce that

$$u \in L^q(-b^*, 0, H^{1+\theta,q}(B(\frac{5}{2}\sigma), \mathbb{R}^N)), \quad \forall \theta \in \left(0, \frac{2}{q}\right) \tag{3.20}$$

and

$$\int_{-b^*}^{0} | Du |_{0,q,B(\frac{5}{2}\sigma)}^q \, dt \le \tag{3.21}$$

$$\leq c(\nu,k,U,\theta,\lambda,\sigma,q,a,b,n)\int_{-b}^{0} dt \int_{B(3\sigma)} (1+\|Du\|)^q dx$$

hence, thanks also to the assumption $u \in C^{0,\lambda}(\overline{Q},\mathbb{R}^N)$, it results for a.e. $t \in (-b^*,0)$

$$u(x,t) \in H^{1+\theta,q}(B(\frac{5}{2}\sigma),\mathbb{R}^N) \cap C^{0,\lambda}(B(\frac{5}{2}\sigma),\mathbb{R}^N), \quad \forall \theta \in \left(0,\frac{2}{q}\right).$$

From Lemma 2.3 (with $\Omega = B(\frac{5}{2}\sigma)$ and $\theta = 1-\frac{\lambda}{2}$) we get for a.e. $t \in (-b^*,0)$

$$u(x,t) \in W^{1,p}(\Omega,\mathbb{R}^N) \quad \text{where } p = 2q + \frac{2q^2\lambda}{2n-\lambda q} \tag{3.22}$$

and

$$\|u\|_{1,p,B(\frac{5}{2}\sigma)} \leq c(\lambda,\sigma,n)\|u\|^{\frac{1}{2}}_{2-\frac{\lambda}{2},q,B(\frac{5}{2}\sigma)} \|u\|^{\frac{1}{2}}_{C^{0,\lambda}(B(\frac{5}{2}\sigma),R^N)}. \tag{3.23}$$

Now, since $p > 2q$, we obtain

$$W^{1,p}(B(\frac{5}{2}\sigma),\mathbb{R}^N) \subset W^{1,2q}(B(\frac{5}{2}\sigma),\mathbb{R}^N)$$

and this is an algebraic and topological inclusion; from which by (3.22) and (3.23), it follows for a.e. $t \in \left(-b^*,-\frac{1}{m}\right)$ that

$$u(x,t) \in W^{1,2q}(B(\frac{5}{2}\sigma),\mathbb{R}^N) \tag{3.24}$$

and

$$\begin{aligned} &\|u\|^{2q}_{1,2q,B(\frac{5}{2}\sigma)} \leq \\ &\leq c(k,U,\lambda,\sigma,n)\|u\|^{q}_{2-\frac{\lambda}{2},q,B(\frac{5}{2}\sigma)} \leq \\ &\leq c(k,U,\lambda,\sigma,n)\left\{1+|u|^q_{1,q,B(\frac{5}{2}\sigma)} + |Du|^q_{1-\frac{\lambda}{2},q,B(\frac{5}{2}\sigma)}\right\}. \end{aligned} \tag{3.25}$$

This estimate holds in particular for a.e. $t \in \left(-b^*, -\frac{1}{m}\right)$; for such t therefore we obtain, $\forall \varepsilon > 0$

$$
\begin{aligned}
&c(k)\int_{B(\frac{5}{2}\sigma)} (1+\|Du\|^2)^{\frac{q}{2}} \left\|\tau_{i,-h}(\psi^2 \tau_{i,h} u)\right\| dx \leq \\
&\leq \left(\int_{B(\frac{5}{2}\sigma)} |\, h\,|^{-2} \left\|\tau_{i,-h}(\psi^2 \tau_{i,h} u)\right\|^2 dx\right)^{\frac{1}{2}} \left(c(k)\int_{B(\frac{5}{2}\sigma)} |\, h\,|^2 \,(1+\|Du\|^2)^q dx\right)^{\frac{1}{2}} \leq \\
&\leq \frac{\varepsilon}{2} |\, h\,|^{-2} \int_{B(\frac{5}{2}\sigma)} \left\|\tau_{i,-h}(\psi^2 \tau_{i,h} u)\right\|^2 dx + c(k,\varepsilon) |\, h\,|^2 \int_{B(\frac{5}{2}\sigma)} (1+\|Du\|^2)^q dx \leq \\
&\leq \frac{\varepsilon}{2} \int_{B(2\sigma)} \left\|D(\psi^2 \tau_{i,h} u)\right\|^2 dx + c(k,\sigma,\varepsilon) |\, h\,|^2 \left\{1 + \int_{B(\frac{5}{2}\sigma)} \|Du\|^{2q} dx\right\} \leq \\
&\leq \varepsilon \int_{B(2\sigma)} \psi^4 \left\|\tau_{i,h} Du\right\|^2 dx + c(\sigma,\varepsilon) \int_{B(2\sigma)} \psi^2 \left\|\tau_{i,h} u\right\|^2 dx + \\
&+ c(k,\sigma,\varepsilon) |\, h\,|^2 \left\{1 + \int_{B(\frac{5}{2}\sigma)} \|Du\|^{2q} dx\right\} \leq \\
&\leq \varepsilon \int_{B(2\sigma)} \psi^2 \left\|\tau_{i,h} Du\right\|^2 (1 + \|Du\| + \left\|\tau_{i,h}\, Du\right\|)^{q-2} dx + \\
&+ c(\sigma,\varepsilon) \int_{B(2\sigma)} \psi^2 \left\|\tau_{i,h} u\right\|^2 (1 + \|Du\| + \left\|\tau_{i,h}\, Du\right\|)^q dx + \\
&+ c(k,\sigma,\varepsilon) |\, h\,|^2 \left\{1 + \|u\|^{2q}_{1,2q,B(\frac{5}{2}\sigma)}\right\}.
\end{aligned}
$$

From this, for $\varepsilon = \frac{\nu}{4}$, it follows that

$$
\begin{aligned}
&c(k)\int_{B(\frac{5}{2}\sigma)} (1+\|Du\|^2)^{\frac{q}{2}} \left\|\tau_{i,-h}(\psi^2 \tau_{i,h} u)\right\| dx \leq \\
&\leq \frac{\nu}{4} \int_{B(2\sigma)} \psi^2 \left\|\tau_{i,h} Du\right\|^2 (1 + \|Du\| + \left\|\tau_{i,h}\, Du\right\|)^{q-2} dx + \\
&+ c(\sigma,\nu) \int_{B(2\sigma)} \psi^2 \left\|\tau_{i,h} u\right\|^2 (1 + \|Du\| + \left\|\tau_{i,h}\, Du\right\|)^q dx + \\
&+ c(k,\sigma,\nu) |\, h\,|^2 \left\{1 + \|u\|^{2q}_{1,2q,B(\frac{5}{2}\sigma)}\right\}
\end{aligned}
$$

and, from which, by multiplicating both members for ρ_m^2 and by integrating with respect to t in $\left(-b^*, -\frac{1}{m}\right)$ we deduce

$$
c(k)\int_{-b^*}^{-\frac{1}{m}} \rho_m^2 dt \int_{B(\frac{5}{2}\sigma)} (1+\|Du\|^2)^{\frac{q}{2}} \left\|\tau_{i,-h}(\psi^2 \tau_{i,h} u)\right\| dx \leq
$$

$$\leq \frac{\nu}{4}\int_{-b^*}^{-\frac{1}{m}} dt \int_{B(2\sigma)} \psi^2 \left\|\tau_{i,h} Du\right\|^2 \rho_m^2 (1+\|Du\|+\left\|\tau_{i,h}\, Du\right\|)^{q-2} dx + \tag{3.26}$$

$$+c(\sigma,\nu)\int_{-b^*}^{-\frac{1}{m}} dt \int_{B(2\sigma)} \left\|\tau_{i,h} u\right\|^2 (1+\|Du\|+\left\|\tau_{i,h}\, Du\right\|)^q dx +$$

$$+c(\nu,k,U,\lambda,\sigma,n)\mid h\mid^2 \int_{-b^*}^{-\frac{1}{m}} \left\{1+\|u\|^{2q}_{1,2q,B(\frac{5}{2}\sigma)}\, dx\right\} dt.$$

Let us consider the penultimate integral that appears at the right hand side of (3.26) and (3.19). Using the Hölder inequality and thanks to Lemma 2.I and the (3.25), we have, for a.e. $t\in(-b^*,0)$, that[2]

$$\int_{B(2\sigma)} (1+\|Du\|+\left\|\tau_{i,h}\, Du\right\|)^q \left\|\tau_{i,h} u\right\|^2 dx \leq$$

$$\leq \left(\int_{B(2\sigma)} (1+\|Du\|+\left\|\tau_{i,h}\, Du\right\|)^p dx\right)^{\frac{q}{p}} \left(\int_{B(2\sigma)} \left\|\tau_{i,h} u\right\|^{\frac{4n}{n+q(\theta+\lambda-1)}} dx\right)^{\frac{n+q(\theta+\lambda-1)}{2n}} \leq$$

$$\leq c(q)\left\{1+\left(\int_{B(\frac{5}{2}\sigma)} \|Du\|^p\, dx\right)\right\}^{\frac{q}{p}} \mid h\mid^2 \left(\int_{B(\frac{5}{2}\sigma)} \mid Du\mid^{\frac{4n}{n+q(\theta+\lambda-1)}} dx\right)^{\frac{n+q(\theta+\lambda-1)}{2n}} \leq$$

$$\leq c(n,q,\sigma,\theta,\lambda)\mid h\mid^2 \left\{1+\left(\int_{B(\frac{5}{2}\sigma)} \|Du\|^p\, dx\right)\right\}^{\frac{q}{p}} \left\{1+\left(\int_{B(\frac{5}{2}\sigma)} \|Du\|^p\, dx\right)\right\}^{\frac{2}{p}} \leq$$

$$\leq c(q,\sigma,\theta,n,\lambda)\mid h\mid^2 \left\{1+\mid Du\mid_{0,p,B(\frac{5}{2}\sigma)}\right\}^{q+2} \leq$$

$$\leq c(n,q,\sigma,\theta,\lambda)\mid h\mid^2 \left\{1+\|u\|^{2q}_{1,p,B(\frac{5}{2}\sigma)}\right\} \leq$$

$$\leq c(n,q,\lambda,\theta,\sigma)\mid h\mid^2 \left\{1+\mid u\mid^q_{1,q,B(\frac{5}{2}\sigma)} + \mid Du\mid^q_{1-\frac{\lambda}{2},q,B(\frac{5}{2}\sigma)}\right\}$$

from which, by integrating with respect to t in $(-b^*,0)$ we have

$$\int_{-b^*}^{0} dt \int_{B(2\sigma)} (1+\|Du\|+\left\|\tau_{i,h}\, Du\right\|)^q \left\|\tau_{i,h} u\right\|^2 dx \leq \tag{3.27}$$

$$\leq c(q,\sigma,n,\theta,\lambda,k,U,a,b)\mid h\mid^2 \left\{1+\int_{-b^*}^{0} \left(\mid u\mid^q_{1,q,B(\frac{5}{2}\sigma)} + \mid Du\mid^q_{1-\frac{\lambda}{2},q,B(\frac{5}{2}\sigma)}\right) dt\right\}$$

From (3.19), (3.25), (3.26), (3.27) and (3.21) (for $\theta = 1-\frac{\lambda}{2}$) we deduce, for each integer i, $1\leq i\leq n$, and for each $|h|<\min\{1,\frac{\sigma}{2}\}$, taking the limit as

[2] $\frac{4n}{n+q(\theta+\lambda-1)} = \frac{2p}{p-q} < p$ since $2 < p-q$

$m \to \infty$, we get

$$\frac{\nu}{4} c(k,q) \int_{-a}^{0} dt \int_{B(\sigma)} \left\| \tau_{i,h} Du \right\|^2 (1 + \|Du\| + \left\| \tau_{i,h} Du \right\|)^{q-2} dx \leq$$

$$\leq c(q,\sigma,n,\nu,\lambda,k,U,a,b) \mid h \mid^2 \left\{ 1 + \int_{-b}^{0} \mid u \mid^q_{1,q,B(3\sigma)} dt \right\} \tag{3.28}$$

and then, $\forall h$ such that $\mid h \mid < \min\left\{1, \frac{\sigma}{2}\right\}$ we have

$$\int_{-a}^{0} dt \int_{B(\sigma)} \left\| \tau_{i,h} Du \right\|^2 dx \leq$$

$$\leq c(q,\sigma,n,\nu,\lambda,k,U,a,b) \mid h \mid^2 \left\{ 1 + \int_{-b}^{0} \mid u \mid^q_{1,q,B(3\sigma)} dt \right\} \tag{3.29}$$

The estimate (3.29) is trivial if $\min\left\{1, \frac{\sigma}{2}\right\} < h < \sigma$ and then (3.29) will be true for each integer i, $1 \leq i \leq n$ and for each $\mid h \mid < \sigma$.
From (3.29) and by lemma 2.2 we have that

$$Du \in L^2(-a,0,H^{1,2}(B(\sigma),\mathbb{R}^N)) \tag{3.30}$$

and then

$$u \in L^2(-a,0,H^2(B(\sigma),\mathbb{R}^N)) \tag{3.31}$$

and moreover

$$\int_{-a}^{0} \mid u \mid^2_{2,B(\sigma)} dt \leq c(\nu,k,U,\sigma,a,b,q,n) \left\{ 1 + \int_{-b}^{0} \mid u \mid^q_{1,q,B(3\sigma)} dt \right\} \tag{3.32}$$

It remains to show that $u \in H^1(-a,0,L^2(B(\sigma),\mathbb{R}^N))$ and that the relative estimate holds. From (3.25) it follows, for a.e. $t \in (-a,0)$

$$\int_{B(\sigma)} \left\| D_i u \right\|^{2q} dx \leq c(k,U,\lambda,\sigma,n) \left\{ 1 + \mid u \mid^q_{1,q,B(\frac{5}{2}\sigma)} + \mid Du \mid^q_{1-\frac{1}{2},q,B(\frac{5}{2}\sigma)} \right\}$$

$i = 1,2,...,n$, from which and from (3.21), by integrating with respect to t in $(-a,0)$ we deduce:

$$D_i u \in L^{2q}(B(\sigma)\times(-a,0),\mathbb{R}^N), \quad i=1,2,..,n.$$

and

$$\int_{-a}^{0} dt \int_{B(\sigma)} \|Du\|^{2q}\, dx \le c(\nu,k,U,\lambda,\sigma,n,a,b)\left\{1+\int_{-b}^{0} |\, u\,|^{q}_{1,q,B(3\sigma)}\, dt\right\} \quad (3.33)$$

Now, by assumption (1.2)

$$\left\|B^0(X,u,Du)\right\| \le M(k,q)(1+\|Du\|^q)$$

and then, from (3.33) we deduce that

$$B^0(X,u,Du) \in L^2(B(\sigma)\times(-a,0),\mathbb{R}^N) \quad (3.34)$$

and

$$\int_{-a}^{0} dt \int_{B(\sigma)} \left\|B^0(X,u,Du)\right\|^2 dx \le c(k,q)\int_{-a}^{0} dt \int_{B(\sigma)} (1+\|Du\|^{2q})dx. \quad (3.35)$$

On the other hand, by assumption (1.3) we have that:

$$D_i a^i(X,u,Du) \in L^2(B(\sigma)\times(-a,0),\mathbb{R}^N), \quad i=1,2,..,n \quad (3.36)$$

and that:

$$\begin{aligned} &\int_{-a}^{0} dt \int_{B(\sigma)} \sum_{i=1}^{n} \left\|D_i a^i(X,u,Du)\right\|^2 dx \le \\ &\le c(k,n,q)\int_{-a}^{0} dt \int_{B(\sigma)} \left[1+\|Du\|^{2q}+\sum_{i,j=1}^{n}\left\|D_{i,j}u\right\|^2\right] dx. \end{aligned} \quad (3.37)$$

Now, taking into account that u is a solution in Q (and then in $B(\sigma)\times(-a,0)$) of the system (1.1), we deduce that, for each $\varphi \in C_0^\infty(B(\sigma)\times(-a,0),\mathbb{R}^N)$

$$\int_{-a}^{0} dt \int_{B(\sigma)} \left(u \left| \frac{\partial \varphi}{\partial t} \right. \right) dx =$$

$$= -\int_{-a}^{0} dt \int_{B(\sigma)} \left(\left(\sum_{i=1}^{n} D_i a^i(X, u, Du) + B^0(X, u, Du) \right) \middle| \varphi \right) dx$$

from which, by (3.34) and (3.36), it results that

$$\exists \frac{\partial u}{\partial t} \in L^2(B(\sigma) \times (-a, 0), \mathbb{R}^N) \tag{3.38}$$

and from (3.35), (3.37) it follows that

$$\int_{-a}^{0} dt \int_{B(\sigma)} \left\| \frac{\partial u}{\partial t} \right\|^2 dx \leq$$

$$\leq c(k, n) \int_{-a}^{0} dt \int_{B(\sigma)} \left[1 + \|Du\|^{2q} + \sum_{i,j=1}^{n} \left\| D_{i,j} u \right\|^2 \right] dx$$

and then, by (3.32), (3.33) we deduce that

$$\int_{-a}^{0} dt \int_{B(\sigma)} \left\| \frac{\partial u}{\partial t} \right\|^2 dx \leq$$

$$\leq c(\nu, k, U, \lambda, \sigma, a, b, n) \left\{ 1 + \int_{-b}^{0} | u |_{1,q,B(3\sigma)}^{q} \, dt \right\} \tag{3.39}$$

Finally we deduce (3.1) and (3.2) from (3.31), (3.32), (3.38), (3.39)[3]

REFERENCES

[1] S. Campanato, Sistemi ellittici in forma di divergenza. Regolarità all'interno, Quaderni Scuola Norm. Sup. Pisa, 1980

[2] S. Campanato, Differentiability of the solutions of nonlinear elliptic systems with natural growth, Annali di Matematica Pura e Applicata, Serie Quarta, Tomo CXXXI, 1982.

[3] L. Fattorusso, Sulla differenziabilità delle soluzioni di sistemi parabolici non lineari del secondo ordine ad andamento quadratico, Bollettino U.M.I. (7) 1.B (1987), 741–764.

[3] Theorem 3.1 can be proved by substituting (1.3) and (1.4) with the monotony assumptions (1.3), (1.4) of [4] and by following the technique in [4] instead of the one in [3]

[4] L. Fattorusso and M. Marino, Differenziabilità locale per sistemi parabolici non lineari del secondo ordine con non linearità $q \geq 2$, Ricerche di Matematica Vol. XLI, fase 1°, (1992), 89–112.

[5] L. Fattorusso, Differenziabilità locale per sistemi parabolici non lineari del secondo ordine con non linearità $1 < q < 2$, Le Matematiche, Vol. XLVIII (1993).

[6] M. Marino and M. Maugeri, Differentiability of weak solutions of non linear parabolic systems with quadratic growth, Le Matematiche, Vol. L (1995), Fase II, pp. 361–377.

AN OPTIMIZATION PROBLEM WITH AN EQUILIBRIUM CONSTRAINT IN URBAN TRANSPORT

P. Ferrari
University School of Engineering of Pisa, Pisa, Italy

Abstract: The paper presents a study of transport in urban areas served by a public transport system, as well as by private vehicles on which road pricing is imposed. It is supposed that the road pricing fare, the ticket price and the frequency of the lines of public transport are established by the Public Administration in such a way that the surplus of users of both the transport modes is maximised, under the conditions that the system is in equilibrium, the budget constraint of the company managing public transport is satisfied, and the private transport demand does not exceed a given threshold for environmental reasons. The theoretical model that has been devised leads to a problem of nonlinear programming, with an equilibrium constraint formulated as a fixed point problem. From an application of the model to an urban area it emerges that, if the proceeds of road pricing are used for financing public transport, the results of road pricing essentially depend on the proportion of demand that is captive to public transport, and on the level of congestion existing on the urban road network before the imposition of road pricing.

Key words: Nonlinear programming. Equilibrium constraint. Road pricing. Urban transport.

1. INTRODUCTION.

Consider (fig. 1) a pair (i,k) of nodes connected by a road. The transport demand x from i to k is constituted by private cars, and is a function of

transport cost y. $x(y)$ is the *demand function*, and $y(x)$ is the *inverse demand function*. $c(x)$ is the *cost function* of the road connecting i and k, which furnishes the cost borne by each driver as a function of traffic volume x on the road. The abscissa $\bar{x}$ of the intersection S of $y(x)$ and $c(x)$ is the *equilibrium demand*. The benefit that users obtain from the transport system at equilibrium is measured by their *surplus*, that is the area under the curve $y(x)$ and above the horizontal US. The demand $\hat{x}$ that maximizes the user surplus is less than $\bar{x}$, and it is the abscissa of the intersection R of $y(x)$ and $C(x)$, the *marginal cost function*, which furnishes the derivative of the total cost borne by users with respect to x, as a function of x. The system does not reach $\hat{x}$ spontaneously, because it is not an equilibrium demand; but it can be transformed into the latter if an additional cost RT is imposed, as *road pricing*, on drivers travelling the road, so that the surplus is given by the area above the horizontal VT. In this way the optimal social welfare is reached, but at the cost of a transfer of money from road users to society as a whole, which is measured by the area $VTRZ$ in fig. 1. A similar situation takes place if one imposes road pricing in order to reduce the demand below the equilibrium value, so that the environmental damages due to traffic remain within acceptable limits. Even in this case there is an increase in social welfare, due to the reduction of traffic pollution, but at expense of drivers who have to pay the amount of road pricing.

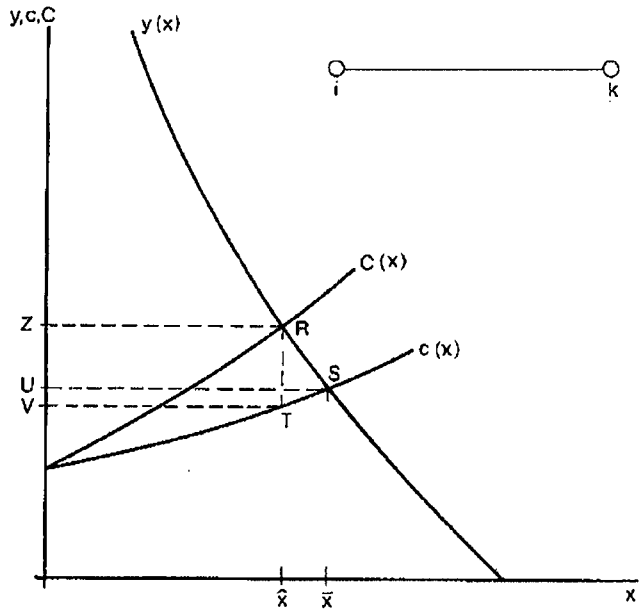


Fig. 1 – The impact of road pricing on users' welfare

These results, illustrated with reference to a very simple example, are common to all road networks traveled by private cars, and show that road pricing increases the welfare of society as a whole, but reduces the welfare of road users (see e.g. Ferrari, 1999; Hearn and Ramana, 1998; Hearn and Yildirim, 2000).

However urban transport systems are constituted in general by two modes of transport, private vehicles and public transport. Moreover transport demand is formed in part by people who have a private vehicle at their disposal, and thus are *free* to choose between the two modes of transport; and in part by people who are *captive* to public transport: these may be city dwellers, and/or visitors arriving by train, coach or plane. Thus it is possible to use the revenue from road pricing for increased funding to public transport, thereby decreasing the cost borne by its users. Such reductions can be obtained, in part, by improving the quality of service, for instance by increasing the frequency of buses, which will reduce waiting times at bus stops, and, in part, by lowering fares.

Were road pricing proceeds used in this way, it would important to know if there are situations in which road pricing imposition could give rise, not only to an increase in social welfare, but even to an increase in welfare of urban transport users, that is to a reduction of the overall transport cost borne by the users of both the transport modes, inclusive of the road pricing cost.

The evaluation of the impact of road pricing on urban transport cost, when its revenue is used for financing public transport, is the purpose of this paper. It is supposed that the road pricing fare, the ticket price and the frequency of the lines of public transport are established by the Public Administration in such a way that the surplus of users of both transport modes, inclusive of the cost of road pricing, is maximized, under the conditions that the system is in equilibrium, the budget constraint of the company managing public transport is satisfied, and the private transport demand does not exceed a threshold suitable to maintain congestion and pollution due to traffic within acceptable limits. The theoretical model that has been used leads to a problem of nonlinear programming, with an equilibrium constraint which is formulated as a fixed point problem.

The paper is organized as follows. Section 2 presents a model of urban transport devised with the specific purpose of answering the questions posed in the paper. Section 3 deals with a method for solving the problem formulated in the model, which is applied to an urban area, considering various combinations of factors on which the functioning of urban transport depends. The results so obtained are discussed in Section 4. Some final considerations are advanced in Section 5.

2. A MODEL OF URBAN TRANSPORT SYSTEM.

Let us consider a square urban area, with sides of length b, served by two means of transport, private automobiles and city buses. Let A be the potential demand for transport, that is, the mean number of trips that would be made between various points in the area over the course of a day if transport costs were nil, or in any event, were perceived as such by users. A is the sum of two terms, A_1 and A_2. A_1 is the potential demand of those who do not have a private auto at their disposal, and who are therefore *captive* to public transport. A_2 is the potential demand of those who instead do have a private vehicle, and are therefore *free* to choose between the two alternative means of transport.

The actual demand, that is the average number of trips that actually are made during a day, is a function of the transport costs perceived by users in each of the two categories:

$$d_1 = A_1 \exp(-T_1 Y_1) \qquad d_2 = A_2 \exp(-T_2 Y_2) \tag{1}$$

where d_1 and d_2 are respectively the daily transport demands of captive and free users, Y_1 the mean cost of a journey to captive users, Y_2 the *inclusive* cost[1] of the two means of transport as perceived by free users, and T_1 and T_2 are two parameters determining the elasticity of demand, which is equal to $T_1 Y_1$ and $T_2 Y_2$ respectively.

From (1) we derive the expressions for the inverse functions of demand:

$$Y_1 = -\frac{1}{T_1} \ln \frac{d_1}{A_1} \qquad Y_2 = -\frac{1}{T_2} \ln \frac{d_2}{A_2} \tag{2}$$

while the user *surplus*, which measures the benefits to the entire set of potential users from the supply of transport in the area when actual demand takes on the values d_1 and d_2, is:

$$S = \int_0^{d_1} Y_1(x)dx - d_1 Y_1(d_1) + \int_0^{d_2} Y_2(x)dx - d_2 Y_2(d_2) = \frac{d_1}{T_1} + \frac{d_2}{T_2} \tag{3}$$

We assume that journeys are distributed uniformly throughout the area, and that the arrangement of the bus lines is represented by the square grid in fig. 2, proposed by Holroyd, 1965, where a is the distance between the bus lines, and ν the bus frequency, which is equal for all lines. This grid is the

[1] The inclusive cost (Domencich and McFadden, 1975,p. 75) is the opposite in sign of the average of the maximum utility values attributed by users to the two means of transport.

optimal network among all rectangular configurations of linear routes for a uniform distribution of origins and destinations (Newell 1979).

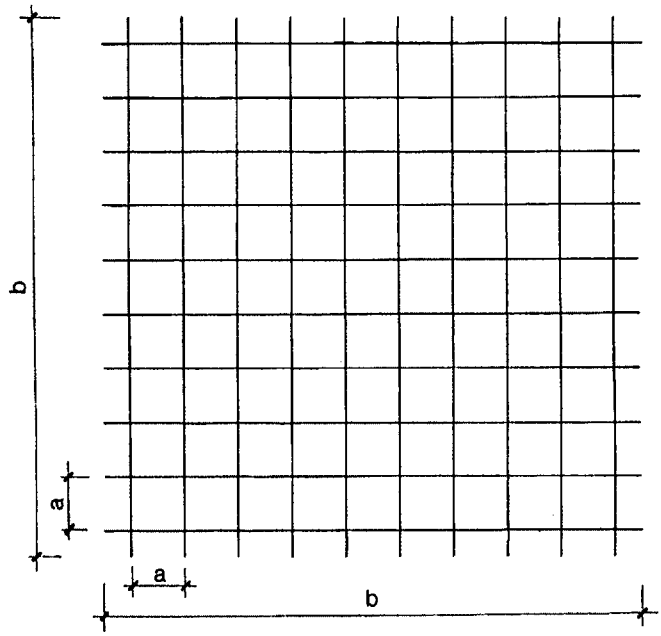


Fig. 2 – A square grid of linear bus routes

Let G be the average cost incurred by the company managing public transport per unit distance travelled by a bus. The number of lines is $2b/a$, and the distance travelled by a bus during a return journey is $2b$, by which, if H is the daily duration of service, the cost to the firm per day is:

$$C = 4\frac{b^2}{a}HGv \tag{4}$$

The public service company defrays this cost by utilising the proceeds from the bus ticket price, the road pricing and the per diem external funding TR received from the Public Administration. Let X_1 and X_2 be the ticket price and the road pricing paid on average for each journey by users of public transport and by private motorists, respectively. If d_1 is the number of daily journeys made by captive users and d_2 those made by free users, of which d_2^b by bus, in order for the bus company to balance its budget, it must hold that:

$$\left(d_1 + d_2^b\right)X_1 + \left(d_2 - d_2^b\right)X_2 + TR = C \tag{5}$$

Substituting (4) into (5), we obtain the following expression for the frequency that satisfies the company's budget:

$$\nu = \frac{\left(d_1 + d_2^b\right)X_1 + \left(d_2 - d_2^b\right)X_2 + TR}{4\dfrac{b^2}{a}HG} \tag{6}$$

The cost of a journey on each of the two means of transport as perceived by users is a random variable distributed around a central value that we assume to be a linear combination of the average attributes of the two transport means. Expressing the cost in monetary terms, the coefficients of this linear combination represent the mean monetary value of each unit attribute.

If buses arrive stops at regular intervals, while passengers arrive randomly, the mean waiting time at a bus stop is $w = 1/(2\nu)$. Let p^a and p^b be the average walking distance covered for each trip by users of private and public transport, respectively. We assume p^b to be equal to the mean distance from the journey's point of departure to the bus stop plus the average distance from the arrival bus stop to destination: $p^b = a/2$.

Having assumed that travel destinations, independently of their origins, are uniformly distributed throughout the area, it is easy to verify that the mean trip length is equal to $2b/3$ (Newell, 1979). This is assumed to be equal (with a good degree of approximation) to the average distance travelled aboard a vehicle by both users of public transport and private motorists.

The burden of transfer from one bus line to another is generally perceived by users as an increase τ in riding time of between 5 and 10 minutes (Newell, 1979). Given V, the mean velocity of a bus, the cost of transferring is equal to an increase of $\tau \cdot V$ in the length of bus trip.

Let G_1 be the mean monetary value of a unit of waiting time, G_2 that of a unit distance travelled on foot, G_3 that of a unit distance travelled aboard a bus, exclusive of ticket price, and G_4 that of a unit distance travelled in a private auto, including both travel time and vehicle operating costs.

G_3 and G_4 depend on the level of congestion of the road network, because congestion increases the journey times and causes stress and discomfort to users. By denoting $d_2^a = d_2 - d_2^b$ the daily demand by private car, we assume:

$$G_3 = G^b\left[1+\alpha\left(\frac{d_2^a}{K}\right)^\beta\right] \qquad G_4 = G^a\left[1+\alpha\left(\frac{d_2^a}{K}\right)^\beta\right] \tag{7}$$

where α, β, K are parameters that depend on the geometric and operating characteristics of the road network, G^a and G^b are the mean monetary values of a unit distance travelled in a private auto and in a bus respectively, in free flow conditions. Thus the mean cost of a journey via private transport is:

$$C^a = G_2 p^a + \frac{2}{3} G_4 b + X_2 \tag{8}$$

As we have assumed that trips from an any origin are equally distributed in all directions, if the distances travelled are large in comparison to that separating the bus lines, only a small proportion of bus riders will have their points of departure and destination near the same bus line. It can therefore be assumed that all bus users will need to change lines during a trip (Newell, 1979), and must therefore sustain the cost of two waiting periods *w*, beyond that of the transfer. The mean cost of a journey by public transport is therefore:

$$C^b = \frac{G_1}{v} + \frac{G_2 a}{2} + G_3 \frac{2b}{3} + G^b \tau V + X_1 \tag{9}$$

Since $Y_1 = C^b$, from (1) we can derive the expression for the demand of captive users:

$$d_1 = A_1 \exp\left(-T_1 C^b\right) \tag{10}$$

Under the assumption that the utilities of a journey on each of the two means of transport are independent Weibull variables, with unit parameter and averages $-C^a$ and $-C^b$, respectively, the inclusive cost of the two means of transport is:

$$Y_2 = -\ln\left[\exp(-C^a) + \exp(-C^b)\right] \tag{11}$$

whence the expression for free users' demand:

$$d_2 = A_2 \exp\left(T_2 \ln\left[\exp(-C^a) + \exp(-C^b)\right]\right) = A_2\left[\exp(-C^a) + \exp(-C^b)\right]^{T_2} \tag{12}$$

while the mean number d_2^b of free users that utilise public transport is:

$$d_2^b = d_2 \frac{\exp(-C^b)}{\exp(-C^a) + \exp(-C^b)} = \frac{d_2}{1 + \exp(C^b - C^a)} \tag{13}$$

where C^a and C^b are expressed by (8) and (9), respectively.

Let $d = \left(d_1, d_2, d_2^b\right)^t$ be the vector of demand, and $X = \left(X_1, X_2\right)^t$, the vector of ticket price and of road pricing. The components of the right-hand sides of expressions (10), (12), (13) are functions of C^a and C^b. Thus, given parameters, T_1 and T_2 of the demand functions (1), the amount of funding *TR*, and the coefficients in the expressions for C^a and C^b, by means of (6), (8), (9), the right hand sides of expressions (10), (12), (13) are functions of d and X. Thus we have:

$$\begin{aligned} d_1 &= \Psi_1(d, X) \\ d_2 &= \Psi_2(d, X) \\ d_2^b &= \Psi_3(d, X) \end{aligned} \tag{14}$$

Now, let $\Psi(d, X) = \left[\Psi_1(d, X), \Psi_2(d, X), \Psi_3(d, X)\right]^t$. The equilibrium demand vector that respects the budget constraint of the company that manages the public transport, for each X is vector $\bar{d}$ for which it holds that:

$$\bar{d} = \Psi(\bar{d}, X) \tag{15}$$

that is, the fixed point of function $\Psi(d, X)$. Since, as we will see in the next Section, the fixed point of $\Psi(d, X)$ in the set of feasible d is unique, equation (15) implicitly defines the demand vector $\bar{d}$ as a function of X: $\bar{d} = \bar{d}(X)$.

We assume that X_1 and X_2 are established by the Public Administration so as to maximize surplus S of users of both the transport modes, under the constraints that the system is in equilibrium, that the budget constraint of the company managing public transport is satisfied, and that the demand $d_2^a = d_2 - d_2^b$ for private transport does not exceed a preset threshold level *CP* in order to respect the physical and environmental

d_1 and X, is denoted by $d_2(d_1, X)$. Function

$$\hat{\Psi}_1(d_1) = d_1 - \Psi_1\left[d_1, d_2(d_1, X), d_2^b(d_1, d_2(d_1, X), X), X\right]$$

is almost linear and has a unique nil point in the interval $[0, A_1]$: this point, which is a function of X, is the component $\bar{d}_1(X)$ of the equilibrium demand vector. Taking account of these results, $\bar{d}_1(X)$ has been computed for each X by applying, in the interval $[0, A_1]$, the bisection method to function $\hat{\Psi}_1(d_1)$, in which $d_2(d_1, X)$ has been computed, for each d_1, by applying the bisection method, in the interval $[0, A_2]$, to function $\hat{\Psi}_2(d_2)$, in which $d_2^b(d_1, d_2, X)$ has been computed, for each pair (d_1, d_2), by applying the bisection method to function $\hat{\Psi}_3(d_2^b)$ in the interval $[0, A_2]$. And, at the same time, $\bar{d}_2$ and $\bar{d}_2^b$ have been computed.

It has been assumed that $G = 4$ €/km (this is the average value borne by the public transport companies in Italy) and that the parameters of the demand function is the same for the two different user categories, $T_1 = T_2 = T$, and two values of T, 0.1 and 0.2, have been considered. We have examined the case in which all users have a private auto at their disposal $(A_1 = 0)$, and the case in which the captives to public transport are one third of total potential demand $(A_1 = 100{,}000)$. It has been supposed that there is no external funding to public transport: $TR = 0$.

Setting $\tau = 0.125$ h as the increase in journey time whose associated cost is perceived on average by users as equivalent to a transfer between two bus lines, and $V = 15$ km/h as the mean velocity of a bus, we have $\tau \cdot V = 1.875$ km.

We have set $\alpha = 1$ and $\beta = 5$ in Eqs (7), while the coefficients in the cost functions (8) and (9) have the following values:

$$G_1 = 12 \text{ €/h} \quad G_2 = 2 \text{ €/km} \quad G^a = 0.35 \text{ €/km} \quad G^b = 0.30 \text{ €/km} \qquad (17)$$

The system has been studied considering various values of parameter K in Eqs. (7), thus different levels of physical capacity of the road network, and the demand constraint CP has been set equal to K. Moreover we have considered the case in which the road network capacity is adequate to satisfy the demand, $K = 300{,}000$, and a constraint CP has been imposed on private demand in order to reduce the environmental damage due to traffic; and different values of CP have been considered.

Problem (16) has been solved for the different situations examined, and for each of them we have computed the optimal values of X_1 and X_2, surplus S, the components of the equilibrium demand vector $\bar{d}$, the frequency ν of the bus lines, the costs Y_1 and Y_2. Some of the results are synthesised in fig. 3,…,6.

4. AN ANALYSIS OF THE RESULTS.

Fig. 3 refers to the case when all users are free, and it shows the percentage variation of user surplus with respect to the situation in which road pricing is absent, as a function of parameter K, thus of the network physical capacity. It can be noted that road pricing produces an increase in user surplus, which is particularly high when demand is rather rigid (T = 0.1), and network capacity is low, so that there would be high congestion in the absence of road pricing. For all values of K we have examined, the optimal surplus has been reached with inactive constraint imposed on private demand. This means that users reach their maximum surplus when a large part of them transfers to public transport, so that the portion of demand that uses private cars is less than the network capacity. This transfer is a consequence of the fact that road pricing imposition causes a great reduction of the cost of public transport, due in part to the fact that the utilisation of road pricing proceeds for financing public transport makes it possible to increase in a substantial manner the bus frequency, in part because the diminution of private transport demand causes a decrease of congestion, thereby of the journey times of buses.

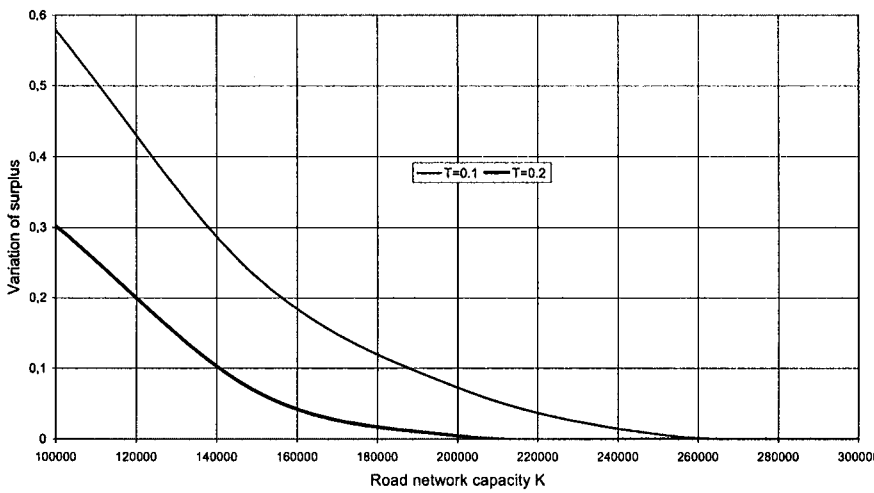


Fig. 3. Percentage variation of surplus as a function of the road network capacity when all users are free

These results of road pricing are not substantially influenced by the presence of a portion of users captive to public transport, even if in this case the increases in surplus have resulted less than the values shown in fig. 3.

Fig. 4 refers to the case in which all users are free, the road network capacity is adequate to satisfy the demand (K = 300,000), and a constraint CP on private demand is imposed in order to maintain the pollution due to traffic below an acceptable threshold. The figure shows the percentage variation of user surplus as a function of CP. For all values of CP less than 250,000 the optimal surplus has been reached with active constraint on private demand: this means that road pricing causes a damage to users, which is measured by the decrease in surplus. This decrease is particularly high when demand is rather elastic (T = 0.2) and the constraint is strict.

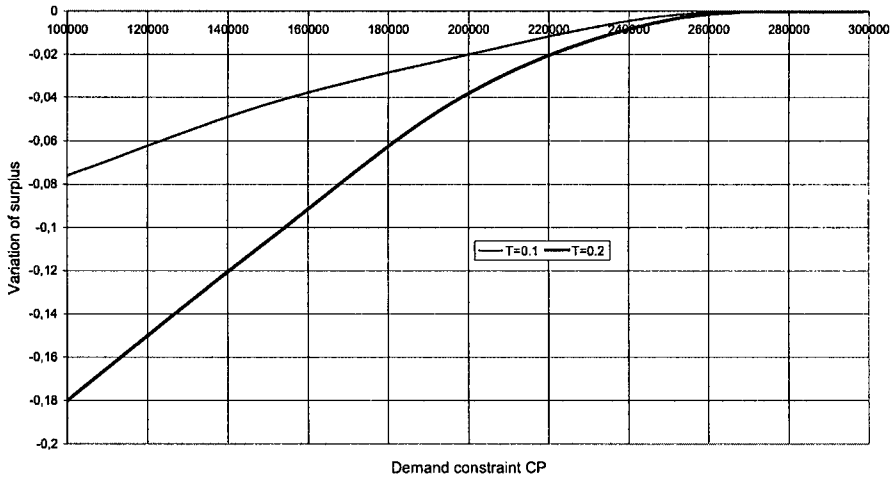


Fig. 4. Percentage variation of surplus as a function of the constraint on private demand when all users are free

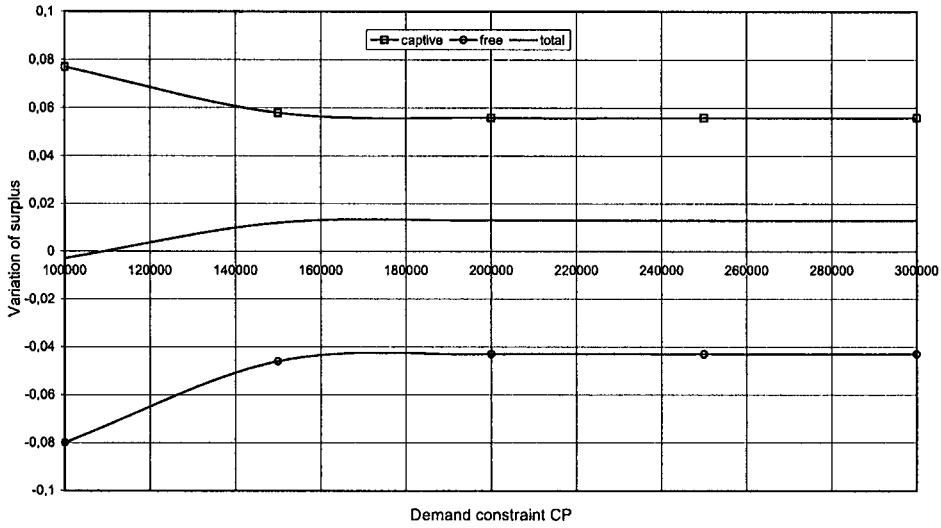


Fig. 5. Percentage variation of surplus as a function of the constraint on private demand when one third of potential demand is captive to public transport (T = 0.1)

The results of road pricing are completely different when the road network capacity is adequate to satisfy the demand and one third of potential demand is captive to public transport, as it is shown in fig. 5 and fig. 6. They show, respectively for $T = 0.1$ and $T = 0.2$, the percentage variation of surplus as a function of constraint *CP* imposed for environmental reasons. The variations of surplus are reported for the entire set of users, and separately for each user category. When *CP* is less than 200,000, the constraint on private demand has resulted always active, but the figures show that the impact of road pricing is in general an increase in user surplus, more substantial when demand is rather elastic ($T = 0.2$). This substantial difference with respect to the case in which all potential demand is free, is due to the fact, shown in the figures, that the loss of surplus on the part of free users is balanced by the increase on the part of captive users, who benefit by the reduction of cost of public transport, financed by the road pricing revenue; so that the results of road pricing are essentially a redistribution of transport costs, and therefore of surplus, between the two categories of users.

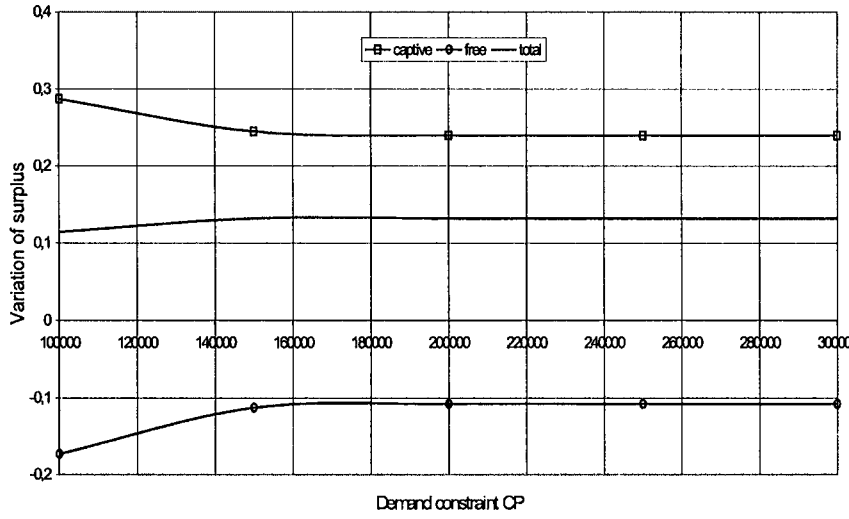


Fig. 6. Percentage variation of surplus as a function of the constraint on private demand when one third of the potential demand is captive to public transport (T = 0.2)

5. CONCLUSIONS.

The impact of road pricing on the cost of urban transport essentially depends on the level of congestion which is present on the road network before the imposition of road pricing. If there is high congestion, because the capacity of road network is inadequate to satisfy transport demand, road pricing causes benefits to both the transport users and the environment. In fact it gives rise to a new equilibrium pattern of the transport system, characterised by the transfer of a substantial portion of users to public transport: traffic pollution decreases substantially and at the same time, as a consequence of the congestion decrease, the costs of transport diminish and the user surplus increases. The characteristics of this phenomenon are not substantially influenced by the proportion of users that are captive to public transport, even if they are more marked when all users have a private auto at their disposal.

The results of road pricing are substantially different when the capacity of road network is adequate to satisfy transport demand, so that users do not bear congestion, but traffic pollution causes damages to environment. In this case the imposition of road pricing in order to maintain the volume of private cars, and then the pollution it produces, below a certain threshold, has

consequences very different according whether there is or not a substantial proportion of users that are captive to public transport (one third in the example considered in this paper). If all the users have a private auto at their disposal, road pricing causes a substantial decrease in their surplus. If instead a portion of users are captive to public transport, they receive an increase in surplus from road pricing, which balances the loss of free users, so that the overall surplus increases, in a more substantial manner if demand is rather elastic.

Some conclusions can be drawn from the analysis carried out in this paper, which seem to have general validity when demand is rather uniformly distributed in the urban area, because they do not depend on the particular configuration of the road network:

- When private transport demand causes congestion and pollution because the road network capacity is inadequate, the imposition of road pricing, and the utilisation of its proceeds for financing public transport, produces not only the reduction of pollution within the preset level, but even a decrease in the transport costs borne by all the user categories, and thus an increase in their surplus, in a more substantial measure the less is the road network capacity, and the greater is the percentage of users that have a private auto at their disposal.
- When the road network capacity is adequate to satisfy transport demand, and the majority of users have a private auto at their disposal, road pricing produces the wanted reduction of pollution, but at the same time causes a remarkable decrease in their surplus. Therefore in this case, when the Public Administration defines the constraint to impose on private demand, it has to balance in a correct way the advantages this constraint causes to environment with the social and economic damages it causes to the city. When instead a remarkable proportion of users is captive to public transport, road pricing produces an increases in surplus, because the damage to the free users is balanced by the benefits to the captive ones.

REFERENCES

[1] Domencich T.A. and McFadden D. (1975) *Urban travel demand*, North Holland/American Elsevier, New York, N.Y.

[2] Ferrari P. (1999) A model of urban transport management, *Transportation Research* B 33, 43-61.

[3] Hearn D.W. and Ramana M.V. (1998) Solving congestion toll-pricing models. In: Marcotte P., Nguyen S. (Eds.) *Equilibrium and Advanced Transportation Modelling*, Kluwer Academic Publisher, Dordrecht, 109-114.

[4] Hearn D.W. and Yildirim M.B. (2000) A toll pricing framework for traffic assignment problems with elastic demand. To be appear in the edited volume: *Current trends in Transportation and Network Analysis – Papers in honor of Michael Florian.*

[5] Holroyd E.M. (1965) The optimal bus service: a theoretical model for a large uniform urban area, *Proceedings of the Third International Symposium on the Theory of Traffic Flow*, New York, 308-328.

[6] Newell G.E. (1979) Some issues relating to the optimal design of bus routes, *Transportation Science* 13, 20-35.

SHARP ESTIMATES FOR GREEN'S FUNCTIONS: SINGULAR CASES

M.G. Garroni
Dept. of Mathematics "Guido Castelnuovo", University of Rome "La Sapienza", Rome, Italy

0. INTRODUCTION

The purpose of this paper is to present a survey of some results contained in a number of papers and two books, over the years 1984-2002, concerning the construction and the properties of the Green function for parabolic second-order operators, under different "non-regular" hypotheses.

In this exposition, I shall consider only the parabolic operators not in divergence form and only boundary operators satisfying the "regular oblique derivative condition". Let Ω be a (bounded) open subset of $\mathbb{R}^N$, $N \geq 2$, we denote by Q_T the cylinder $\Omega \times (0,T)$, $0 < T < +\infty$, and by $\Sigma_T = \partial\Omega \times [0,T]$ its lateral boundary. Consider the following parabolic problem

$$\begin{cases} Lu(x,t) = \partial_t u(x,t) + Au(x,t) = f(x,t), & (x,t) \in Q_T \\ Bu(x,t) = 0, & (x,t) \in \Sigma_T \\ u(x,0) = 0, & x \in \Omega \end{cases} \tag{0.1}$$

where[1]

[1] Throughout this paper we omit the summation symbol whenever it refers to an index that occurs twice and we use the same symbol c or C for different constants (depending on prescribed sets of arguments).

$$
\begin{aligned}
&A \equiv A(x,t,\partial_x) = a_{ij}(x,t)\partial_{x_i x_j} + a_i(x,t)\partial_{x_i} + a_0(x,t),\\
&\qquad a_{ij}\lambda_i\lambda_j \geq \mu\lambda^2, \quad \forall (x,t) \in Q_T, \quad \mu > 0.\\
&B \equiv B(x,t,\partial_x) = b_i(x,t)\partial_{x_i} + b_0(x,t)\\
&\qquad b_i(x,t)n_i(x) \geq \nu, \quad \forall (x,t) \in \Sigma_T, \quad \nu > 0
\end{aligned}
\tag{0.2}
$$

where $n = (n_i(x), i = 0,...,N)$ is the unit outward normal vector to $\partial\Omega$ at the point $x \in \partial\Omega$.

For this type of problems the classical hypotheses are the following:

$$
\begin{cases}
i) \quad a_{ij} \in C^{\alpha,\frac{\alpha}{2}}(\overline{Q}_T) & ,0 < \alpha < 1\\
ii) \quad a_i, a_0 \in C^{\alpha,\frac{\alpha}{2}}(\overline{Q}_T)
\end{cases}
\tag{0.3}
$$

$$
b_i, b_0 \in C^{1+\alpha,\frac{1+\alpha}{2}}(\Sigma_T) \tag{0.4}
$$

$$
\partial\Omega \in C^{2+\alpha}. \tag{0.5}
$$

It is well known (see f.i. [21] and [22]) that under these hypotheses for any $f \in C^{\alpha,\frac{\alpha}{2}}(\overline{Q}_T)$ there exists a unique solution of problem (0.1) in $C^{2+\alpha,\frac{2+\alpha}{2}}(\overline{Q}_T)$ and there exists a unique Green function $G(x,y,t,\tau)$ satisfying the heat kernel type estimates:

$$
\begin{aligned}
&\left|D_t^r D_x^s G(x,y,t,\tau)\right| \leq C(t-\tau)^{-\frac{N+2r+s}{2}} \exp\left(-C\frac{|x-y|^2}{t-\tau}\right), \qquad 2r+s \leq 2\\
&\left|D_t^r D_x^s G(x,y,t,\tau) - D_t^r D_{x'}^s G(x',y,t,\tau)\right| \leq\\
&\qquad \leq C|x-x'|^{\alpha}(t-\tau)^{-\frac{N+2+\alpha}{2}} \exp\left(-C\frac{|x''-y|^2}{t-\tau}\right), \qquad 2r+s=2,\\
&\left|D_t^r D_x^s G(x,y,t,\tau) - D_{t'}^r D_x^s G(x,y,t',\tau)\right| \leq\\
&\qquad \leq C(t-t')^{\frac{2-2r-s+\alpha}{2}}(t'-\tau)^{-\frac{N+2+\alpha}{2}} \exp\left(-C\frac{|x-y|^2}{t-\tau}\right), \qquad 2r+s=1,2.
\end{aligned}
\tag{0.6}
$$

Here $|x''-y| = \min(|x-y|,|x'-y|)$, $t > t' > \tau$, the constants C are independent of (x,y,t,τ).

We shall present three cases. In the first case, studied in [9], we replace the hypothesis (0.4) by

$$b_i, b_0 \in C^{\alpha,\frac{\alpha}{2}}(\Sigma_T). \tag{0.4*}$$

In the second case, studied in [5], [6] and [8], we add to L an integral operator I, called jump operator.

In the third case the heat equation with oblique constant conditions in a dihedral angle (i.e. a set which is neither bounded nor regular) is studied, see [12], [13], [14] and [15].

For all these problems we construct a Green function i.e. a function such that, for any "convenient" function $f(y,\tau)$, the domain potential

$$u(x,t) = \int_{-\infty}^{t} d\tau \int_{\Omega} G(x,y,t,\tau) f(y,\tau) dy \tag{0.7}$$

is a "strong" solution of (0.1). One of the questions is in what kind of space we have to fix f and in what space the solution of (0.1) exists. As usually the Green function is constructed in the form of a sum of two terms: the principal term $G_0(x,y,t,\tau)$ with the highest singularity for $x=y$, $t=\tau$ (of the heat kernel type) and an additional term $G_1(x,y,t,\tau)$ (in general less singular for $x=y$, $t=\tau$) which presents new singularities due to non "regular" hypotheses.

1. PROBLEMS WITH HÖLDER CONTINUOUS COEFFICIENTS

We study problem (0.1) when the coefficients of the boundary operator have a fairly low regularity (Hölder-continuity only). Under these weakened hypotheses it is not possible to obtain the Schauder estimates with global Hölder norms for upper order derivatives and to expect solutions of problem (0.1) in $C^{2+\alpha,\frac{2+\alpha}{2}}(\overline{Q}_T)$. It is necessary to introduce weighted Hölder spaces. For any $(x,t) \in Q_T$, let $\rho(x)$ be the distance of x to $\partial\Omega$ and

$$K(x,t) = \left\{ (y,\tau) \in Q_T \,/\, |x-y| < \frac{1}{2}\rho(x),\ t - \frac{1}{4}\rho^2(x) < \tau < t \right\}.$$

We define $C_b^{\ell,\frac{\ell}{2}}(\overline{Q}_T)$, $b \le \ell$, with a non integer ℓ, as a set of functions $u(x,t)$, $(x,t) \in Q_T$, with a finite norm:

$$|u|_{C_b^{\ell,\frac{\ell}{2}}(\overline{Q}_t)} = \sum_{o\vee b<2r+s\le[\ell]} \left|\rho(x)^{2r+s-b} D_t^r D_x^s u\right|_{Q_T}^{(0)} + \sup_{(x,t)\in Q_T} \rho(x)^{l-b} <u>_{K(x,t)}^{(\ell)} +$$

$$+\begin{cases} |u|_{Q_T}^{(b)}, & \text{if} \quad b\ge 0 \\ 0, & \text{if} \quad b<0, \end{cases}$$

where $<u>^{(\ell)}$ and $|u|^{(b)}$ are the usual Hölder seminorms and norms respectively in the parabolic spaces.

The following theorem is proved:

Theorem 1.1 *Suppose that the conditions (0,2), (0.3), (04*) and (0.5) hold. Then for every* $f\in C_{\alpha-l}^{\alpha,\frac{\alpha}{2}}(\overline{Q}_T)$ *the problem (0.1) has a unique solution* $u\in C_{l+\alpha}^{2+\alpha,\frac{2+\alpha}{2}}(\overline{Q}_T)$ *and*

$$|u|_{C_{l+\alpha}^{2+\alpha}(\overline{Q}_T)} \le C\,|f|_{C_{\alpha-l}^{\alpha,\frac{\alpha}{2}}(\overline{Q}_T)},$$

see [9] for more general results and [18] for the analogous Dirichlet elliptic problems.

We also prove local Schauder estimates for the solution near the boundary Σ_T. Finally we construct the Green function.

The principal term $G_0(x,y,t,\tau)$ is explicitely constructed, $G_1(x,y,t,\tau)$ is obtained as a solution of a parabolic problem in a Sobolev space $W_p^{2,1}(Q_T), \forall p\in\left(1,\frac{N+2}{N+2-\alpha}\right)$. Then G_1 can be estimated by using a convenient auxiliary L_p-estimate of solutions of our problem with an exponential weight.

The main result is the following:

Theorem 1.2 *(Green function) Suppose that conditions (0.2), (0.3), (0.4*) and (0.5) hold. Then there exists a unique Green function* $G(x,y,t,\tau)=G_0(x,y,t,\tau)+G_1(x,y,t,\tau)$. G_0 *satisfies the inequalities (0.6) and* G_1 *the following ones:*

$$\left|D_x^s G_1(x,y,t,\tau)\right| \le C(t-\tau)^{-\frac{N+s-\alpha}{2}} \exp\left(-C\frac{|x-y|^2}{t-\tau}\right), \quad s=0,1$$

$$\left|G_1(x,y,t,\tau)-G_1(x,y,t',\tau)\right| \le C(t-t')^{\frac{1+\alpha}{2}}(t'-\tau)^{-\frac{N+1}{2}} exp\left(-C\frac{|x-y|^2}{t-\tau}\right),$$

$$\left|\nabla_x G_1(x,y,t,\tau)-\nabla_x G_1(x,y,t',\tau)\right| \le C(t-t')^{\frac{\alpha}{2}}(t'-\tau)^{-\frac{N+1}{2}} exp\left(-C\frac{|x-y|^2}{t-\tau}\right),$$

$$\left|\nabla_x G_1(x,y,t,\tau)-\nabla_{x'} G_1(x',y,t,\tau)\right| \le C\,|x-x'|^{\alpha}\,(t-\tau)^{-\frac{N+1}{2}} exp\left(-C\frac{|x''-y|^2}{t-\tau}\right)$$

where $t>t'>\tau$, $|x''-y|=|x-y|\wedge|x'-y|$. *Moreover,*

$$\left|\nabla_x G_1(x,y,t,\tau)-\nabla_x G_1(x,y,t',\tau)\right| \le$$
$$\le C(t-t')^{\frac{1+\alpha}{2}}(t'-\tau)^{-\frac{N+1}{2}}\left(\rho^{-1}(x)\vee(t'-\tau)^{-\frac{1}{2}}\right)exp\left(-C\frac{|x-y|^2}{t-\tau}\right),$$

and the higher derivatives $D_t^r D_x^s G_1$, $2r+s=2$, *satisfy the inequalities*

$$\left|D_t^r D_x^s G_1(x,y,t,\tau)\right| \le C(t-\tau)^{-\frac{N+1}{2}}\left(\rho^{\alpha-1}(x)\vee(t-\tau)^{\frac{\alpha-1}{2}}\right)exp\left(-C\frac{|x-y|^2}{t-\tau}\right),$$

$$\left|D_t^r D_x^s G_1(x,y,t,\tau)-D_t^r D_x^s G_1(x,y,t',\tau)\right| \le$$
$$\le C(t-t')^{\frac{\alpha}{2}}(t'-\tau)^{-\frac{N+1}{2}}\left(\rho^{-1}(x)\vee(t'-\tau)^{-\frac{1}{2}}\right)exp\left(-C\frac{|x-y|^2}{t-\tau}\right),$$

$$\left|D_t^r D_x^s G_1(x,y,t,\tau)-D_t^r D_x^s G_1(z,y,t,\tau)\right| \le$$
$$\le C\,|x-z|^{\alpha}\,(t-\tau)^{-\frac{N+1}{2}}\left(\rho^{-1}(x)\vee(t-\tau)^{-\frac{1}{2}}\right)exp\left(-C\frac{|x-y|^2}{t-\tau}\right),$$

if $|z-x|<\frac{1}{2}\rho(x)$.

The use of this Green function allows one to solve different asymptotic problems motivated by the control theory of stochastic processes when the non-divergence structure of A and the fairly low regularity of the coefficients do not allow a Fredholm alternative approach see [2].

2. PROBLEMS WITH INTEGRO-DIFFERENTIAL OPERATORS

Consider the case when in the problem (0,1) L is replaced by an integro-differential operator $L-I$ associated with diffusion processes with jumps. Some basic material on this subject can be found in the books [1] and [17].

For each x in $\mathbb{R}^N$, $t\in[0,T]$, $T<+\infty$, a Radon measure $M(x,t,dz)$ on $\mathbb{R}^N_* = \mathbb{R}^N - \{0\}$, such that

$$\int_{|z|<1} |z|^\gamma M(x,t,dz) + \int_{|z|\geq 1} M(x,t,dz) \leq C_0 < +\infty, \quad 0\leq\gamma\leq 2, \qquad (2.1)$$

determines operator I.

We need to describe the dependency of the variable x in the Levy kernel $M(x,t,dz)$, see [5] and [6]. Suppose that there exist a σ-finite measure space $(F,\mathcal{F},\pi)$, two Borel measurable functions $j(x,t,\zeta)$ and $m(x,t,\zeta)$ from $\mathbb{R}^N\times[0,T]\times F$ into $\mathbb{R}^N_*$ and $[0,\infty)$, respectively, such that

$$M(x,t,A) = \int_{\{\zeta: j(x,t,\zeta)\in A\}} m(x,t,\zeta)\pi(d\zeta), \qquad (2.2)$$

for any Borel measurable subset A of $\mathbb{R}^N_*$. The functions $j(x,t,\zeta)$ and $m(x,t,\zeta)$ are called the *jump size (or amplitude)* and the *jump density (or intensity)*, respectively. The condition (2.1) on the singularity at the origin of the Levy kernel $M(x,t,dz)$ will be assumed to hold uniformly in x, so that for some measurable function $\overline{j}(\zeta)$ from F into $(0,\infty)$ and some constant $C_0>0$ we have

$$\begin{cases} |j(x,t,\zeta)| \leq \overline{j}(\zeta), \qquad 0\leq m(x,t,\zeta)\leq 1, \\ \\ \displaystyle\int_{\{\overline{j}<1\}} |\overline{j}(\zeta)|^\gamma \pi(d\zeta) + \int_{\{\overline{j}\geq 1\}} \pi(d\zeta) \leq C_0, \end{cases} \qquad (2.3)$$

where $0\leq\gamma\leq 2$ is the order of the Levy kernel. Actually, we may allow $0\leq m(x,\zeta)\leq C$ if we re-define the measure $\pi(d\zeta)$.

Thus for any smooth function φ the integro-differential operator has the form

$$\begin{cases} I\varphi = \int_F [\varphi(\cdot + j(\cdot,\cdot,\zeta),\cdot) - \varphi] m(\cdot,\cdot,\zeta)\pi(d\zeta), & 0 \le \gamma < 1 \\ I\varphi = \int_F [\varphi(\cdot + j(\cdot,\cdot,\zeta),\cdot) - \varphi - j(\cdot,\cdot,\zeta)\nabla\varphi_{1_{\{|\zeta|<1\}}}] m(\cdot,\cdot,\zeta)\pi(d\zeta), & 1 \le \gamma \le 2 \end{cases} \tag{2.4}$$

It is convenient decompose I in such way

$$I\varphi = \int_0^1 d\theta \int_{\{\bar{j}<1\}} j(\cdot,\cdot,\zeta)\cdot\nabla\varphi(\cdot + \theta j(\cdot,\cdot,\zeta)) m(\cdot,\cdot,\zeta)\pi(d\zeta) + \\ + \int_{\{\bar{j}\ge 1\}} [\varphi(\cdot + j(\cdot,\cdot,\zeta),\cdot) - \varphi] m(\cdot,\cdot,\zeta)\pi(d\zeta); \qquad \text{for} \quad 0 < \gamma < 1 \tag{2.5}$$

$$I\varphi = \int_0^1 (1-\theta) d\theta \int_{\{\bar{j}<1\}} j(\cdot,\cdot,\zeta)\cdot\nabla^2\varphi(\cdot + \theta j(\cdot,\cdot,\zeta),\cdot) j(\cdot,\cdot,\zeta) m(\cdot,\cdot,\zeta)\pi(d\zeta) + \\ + \int_{\{\bar{j}\ge 1\}} [\varphi(\cdot + j(\cdot,\cdot,\zeta),\cdot) - \varphi] m(\cdot,\cdot,\zeta)\pi(d\zeta); \qquad \text{for} \quad 1 \le \gamma \le 2, \tag{2.6}$$

see [6] and [8] for examples.

In order to study this integro-differential operator as acting on Lebesgue (Sobolev) spaces, we will need to perform a change of variables. Assume that the jump amplitude function $j(x,t,\zeta)$ is continuously differentiable in x for any fixed t in $[0,T]$ and ζ in F, and that there exists a constant $c_0 > 0$ such that for any x, x' and $0 \le \theta \le 1$ we have

$$c_0 |x - x'| \le |x - x' + \theta[j(x,t,\zeta) - j(x',t,\zeta)]| \le c_0^{-1} |x - x'|. \tag{2.7}$$

This implies that the change of variables $X = x + \theta j(x,t,\zeta)$ is a diffeomorphism of class C^1 in $\mathbb{R}^N$, for any θ in $[0,1]$, t in $[0,T]$ and ζ in F.

The integro-differential operator is defined **a priori** for functions $\varphi(x,t)$, with x in the whole space $\mathbb{R}^N$. However, we want to consider equations on a domain Ω of $\mathbb{R}^N$ and here we want to treat only the homogeneous oblique boundary condition. For this case will use a condition on the jumps **only interior jumps are allowed**, i.e.

$$\begin{cases} \text{if } m(x,t,\zeta) \ne 0, \quad x \in \Omega, \quad t \in [0,T], \quad \zeta \in F \\ \quad \text{then } x + \theta j(x,t,\zeta) \in \overline{\Omega}, \quad \forall \theta \in [0,1]. \end{cases} \tag{2.8}$$

Remark 2.1 *Condition (2.8) is assumed for the sake of simplicity. Actually, it suffices to impose conditions (2.7) and (2.8) only for* $\theta = 1$, *plus a (locally) convex condition on* Ω. *For a complete discussion see [6]. Much more complicated is the Dirichlet problem; to study this problem we have introduced a "convenient"* **localization** *of the operator* I, *see section 2.3 of [8].*

Problem (0.1) becomes

$$\begin{cases} (L-I)u(x,t) = f(x,t), & (x,t) \in Q_T \\ Bu(x,t) = 0, & (x,t) \in \Sigma_T \\ u(x,0) = 0, & x \in \Omega. \end{cases} \tag{2.9}$$

Two basic assumptions are used for the construction of the Green function. The first hypothesis allows us to work in Sobolev spaces and the second one on Hölder spaces.

Hypothesis 2.1 (L^p) *Let* Ω *be a bounded domain in* $\mathbb{R}^N, L, I$ *and* B *be operators as above satisfying (0.2), (0.3)* i, *(0.4***), (0.5), (2.1), ..., (2.8), with* $0 \le \gamma < 2 - \alpha$.

The condition (0.3) ii, *is weakened in*

$$a_i, a_0 \in L^\infty.$$

Hypothesis 2.2 (C^α) *Assume Hypothesis 2.1, the smoothness conditions on the coefficients* a_i, a_0, b_i, b_0 *(0.3)* ii, *(0.4) and that there exist a measurable function (again denoted by* $j(\cdot)$*) from* F *into* $(0,\infty)$ *and some constant* $M_0 > 0$ *s.t.*

$$\begin{cases} |j(x,t,\zeta) - j(x',t',\zeta)| \le \bar{j}(\zeta)[|x-x'|^\alpha + |t-t'|^{\frac{\alpha}{2}}] \\ |m(x,t,\zeta) - m(x',t',\zeta)| \le M_0[|x-x'|^\alpha + |t-t'|^{\frac{\alpha}{2}}], \\ \qquad \text{for any } x, x', t, t' \text{ and } \zeta. \end{cases} \tag{2.10}$$

Notice that these are not minimal assumptions for the existence and uniqueness of solutions of (2.9), but are sufficient to ensure the construction of the Green function, as proved in [6].

Also in this case the Green function is sought as the sum of two terms: a principal term (with the highest singularity) and an additional one. It is well known (see f.i. [21]) that for differential operators, the exponential factor in a kernel of the heat type plays an essential role in constructing this additional

term and in establishing its sharp estimates. As pointed out in Chapter II of [6] in a simple example, estimates of the heat-kernel type cannot exist. The integro-differential operator propagates the classic singularity at the origin. Starting with the principal term equal to the Green function of the differential operator L we use the method of successive approximations to obtain the additional term corresponding to the integro-differential operator. This method of successive approximations involves the solution of a Volterra type equation. Since we cannot make use of heat-kernel type estimates, we are forced to identify the key properties needed to carry over these successive approximations. These properties have the form of following seminorms of the L^∞, L^1, or $C^{\alpha,\frac{\alpha}{2}}$ type which will be used to define a decreasing family of Banach spaces.

For any kernel $\varphi(x,t,\xi)$ with $x,\xi \in \Omega, t \in (0,T], k \geq 0$ and $0 < \alpha < 1$, we define

$$C(\varphi,k) = inf\{C \geq 0 : |\varphi(x,t,\xi)| \leq Ct^{-1+\frac{k-N}{2}}, \forall x,t,\xi\}, \tag{2.11}$$

$$K(\varphi,k) = K_1(\varphi,k) + K_2(\varphi,k), \tag{2.12}$$

$$\begin{cases} K_1(\varphi,k) = inf\{K_1 \geq 0 : \int_\Omega |\varphi(x,t,\xi)|\, d\xi \leq K_1 t^{-1+\frac{k}{2}}, \forall x,t\}, \\ K_2(\varphi,k) = inf\{K_2 \geq 0 : \int_\Omega |\varphi(x,t,\xi)|\, dx \leq K_1 t^{-1+\frac{k}{2}}, \forall t,\xi\}, \end{cases}$$

$$M(\varphi,k,\alpha) = M_1(\varphi,k,\alpha) + M_2(\varphi,k,\alpha) + M_3(\varphi,k,\alpha), \tag{2.13}$$

$$\begin{cases} M_1(\varphi,k,\alpha) = inf\big\{M_1 \geq 0 : \ |\varphi(x,t,\xi) - \varphi(x',t,\xi)| \leq M_1 |x-x'|^\alpha \times \\ \qquad\qquad \times t^{-1+\frac{k-N-\alpha}{2}}, \forall x,x',t,\xi\big\}, \end{cases}$$

$$\begin{cases} M_2(\varphi,k,\alpha) = inf\big\{M_2 \geq 0 : \ |\varphi(x,t,\xi) - \varphi(x,t',\xi)| \leq M_2 |t-t'|^{\frac{\alpha}{2}} \times \\ \qquad\qquad \times \Big[t^{-1+\frac{k-N-\alpha}{2}} \vee t'^{-1+\frac{k-N-\alpha}{2}}\Big], \forall x,t,t',\xi\big\}, \end{cases}$$

$$\begin{cases} M_3(\varphi,k,\alpha) = inf\big\{M_3 \geq 0 : \ |\varphi(x,t,\xi) - \varphi(x,t,\xi')| \leq M_3 |\xi-\xi'|^\alpha \times \\ \qquad\qquad \times t^{-1+\frac{k-N-\alpha}{2}}, \forall x,t,\xi,\xi'\big\}, \end{cases}$$

$$N(\varphi,k,\alpha) = N_1(\varphi,k,\alpha) + N_2(\varphi,k,\alpha) + N_3(\varphi,k,\alpha) + N_4(\varphi,k,\alpha), \tag{2.14}$$

$$\begin{cases} N_1(\varphi,k,\alpha) = inf\{N_1 \geq 0: \quad \int_\Omega |\varphi(x,t,\xi) - \varphi(x',t,\xi)|\, d\xi \leq \\ \qquad\qquad \leq N_1 |x-x'|^\alpha\, t^{-1+\frac{k-\alpha}{2}}, \forall x,x',t\}, \end{cases}$$

$$\begin{cases} N_2(\varphi,k,\alpha) = inf\{N_2 \geq 0: \quad \int_\Omega |\varphi(x,t,\xi) - \varphi(x,t',\xi)|\, d\xi \leq N_2 |t-t'|^{\frac{\alpha}{2}} \times \\ \qquad\qquad \times\left[t^{-1+\frac{k-\alpha}{2}} \vee t'^{-1+\frac{k-\alpha}{2}}\right], \forall x,t,t'\}, \end{cases}$$

$$\begin{cases} N_3(\varphi,k,\alpha) = inf\{N_3 \geq 0: \quad \int_\Omega |\varphi(x,t,\xi) - \varphi(x,t',\xi)|\, dx \leq N_3 |t-t'|^{\frac{\alpha}{2}} \times \\ \qquad\qquad \times\left[t^{-1+\frac{k-\alpha}{2}} \vee t'^{-1+\frac{k-\alpha}{2}}\right], \forall t,t',\xi\}, \end{cases}$$

$$\begin{cases} N_4(\varphi,k,\alpha) = inf\{N_4 \geq 0: \quad \int_\Omega |\varphi(x,t,\xi) - \varphi(x,t,\xi')|\, dx \leq \\ \qquad\qquad \leq N_4 |\xi-\xi'|^\alpha\, t^{-1+\frac{k-\alpha}{2}}, \forall t,\xi,\xi'\}, \end{cases}$$

$$R(\varphi,k,\alpha) = R_1(\varphi,k,\alpha) + R_2(\varphi,k,\alpha), \tag{2.15}$$

$$\begin{cases} R_1(\varphi,k,\alpha) = inf\{R_1 \geq 0: \quad \int_\Omega |\varphi(Z,t,\xi) - \varphi(Z',t,\xi)|\, J_\eta(Z,Z')dz \leq \\ \qquad\qquad \leq R_1 \eta^\alpha t^{-1+\frac{k-\alpha}{2}}, \forall Z,Z',t,\xi \text{ and } \eta > 0\}, \end{cases}$$

$$\begin{cases} R_2(\varphi,k,\alpha) = inf\{R_2 \geq 0: \quad \int_\Omega |\varphi(x,t,Z) - \varphi(x,t,Z')|\, J_\eta(Z,Z')dz \leq \\ \qquad\qquad \leq R_2 \eta^\alpha t^{-1+\frac{k-\alpha}{2}}, \forall x,t,Z,Z' \text{ and } \eta > 0\}, \end{cases}$$

where the change of variables $Z(z)$ and $Z'(z)$ are diffeomorphisms of class C^1 in $\mathbb{R}^N$, and the Jacobian

$$J_\eta(Z,Z') = \begin{cases} |det(\nabla Z)| \wedge |det(\nabla Z')| & \text{if } |Z-Z'| \leq \eta \text{ and } Z,Z' \in \overline{\Omega}, \\ 0 & \text{otherwise,} \end{cases}$$

where ∇Z, $\nabla Z'$ stand for the matrices of the first partial derivatives of $Z(z)$, $Z'(z)$ with respect to the variable z, and $\wedge, \vee$ denote the minimum, maximum (resp.) between two real numbers.

Definition 2.1 (Green function spaces). *Let us denote by* $\mathcal{G}_k^{\alpha,\frac{\alpha}{2}}$, $k \geq 0$ *and* $0 < \alpha < 1$, *the space of all continuous functions (or kernels) defined for* x, ξ *in* $\Omega \subset \mathbb{R}^N$ *and* $0 < t \leq T$, *with values in* $\mathbb{R}$ *and such that the above infima (semi-norms) (2.11),...,(2.15) (of order* k*) are finite. Thus the maximum of the quantities (2.11),...,(2.15), denoted by*

$$[\cdot]_{k,\alpha} = [\cdot]_{\mathcal{G}_k^{\alpha,\frac{\alpha}{2}}},$$

is the norm of the Banach space $\mathcal{G}_k^{\alpha,\frac{\alpha}{2}}$. *When* $\alpha = 0$ *we denote by* $\mathcal{G}_k^0$, $k \geq 0$, *the space of all measurable functions (or kernels)* $\varphi(x,t,\xi)$ *defined for* x, ξ *in* $\Omega \subset \mathbb{R}^N$ *and* $0 < t \leq T$, *with values in* $\mathbb{R}$ *and such that the two infima (2.11) and (2.12) (of order* k*) are finite, with the norm*

$$[\cdot]_{k,0} = [\cdot]_{\mathcal{G}_k^0}.$$

Remark 2.2 *Note that* $\mathcal{G}_k^{\alpha,\frac{\alpha}{2}} \subset \mathcal{G}_\ell^{\alpha,\frac{\alpha}{2}}$ *if* $\ell < k$.

Definition 2.2 *Let us denote by* $\mathcal{G}_k^{i+\alpha,\frac{i+\alpha}{2}}$, *for* $i = 1,2$, *the subspace of* $\mathcal{G}_k^{\alpha,\frac{\alpha}{2}}$ *of functions* $\varphi(x,t,y,s)$ *such that* $\partial^\ell \varphi(x,t,y,s)$ *belongs to* $\mathcal{G}_{k-\ell}^{\alpha,\frac{\alpha}{2}}$ *for* $\ell = 0,...,i$ *and,* $M_2(\partial^\ell \varphi, k-\ell, 1+\alpha)$, $N_2(\partial^\ell \varphi, k-\ell, 1+\alpha)$ *for* $\ell = i-1$ *are finite. Recall that* ∂^ℓ *means the derivatives of (parabolic) order equals to* ℓ *in the first variables, i.e.* x,t. *This is similar to the definition of the Hölder spaces* $C^{i+\alpha,\frac{i+\alpha}{2}}$.

Remark 2.3 *Notice that all the Green functions for "regular" parabolic differential Dirichlet or oblique problems of second order belong to* $\mathcal{G}_2^{2+\alpha,\frac{2+\alpha}{2}}$.

Let G_L be the Green function associated with the differential operator L. As mentioned before, to construct the Green function G associated with the integro-differential operator $L - I$, we solve a Volterra equation

$$\begin{cases} \text{either} & \text{find } Q_I \text{ such that} & Q_I = Q_L + Q_L \bullet Q_I, \\ \text{or} & \text{find } G \text{ such that} & G = G_L + G_L \bullet IG, \end{cases}$$

with the relations $Q_L = IG_L$ and $G = G_L + G_L \bullet Q_I$, where $\bullet$ denotes the "kernel convolution". Actually, we express Q_I as the following series

$$Q_I = \sum_{n=1}^{\infty} Q_n, \quad Q_0 = Q_L, \quad Q_n = Q_L \bullet Q_{n-1}\ n = 1,2,..., \qquad . \qquad (2.16)$$

Each term of this series belongs to a Green function space of on appropriate order and the convergence is in the sense of the above Green spaces. The final result is the following, see pp. 335–408 of [6]:

Theorem 2.3 *(Green function) Under Hypothesis 2.2 there exists a unique strong Green function* $G(x,y,t,\tau)$ *for the problem (2,9):* $G = G_L + G_1$,

1) $G \geq 0$, $\int_\Omega G(x,y,t,\tau)dy \leq 1$, $\forall x,t,\tau$, *with* $\tau < t$,

2) $G_L \in \mathcal{G}_2^{2+\alpha,\frac{2+\alpha}{2}}$, $G_1 \in \mathcal{G}_{(4-\gamma)\wedge 3}^{2+\alpha,\frac{2+\alpha}{2}}$,

3) the Chapman-Kolmogorov equation

$$G(x,y,t,\tau) = \int_\Omega G(x,\xi,t,s)G(\xi,y,s,\tau)d\xi \quad \forall x,y,t,\tau, \quad \tau < s < t,$$

is satisfied,

4) if $a_0 \equiv 0$, $b_0 \equiv 0$, *then*

$$\int_\Omega G(x,y,t,\tau)dy = 1, \quad \forall x,t,\tau, \quad \text{with} \quad \tau < t.$$

Remark 2.4 G_L *has the unique singularity for* $t = \tau$ *and* $x = y$, *whereas* G_1 *can have infinitely many singularities of lower order for* $t = \tau$ *and* x *depending on the nature of the integral operator. Anyway the behavior in the singular points is controlled by the norms in* $\mathcal{G}_2^{2+\alpha,\frac{2+\alpha}{2}}$ *and* $\mathcal{G}_{(4-\gamma)\wedge 3}^{2+\alpha,\frac{2+\alpha}{2}}$ *respectively.*

Remark 2.5. *Property 4) has a probabilistic interpretation. Let* $\{w(t), t \geq 0, w(0) = x\}$ *be the diffusion process with jumps reflected at the boundary associated to the operators* $L - I$ *and* B *and let* $P(x,t,E,0)$ *be its probability of transition i.e. for any Borel set* E

$$P(x,t,E,0) = P\{w(t) \in E / w(0) = x, \text{ for any } t \geq 0, x \in E\},$$

then G *is the transition probability density function i.e.*

$$P(x,t,E,0) = \int_E G(x,y,t,0)dy \ , \ \forall x \in E \ , \ t > 0.$$

Remark 2.6 *If in the above theorem we replace the assumption (0.4) by (0.4*) we can still construct the Green function for* $0 \leq \gamma \leq 1$, *see [5]. Furthermore for any function* f *in* $C^{\alpha,\frac{\alpha}{2}}(\overline{Q}_T)$ *the classic solution given by*

(0.7) belongs to $C_{1+\alpha}^{2+\alpha,\frac{2+\alpha}{2}}(\overline{Q}_T)$ *(see theorem 1.1) and properties 1), 2), 3), 4) still hold, see [5] and [6].*

Remark 2.7 *Under Hypothesis 2.1, with* $0 \le \gamma < 2$, *the problem (2.9) has solutions only in the Sobolev spaces* $W_p^{2,1}(Q_T)$, $1 < p < \infty$. *Under this hypothesis we can construct a (weak) Green function, i.e. G is such that the potential as (0.7) is the unique solution of (2.9). G belongs to the space* $\mathcal{G}_2^0$, *where only the seminorms C (control in* L^∞*) and K (control in* L^1*) are involved. Moreover properties 1),...,4) of theorem 2.3 still hold. In this case each term of (2.16) belongs to a space of type* $\mathcal{G}_k^0$.

The Green functions constructed in [5] and in [6] are the essential tools in many applications to linear and non linear elliptic or parabolic problems see [4], [7], [10], [11] and [16].

3. PROBLEMS IN A DIHEDRAL ANGLE

In this case L is the heat operator. We construct the Green function of the oblique boundary value problem in a dihedral angle $D_\theta \subset \mathbb{R}^N$, with the opening angle $\theta \in (0, 2\pi)$. We assume that

$$D_\theta = d_\theta \times \mathbb{R}^{N-2} = \{x \in \mathbb{R}^N : x' = (x_1, x_2) \in d_\theta, x'' = (x_3, ..., x_N) \in \mathbb{R}^{N-2}\},$$

where d_θ is an infinite plane sector which can be given in the polar coordinates $r > 0$, $0 < \varphi < \theta$, where $x_1 = r\cos\varphi, x_2 = r\sin\varphi$. We denote by Γ_0 and Γ_1 the faces of D_θ:

$$\Gamma_0 = \gamma_0 \times \mathbb{R}^{N-2}, \quad \Gamma_1 = \gamma_1 \times \mathbb{R}^{N-2},$$

where $\gamma_0 = \{\varphi = 0, r \ge 0\}$ and $\gamma_1 = \{\varphi = \theta, r \ge 0\}$ are boundary lines of d_θ.

The problem (0.1) becames

$$\begin{cases} (i)\ \partial_t u - \Delta_x u = f & \text{in } D_{0,T} \equiv D_\theta \times (0,T) \\ (ii)\ B_0 u \equiv \dfrac{\partial u}{\partial n} + h_0 \dfrac{\partial u}{\partial r} = 0 & \text{in } \Gamma_{0,T} \equiv \Gamma_0 \times (0,T] \\ (iii)\ B_1 u \equiv \dfrac{\partial u}{\partial n} + h_1 \dfrac{\partial u}{\partial r} = 0 & \text{in } \Gamma_{1,T} \equiv \Gamma_1 \times (0,T] \\ (iv)\ u(\cdot,0) = 0 & \text{in } D_\theta, \end{cases} \tag{3.1}$$

where $\frac{\partial}{\partial n}$ is the derivative in the direction of the exterior normal to the boundary of D_θ, h_0 and h_1 are real numbers.

It is well known, (see e.g. [20] and [24]) that, because of the presence of the angle, the solutions of such problems (also with Dirichlet or Neuman conditions) and regular data have, in general, pole singularities at the edge of the dihedral angle.

The solution is thus considered in special wheighted Sobolev spaces where the weight is the distance from the vertex with an appropriate exponent. It is necessary to introduce two different weighted spaces: one for $h_0 + h_1 > 0$ and one for $h_0 + h_1 \le 0$. Fix the real number $\mu \ge 0$ and the integer $k \ge 0$. By $H^{k,\frac{k}{2}}_{0,\mu}(D_{\theta,T})$ and $L^{k,\frac{k}{2}}_{0,\mu}(D_{\theta,T})$ we mean the closure of the set of smooth functions, defined in $D_\theta \times (-\infty, T)$ and vanishing for $t \le 0$, near the vertex of the angle and for large $|x|$, with respect to the norms:

$$\|u\|_{H^{k,\frac{k}{2}}_{0,\mu}(D_{\theta,T})}$$
$$= [\sum_{|\alpha|+2a\le k} \int_0^T dt \int_{D_\theta} |x'|^{2\mu-2k+2(|\alpha|+2a)} |D_x^\alpha D_t^a u(x,t)|^2 dx$$
$$+ \sum_{|\alpha|+2a=k-1} \int_{D_\theta} |x'|^{2\mu} dx \int_{-\infty}^T dt \int_{-\infty}^T \frac{|D_x^\alpha D_t^a u(x,t) - D_x^\alpha D_s^a u(x,s)|^2}{|t-s|^2} ds]^{\frac{1}{2}};$$

$$\|u\|_{L^{k,\frac{k}{2}}_{0,\mu}(D_{\theta,T})}$$
$$= [\sum_{|\alpha|+2a=k} \int_0^T dt \int_{D_\theta} |x'|^{2\mu} |D_x^\alpha D_t^a u(x,t)|^2 dx$$
$$+ \sum_{|\alpha|+2a=k-1} \int_{D_\theta} |x'|^{2\mu} dx \int_{-\infty}^T dt \int_{-\infty}^T \frac{|D_x^\alpha D_t^a u(x,t) - D_x^\alpha D_s^a u(x,s)|^2}{|t-s|^2} ds]^{\frac{1}{2}};$$

The most obvious difference between these two spaces is that functions belonging to the spaces $H^{k,\frac{k}{2}}_{0,\mu}(d_{\theta,T})$ vanish at the origin, while the functions belonging in $L^{k,\frac{k}{2}}_{0,\mu}(d_{\theta,T})$ do not necessarily vanish. This fact, connected with the sign of $h_0 + h_1$, becomes clear in Section 4 and 5 of the [13].

For $k > 0$, the elements of the previous spaces have traces on half hyperplane $\Gamma_{\theta_1} = \gamma_{\theta_1} \times \mathbb{R}^{N-2}$, $\gamma_{\theta_1} = \{\varphi = \theta_1, r \ge 0\}$ $(\theta_1 \in [0, \theta])$, belonging to $H^{k-\frac{1}{2},\frac{k}{2}-\frac{1}{4}}_{0,\mu}(\Gamma_{\theta_1,T})$ and $L^{k-\frac{1}{2},\frac{k}{2}-\frac{1}{4}}_{0,\mu}(\Gamma_{\theta_1,T})$ respectively, see [23] and [24] for the

corresponding norms. We denote $H^{0,0}_{0,\mu}$ and $L^{0,0}_{0,\mu}$ by $L_{2\mu}$ and we set for any subset A of D_θ

$$\|u\|^2_{L_{2,\mu}(A\times(0,T))} = \int_0^T dt \int_A |u(x,t)|^2 |x'|^{2\mu}\, dx.$$

Consider problem (3.1) with nonhomogeneous data Φ_i at the boundary $\Gamma_{i,T}, i = 0,1$. In [13] we prove the following solvability and regularity results under appropriate condition for μ and k with respect to the aperture θ (see following conditions (3.2) and (3.3)). These conditions are, as usual, connected with well-known Kondrat'ev results [19].

Theorem 3.1 *Let* $\mu \geq 0, \beta_i = \arctan h_i \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right), h_0 + h_1 > 0$ *and*

$$0 < 1 + k - \mu < \frac{\beta_0 + \beta_1}{\theta}. \tag{3.2}$$

For arbitrary $f \in H^{k,\frac{k}{2}}_{0,\mu}(D_{\theta,T})$ *and* $\Phi_i \in H^{k+1-\frac{1}{2},\frac{k+1}{2}-\frac{1}{4}}_{0,\mu}(\Gamma_{i,T}), i = 0,1$, *problem (3.1) has a unique solution* $u \in H^{k+2,\frac{k+2}{2}}_{0,\mu}(D_{\theta,T})$ *and*

$$\|u\|_{H^{k+2,\frac{k+2}{2}}_{0,\mu}(D_{\theta,T})} \leq c\left\{\|f\|_{H^{k,\frac{k}{2}}_{0,\mu}(D_{\theta,T})} + \sum_{i=0}^{1} \|\Phi_i\|_{H^{k+1-\frac{1}{2},\frac{k+1}{2}-\frac{1}{4}}_{0,\mu}(D_{\theta,T})}\right\},$$

where c *is a positive constant independent of* f, Φ_i *and* T.

Theorem 3.2 *Let* $\mu \geq 0, \beta_i = \arctan h_i \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right), h_0 + h_1 \leq 0$ *and*

$$0 < 1 + k - \mu < \frac{\pi + \beta_0 + \beta_1}{\theta}. \tag{3.3}$$

Then for arbitrary $f \in L^{k,\frac{k}{2}}_{0,\mu}(D_{\theta,T})$ *and* $\Phi_i \in L^{k+1-\frac{1}{2},\frac{k+1}{2}-\frac{1}{4}}_{0,\mu}(D_{\theta,T}), i = 0,1$, *problem (3.1) has a unique solution* $u \in L^{k+2,\frac{k+2}{2}}_{0,\mu}(D_{\theta,T})$ *and*

$$\|u\|_{L^{k+2,\frac{k+2}{2}}_{0,\mu}(D_{\theta,T})} \leq c\left\{\|f\|_{L^{k,\frac{k}{2}}_{0,\mu}(D_{\theta,T})} + \sum_{i=0}^{1} \|\Phi_i\|_{L^{k+1-\frac{1}{2},\frac{k+1}{2}-\frac{1}{4}}_{0,\mu}(\Gamma_{i,T})}\right\}$$

where c *is a positive constant independent of* f, Φ_i *and* T.

The results of the above theorems are essential to obtain the following final results.

Take $\beta_i = \arctan h_i, i = 0,1$, $G(x,y,t,0) \equiv G(x,y,t)$.

Theorem 3.3 *(Green function) Suppose* $\beta_0 + \beta_1 > 0$ $[\beta_0 + \beta_1 \le 0]$, *there exists a unique Green function* $G(x,y,t)$ *of problem (3.1), i.e. a function such that:*

1) for any $0 < 1 - \mu < \frac{\beta_0 + \beta_1}{\theta}$ $[0 < 1 - \mu < \frac{\pi + \beta_0 + \beta_1}{\theta}]$ *and for any* $f \in L_{2,\mu}(D_{\theta T})$

$$u(x,t) = \int_{-\infty}^{t} d\tau \int_{D_\theta} G(x,y,t-\tau) f(y,\tau) dy \quad (3.4)$$

is the solution of problem (3.1) in $H_{0,\mu}^{2,1}$ $[L_{0,\mu}^{2,1}]$;

2) G is infinitely differentiable relative to all its arguments for $x,y,t \in \bar{D}_{\theta,T} \setminus \{x'' = 0\}, x \ne y, t \ne 0$,

3) the function $x,t \to \eta(|x||y|^{-1})\,\eta(|1-t|)\,G(x,y,t)$ *lies in the space* $H_{0,\mu}^{k+2,\frac{k+2}{2}}(D_{\theta T})$ $[L_{0,\mu}^{k+2,\frac{k+2}{2}}(D_{\theta T})]$, $k = 0,1,\ldots$ *for each fixed* $y \in D_{\theta T}$.

Here $\eta \in C^\infty(0,+\infty)$, $\eta(s) = \begin{cases} 1 & s < \frac{1}{2} \text{ and } s > 2, \\ 0 & \frac{3}{4} < s < 1. \end{cases}$

Moreover the following estimates hold:

4) For $x,y \in D_\theta, t > 0$ *and any*

$$\alpha \equiv (\alpha_1, \alpha_2, \ldots, \alpha_N), \gamma \equiv (\gamma_1, \gamma_2, \ldots, \gamma_N), \alpha_i, \gamma_i \in \mathbb{N} \cup \{0\}, i = 1,\ldots,N, b \in \mathbb{N} \cup \{0\}.$$

$$|D_x^\alpha D_y^\gamma D_t^b G(x,y,t)| \le C(\alpha,\gamma,b,\theta) \frac{e^{-c\frac{|x-y|^2}{t}}}{(|x-y|^2 + t)^{\frac{N+|\alpha|+|\gamma|+2b}{2}}} \cdot \left(\frac{|x'|}{|x'|+|x-y|+\sqrt{t}}\right)^{\lambda_1(|\alpha|)} \cdot \left(\frac{|y'|}{|y'|+|x-y|+\sqrt{t}}\right)^{\lambda_2(|\gamma'|)},$$

where

$$\lambda_1(|\alpha'|) = \begin{cases} \dfrac{\beta_0+\beta_1}{\theta} - |\alpha'| - \varepsilon_1, & \text{if } \beta_0 + \beta_1 > 0 \\ \min\left\{0, \dfrac{\pi+\beta_0+\beta_1}{\theta} - |\alpha'| - \varepsilon_2\right\}, & \text{if } \beta_0 + \beta_1 \le 0 \end{cases}$$

$$\lambda_2(|\gamma'|)=\begin{cases} min\left\{0,\dfrac{\pi-\beta_0-\beta_1}{\theta}-|\gamma'|-\varepsilon_2\right\}, & \text{if } \beta_0+\beta_1\geq 0 \\ -\dfrac{\beta_0+\beta_1}{\theta}-|\gamma'|-\varepsilon_1, & \text{if } \beta_0+\beta_1<0, \end{cases}$$

$$\varepsilon_i>0, i=1,2, \quad c, C(\alpha,\gamma,b,\theta)>0, \quad \alpha'=(\alpha_1,\alpha_2), \gamma'=(\gamma_1,\gamma_2).$$

Remark 3.1 *A similar construction is done in [23] for the Green function of the Neuman problem (i.e. for $h_0=h_1=0$). As in [23] problem (3.1) can be solved also in convenient weighted Hölder spaces and the solution can be expressed by (3.4).*

Remark 3.2 *For $N=2$ our estimates are analogous to the estimates obtained by Kozlov in a cone [20].*

REFERENCES

[1] A. Bensoussan, J.L. Lions, *Impulse control and Quasi-Variational Inequalities*, Gauthier-Villard, Paris, 1984.

[2] I. Dolcetta, M.G. Garroni, *Oblique Derivative Problems and Invariant Measure*, in "Annali della Sc. Norm. Sup. di Pisa", vol XIII, n. 4, 1986, pp. 689-720.

[3] M.G. Garroni, J.L. Menaldi, *Green's Function and Asymptotic Behaviour of the Solution of Some Oblique...*, in "Pitman Research Notes in Mathematics Series", 149, Longman Scient. Techn., 1987, pp. 114-119.

[4] M.G. Garroni, J.L. Menaldi, *On Asymptotic Behaviour of Solutions of Integro-Differential Inequalities*, in "Ricerche di Matematica", vol.36 (in onore di C.Miranda), 1987, pp. 149-171.

[5] M.G. Garroni, J.L. Menaldi, *Green's Functions and Invariant Density for Integro-Differential Operator of Second Order*, in "Annali di Matematica Pura e Applicata", (IV), vol. CLIV, 1989, pp. 147-222.

[6] M.G. Garroni, J.L. Menaldi, *Green's Functions for Second Order Parabolic Integro-Differential Problems*, "Pitman Research Notes in Mathematics Series", Longman, London, 1992, pp. 423.

[7] M.G. Garroni, J.L. Menaldi, *Regularizing Effect for Integro-Differential Parabolic Equations*, in "Comm. P.D.E.", vol. 18, 1993, pp. 2023-2050.

[8] M.G. Garroni, J.L. Menaldi, *Second Order Elliptic Integro-Differential Problems*, "Chapman & Hall CRC", Bocaraton, London, 2002, pp. 221.

[9] M.G. Garroni, V.A. Solonnikov, *On Parabolic Oblique Derivative Problem with Hölder Continuous Coefficients*, in "Communications in Partial Differential Equations", 9, 14, 1984, pp. 1323-1372.

[10] M.G. Garroni, V.A. Solonnikov, M.A. Vivaldi, *Quasi-Linear Integro-Differential Parabolic Problems with Non Homogeneous Conditions*, in "Houston Journal of Mathematics", 18 (4), 1992, pp. 481-532.

[11] M.G. Garroni, V.A. Solonnikov, M.A. Vivaldi, *Fully Non-Linear Boundary Conditions for Quasi-Linear Integro-Differential Operators*, in "Nonlinear Partial Differential Equations and their Applications - Collège de France Seminar", vol. XI, Pitman Researches Notes, in Math. Series, vol. 299, 1994, pp. 97-117.

[12] M.G. Garroni, V.A. Solonnikov, M.A. Vivaldi, *On the Oblique Derivative Problem in an Infinite Angle*, in "Topological Meth. in Nonlinear Anal., Journal of the J. Schauder Center", vol. VII (in honour of L. Nirenberg), n. 2, 1996, pp. 299-325.

[13] M.G. Garroni, V.A. Solonnikov, M.A. Vivaldi, *Existence and Regularity Results for Oblique Derivative Problems for Heat Equations in an Angle*, in "Proc. Roy. Soc. Edinburgh", 128A, 1998, pp. 47-79.

[14] M.G. Garroni, V.A. Solonnikov, M.A. Vivaldi, *Green Function for the Heat Equation with Oblique Boundary Conditions in an Angle*, in "Annali della Scuola Norm. Sup. di Pisa", XXVII (vol. in onore di E. De Giorgi), 1998, pp. 455-485.

[15] M.G. Garroni, V.A. Solonnikov, M.A. Vivaldi, *The exponential behavior of the Green function in a behavior angle*, "Communications in Contemporary Math.", vol. III, n. 4, 2001, pp. 571-592.

[16] M.G. Garroni, M.A. Vivaldi, *Quasi-Linear, Parabolic, Integro-Differential Problems with Nonlinear Oblique Boundary Conditions*, in "Nonlinear Analysis", vol. XVI, n. 12, 1991, pp. 1089-1116.

[17] I. Gikhman, A.V. Skorokhod, *Stochastic Differential Equations*, "Springer-Verlag", Berlin, 1972.

[18] D. Gilbarg, L. Hörmander, *Intermediate Schauder estimates*, "Arch. Rat. Mech.Anal.",74 , 1980, pp. 297-318.

[19] V.A. Kondrat'ev, *Boundary-value problems for elliptic equation in domains with conical or angular point*, Tr. Mosk Mat. Obshch 16, pp. 209-292; English Translation in Trans. Moscow Math. Soc. , 1967, pp. 227-314.

[20] V.Kozlov, *On the asymptotic of the Green function and Poisson kernels for the mixed parabolic problem in a cone*, I and II, Zeithschrift für Analysis und ihre Andwendungen (Russian) 8 (2), 1989, pp. 131-151 and 10(1), 1991, pp. 27-42.

[21] O.A. Ladyzěnskaja, V.A. Solonnikov and N.N. Uralceva, *Linear and Quasilinear Equations of Parabolic Type*, Ann. Math. Soc., Providence, 1968.

[22] V.A. Solonnikov, *On boundary value problems for linear general parabolic systems of differential equations*, Trudy Math. Inst. Steklov, 83, 1965.

[23] V.A. Solonnikov, *Solvability of the classical initial-boundary-value problems for the heat-condition equation in a dihedral angle*, J. Sov. Math. 32, 1986, pp. 526-546.

[24] V.A. Solonnikov and E.V. Frolova, *On a problem with the third boundary condition equation in a plane angle and its applications to parabolic problems*, Algebra Analiz. 2 , 1990, pp. 213-241; English Translation in Leningrad Math. J.2 , 1991, pp. 891-916.

[25] V.A. Solonnikov and E.V. Frolova, *On a certain nonstationary problem in a dihedral angle II*, J. Soviet Math. 70(3), 1994, pp. 1841-1846.

FIRST-ORDER CONDITIONS FOR $C^{0,1}$ CONSTRAINED VECTOR OPTIMIZATION

I. Ginchev,[1] A. Guerraggio,[2] and M. Rocca[2]
Technical University of Varna, Department of Mathematics, Varna, Bulgaria;[1] University of Insubria, Department of Economics, Varese, Italy[2]

Abstract: For a Fritz John type vector optimization problem with $C^{0,1}$ data we give scalar characterizations of its solutions applying the so called oriented distance and give necessary and sufficient first order optimality conditions in terms of the Dini derivative. While establishing the sufficiency, we introduce new type of efficient points referred to as isolated minimizers of first order. We show that the obtained necessary conditions are necessary for weak efficiency, and the sufficient conditions are sufficient and under Kuhn-Tucker type constraint qualification also necessary for a point to be an isolated minimizer of first order.

Key words: Vector optimization, Nonsmooth optimization, $C^{0,1}$ functions, Dini derivatives, First-order optimality conditions, Lagrange multipliers.

Math. Subject Classification: 90C29, 90C30, 49J52.

1. INTRODUCTION

In this paper we consider the vector optimization problem

$$\min_C f(x), \quad g(x) \in -K, \tag{1}$$

where $f:\mathbb{R}^n \to \mathbb{R}^m$, $g:\mathbb{R}^n \to \mathbb{R}^p$. Here n, m and p are positive integers, $K \subset \mathbb{R}^p$ is a closed convex cone and we assume that a partial ordering on $\mathbb{R}^m$ is induced by a cone C which we assume to be closed, convex and pointed.

Problem (1) generalizes from scalar to vector optimization the Fritz John problem [10].

There are different type of solutions of problem (1). Usually the solutions are called points of efficiency. We prefer, as in scalar optimization, to call them minimizers. In Section 2 we define different type of minimizers and give their scalar characterizations applying the so called oriented distance.

We assume that the functions f and g are $C^{0,1}$, that is f and g are locally Lipschitz. The purpose of the paper is to give necessary and sufficient first-order optimality conditions in terms of Dini directional derivatives. For this purpose we introduce new type of efficient points referred to as isolated minimizers of first order. We show that the necessary conditions are necessary for weakly efficiency, and the sufficient conditions are sufficient and under Kuhn-Tucker type constraint qualification also necessary for a point to be an isolated minimizer of first order.

We confine to functions f, g defined on the whole space $\mathbb{R}^n$. Usually in optimization functions on open subsets are considered, but such a more general assumption does not introduce new features in the problem.

The present paper is a part of a project, whose aim is to establish first and higher-order optimality conditions for $C^{k,1}$ vector optimization problems in terms of Dini derivatives. Recall that a vector function is said to be of class $C^{k,1}$ if it is k-times Fréchet differentiable with locally Lipschitz k-th derivative. The functions from the class $C^{0,1}$ are the locally Lipschitz ones. The $C^{1,1}$ functions in optimization and second-order optimality conditions have been introduced in Hiriart-Urruty, Strodiot, Hien Nguen [7]. Thereafter Klatte, Tammer [11], Yang, Jeyakumar [20], Yang [21] and others have studied various aspects of $C^{1,1}$ functions. For Taylor expansion formula and other properties of $C^{k,1}$ functions with arbitrary k see Luc [14]. Dini derivatives and second-order necessary and sufficient conditions for $C^{1,1}$ Fritz John type vector optimization problems have been considered by Liu, Neittaanmäki, Křížek [13]. While they work with polyhedral cones and the sufficiency concerns Pareto efficiency, in [4] applying new tools, namely, isolated minimizers of second order and the oriented distance, for the case of unconstrained problems we succeeded to improve this result. The result in [4] is related to an arbitrary closed convex ordering cone, the sufficient conditions are sufficient, and moreover necessary, the reference point to be an isolated minimizer of second order, a stronger notion of efficiency than the notion of a Pareto efficient point. In [4] also a comparison with Guerraggio, Luc [6] has been done from which it is seen that the results

based on Dini derivatives are stronger than those based on Clarke derivatives.

The present paper studies the first-order case, as a step toward generalization to constrained problems of the results from [4]. First-order optimality conditions in vector optimization are known, see e.g. Amahroq, Taa [1], Ciligot-Travain [3], Pappalardo, Stocklin [18] and the references therein. Nevertheless, the obtained here first-order conditions for $C^{0,1}$ functions in terms of Dini derivatives present new features.

In Section 2 we introduce different concepts of optimality and their scalarization. In Section 3 we prove our main result. In Section 4 we apply it to prove some relation of the isolated minimizers of first order and the properly efficient points.

2. CONCEPTS OF OPTIMALITY AND SCALAR CHARACTERIZATIONS

We denote the unit sphere and the open unit ball in $\mathbb{R}^n$ respectively by $S=\{x\in\mathbb{R}^n \,|\, \|x\|=1\}$ and $B=\{x\in\mathbb{R}^n \,|\, \|x\|<1\}$. For the norm and the scalar product in the considered finite-dimensional spaces we write $\|\,.\,\|$ and $\langle .,. \rangle$. From the context it should be clear to exactly which spaces these notations are applied.

Let us consider problem (1). The point x is said to be feasible when $g(x)\in -K$ (equivalently $x\in g^{-1}(-K)$). There are different concepts of solutions for problem (1). In any case a solution x^0 should be a feasible point, which is assumed in the following definitions.

The feasible point x^0 is said to be a weakly efficient (efficient) point, if there is a neighbourhood U of x^0, such that if $x\in U\cap g^{-1}(-K)$ then $f(x)-f(x^0)\notin -\operatorname{int} C$ (respectively $f(x)-f(x^0)\notin -(C\setminus\{0\})$). The feasible point x^0 is said to be properly efficient if there exists a closed convex cone $\tilde{C}\subset\mathbb{R}^n$, such that $C\setminus\{0\}\subset \operatorname{int}\tilde{C}$ and x^0 is weakly efficient point with respect to $\tilde{C}$ (that is x^0 is weakly efficient for the problem $\min_{\tilde{C}} f(x)$, $g(x)\in -K$). In this paper the weakly efficient, the efficient and the properly efficient points for problem (1) are called respectively *w*-minimizers, *e*-minimizers and *p*-minimizers. Finally, we call x^0 a strong *e*-minimizer, if there is a neighbourhood U of x^0, such that $f(x)-f(x^0)\notin -C$ for $x\in(U\setminus\{x^0\})\cap g^{-1}(-K)$. Obviously, each strong *e*-minimizer is *e*-minimizer.

The unconstrained problem

$$\min_C f(x) \tag{2}$$

should be considered as a particular case of problem (1).

We recall that each *p*-minimizer is *e*-minimizer and each *e*-minimizer is *w*-minimizer.

For the cone $M \subset \mathbb{R}^k$ its positive polar cone M' is defined by $M' = \{\zeta \in \mathbb{R}^k \,|\, \langle \zeta, \phi \rangle \geq 0 \text{ for all } \phi \in M\}$. The cone M' is closed and convex. It is well known that $M'' := (M')' = \mathrm{cl}\,\mathrm{conv}\, M$, see e.g. Rockafellar [19] (here conv A denotes the convex hull of the set A). In particular, when M is a closed convex cone we have $M' = \{\zeta \in \mathbb{R}^k \,|\, \langle \zeta, \phi \rangle \geq 0 \text{ for all } \phi \in M\}$ and $M = M'' = \{\phi \in \mathbb{R}^k \,|\, \langle \zeta, \phi \rangle \geq 0 \text{ for all } \zeta \in M'\}$.

The linear span of the cone $M \subset \mathbb{R}^k$, that is the smallest subspace of $\mathbb{R}^k$ containing M, is denoted $\mathbf{L}_M$. The positive polar cone of M related to the linear span of M is

$$M'_{\mathbf{L}_M} = \{\zeta \in \mathbf{L}_M \,|\, \langle \zeta, \phi \rangle \geq 0 \text{ for all } \phi \in M\} = M' \cap \mathbf{L}_M .$$

The relative interior $\mathrm{ri}\, M$ of M is defined as the interior of M with respect to the relative topology of the linear span $\mathbf{L}_M \subset \mathbb{R}^k$ of M, that is $\mathrm{ri}\, M = \mathrm{int}_{\mathbf{L}_M} M$.

When M is a closed convex cone, then we have

$$M = \{\phi \in \mathbf{L}_M \,|\, \langle \zeta, \phi \rangle \geq 0 \text{ for all } \zeta \in M'_{\mathbf{L}_M}\},$$
$$\mathrm{ri}\, M = \{\phi \in \mathbf{L}_M \,|\, \langle \zeta, \phi \rangle > 0 \text{ for all } \zeta \in M'_{\mathbf{L}_M}\}.$$

We recall that for every nonempty convex set $A \subseteq \mathbb{R}^k$, we have $\mathrm{ri}\, A \neq \varnothing$.

If $\phi \in -\mathrm{cl}\,\mathrm{conv}\, M$, then $\langle \zeta, \phi \rangle \leq 0$ for all $\zeta \in M'$. We set $M'(\phi) = \{\zeta \in M' \,|\, \langle \zeta, \phi \rangle = 0\}$. Then $M'(\phi)$ is a closed convex cone and $M'(\phi) \subset M'$. Consequently its positive polar cone $M(\phi) = (M'(\phi))'$ is a closed convex cone, $M \subset M(\phi)$ and its positive polar cone satisfies $(M(\phi))' = M'(\phi)$. In this paper we apply this notation for $M = K$ and $\phi = g(x^0)$. Then we write for short $K'(x^0)$ instead of $K'(g(x^0))$ (and call this cone the index set of problem (1) at x^0) and $K(x^0)$ instead of $K(g(x^0))$. We find this abbreviation convenient and not ambiguous, since further this is the unique case, in which we make use of the cones $M'(\phi)$ and $M(\phi)$.

For the closed convex cone M' we apply in the sequel the notations $\Gamma_{M'} = \{\zeta \in M' \,|\, \|\zeta\| = 1\}$ and

$\Gamma_{M'\cap \mathbf{L}_M} = \{\zeta \in M' \cap \mathbf{L}_M \,|\, \|\zeta\| = 1\} = \{\zeta \in M'_{\mathbf{L}_M} \,|\, \|\zeta\| = 1\}$. The sets $\Gamma_{M'}$ and $\Gamma_{M'\cap \mathbf{L}_M}$ are compact, since they are closed and bounded.

Further we make use of the orthogonal projection. Let $\mathbf{L} \subset \mathbb{R}^k$ be a given subspace of $\mathbb{R}^k$. The orthogonal projection is a linear function $\pi_{\mathbf{L}} : \mathbb{R}^k \to \mathbf{L}$ determined by $\pi_{\mathbf{L}}\phi \in \mathbf{L}$ and $\langle \zeta, \phi - \pi_{\mathbf{L}}\phi \rangle = 0$, that is $\langle \zeta, \phi \rangle = \langle \zeta, \pi_{\mathbf{L}}\phi \rangle$ for all $\zeta \in \mathbf{L}$. It follows easily from the Cauchy inequality that $\|\pi_{\mathbf{L}}\| := \max_{\phi \in S} \|\pi_{\mathbf{L}}\phi\| = 1$ if $\mathbf{L} \neq \{0\}$ and $\|\pi_{\mathbf{L}}\| = 0$ if $\mathbf{L} = \{0\}$ (here S denotes the unit sphere in $\mathbb{R}^k$).

A relation of the vector optimization problem (1) to some scalar optimization problem can be obtained in terms of positive polar cones.

Proposition 1. *Let* $\varphi(x) = \max\left\{ \langle \xi, f(x) - f(x^0) \rangle \,\middle|\, \xi \in C', \|\xi\| = 1 \right\}$. *The feasible point* $x^0 \in \mathbb{R}^n$ *is a w-minimizer for problem (1), if and only if* x^0 *is a minimizer for the scalar problem*

$$\min \varphi(x), \quad g(x) \in -K. \tag{3}$$

Proof 1⁰ Let $\operatorname{int} C = \varnothing$. Then each feasible point x^0 is a *w*-minimizer. At the same time C is contained in some hyperplane $H = \{z \in \mathbb{R}^m \,|\, \langle \xi^0, z \rangle = 0\}$ with $\xi^0 \in \mathbb{R}^m$, $\|\xi^0\| = 1$. Then both $\xi^0 \in C'$ and $-\xi^0 \in C'$, whence

$$\varphi(x) \geq \max\left(\langle \xi^0, f(x) - f(x^0) \rangle, -\langle \xi^0, f(x) - f(x^0) \rangle \right) =$$

$$= \left| \langle \xi^0, f(x) - f(x^0) \rangle \right| \geq 0 = \varphi(x^0),$$

which shows that each feasible point x^0 is a minimizer of the corresponding scalar problem (3).

2⁰. Let $\operatorname{int} C \neq \varnothing$. Suppose x^0 is a *w*-minimizer of problem (1). Let U be the neighbourhood from the definition of a *w*-minimizer and fix $x \in U \cap g^{-1}(-K)$. Then $f(x) - f(x^0) \notin -\operatorname{int} C \neq \varnothing$. From the well known Separation Theorem there exists $\xi^x \in \mathbb{R}^m$, $\|\xi^x\| = 1$, such that $\langle \xi^x, f(x) - f(x^0) \rangle \geq 0$ and $\langle \xi^x, -y \rangle = -\langle \xi^x, y \rangle \leq 0$ for all $y \in C$. The latter inequality shows that $\xi^x \in C'$ and the former one shows that $\varphi(x) \geq \langle \xi^x, f(x) - f(x^0) \rangle \geq 0 = \varphi(x^0)$. Thus $\varphi(x) \geq \varphi(x^0)$, $x \in U \cap g^{-1}(-K)$, and therefore x^0 is a minimizer of the scalar problem (3).

Let now x^0 be a minimizer of the scalar problem (3). Choose the neighbourhood U of x^0, such that $\varphi(x) \geq \varphi(x^0)$ for all $x \in U \cap g^{-1}(-K)$ and fix one such x. Then there exists $\xi^x \in C'$, $\|\xi^x\| = 1$, such that $\varphi(x) = \langle \xi^x, f(x) - f(x^0) \rangle \geq \varphi(x^0) = 0$ (here we use the compactness of the

set $\{\xi \in C' \mid \|\xi\| = 1\}$). From $\xi^x \in C'$ it follows $\langle \xi^x, -y \rangle < 0$ for $y \in \operatorname{int} C$. Therefore $f(x) - f(x^0) \notin -\operatorname{int} C$ and consequently x^0 is a *w*-minimizer of problem (1). □

If $\operatorname{int} C = \varnothing$, then each feasible point x^0 of problem (1) is *w*-minimizer. For this case the concept of a relatively weakly efficient point (*rw*-minimizer) turns to be reacher in content. We use in the sequel the concept of *rw*-minimizer instead of *w*-minimizer in some of the results for the case when $\operatorname{int} C = \varnothing$. Let us say in advance that if $\operatorname{int} C \neq \varnothing$ the concepts of *rw*-minimizer and *w*-minimizer coincide.

In order to define a *rw*-minimizer we consider the problem

$$\min_C \overline{f}(x), \quad \overline{g}(x) \in -K, \tag{4}$$

where $\overline{f} = \pi_{\mathbf{L}_C} \circ f$ and $\overline{g} = \pi_{\mathbf{L}_K} \circ g$. Then we call the feasible point x^0 (i.e. $\overline{g}(x^0) \in -K$) a *rw*-minimizer for problem (1), if there exists a neighbourhood U of x^0 such that $\overline{f}(x) - \overline{f}(x^0) \notin -\operatorname{ri} C$ for $x \in U \cap \overline{g}^{-1}(-K)$. The following proposition characterizes the *rw*-minimizers.

Proposition 2.
Let $\psi(x) = \max\left\{\left\langle \xi, f(x) - f(x^0) \right\rangle \middle| \xi \in C'_{\mathbf{L}_C} = C' \cap \mathbf{L}_C, \|\xi\| = 1\right\}$. The feasible point x^0 is a *rw*-minimizer for problem (1), if and only if x^0 is a minimizer for the scalar problem

$$\min \psi(x), \quad \overline{g}(x) \in -K. \tag{5}$$

Proof Due to $\langle \xi, f(x) \rangle = \langle \xi, \overline{f}(x) \rangle$ and $\langle \xi, f(x^0) \rangle = \langle \xi, \overline{f}(x^0) \rangle$ for $\xi \in \mathbf{L}_C$, we have $\langle \xi, f(x) - f(x^0) \rangle = \langle \xi, \overline{f}(x) - \overline{f}(x^0) \rangle$.

Let x^0 be a minimizer of problem (5). Then there exists a neighbourhood U of x^0, such that $\psi(x) \geq \psi(x^0)$ for $x \in U \cap \overline{g}^{-1}(-K)$. Fix one such x. From the definition of ψ and the compactness of $\Gamma_{C' \cap \mathbf{L}_C}$, there exists $\xi^0 \in \Gamma_{C' \cap \mathbf{L}_C}$, such that $\psi(x) = \langle \xi^0, \overline{f}(x) - \overline{f}(x^0) \rangle \geq \psi(x^0) = 0$, whence $\overline{f}(x) - \overline{f}(x^0) \notin -\operatorname{ri} C$ and consequently x^0 is a *rw*-minimizer.

Conversely, let x^0 be a *rw*-minimizer and let U be the neighbourhood from the definition of the *rw*-minimizer. Fix $x \in U \cap \overline{g}^{-1}(-K)$. Since $\overline{f}(x) - \overline{f}(x^0) \notin -\operatorname{ri} C \neq \varnothing$, there exists $\xi^0 \in \Gamma_{C' \cap \mathbf{L}_C}$, such that

$\langle \xi^0, \bar{f}(x) - \bar{f}(x^0) \rangle \geq 0$. Then $\psi(x) \geq \langle \xi^0, \bar{f}(x) - \bar{f}(x^0) \rangle \geq 0 = \psi(x^0) = 0$. Therefore x^0 is a minimizer of problem (5).

We see that the proof of Proposition 2 repeats in some sense the proof of Proposition 1, and is even simpler, since $\operatorname{ri} C$ in Proposition 2, being an analogue of $\operatorname{int} C$ from Proposition 1, is never empty. While the phase space in Proposition 1 is $\mathbb{R}^m$, in Proposition 2 it is $\mathbf{L}_C$.

After Proposition 2 the following definitions look natural. We call the feasible point x^0 a relatively efficient point for problem (1), for short *re*-minimizer, (relatively properly efficient point, for short *rp*-minimizer) if x^0 is an efficient (properly efficient) point for problem (4).

We call x^0 a strong *re*-minimizer, if there is a neighbourhood U of x^0, such that $\bar{f}(x) - \bar{f}(x^0) \notin -C$ for $x \in (U \setminus \{x^0\}) \cap \bar{g}^{-1}(-K)$. Obviously, each strong *re*-minimizer is *e*-minimizer. The following characterization of the strong *e*-minimizers (strong *re*-minimizers) holds. The proof is omitted, since it nearly repeats the one from Proposition 1 (Proposition 2).

Proposition 3. The feasible point x^0 is a strong *e*-minimizer (strong *re*-minimizer) of problem (1) with C and K closed convex cones, if and only if x^0 is a strong minimizer of problem (3) (problem (5)).

Proposition 1 claims that the statement x^0 *is a w-minimizer of problem (1)* is equivalent to the statement x^0 *is a minimizer of the scalar problem (3)*. Applying some first or second-order sufficient optimality conditions to check the latter, we usually get more, namely that x^0 *is an isolated minimizer respectively of first and second order of (3)*. Recall, that the feasible point x^0 is said to be an isolated minimizer of order κ (κ positive) of problem (3) if there is a constant $A > 0$ such that $\varphi(x) \geq \varphi(x^0) + A \| x - x^0 \|^\kappa$ for all $x \in U \cap g^{-1}(-K)$. The concept of an isolated minimizer has been popularized by Auslender [2].

It is natural to introduce the following concept of optimality for the vector problem (1):

Definition 1. We say that the feasible point x^0 is an isolated minimizer of order κ for vector problem (1) if it is an isolated minimizer of order κ for scalar problem (3).

Obviously, also a "relative" variant of an isolated minimizer, and as well for other type of efficient points, does exist. From here on we skip such definitions.

To interpret geometrically the property that x^0 is a minimizer of problem (1) of certain type we introduce the so called oriented distance. Given a set $A \subset R^k$, then the distance from $y \in R^k$ to A is given by $d(y, A) = \inf\{\| a - y \| \mid a \in A\}$. The oriented distance from y to A is defined by $D(y, A) = d(y, A) - d(y, \mathbb{R}^k \setminus A)$. The function D is introduced in Hiriart-

Urruty [8,9] and is used later in Ciligot-Travain [3], Amahroq, Taa [1], Miglierina [16], Miglierina, Molho [17]. Zaffaroni [22] gives different notions of efficiency and uses the function D for their scalarization and comparison. Ginchev, Hoffmann [5] use the oriented distance to study approximation of set-valued functions by single-valued ones and in case of a convex set A show the representation $D(y,A)=\sup_{\|\xi\|=1}\left(\inf_{a\in A}\langle\xi,a\rangle-\langle\xi,y\rangle\right)$. From this representation, if C is a convex cone and taking into account

$$\inf_{a\in C}\langle\xi,a\rangle=\begin{cases}0 & , \quad \xi\in C',\\ -\infty & , \quad \xi\notin C',\end{cases}$$

we get easily $D(y,-C)=\sup_{\|\xi\|=1,\xi\in C'}\left(\langle\xi,y\rangle\right)$. In particular the function φ in (3) is expressed by $\varphi(x)=D(f(x)-f(x^0),-C)$. Propositions 1 and 3 are easily reformulated in terms of the oriented distance, namely:

$$x^0 \ w-\text{minimizer} \quad \Leftrightarrow \quad \begin{cases} D(f(x)-f(x^0),-C)\geq 0\\ \text{for } x\in U\cap g^{-1}(-K),\end{cases}$$

$$x^0 \text{ strong } e-\text{minimizer} \quad \Leftrightarrow \quad \begin{cases} D(f(x)-f(x^0),-C)>0\\ \text{for } x\in (U\setminus\{x^0\})\cap g^{-1}(-K).\end{cases}$$

The definition of the isolated minimizers gives

$$x^0 \text{ isolated minimizer of order } \kappa \quad \Leftrightarrow \quad \begin{cases} D(f(x)-f(x^0),-C)\geq A\left\|x-x^0\right\|^{\kappa},\\ \forall x\in U\cap g^{-1}(-K).\end{cases}$$

We see, that the isolated minimizers (of a positive order) are strong e-minimizers. There is a relation between the p-minimizers and the isolated minimizers of first order, which for the unconstrained case is illustrated in the next proposition.

Proposition 4. Let in problem (2) f be Lipschitz in a neighbourhood of x^0 and let x^0 be an isolated minimizer of first order. Then x^0 is p-minimizer of the unconstrained problem (2).

Proof Assume in the contrary, that x^0 is isolated minimizer of first order, but not p-minimizer. Let f be Lipschitz with constant L in $x^0 + r\,\mathrm{cl}\,B$. Take sequences $\delta_k \to +0$ and $\varepsilon_k \to +0$ and define the cones $\tilde{C}_k = \mathrm{cone}\{y \in \mathbb{R}^m \,|\, D(y,C) \le \varepsilon_k, \|y\| = 1\}$ (here $\mathrm{cone}\,A$ denotes the cone generated by the set A). It holds $\mathrm{int}\,\tilde{C}_k \supset C \setminus \{0\}$. From our assumption, there exists a sequence of points $x^k \in (x^0 + \delta_k B)$, such that $f(x^k) - f(x^0) \in -\mathrm{int}\,\tilde{C}_k$, and in particular $f(x^k) - f(x^0) \neq 0$. From the definition of $\tilde{C}_k$ we get

$$D(f(x^k) - f(x^0), -C) \le \varepsilon_k \left\| f(x^k) - f(x^0) \right\| \le \varepsilon_k \left\| L x^k - x^0 \right\|,$$

which contradicts to x^0 isolated minimizer of first order. □

Developing second-order optimality conditions for $C^{1,1}$ functions, we meet with isolated minimizers of second order. Though this trend is not developped in the present paper, let us mention that then the property x^0 isolated minimizer of second order can be considered as some refinement of the property x^0 is p-minimizer, compare with Ginchev, Guerraggio, Rocca [4].

Let C be a closed convex pointed cone with $\mathrm{int}\,C \neq \varnothing$. Then its positive polar C' is a pointed closed convex cone. Recall that the set Ξ is a base for C', if Ξ is convex with $0 \notin \Xi$ and $C' = \mathrm{cone}\,\Xi := \{y \,|\, y = \lambda\xi, \lambda \ge 0, \xi \in \Xi\}$. The property C' pointed closed convex cone in $\mathbb{R}^m$ implies that C' possesses a compact base Ξ and

$$0 < \alpha = \min\{\|\xi\| \,|\, \xi \in \Xi\} \le \max\{\|\xi\| \,|\, \xi \in \Xi\} = \beta < +\infty. \tag{6}$$

Further we assume that Ξ_0 is a compact set with $\Xi = \mathrm{conv}\,\Xi_0$. With the help of Ξ_0 we define the function $\varphi_0(x) = \max\left\{\left\langle \xi, f(x) - f(x^0)\right\rangle \middle| \xi \in \Xi_0\right\}$ and consider the problem

$$\min \varphi_0(x), \quad g(x) \in -K. \tag{7}$$

Proposition 5. *Let Ξ be a base of C' satisfying (6), φ be the function in (3) and*

$$\varphi_\Xi(x) = \max\{\left\langle \xi, f(x) - f(x^0)\right\rangle \} \,|\, \xi \in \Xi\}.$$

Then $\alpha\,\varphi(x) \le \varphi_\Xi(x) \le \beta\,\varphi(x)$.

Proof If $\xi \in \Gamma_{C'} = \{\xi \in \mathbb{R}^m \,|\, \xi \in C', \|\xi\| = 1\}$, then there exists $\lambda_\xi > 0$, such that $\lambda_\xi \xi \in \Xi$. In fact, $\lambda_\xi = \|\lambda_\xi \xi\|$, whence from inequality (6) we have $0 < \alpha \le \lambda_\xi = \|\lambda_\xi \xi\| \le \beta$.

Fix $x \in \mathbb{R}^n$. From the compactness of $\Gamma_{C'}$ there exists $\xi^x \in \Gamma_{C'}$, such that

$$\varphi(x) = \langle \xi^x, f(x) - f(x^0) \rangle = \frac{1}{\lambda_{\xi^x}} \langle \lambda_{\xi^x} \xi^x, f(x) - f(x^0) \rangle \le \frac{1}{\lambda_{\xi^x}} \varphi_\Xi(x) \le \frac{1}{\alpha} \varphi_\Xi(x),$$

whence $\alpha \varphi(x) \le \varphi_\Xi(x)$. For the other inequality, from the compactness of Ξ there exists $\eta^x \in \Xi$, such that $\varphi_\Xi(x) = \langle \eta^x, f(x) - f(x^0) \rangle$. Put $\lambda = \lambda_{\eta^x / \|\eta^x\|}$. Then

$$\varphi_\Xi(x) = \langle \eta^x, f(x) - f(x^0) \rangle = \lambda \left\langle \frac{\eta^x}{\lambda}, f(x) - f(x^0) \right\rangle \le \lambda \varphi(x) \le \beta \varphi(x).$$

□

Proposition 6. *Propositions 1 and 3, and Definition 1 remain true, if in their formulation* problem *(3) is replaced by* problem *(7).*

Proof We show first, that $\varphi_0(x) = \varphi_\Xi(x)$, where $\varphi_\Xi(x)$ is the function from Proposition 5.

The inequality $\varphi_0(x) \le \varphi_\Xi(x)$ follows directly from $\Xi_0 \subset \Xi$. To prove the converse inequality, fix x and let $\varphi_\Xi(x) = \langle \xi^x, f(x) - f(x^0) \rangle$, $\xi^x \in \Xi$. Let ξ^x be the convex combination $\xi^x = \sum_j \lambda_j \xi^j$, where $\xi^j \in \Xi_0$, $\sum_j \lambda_j = 1$, $\lambda_j \ge 0$. Then

$$\varphi_\Xi(x) = \langle \xi^x, f(x) - f(x^0) \rangle = \sum_j \lambda_j \langle \xi^j, f(x) - f(x^0) \rangle \le \sum_j \lambda_j \varphi_0(x) = \varphi_0(x).$$

A consequence of the proved equality and Proposition 5 is the inequality $\alpha \varphi(x) \le \varphi_0(x) \le \beta \varphi(x)$. In order to prove the proposition, we have to show that x^0 is a (strong) minimizer of problem (3) if and only if it is a (strong) minimizer of (7). Assume x^0 is a minimizer of (3) so that $\varphi(x) \ge \varphi(x^0)$ for $x \in U \cap g^{-1}(-K)$. Then $\varphi_0(x) \ge \alpha \varphi(x) \ge \alpha \varphi(x^0) = 0 = \varphi_0(x)$, whence x^0 is a minimizer of (7). Conversely, if x^0 is a minimizer of (7), then $\varphi(x) \ge \frac{1}{\beta} \varphi_0(x) \ge \frac{1}{\beta} \varphi_0(x^0) = 0 = \varphi(x^0)$. The same proof applies to strong minimizers. □

Corollary 1. *In the important case $C = \mathbb{R}^n_+$ (and suitable choice of Ξ) the function φ_0 in (7) transforms into*

$$\varphi_0(x) = \max_{1 \le i \le n} \left(f_i(x) - f_i(x^0) \right). \tag{8}$$

Proof Clearly, $C^* = R^n_+$ has a base $\Xi = \operatorname{conv} \Xi_0$, where $\Xi_0 = \{e^1, ..., e^n\}$ are the unit vectors on the coordinate axes. With this set we get immediately that the function φ_0 in (7) transforms into that in (8). □

More generally, the cone C is said to be polyhedral, if $C^* = \operatorname{cone} \Xi_0$ with some finite set of nonzero vectors $\Xi_0 = \{\xi^1, ..., \xi^k\}$. In this case, similarly to Corollary 1 the function φ_0 in (7) transforms into the maximum of the finite number of functions

$$\varphi_0(x) = \max_{1 \le i \le k} \left\langle \xi^i, f_i(x) - f_i(x^0) \right\rangle.$$

3. FIRST-ORDER CONDITIONS FOR $C^{0,1}$ PROBLEMS

In this section we investigate problem (1) under the assumption that f and g are $C^{0,1}$ functions. We obtain optimality conditions in terms of the first-order Dini directional derivative.

Given a $C^{0,1}$ function $\Phi : \mathbb{R}^n \to \mathbb{R}^k$ we define the Dini directional derivative (we use to say just Dini derivative) $\Phi'_u(x^0)$ of Φ at x^0 in direction $u \in \mathbb{R}^n$ as the set of the cluster points of $(1/t)(\Phi(x^0 + tu) - \Phi(x^0))$ as $t \to +0$, that is as the Kuratowski limit

$$\Phi'_u(x^0) = \operatorname{L\,limsup}_{t \to +0} \frac{1}{t} \left(\Phi(x^0 + tu) - \Phi(x^0) \right).$$

If Φ is Fréchet differentiable at x^0 then the Dini derivative is a singleton, coincides with the usual directional derivative and can be expressed in terms of the Fréchet derivative $\Phi'(x^0)$ (called sometimes the Jacobian of Φ at x^0) by

$$\Phi'_u(x^0) = \lim_{t \to +0} \frac{1}{t} \left(\Phi(x^0 + tu) - \Phi(x^0) \right) = \Phi'(x^0)u.$$

In connection with problem (1) we deal with the Dini directional derivative of the function $\Phi:\mathbb{R}^n \to \mathbb{R}^{m+p}$, $\Phi(x)=(f(x),g(x))$ and then we use to write $\Phi_u'(x^0)=(f(x^0),g(x^0))_u'$. If at least one of the derivatives $f_u'(x^0)$ and $g_u'(x^0)$ is a singleton, then $(f(x^0),g(x^0))_u'=(f_u'(x^0),g_u'(x^0))$. Let us turn attention that always $(f(x^0),g(x^0))_u' \subset f_u'(x^0)\times g_u'(x^0)$, but in general these two sets do not coincide.

In the following B denotes the open unit ball in $\mathbb{R}^n$.

Lemma 1. *Let $\Phi:\mathbb{R}^n \to \mathbb{R}^k$ be Lipschitz with constant L in $x^0+r\,\mathrm{cl}B$, where $x^0 \in R^n$ and $r>0$. Then for $u,v \in \mathbb{R}^n$ and $0<t<r/\max(\|u\|,\|v\|)$ it holds*

$$\left\|\frac{1}{t}\left(\Phi(x^0+tv)-\Phi(x^0)\right)-\frac{1}{t}\left(\Phi(x^0+tu)-\Phi(x^0)\right)\right\| \le L\|v-u\|, \tag{9}$$

In particular for $v=0$ and $0<t<r/\|u\|$ we get

$$\left\|\frac{1}{t}\left(\Phi(x^0+tu)-\Phi(x^0)\right)\right\| \le L\|u\|. \tag{10}$$

Proof The left hand side of (9) is obviously transformed and estimated by

$$\left\|\frac{1}{t}\left(\Phi(x^0+tv)-\Phi(x^0+tu)\right)\right\| \le L\|v-u\|.$$

□

Lemma 2. *Let $\Phi:\mathbb{R}^n \to \mathbb{R}^k$ be Lipschitz with constant L in $x^0+r\,\mathrm{cl}B$, where $x^0 \in \mathbb{R}^n$ and $r>0$. Then $\Phi_u'(x^0)$, $u \in \mathbb{R}^n$, is non-empty compact set, bounded by $\sup\{\|\phi\| \mid \phi \in \Phi_u'(x^0)\} \le L\|u\|$. For each $u,v \in \mathbb{R}^n$ and $\phi_u \in \Phi_u'(x^0)$, there exists a point $\phi_v \in \Phi_v'(x^0)$, such that $\|\phi_v-\phi\|_u \le L\|v-u\|$. Consequently, the set-valued function $u \to \Phi_u'(x^0)$ is Lipschitz with constant L (and hence continuous) with respect to the Hausdorff distance in $\mathbb{R}^k$.*

Proof The closedness of $\Phi_u'(x^0)$ follows from the definition of the Dini derivative. Estimation (10) shows that $\Phi_u'(x^0)$ is not empty and $\|\phi_u\| \le L\|u\|$ for each $\phi_u \in \Phi_u'(x^0)$. Let $\phi_u=\lim_k(1/t_k)\left(\Phi(x^0+t_ku)-\Phi(x^0)\right)$. Passing to a subsequence we may assume that $\phi_v=\lim_k(1/t_k)\left(\Phi(x^0+t_kv)-\Phi(x^0)\right)$ (to make this conclusion we use also the boundedness expressed in (10)). A

passing to a limit in (9) gives $\|\phi_v - \phi_u\| \le L\|v-u\|$. Now the Lipschitz property of the set-valued function $u \to \Phi_u'(x^0)$ becomes obvious. □

If x^0 is a feasible point for problem (1), then $g(x^0) \in -K$, which gives $\langle \eta, g(x^0)\rangle \le 0$ for all $\eta \in K'$. Recall that the index set is defined by $K'(x^0) = \{\eta \in K' \mid \langle \eta, g(x^0)\rangle = 0\}$ and that we put $K(x^0) = (K'(x^0))'$. Then $K'(x^0)$ is the positive polar cone of the cone $K(x^0)$, and $K \subset K(x^0)$ (the latter follows from $K'(x^0) \subset K'$).

Lemma 3. *Let f, g be $C^{0,1}$ functions and consider problem (1). If x^0 is a w-minimizer and $(y^0, z^0) \in (f(x^0), g(x^0))_u'$, then $(y^0, z^0) \notin -(\operatorname{int} C \times \operatorname{int} K(x^0))$.*

Proof. Suppose that $(y^0, z^0) \in (f(x^0), g(x^0))_u'$ and $(y^0, z^0) \in -\operatorname{int}(C \times K(x^0)) = -(\operatorname{int} C \times \operatorname{int} K(x^0))$. Let

$$y^0 = \lim_k \frac{1}{t_k}\big(f(x^0 + t_k u) - f(x^0)\big), \quad z^0 = \lim_k \frac{1}{t_k}\big(g(x^0 + t_k u) - g(x^0)\big). \quad (11)$$

Without loss of generality, we may assume that $0 < t_k < r/\|u\|$ for all k and that f and g are Lipschitz with constant L in $x^0 + r\,\mathrm{cl}\,B$.

We show now that there exists k_0, such that $g(x^0 + t_k u) \in -\operatorname{int} K \subset -K$ for $k > k_0$, that is, $x^0 + t_k u$ is feasible for $k > k_0$. Recall the notation $\Gamma_{K'} = \{\eta \in K' \mid \|\eta\| = 1\}$ and $\Gamma_{K'(x^0)} = \{\eta \in K'(x^0) \mid \|\eta\| = 1\}$. The sets $\Gamma_{K'}$ and $\Gamma_{K'(x^0)}$ are compact as being closed and bounded sets in an Euclidean space.

Let $\overline{\eta} \in \Gamma_{K'}$. We show that there exists a positive integer $k(\overline{\eta})$ and a neighbourhood $V(\overline{\eta})$ of $\overline{\eta}$ in $\Gamma_{K'}$, such that $\langle \eta, g(x^0 + t_k u)\rangle < 0$ for $k > k(\overline{\eta})$ and $\eta \in V(\overline{\eta})$.

1^0. Let $\overline{\eta} \in \Gamma_{K'(x^0)}$. From our assumption, we have $\langle \overline{\eta}, z^0\rangle < -\delta < 0$ for some $\delta = \delta(\overline{\eta}) > 0$. Then

$$\lim_k \frac{1}{t_k}\langle \overline{\eta}, g(x^0 + t_k u) - g(x^0)\rangle = \langle \overline{\eta}, z^0\rangle < 0,$$

whence there exists $k(\overline{\eta})$, such that for all $k > k(\overline{\eta})$ it holds

$$\langle \overline{\eta}, g(x^0 + t_k u)\rangle < \langle \overline{\eta}, g(x^0)\rangle = 0.$$

Let $\langle \overline{\eta}, g(x^0 + t_k u)\rangle < -\varepsilon < 0$ for some $\varepsilon = \varepsilon(\overline{\eta}) > 0$. Then

$$\begin{aligned}\langle \eta, g(x^0+t_k u)\rangle &= \langle \overline{\eta}, g(x^0+t_k u)\rangle + \langle \eta-\overline{\eta}, g(x^0+t_k u)\rangle \\ &< -\varepsilon + \|\eta-\overline{\eta}\|\left(\|g(x^0+t_k u)-g(x^0)\| + \|g(x^0)\|\right) \\ &\le -\varepsilon + \|\eta-\overline{\eta}\|\left(Lr\|u\| + \|g(x^0)\|\right) < -\varepsilon + \frac{1}{2}\varepsilon = -\frac{1}{2}\varepsilon < 0\end{aligned}$$

as far as $\|\eta-\overline{\eta}\| < \varepsilon/(2(Lr\|u\| + \|g(x^0)\|))$ (which determines $V(\overline{\eta})$).

2^0. Let $\overline{\eta} \in \Gamma_{K^+} \setminus \Gamma_{K^+(x^0)}$. We have $\langle \overline{\eta}, g(x^0)\rangle < -\varepsilon < 0$ for some $\varepsilon = \varepsilon(\overline{\eta}) > 0$. Then

$$\begin{aligned}\langle \eta, g(x^0+t_k u)\rangle &= \langle \overline{\eta}, g(x^0)\rangle + \langle \eta, g(x^0+t_k u)-g(x^0)\rangle + \langle \eta-\overline{\eta}, g(x^0\rangle \\ &< -\varepsilon + \|g(x^0+t_k u)-g(x^0)\| + \|\eta-\overline{\eta}\|\|g(x^0)\| \\ &< -\varepsilon + Lt_k\|u\| + \|\eta-\overline{\eta}\|\|g(x^0)\| < -\varepsilon + \frac{1}{3}\varepsilon + \frac{1}{3}\varepsilon = -\frac{1}{3}\varepsilon < 0\end{aligned}$$

as far as $t_k < \varepsilon/(3L\|u\|)$ (we choose $k(\overline{\eta})$ in a way that this inequality holds for $k > k(\overline{\eta})$) and $\|\eta-\overline{\eta}\| < \varepsilon/(3\|g(x^0)\|)$ (which determines $V(\overline{\eta})$).

Since Γ_{K^+} is compact, we can find $\eta_1,\ldots,\eta_s \in \Gamma_{K^+}$ such that $\Gamma_{K^+} \subset V(\overline{\eta}_1) \cup \ldots \cup V(\overline{\eta}_s)$. Let $k_0 = \max(k(\overline{\eta}_1) \cup \ldots \cup k(\overline{\eta}_s))$. For $k > k_0$ we have $\langle \eta, g(x^0+t_k u)\rangle < 0$ for all $\eta \in \Gamma_{K^+}$ (and hence for all $\eta \in K^+$). This shows that $g(x^0+t_k u) \in -\operatorname{int} K \subset -K$, in other words the points $x^0+t_k u$ for $k > k_0$ are feasible.

According to the made assumption $y^0 \in -\operatorname{int} C$. Since $y^0 = \lim_k (1/t_k)(f(x^0+t_k u) - f(x^0))$, we see that $f(x^0+t_k u) - f(x^0) \in -\operatorname{int} C$ for all sufficiently large k. This fact, together with $x^0+t_k u$ feasible, contradicts the assumption that x^0 is a w-minimizer. □

The following constraint qualification appears in the Sufficient Conditions part of Theorem 1.

$$\mathbb{Q}_{0,1}(x^0): \quad \begin{aligned}&\text{If } g(x^0) \in -K \text{ and } \frac{1}{t_k}\left(g(x^0+t_k u^0) - g(x^0)\right) \to z^0 \in -K(x^0) \\ &\text{then } \exists u^k \to u^0 : \exists k_0 \in N : \forall k > k_0 : g(x^0+t_k u^k) \in -K.\end{aligned}$$

The next theorem is our main result.

Theorem 1. First-order conditions *Let f, g be $C^{0,1}$ functions and consider problem (1).*

(Necessary Conditions) *Let x^0 be a w-minimizer of problem (1). Then for each $u \in S = \{v \in \mathbb{R}^n \mid \|v\| = 1\}$ the following condition is satisfied:*

$$\mathbb{N}'_{0,l}: \quad \begin{array}{c} \forall (y^0, z^0) \in (f(x^0), g(x^0))'_u : \exists (\xi^0, \eta^0) \in C' \times K' : \\ (\xi^0, \eta^0) \neq (0,0), \quad \langle \eta^0, g(x^0) \rangle = 0 \quad \text{and} \quad \langle \xi^0, y^0 \rangle + \langle \eta^0, z^0 \rangle \geq 0. \end{array}$$

(Sufficient Conditions) *Let $x^0 \in \mathbb{R}^n$ and suppose that for each $u \in S$ the following condition is satisfied:*

$$\mathbb{S}'_{0,l}: \quad \begin{array}{c} \forall (y^0, z^0) \in (f(x^0), g(x^0))'_u : \exists (\xi^0, \eta^0) \in C' \times K' : \\ (\xi^0, \eta^0) \neq (0,0), \quad \langle \eta^0, g(x^0) \rangle = 0 \quad \text{and} \quad \langle \xi^0, y^0 \rangle + \langle \eta^0, z^0 \rangle > 0. \end{array}$$

Then x^0 is an isolated minimizer of first order for problem (1).

Conversely, if x^0 is an isolated minimizer of first order for problem (1) and the constraint qualification $\mathbb{Q}_{0,l}(x^0)$ holds, then condition $\mathbb{S}'_{0,l}$ is satisfied.

Proof of the Necessary Conditions

Let $u \in S$ and $(y^0, z^0) \in (f(x^0), g(x^0))'_u$. According to Lemma 3 we have $(y^0, z^0) \notin -\operatorname{int}(C \times K(x^0)) = -(\operatorname{int}(C) \times \operatorname{int}(K(x^0)))$, whence there exists

$$(\xi^0, \eta^0) \in (C \times K(x^0))' \setminus \{(0,0)\} = C' \times K'(x^0) \setminus \{(0,0)\},$$

such that $(\xi^0, \eta^0)(y^0, z^0) = (\xi^0, y^0) + (\eta^0, z^0) \geq 0$, which proves $\mathbb{N}'_{0,l}$ (let us underline that $\eta^0 \in K'(x^0)$ is equivalent to $\eta^0 \in K'$ and $\langle \eta^0, g(x^0) \rangle = 0$). □

Proof of the Sufficient Conditions

Assume on the contrary, that x^0 is not an isolated minimizer of first order and choose a monotone decreasing sequence $\varepsilon_k \to +0$. From the assumptions, there exist sequences $t_k \to +0$ and $u^k \in S$, such that $g(x^0 + t_k u^k) \in -K$ and

$$D(f(x^0 + t_k u^k) - f(x^0), -C) = \max_{\xi \in \Gamma_{C'}} \langle \xi, f(x^0 + t_k u^k) - f(x^0) \rangle < \varepsilon_k t_k.$$

We may assume that $0 < t_k < r$ and both f and g are Lipschitz with constant L in $x^0 + r\,\mathrm{cl}\,B$. Passing to a subsequence, we may assume also that $u^k \to u^0$ and that equalities (11) hold with $u = u^0$. From them we have $(y^0, z^0) \in (f(x^0), g(x^0))'_{u^0}$.

Denote $z^k = (1/t_k)(g(x^0 + t_k u^k) - g(x^0))$ and $z^{0,k} = (1/t_k)(g(x^0 + t_k u^0) - g(x^0))$. We show that $z^k \to z^0$. This follows from the estimation

$$\left\|z^k - z^0\right\| \le \frac{1}{t_k}\left\|g(x^0 + t_k u^k) - g(x^0 + t_k u^0)\right\| + \left\|z^{0,k} - z^0\right\| \le L\left\|u^k - u^0\right\| + \left\|z^{0,k} - z^0\right\|.$$

We show that $z^0 \in -K(x^0)$. For this purpose we must check that $\langle \eta, z^0 \rangle \le 0$ for $\eta \in K^{'}(x^0)$. We observe that $x^0 + t_k u^k$ feasible and $\eta \in K^{'}(x^0)$ give $\langle \eta, g(x^0 + t_k u^k) \rangle \le 0$, whence

$$\left\langle \eta, \frac{1}{t_k}(g(x^0 + t_k u^k) - g(x^0)) \right\rangle = \frac{1}{t_k}\langle \eta, g(x^0 + t_k u^k) \rangle \le 0.$$

A passing to a limit gives $\langle \eta, z^0 \rangle \le 0$.

In order to obtain a contradiction, we show that $\mathbb{S}^{'}_{0,1}$ is not satisfied at x^0 for $u = u^0$ and (y^0, z^0) as above. Denote $y^k = (1/t_k)(f(x^0 + t_k u^k) - f(x^0))$ and $y^{0,k} = (1/t_k)(f(x^0 + t_k u^0) - f(x^0))$. We have $y^k \to y^0$, which follows from the estimation

$$\begin{aligned} &\left\|y^k - y^0\right\| \le \frac{1}{t_k}\left\|f(x^0 + t_k u^k) - f(x^0 + t_k u^0)\right\| + \left\|y^{0,k} - y^0\right\| \\ &\le L\left\|u^k - u^0\right\| + \left\|y^{0,k} - y^0\right\|. \end{aligned} \tag{12}$$

Let $\overline{\xi} \in \Gamma_{C^{'}}$. Then

$$\begin{aligned} \langle \overline{\xi}, y^k \rangle &= \frac{1}{t_k}\langle \overline{\xi}, f(x^0 + t_k u^k) - f(x^0) \rangle \le \frac{1}{t_k}\max_{\xi \in \Gamma_{C^{'}}} \langle \xi, f(x^0 + t_k u^k) - f(x^0) \rangle \\ &< \frac{1}{t_k}\varepsilon_k t_k = \varepsilon_k. \end{aligned}$$

Passing to a limit with $k \to \infty$ we get $\langle \overline{\xi}, y^0 \rangle \le 0$ for arbitrary $\overline{\xi} \in \Gamma_{C^{'}}$. Therefore $\langle \xi, y^0 \rangle \le 0$ for arbitrary $\xi \in C^{'}$. The latter for $\xi \ne 0$ follows from $\langle \xi, y^0 \rangle = \|\xi\| \langle (\xi/\|\xi\|), y^0 \rangle \le 0$. At the same time $\langle \eta, z^0 \rangle \le 0$ for all $\eta \in K^{'}(x^0)$. Therefore for all $\xi \in C^{'}$ and $\eta \in K^{'}(x^0)$ we have $\langle \xi, y^0 \rangle + \langle \eta, , z^0 \rangle \le 0$, whence the opposite strong inequality from $\mathbb{S}^{'}_{0,1}$ cannot have place. □

Reversal of the Sufficient Conditions Let x^0 be an isolated minimizer of first order for problem (1), which means that $g(x^0) \in -K$ and there exists $r > 0$ and $A > 0$ such that $g(x) \in -K$ and $\|x - x^0\| \le r$ implies

$$D(f(x) - f(x^0), -C) = \max_{\xi \in \Gamma_{C'}} \langle \xi, f(x) - f(x^0) \rangle \ge A \|x - x^0\|. \tag{13}$$

Let $u^0 \in S$ and $(y^0, z^0) \in (f(x^0), g(x^0))'_u$ be determined by (11) with $u = u^0$. We may assume that $0 < t_k < r$ and that f and g are Lipschitz with constant L on $x^0 + r \operatorname{cl} B$.

One of the following two cases has place:

1°. $z^0 \notin -K(x^0)$. Then there exists $\eta^0 \in K'(x^0)$, such that $\langle \eta^0, z^0 \rangle > 0$ (obviously, the strong inequality gives $\eta^0 \ne 0$). Putting $\xi^0 = 0$, we get the pair (ξ^0, η^0) satisfying condition $\mathbb{S}'_{0,l}$.

2°. $z^0 \in -K(x^0)$. Then from the constraint qualification $\mathbb{Q}_{0,l}(x^0)$ it follows $g(x^0 + t_k u^k) \in -K$ for some sequence $u^k \to u^0$ and all sufficiently large k. Taking a subsequence, we may assume that this holds for all k. From inequality (13) we get that for every k there exists $\xi^{0,k} \in \Gamma_{C'}$ (and hence $\xi^{0,k} \in C'$, $\xi^{0,k} \ne 0$), such that

$$\left\langle \xi^{0,k} \frac{1}{t_k} \left(f(x^0 + t_k u^k) - f(x^0) \right) \right\rangle \ge A \|u^k\|.$$

Putting $y^k = (1/t_k)(f(x^0 + t_k u^k) - f(x^0))$ and $y^{0,k} = (1/t_k)(f(x^0 + t_k u^0) - f(x^0))$, we have $y^k \to y^0$, which follows from (12). Passing to the limit we can assume $\xi^{0,k} \to \xi^0 \in \Gamma_{C'}$ and we get $\langle \xi^0, y^0 \rangle \ge A > 0$. Putting $\eta^0 = 0$, we get the pair (ξ^0, η^0) satisfying condition $\mathbb{S}'_{0,l}$. □

Obviously, the proved theorem is valid also for the unconstrained problem (2). We give this case, since then some of the conditions simplify.

Theorem 2. Let f be a $C^{0,l}$ function and consider problem (2).

(Necessary Conditions) *Let x^0 be w-minimizer of problem (2). Then for each $u \in S$ and $y^0 \in f'_u(x^0)$ there exists $\xi^0 \in C' \setminus \{0\}$ such that $\langle \xi^0, y^0 \rangle \ge 0$.*

(Sufficient Conditions) *Let $x^0 \in \mathbb{R}^n$. Suppose that for each $u \in S$ and*

$y^0 \in f_u'(x^0)$ there exists $\xi^0 \in C' \setminus \{0\}$ such that $\langle \xi^0, y^0 \rangle > 0$. Then x^0 is an isolated minimizer of first order for problem (2).

Conversely, the given condition is not only sufficient, but also necessary for the point x^0 to be an isolated minimizer of first order.

The following simple example illustrates Theorem 1 in practice.

Example 1. *Consider the unconstrained problem (2) with*

$$f: \mathbb{R} \to \mathbb{R}^2, \quad f(x) = \begin{cases} (x, -2x) & , \quad x \geq 0, \\ (2x, -x) & , \quad x < 0, \end{cases}$$

and $C = \mathbb{R}^2_+$. The function f is $C^{0,1}$ but not C^1 and the point $x^0 = 0$ is both p-minimizer and isolated minimizer of first order. The latter can be established on the base of the Sufficient Conditions of Theorem 1.

Here the positive polar cone is $C' = \mathbb{R}^2_+$. For $u = 1$ we have $y^0 = f_u'(x^0) = (1, -2)$ and $\langle \xi^0, y^0 \rangle = \xi_1^0 - 2\xi_2^0 > 0$ if we choose $\xi^0 = (1,0) \in \mathbb{R}^2_+ \setminus \{(0,0)\}$. For $u = -1$ we have $y^0 = f_u'(x^0) = (-2,1)$ and $\langle \xi^0, y^0 \rangle = -2\xi_1^0 + \xi_2^0 > 0$ if we choose $\xi^0 = (0,1) \in \mathbb{R}^2_+ \setminus \{(0,0)\}$.

The constraint qualification $\mathbb{Q}_{0,1}(x^0)$ is of Kuhn-Tucker type [12]. One may be astonished, that in the hypothesis of $\mathbb{Q}_{0,1}(x^0)$ we have $z^0 \in -K(x^0)$, while in the conclusion $g(x^0 + t_k u^0) \in -K$ it stands K instead of $K(x^0)$. If the cone K is polyhedral, we may take in the conclusion $g(x^0 + t_k u^0) \in -K(x^0)$, but in general with such a weaker conclusion the reversal of the Sufficient Conditions of Theorem 1 is not true. This is shown in the next example.

Example 2. *Let $f: \mathbb{R} \to \mathbb{R}$, $g: \mathbb{R} \to \mathbb{R}^3$ with $C = \mathbb{R}_+$, $K = \{z \in \mathbb{R}^3 \mid z_3^2 \geq z_1^2 + z_2^2\}$ and $f(x) = x^2$, $g(x) = (x|x|, -1, -1)$. Then f and g are C^1 functions, $x^0 = 0$ is an isolated minimizer of first order, $\mathbb{Q}_{0,1}(x^0)$ does not hold, but we have similar condition with $g(x^0 + t_k u^0) \in -K(x^0)$ in the conclusion, instead of $g(x^0 + t_k u^0) \in -K$. At the same time, whatever $u \in \mathbb{R}$ be, there is no pair $(\xi^0, \eta^0) \in C' \times K'(x^0)$ for which $\langle \xi^0, f'(x^0)u \rangle + \langle \eta^0, g'(x^0)u \rangle > 0$.*

Here x^0 is the only feasible point, and according to the definition x^0 is an isolated minimizer of first order. (This means $D(f(x) - f(x^0), -C) \geq A\|x - x^0\|$ for $x \in U \cap g^{-1}(-K)$, which is true, since $U \cap g^{-1}(-K) = \{x^0\}$). The index set $K(x^0)$ is a half-space determined by the unique tangent plane to the cone $-K$ at $g(x^0)$, whence the modified constraint qualification is checked immediately. More precisely, $-K(x^0) = \{z \in \mathbb{R}^3 \mid -z_2 + z_3 \geq 0\}$. For any $u \in \mathbb{R}$ we have $\lim_k (1/t_k)(g(x^0 + t_k u) - g(x^0)) = (0,0,0) \in -K(x^0)$. At the same time

$g(x^0 + t_k u) = (t_k^2 u \,|\, u\,|, -1, -1) \notin -K$, but $g(x^0 + t_k u) \in -K(x^0)$. Now, for any $u \in \mathbb{R}$ we have $f'(x^0)u = 0$, $g'(x^0)u = (0,0,0)$ and therefore $\langle \xi^0, f'(x^0)u \rangle + \langle \eta^0, g'(x^0)u \rangle = 0$ for all pairs (ξ^0, η^0).

If g is Fréchet differentiable at x^0, then instead of the constraint qualification $\mathbb{Q}_{0,1}(x^0)$ we may consider the constraint qualification $\mathbb{Q}_1(x^0)$ given below.

$\mathbb{Q}_1(x^0)$: If $g(x^0) \in -K$ and $g'(x^0)u^0 = z^0 \in -K(x^0)$ then there exists $\delta > 0$ and a differentiable injective function $\varphi : [0, \delta] \to -K$ such that $\varphi(0) = x^0$ and $\varphi'(0) = g'(x^0)u^0$.

In the case of a polyhedral cone K in $\mathbb{Q}_1(x^0)$ the requirement $\varphi : [0, \delta] \to -K$ can be replaced by $\varphi : [0, \delta] \to -K(x^0)$. This condition coincides with the classical Kuhn-Tucker constraint qualification (compare with Mangasarian [15, p. 102]).

The next theorem is a reformulation of Theorem 1 for C^1 problems, that is problems with f and g being C^1 functions.

Theorem 3. *Let f, g be C^1 functions and consider problem (1).*

(Necessary Conditions) *Let x^0 be a w-minimizer of problem (1). Then for each $u \in S$ the following condition is satisfied:*

$\mathbb{N}_1'$: $\exists (\xi^0, \eta^0) \in C' \times K' \setminus \{(0,0)\}$:
$\langle \eta^0, g(x^0) \rangle = 0$ and $\langle \xi^0, f'(x^0)u \rangle + \langle \eta^0, g'(x^0)u \rangle \geq 0$.

(Sufficient Conditions) *Let $x^0 \in \mathbb{R}^n$. Suppose that for each $u \in S$ the following condition is satisfied:*

$\mathbb{S}_1'$: $\exists (\xi^0, \eta^0) \in C' \times K' \setminus \{(0,0)\}$:
$\langle \eta^0, g(x^0) \rangle = 0$ and $\langle \xi^0, f'(x^0)u \rangle + \langle \eta^0, g'(x^0)u \rangle > 0$.

Then x^0 is an isolated minimizer of first order for problem (1).

Conversely, if x^0 is an isolated minimizer of first order for problem (1) and the constraint qualification $\mathbb{Q}_1(x^0)$ holds, then condition $\mathbb{S}_1'$ is satisfied.

We underline without proof, that Theorem 3 remains true assuming for f and g only Fréchet differentiable at x^0, instead of being C^1.

The pairs of vectors (ξ^0, η^0) are usually referred to as the Lagrange multipliers. Here we have different Lagrange multipliers for different $u \in S$ (and different $(y^0, z^0) \in (f(x^0), g(x^0))'_u$). The natural question arises,

whether a common pair (ξ^0,η^0) can be chosen to all directions. The next example shows that the answer is negative even for C^1 problems.

Example 3. *Let* $f:\mathbb{R}^2\to\mathbb{R}^2$, $f(x_1,x_2)=(x_1,x_1^2+x_2^2)$, *and* $g:\mathbb{R}^2\to\mathbb{R}^2$, $g(x_1,x_2)=(x_1,x_2)$. *Define* $C=\{y\in(y_1,y_2)\in\mathbb{R}^2\mid y_1=0\}$, $K=\mathbb{R}^2$. *Then* f *and* g *are* C^1 *functions and the point* $x^0=(0,0)$ *is a* w*-minimizer of problem (1) (in fact* x^0 *is also isolated minimizer of second order, but not isolated minimizer of first order). At the same time the only pair* $(\xi^0,\eta^0)\in C^{'}\times K^{'}$ *for which* $\langle\xi^0,f^{'}(x^0)u\rangle+\langle\eta^0,g^{'}(x^0)u\rangle\geq 0$ *for all* $u\in S$ *is* $\xi^0=(0,0)$ *and* $\eta^0=(0,0)$.

The point x^0 is a w-minimizer, since $\operatorname{int}C=0$, whence each feasible point is w-minimizer. We have $f^{'}(x)u=(u_1,2x_1u_1+2x_2u_2)$, so that $f^{'}(x^0)u=(u_1,0)$, and $g^{'}(x^0)u=u$. The positive polar cones are $C^{'}=\{\xi\in\mathbb{R}^2\mid\xi_2=0\}$ and $K^{'}=\{0\}$. If $\xi^0=(\xi_1^0,\xi_2^0)\in C^{'}$ and $\eta^0=(\eta_1^0,\eta_2^0)\in K^{'}$ satisfy the desired inequality, then $\eta^0=(0,0)$, $\xi^0=(\xi_1^0,0)$ and the inequality turns into $\xi_1^0u_1\geq 0$, which should be true for all $u_1\in\mathbb{R}$. This gives $\xi_1^0=0$ and finally $\xi^0=(0,0)$ and $\eta^0=(0,0)$.

The next Theorem 4 guarantees, that in the case when x^0 is a *rw*-minimizer of the C^1 problem (1), a nonzero pair (ξ^0,η^0) exists, which satisfies the Necessary Conditions of Theorem 1 and which is common for all directions. In order to prepare the proof, we need the following two lemmas.

Lemma 4. *Let* $f:\mathbb{R}^n\to\mathbb{R}^m$ *be a* $C^{0,1}$ *function and let* $\mathbf{L}\subset\mathbb{R}^m$ *be a subspace. Denote* $\overline{f}=\pi_{\mathbf{L}}\circ f$. *Then* $\overline{f}$ *is a* $C^{0,1}$ *function and* $\overline{f}_u^{'}(x^0)=\pi_{\mathbf{L}}\circ f_u^{'}(x^0)$. *Similarly, if* f *is a* C^1 *function, then* $\overline{f}$ *is a* C^1 *function and* $\overline{f}^{'}(x^0)u=\pi_{\mathbf{L}}\circ f^{'}(x^0)u$.

Proof The function $\overline{f}$ is locally Lipschitz, hence $C^{0,1}$, as a composition of a bounded linear function and a locally Lipschitz function.

Let $y^0\in f_u^{'}(x^0)$ and $y^0=\lim_k(1/t_k)(f(x^0+t_ku)-f(x^0))$. Since the projection commutes with the passing to a limit and with the linear operations, we see that

$$\pi_{\mathbf{L}}\circ y^0=\lim_k\frac{1}{t_k}((\pi_{\mathbf{L}}\circ f)(x^0+t_ku)-(\pi_{\mathbf{L}}\circ f)(x^0))\in\overline{f}_u^{'}(x^0).$$

Conversely, let $\overline{y}^0=\lim_k(1/t_k)(\overline{f}(x^0+t_ku)-\overline{f}(x^0))$. From f locally Lipschitz, it follows that there exists a subsequence $\{t_{k'}\}$ of $\{t_k\}$, such that $\lim_{k'}(1/t_{k'})(f(x^0+t_{k'}u)-f(x^0))=y^0$. Now $y^0\in f_u^{'}(x^0)$ and $\overline{y}^0=\pi_{\mathbf{L}}\circ y^0\in\pi_{\mathbf{L}}\circ f_u^{'}(x^0)$.

The case of $f\in C^1$ is treated similarly. □

Lemma 5. *Consider problem (1) with f and g being $C^{0,1}$ functions. If x^0 is a rw-minimizer and $(y^0,z^0)\in(\overline{f}(x^0),\overline{g}(x^0))'_u$ (here $\overline{f}=\pi_{\mathbf{L}_C}\circ f$ and $\overline{g}=\pi_{\mathbf{L}_K}\circ g$), then $(y^0,z^0)\notin -(\mathrm{ri}\,C\times\mathrm{ri}(K(x^0)\cap\mathbf{L}_K))$.*

The proof is omitted, since it nearly repeats that of Lemma 3, but relating the considerations to the phase space $\mathbf{L}_C\times\mathbf{L}_K$ instead of $\mathbb{R}^m\times\mathbb{R}^p$.

Theorem 4. Necessary Conditions *Let f, g be C^1 functions and let x^0 be a rw-minimizer of problem (1). Then there exists a pair $(\xi^0,\eta^0)\in C'_{\mathbf{L}_C}\times K'_{\mathbf{L}_K}\setminus\{(0,0)\}$ such that $\langle\eta^0,g(x^0)\rangle=0$ and $\langle\xi^0,f'(x^0)u\rangle+\langle\eta^0,g'(x^0)u\rangle=0$ for all $u\in\mathbb{R}^n$. The latter equality could be written also as $\xi^0 f'(x^0)+\eta^0 g'(x^0)=0$.*

Proof Put $\overline{f}=\pi_{\mathbf{L}_C}\circ f$ and $\overline{g}=\pi_{\mathbf{L}_K}\circ g$. According to Lemma 5, $(\overline{f}'(x^0)u,\overline{g}'(x^0)u)\notin -(\mathrm{ri}\,C\times\mathrm{ri}(K(x^0)\cap\mathbf{L}_K))\neq\varnothing$ for all $u\in\mathbb{R}^n$. Therefore the convex set $M=\{(\overline{f}'(x^0)u,\overline{g}'(x^0)u)\,|\,u\in\mathbb{R}^n\}\subset\mathbf{L}_C\times\mathbf{L}_K$ does not intersect the non-empty interior (relative to $\mathbf{L}_C\times\mathbf{L}_K$) of the convex set $-C\times(K(x^0)\cap\mathbf{L}_K)$. From the Separation Theorem there exists a nonzero pair $(\xi^0,\eta^0)\in C'_{\mathbf{L}_C}\times K'_{\mathbf{L}_K}$ such that $\langle\xi^0,\overline{f}'(x^0)u\rangle+\langle\eta^0,\overline{g}'(x^0)u\rangle\geq 0$ for all $u\in\mathbb{R}^n$. This leads to an equality, since

$$0\leq\langle\xi^0,\overline{f}'(x^0)(-u)\rangle+\langle\eta^0,\overline{g}'(x^0)(-u)\rangle=-\left(\langle\xi^0,\overline{f}'(x^0)u\rangle+\langle\eta^0,\overline{g}'(x^0)u\rangle\right)\leq 0.$$

Since $\xi^0\in\mathbf{L}_C$ we have $\langle\xi^0,\overline{f}'(x^0)u\rangle=\langle\xi^0,f'(x^0)u\rangle$. Indeed, applying Lemma 4, we get

$$\langle\xi^0,\overline{f}'(x^0)u\rangle=\langle\xi^0,(\pi_{\mathbf{L}_C}\circ f)'(x^0)u\rangle=\langle\xi^0,\pi_{\mathbf{L}_C}\circ f'(x^0)u\rangle=\langle\xi^0,f'(x^0)u\rangle.$$

Similarly, since $\eta^0\in\mathbf{L}_K$, we get $\langle\xi^0,\overline{g}'(x^0)u\rangle=\langle\xi^0,g'(x^0)u\rangle$. Finally $\eta^0\in(K(x^0)\cap\mathbf{L}_K)'_{\mathbf{L}_K}$ gives $0=\langle\eta^0,\overline{g}(x^0)\rangle=\langle\eta^0,\pi_{\mathbf{L}_K}\circ g(x^0)\rangle=\langle\eta^0,g(x^0)\rangle$. □

If $\mathrm{int}\,C=\varnothing$ each feasible point of problem (1) is a w-minimizer and the Necessary Conditions are trivially satisfied. In this case a more essential information is that x^0 is a rw-minimizer. The next Theorem 5 generalizes the Necessary Conditions part of Theorem 1 to relative concepts. Obviously, the Sufficient Conditions part admits also a generalization, which is not given here.

Theorem 5. (First-order conditions) *Consider problem (1) with f, g being $C^{0,1}$ functions and C and K closed convex cones.*

(Necessary Conditions) *Let x^0 be rw-minimizer of problem (1). Then for each $u \in S$ the following condition is satisfied:*

$$\forall (y^0, z^0) \in (f(x^0), g(x^0))'_u : \exists (\xi^0, \eta^0) \in C'_{\mathbf{L}_C} \times K'_{\mathbf{L}_K} :$$
$$(\xi^0, \eta^0) \neq (0,0), \quad \langle \eta^0, g(x^0) \rangle = 0 \quad \text{and} \quad \langle \xi^0, y^0 \rangle + \langle \eta^0, z^0 \rangle \geq 0.$$

We omit the proof. In principle it repeats the proof of the Necessary Conditions of Theorem 1 replacing the phase space from $\mathbb{R}^m \times \mathbb{R}^p$ to $\mathbf{L}_C \times \mathbf{L}_K$, replacing the considered problem from (1) to (4) and making use of Lemma 4.

4. ISOLATED MINIMIZERS AND PROPER EFFICIENCY

Consider the unconstrained problem (2) with a $C^{0,1}$ function f. According to Proposition 4 if x^0 is an isolated minimizer of first order, then x^0 is p-minimizer. It is natural to ask, whether the converse is true. Example 4 gives a negative answer of this question. We conclude the paper with Proposition 7, which as an application of Theorem 1 reverts the result of Proposition 4 under some additional assumption.

Example 4. *Let $t_k \to +0$, $k = 0,1,\ldots$, be a strictly decreasing sequence with $t_0 = +\infty$. Define the function $h : \mathbb{R} \to \mathbb{R}$,*

$$h(t) = \begin{cases} \min\left(t_{k-1} - |t|, |t| - t_k\right) & , \quad t_k \leq |t| \leq t_{k-1}, \\ 0 & , \quad t = 0. \end{cases}$$

Consider the unconstrained problem (2) with $f : \mathbb{R} \to \mathbb{R}^2$, $f(x) = (h(x), h(x))$ and $C = \mathbb{R}^2_+$. Then $x^0 = 0$ is p-minimizer, but not an isolated minimizer of first order.

The function f is $C^{0,1}$, since h is $C^{0,1}$. The latter follows by the easy-to-prove inequality $|h(t') - h(t'')| \leq |t' - t''|$, $t', t'' \in \mathbb{R}$.

According to Proposition 6 and Corollary 1,if x^0 is an isolated minimizer of first order for (2), we should have, that x^0 is an isolated minimizer of first order for the function $\varphi_0(x) = \min(f_1(x) - f_1(x^0), f_2(x) - f_2(x^0)) = h(x)$.

However, this is not the case, since for $x^k = t_k \to x^0 = 0$ we have $\varphi_0(x^k) = h(t_k) = 0$.

The point x^0 is p-minimizer. Indeed, let $\tilde{C} = \{y \in \mathbb{R}^2 \mid y_1 + y_2 \geq 0\}$. Then $\operatorname{int}\tilde{C} = \{y \in \mathbb{R}^2 \mid y_1 + y_2 > 0\} \supset C \setminus \{0\} = R_+^2 \setminus \{(0,0)\}$, $f(x) = (h(x), h(x)) \in \mathbb{R}_+^2 = C$ and $\mathbb{R}_+^2$ is disjoint from $-\operatorname{int}\tilde{C} = \{y \in \mathbb{R}^2 \mid y_1 + y_2 < 0\}$.

By a slight modification of this example we can see, that even the additional assumption x^0 strong e-minimizer does not guarantee that x^0 is an isolated minimizer of first order.

Example 5. *Let h be as in Example 4. Consider problem (2) with $f: \mathbb{R}^2 \to \mathbb{R}$, $f(x) = (h(x) + x^2, h(x) + x^2)$ and $C = \mathbb{R}_+^2$. Then f is $C^{0,1}$, $x^0 = 0$ is both strong e-minimizer and p-minimizer, but not an isolated minimizer of first order.*

Here $\varphi_0(x) = h(x) + x^2$ has $x^0 = 0$ as a strong minimizer, but not as an isolated minimizer of first order.

Proposition 7. *Let f be a $C^{0,1}$ function and consider the unconstrained problem (2). Let x^0 be a p-minimizer, which has the property that $y^0 \neq 0$ for each $y^0 \in f_u'(x^0)$ and arbitrary $u \in S$. Then x^0 is an isolated minimizer of first order.*

Proof Since x^0 is a p-minimizer, therefore there exists a closed convex cone $\tilde{C}$, such that $\operatorname{int}\tilde{C} \supset C \setminus \{0\}$ and $f(x) - f(x^0) \notin -\operatorname{int}\tilde{C}$. According to the Necessary Conditions of Theorem 1 (and Theorem 2), this means, that for each $u \in S$ and $y^0 \in f_u'(x^0)$, there exists $\tilde{\xi}^0 \in \tilde{C}' \setminus \{0\}$, such that $\langle \tilde{\xi}^0, y^0 \rangle \geq 0$. This inequality, together with the made assumptions shows that $y^0 \notin -\operatorname{int}\tilde{C} \cup \{0\}$. Since $C \subset \operatorname{int}\tilde{C} \cup \{0\}$, we see that $y^0 \notin -C$. This implies, that there exists $\xi^0 \in C'$, such that $\langle \xi^0, y^0 \rangle > 0$. According to the Sufficient Conditions of Theorem 1 (and Theorem 2), the point x^0 is an isolated minimizer of first order. □

REFERENCES

[1] T. Amahroq, A. Taa: On Lagrange-Kuhn-Tucker multipliers for multiobjective optimization problems. Optimization 41 (1997), 159–172.

[2] A. Auslender: Stability in mathematical programming with nondifferentiable data. SIAM J. Control Optim. 22 (1984), 239–254.

[3] M. Ciligot-Travain: On Lagrange-Kuhn-Tucker multipliers for Pareto optimization problems. Numer. Funct. Anal. Optim. 15 (1994), 689–693.

[4] I. Ginchev, A. Guerraggio, M. Rocca: From scalar to vector optimization. Appl. Math., to appear.

[5] I. Ginchev, A. Hoffmann: Approximation of set-valued functions by single-valued one. Discussiones Mathematicae, Differential Inclusions, Control and Optimization 22 (2002), 33–66.

[6] A. Guerraggio, D. T. Luc: Optimality conditions for $C^{1,1}$ vector optimization problems. J. Optim. Theory Appl. 109 No. 3 (2001), 615–629.

[7] J.-B. Hiriart-Urruty, J.-J Strodiot, V. Hien Nguen: Generalized Hessian matrix and second order optimality conditions for problems with $C^{1,1}$ data. Appl. Math. Optim. 11 (1984), 169–180.

[8] J.-B. Hiriart-Urruty: New concepts in nondifferentiable programming. Analyse non convexe, Bull. Soc. Math. France 60 (1979), 57–85.

[9] J.-B. Hiriart-Urruty: Tangent cones, generalized gradients and mathematical programming in Banach spaces. Math. Oper. Res. 4 (1979), 79–97.

[10] F. John: Extremum problems with inequalities as subsidiary conditions. In: K. O. Friedrichs, O. E. Neugebauer, J. J. Stroker (eds.), Studies and Essays, Courant Anniversary Volume, pp. 187–204, Interscience Publishers, New York, 1948.

[11] D. Klatte, K. Tammer: On the second order sufficient conditions to perturbed $C^{1,1}$ optimization problems. Optimization 19 (1988), 169–180.

[12] H. W. Kuhn, A. W. Tucker: Nonlinear programming. In: J. Neyman (Ed.), Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, pp. 481–492, University of California Press. Berkeley, California, 1951

[13] L. Liu, P. Neittaanmäki, M. Křížek: Second-order optimality conditions for nondominated solutions of multiobjective programming with $C^{1,1}$ data. Appl. Math. 45 (2000), 381–397.

[14] D. T. Luc: Taylor's formula for $C^{k,1}$ functions. SIAM J. Optimization 5 No. 3 (1995), 659–669.

[15] O. L. Mangasarian: Nonlinear programming. Society for Industrial and Applied Mathematics, Philadelphia, 1994.

[16] E. Miglierina: Characterization of solutions of multiobjective optimization problems. Rendiconti Circolo Matematico di Palermo, 50 (2001), 153–164.

[17] E. Miglierina, E. Molho: Scalarization and its stability in vector optimization. J. Optim. Theory Appl., 114 (2002), 657–670.

[18] M. Pappalardo, W. Stocklin: Necessary optimality conditions in nondifferentiable vector optimization. Optimization 50 (2001), 233–251.

[19] R. T. Rockafellar: Convex analysis. Princeton University Press, Princeton, 1970.

[20] X. Q. Yang , V. Jeyakumar: Generalized second-order directional derivatives and optimization with $C^{1,1}$ functions. Optimization 26 (1992), 165–185.

[21] X. Q. Yang: Second-order conditions in $C^{1,1}$ optimization with applications. Numer. Funct. Anal. Optim. 14 (1993), 621–632.

[22] A. Zaffaroni: Degrees of efficiency and degrees of minimality. SIAM J. Control Optim. 42 No. 3 (2003), 1071-1086.

GLOBAL REGULARITY FOR SOLUTIONS TO DIRICHLET PROBLEM FOR ELLIPTIC SYSTEMS WITH NONLINEARITY $q \geq 2$ AND WITH NATURAL GROWTH

S. Giuffrè[1] and G. Idone[2]
D.I.M.E.T., Faculty of Engineering, University of Reggio Calabria, Reggio Calabria, Italy[1,2]

Abstract: Hölder regularity up to the boundary of the solutions to a nonhomogeneous Dirichlet problem for second order discontinuous elliptic systems with nonlinearity $q \geq 2$ and with natural growth is proved when $n = q$.

2000 Mathematics Subject Classification. 35J65; 35J55.

Key words: nonlinear elliptic systems, global Hölder regularity, higher gradient summability.

1. INTRODUCTION

In this paper we study the global Hölder continuity in $\overline{\Omega}$ of a solution $u \in H^{1,q}(\Omega) \cap L^{\infty}(\Omega)$ to the following Dirichlet problem

$$\begin{cases} u - g \in H_0^{1,q}(\Omega) \cap L^{\infty}(\Omega) \\ \sum_{i=1}^{n} D_i a^i(x,u,Du) = -B(x,u,Du) \quad \text{in } \Omega \end{cases} \tag{1.1}$$

where Ω is a bounded open set in $\mathbb{R}^n$, $n \geq 2$, q is a real number ≥ 2 and $g \in H^{1,s}(\Omega) \cap L^\infty(\Omega)$, $s > q$.

For solution u to (1.1) we mean that $u = g + w$, where $w \in H_0^{1,q}(\Omega) \cap L^\infty(\Omega)$ is such that

$$\int_\Omega \sum_{i=1}^{n} \left(a^i(x, w+g, Dw+Dg) \mid D_i\varphi\right) dx = \int_\Omega \left(B(x, w+g, Dw+Dg) \mid \varphi\right) dx, \quad \forall \varphi \in H_0^{1,q}(\Omega) \cap L^\infty(\Omega). \tag{1.2}$$

If $u, v \in \mathbb{R}^N$, $(u \mid v)$ denotes the inner product in $\mathbb{R}^N$. We set $p = (p^1, ..., p^n)$, with $p^i \in \mathbb{R}^N$; p is a typical vector of $\mathbb{R}^{nN}$.

For every $p \in \mathbb{R}^K$, $K \geq 1$, we set

$$V(p) = \left(1 + \|p\|^2\right)^{\frac{1}{2}} \quad \text{and} \quad W(p) = V^{\frac{q-2}{2}}(p)\, p. \tag{1.3}$$

Let $a^i(x,u,p)$, $i = 1,2,...,n$, be vectors of $\mathbb{R}^N$, defined on $\Omega \times \mathbb{R}^N \times \mathbb{R}^{nN}$, such that $a^i(x,u,p)$ are measurable in x and continuous in u, p, and $a^i(x,u,0) = 0 \quad \forall x \in \Omega, \ \forall u \in \mathbb{R}^N$. We assume that there exist two positive constants M, ν such that for all $x \in \Omega$, $u \in \mathbb{R}^N$, $p \in \mathbb{R}^{nN}$ it results

$$\left\|a^i(x,u,p)\right\| \leq M V^{q-2}(p) \|p\|, \qquad i = 1,...,n \tag{1.4}$$

$$\sum_{i=1}^{n} \left(a^i(x,u,p) \mid p^i\right) \geq \nu V^{q-2}(p) \|p\|^2. \tag{1.5}$$

Moreover we suppose that there exist two positive constants a, b, such that for all $x \in \Omega$, $u \in \mathbb{R}^N$, $p \in \mathbb{R}^{nN}$ it results

$$\|B(x,u,p)\| \leq a + b \|W(p)\|^2, \tag{1.6}$$

and, if u is a solution to Problem (1.1), also the following smallness condition holds

$$2b \|u - g\|_{\infty,\Omega} < \nu, \tag{1.7}$$

(for the notations see section 2).

Condition (1.6) is called natural growth condition and the aim of this paper is to study the global Hölder continuity in $\overline{\Omega}$ of the solutions to the

Dirichlet problem (1.1) under the above assumptions. A smallness condition of type (1.7) is necessary in order to obtain the regularity result of the solutions, in virtue of the counter-example provided by [6]; however it seems that the optimal smallness condition could be $b\|u-g\|_{\infty,\Omega}<\nu$. Moreover it is well known that it is not possible to obtain the global Hölder continuity in $\overline{\Omega}$ for each value of the dimension as the counter-examples in [6], [9], [11], [16] show. Therefore, taking into account the above counter-examples and the general form of the coefficients $a^i(x,u,Du)$ of problem (1.1), we can expect the global Hölder continuity of solutions to the Dirichlet problem (1.1) only for $n \le q$. As for $n<q$ the desired regularity derives from the Sobolev imbedding theorems, our goal remains only in proving the regularity up to the boundary for $n=q$. We achieve this result by means of the following theorem on higher global integrability of the gradient:

Theorem 1.1 *Assume that conditions (1.4), (1.5), (1.6) and (1.7) are fulfilled. Let* $\partial\Omega$ *be of class* C^2 *and* $g \in H^{1,s}(\Omega)\cap L^\infty(\Omega)$, *with* $s>q\ge n$. *If* $u \in H^{1,q}(\Omega)\cap L^\infty(\Omega)$ *is a solution to Dirichlet Problem (1.1), then there exist a number* $r>1$ *such that*

$$u \in H^{1,qr}(\Omega).$$

From Theorem 1.1 we immediately derive the following corollary.

Corollary 1.1 *Under the same assumptions of Theorem 1.1, a solution* u *to Dirichlet Problem (1.1), for* $q=n$, *belongs to* $C^{0,\alpha}(\overline{\Omega})$, *with* $\alpha = 1-\frac{1}{r}$.

An essential tool in order to achieve the global higher summability of Du is to get the so called "Caccioppoli type inequality", both in the interior case and near the boundary.

In the general case the result we can expect if $q>n$ is only the so called "partial Hölder regularity", namely there exists a closed singular set Ω_0 such that u is Hölder continuous in $\Omega \setminus \Omega_0$ and, even if the trace of u on $\partial\Omega$ is smooth, there exists a closed singular set Σ_0 on $\partial\Omega$ such that u is Hölder continuous up to the boundary except for the points of Σ_0 (see [3], [4]; for nonlinearity $q=2$ see [1], [2], [5], [10], [12], [15], [18]). Moreover in particular cases it is possible to estimate the Hausdorff dimension of both the singular sets Ω_0 and Σ_0 (see for example [3], [12], [15]). This behaviour is analogous to the one we meet when we consider elliptic nonvariational systems, namely we obtain global Hölder continuity up to the boundary only for low values of n and partial Hölder continuity in the general case (see [7], [14]).

Finally we recall that in [17] the author obtains the global Hölder continuity up to the boundary for $n=q$ in the case of term B fulfilling a growth of the type $\|p\|^{q-1}$.

2. PRELIMINARY RESULTS

We define

$$B(x^0,\sigma)=\{x:\|x-x^0\|<\sigma\}; \tag{2.1}$$

moreover, if $x_n^0=0$,

$$B^+(x^0,\sigma)=\{x\in B(x^0,\sigma):x_n>0\}, \tag{2.2}$$

$$\Gamma(x^0,\sigma)=\{x\in B(x^0,\sigma):x_n=0\}. \tag{2.3}$$

We will simply write $B^+(\sigma)$, $\Gamma(\sigma)$ and Γ instead of $B^+(0,\sigma)$, $\Gamma(0,\sigma)$ and $\Gamma(0,1)$, respectively.

Through the present paper, Ω will denote a bounded open set of $\mathbb{R}^n$ with diameter d_Ω and with boundary $\partial\Omega$ of class C^2.

The notation $B(x^0,\sigma)\subset\subset\Omega$ means that $B(x^0,\sigma)\subset\Omega$.

Moreover if $u\in L^1(\mathcal{B})$ and $\mathcal{B}$ is a measurable set with $\operatorname{meas}\mathcal{B}\neq 0$, then

$$u_{\mathcal{B}}=\fint_{\mathcal{B}}u(x)dx=\frac{1}{\operatorname{meas}\mathcal{B}}\int_{\mathcal{B}}u(x)dx. \tag{2.4}$$

If $u\in L^\infty(\Omega)$, we define

$$\|u\|_{\infty,\Omega}=\operatorname{ess\,sup}_{\Omega}\|u(x)\|. \tag{2.5}$$

If $u\in C^{0,\alpha}\bar{\Omega})$, $0<\alpha\leq 1$, we set

$$[u]_{\alpha,\bar{\Omega}}=\sup_{x,y\in\Omega}\frac{\|u(x)-u(y)\|}{\|x-y\|^\alpha} \tag{2.6}$$

and we will say that $u\in C^{0,\alpha}(\Omega)$ if $u\in C^{0,\alpha}(K)$ for every compact subset $K\subset\Omega$.

For what follows we need the following Gehring-Giaquinta-Modica's Lemma.

Lemma 2.1 *If U and G are nonnegative functions on Ω such that*

$$U \in L^r(\Omega), \quad G \in L^s(\Omega), \quad 1 < r < s$$

and if, for every $B(x^0,\sigma) \subset B(x^0,2\sigma) \subset \Omega$, it results

$$\int_{B(x^0,\sigma)} U^r dx \le c\left\{\left[\int_{B(x^0,2\sigma)} U dx\right]^r + \int_{B(x^0,2\sigma)} G^r dx\right\}, \quad c > 1$$

then there exists $\varepsilon > 0$ such that $U \in L^t_{loc}(\Omega)$, $\forall t \in [r, r+\varepsilon)$ and

$$\left(-\int_{B(x^0,\sigma)} U^t dx\right)^{\frac{1}{t}} \le k\left\{\left[-\int_{B(x^0,2\sigma)} U^r dx\right]^{\frac{1}{r}} + \left[-\int_{B(x^0,2\sigma)} G^t dx\right]^{\frac{1}{t}}\right\}$$

where k and ε are positive constants depending only on c, r, s and n (see [1] p.125).

The estimate contained in the next Lemma will be also useful in the sequel.

Lemma 2.2 *There exists a positive constant $c(q)$ such that, $\forall p, \tilde{p} \in \mathbb{R}^N$ it results*

$$V^{q-2}(p+\tilde{p})\|p+\tilde{p}\| \le c(q)\,[V^{q-2}(p)\|p\| + V^{q-2}(\tilde{p})\|\tilde{p}\|] \tag{2.7}$$

Proof.

$$V^{q-2}(p+\tilde{p})\|p+\tilde{p}\| = \left(1+\|p+\tilde{p}\|^2\right)^{\frac{q-2}{2}}\|p+\tilde{p}\|$$

If $\|p\| \le \|\tilde{p}\|$, then

$$\left(1+\|p+\tilde{p}\|^2\right)^{\frac{q-2}{2}}\|p+\tilde{p}\| \le 2\left(1+4\|\tilde{p}\|^2\right)^{\frac{q-2}{2}}\|\tilde{p}\|$$

$$\le 2\,4^{\frac{q-2}{2}}\left(1+\|\tilde{p}\|^2\right)^{\frac{q-2}{2}}\|\tilde{p}\| = c(q)V^{q-2}(\tilde{p})\|\tilde{p}\|.$$

In a similar way, if $\|\tilde{p}\| \le \|p\|$, we get

$$\left(1+\left\|p+\tilde{p}\right\|^2\right)^{\frac{q-2}{2}}\left\|p+\tilde{p}\right\| \le c(q)V^{q-2}(p)\|p\|.$$

Then, by summing the previous inequalities, we obtain (2.7).

Finally we recall that, if $G \in L^q(\Omega)$, it results

$$\|G\|^q_{L^q(\Omega)} \le \|W(G)\|^2_{L^2(\Omega)}. \tag{2.8}$$

3. GLOBAL HIGHER SUMMABILITY OF THE GRADIENT

In order to obtain the global higher summability of the gradient, we prove in a first step the interior higher summability of the gradient. To this end a crucial step is the following "Caccioppoli's type" inequality.

Theorem 3.1 *Assume that conditions (1.4), (1.5), (1.6) and (1.7) are fulfilled and* $g \in H^{1,s}(\Omega) \cap L^\infty(\Omega)$, *with* $s > q \ge n$. *Let* $w \in H^{1,q}_0(\Omega) \cap L^\infty(\Omega)$ *be a solution of the strongly elliptic system:*

$$\int_\Omega \sum_{i=1}^n \left(a^i(x, w+g, Dw+Dg) \mid D_i\varphi\right)dx = \int_\Omega \left(B(x, w+g, Dw+Dg) \mid \varphi\right)dx, \quad \forall \varphi \in H^{1,q}_0(\Omega) \cap L^\infty(\Omega). \tag{3.1}$$

Then for every couples of concentric balls $B(\sigma) \subset B(2\sigma) \subset \Omega$, *it results*

$$\int_{B(\sigma)} \|Dw\|^q dx \le c\,\sigma^{-q} \int_{B(2\sigma)} \left\|w - w_{B(2\sigma)}\right\|^q dx + c_1 \int_{B(2\sigma)} (1+\|Dg\|)^q dx \tag{3.2}$$

where c, c_1 *depend on* q, M, ν, a, b, $\|u-g\|_{\infty,\Omega}$.

Proof. Let us fix $B(2\sigma) \subset \Omega$ and let $\theta \in C^\infty_0(\mathbb{R}^n)$ be a function with these properties

$$0 \le \theta \le 1, \quad \theta = 1 \text{ in } B(\sigma), \quad \theta = 0 \text{ in } \mathbb{R}^n \setminus B(2\sigma), \quad \|D\theta\| \le C\sigma^{-1},$$

with C numerical constant. Let us assume in (3.1) $\varphi = \theta^q (w - w_{B(2\sigma)})$. Then (3.1) becomes:

$$\begin{aligned} &\int_\Omega \sum_{i=1}^n \left(a^i(x, w+g, Dw+Dg)|\theta^q D_i w\right) dx \\ &= -q \int_\Omega \sum_{i=1}^n a^i(x, w+g, Dw+Dg)|\theta^{q-1} D_i \theta\,(w - w_{B(2\sigma)})\, dx \\ &\quad + \int_\Omega \left(B(x, w+g, Dw+Dg)|\theta^q\,(w - w_{B(2\sigma)})\right) dx. \end{aligned} \tag{3.3}$$

We may rewrite (3.3) in the equivalent way

$$\begin{aligned} &\int_\Omega \sum_{i=1}^n \left(a^i(x, w+g, Dw+Dg)|\theta^q\,(D_i w + D_i g)\right) dx \\ &\quad = \int_\Omega \sum_{i=1}^n \left(a^i(x, w+g, Dw+Dg)|\theta^q\, D_i g\right) dx \\ &-q \int_\Omega \sum_{i=1}^n a^i(x, w+g, Dw+Dg)|\theta^{q-1} D_i \theta\,(w - w_{B(2\sigma)})\, dx \\ &+ \int_\Omega \left(B(x, w+g, Dw+Dg)|\theta^q\,(w - w_{B(2\sigma)})\right) dx = A + C + D. \end{aligned} \tag{3.4}$$

As it concerns the left hand side of (3.4), in virtue of the strong ellipticity condition (1.5) it results

$$\begin{aligned} &\int_\Omega \sum_{i=1}^n \left(a^i(x, w+g, Dw+Dg)|\theta^q\,(D_i w + D_i g)\right) dx \\ &\geq \nu \int_\Omega \|Dw+Dg\|^2\, V^{q-2}(Dw+Dg)\,\theta^q dx = \nu \int_\Omega \|W(Dw+Dg)\|^2\,\theta^q dx. \end{aligned} \tag{3.5}$$

Let us examine the terms in the right hand side of (3.4) and let us start with the first term A. By condition (1.4), taking into account Lemma 2, we get

$$
\begin{aligned}
|A| &\le \int_\Omega \sum_{i=1}^{n} \left\| a^i(x, w+g, Dw+Dg) \right\| \theta^q \left\| D_i g \right\| dx \\
&\le M \int_\Omega V^{q-2}(Dw+Dg) \left\| Dw+Dg \right\| \theta^q \left\| Dg \right\| dx \\
&\le c(q) M \int_\Omega V^{q-2}(Dw) \left\| Dw \right\| \theta^q \left\| Dg \right\| dx + c(q) M \int_\Omega V^{q-2}(Dg) \left\| Dg \right\|^2 \theta^q \, dx \\
&\le c(q) M \int_\Omega \theta^q \left(1 + \left\| Dw \right\|^{q-2}\right) \left\| Dw \right\| \left\| Dg \right\| dx \\
&\quad + c(q) M \int_\Omega \theta^q \left(1 + \left\| Dg \right\|^2\right)^{\frac{q-2}{2}} \left(1 + \left\| Dg \right\|^2\right) dx \\
&\le c(q) M \int_\Omega \theta^q \left(\left\| Dw \right\| \left\| Dg \right\| + \left\| Dw \right\|^{q-1} \left\| Dg \right\| \right) dx + c(q) M \int_\Omega \theta^q \left(1 + \left\| Dg \right\|^2\right)^{\frac{q}{2}} dx.
\end{aligned}
\tag{3.6}
$$

Using Young's inequality in the first integral in the last line of (3.6), it follows

$$
\begin{aligned}
|A| &\le c(q) M \varepsilon \int_\Omega \theta^q \left\| Dw \right\|^q dx \\
&+ M c(\varepsilon, q) \int_\Omega \left(1 + \left\| Dg \right\|\right)^q dx + c(q) M \int_\Omega \theta^q \left(1 + \left\| Dg \right\|^q\right) dx \\
&\le c(q) \varepsilon M \int_\Omega \theta^q \left\| Dw \right\|^q dx + M c(q, \varepsilon) \int_{B(2\sigma)} \left(1 + \left\| Dg \right\|\right)^q dx.
\end{aligned}
\tag{3.7}
$$

Let us consider the second term C. In virtue of condition (1.4)

$$
\begin{aligned}
|C| &\le q \int_\Omega \sum_{i=1}^{n} \left\| a^i(x, w+g, Dw+Dg) \right\| \theta^{q-1} \left| D_i \theta \right| \left\| w - w_{B(2\sigma)} \right\| dx \\
&\le qM \int_\Omega V^{q-2}(Dw+Dg) \left\| Dw+Dg \right\| \theta^{q-1} \left\| D\theta \right\| \left\| w - w_{B(2\sigma)} \right\| dx \\
&\le qM \int_\Omega \left(1 + \left\| Dw+Dg \right\|\right)^{q-2} \left\| Dw+Dg \right\| \theta^{q-1} \left\| D\theta \right\| \left\| w - w_{B(2\sigma)} \right\| dx \\
&\le qM \int_\Omega \left(1 + \left\| Dw \right\| + \left\| Dg \right\|\right)^{q-1} \theta^{q-1} \left\| D\theta \right\| \left\| w - w_{B(2\sigma)} \right\| dx.
\end{aligned}
\tag{3.8}
$$

Applying Hölder inequality, we get

$$
\begin{aligned}
&|C| \\
&\le c(q,M)\varepsilon\int_\Omega (1+\|Dw\|+\|Dg\|)^q\,\theta^q\,dx + c(q,M,\varepsilon)\int_\Omega \|D\theta\|^q \|w-w_{B(2\sigma)}\|^q\,dx \\
&\le c(q,M)\varepsilon\int_\Omega (1+\|Dw\|)^q\,\theta^q\,dx \\
&+c(q,M)\varepsilon\int_{B(2\sigma)} \|Dg\|^q\,dx + c(q,M,\varepsilon)\sigma^{-q}\int_{B(2\sigma)} \|w-w_{B(2\sigma)}\|^q\,dx \\
&\le c(q,M)\varepsilon\int_{B(2\sigma)} \theta^q\,dx + c(q,M)\varepsilon\int_\Omega \|Dw\|^q\,\theta^q\,dx \\
&+c(q,M)\varepsilon\int_{B(2\sigma)} (1+\|Dg\|)^q\,dx + c(q,M,\varepsilon)\sigma^{-q}\int_{B(2\sigma)} \|w-w_{B(2\sigma)}\|^q\,dx.
\end{aligned}
\tag{3.9}
$$

Finally for the last term D, we have from condition (1.6)

$$
\begin{aligned}
|D| &\le \int_\Omega \|B(x,w+g,Dw+Dg)\|\,\theta^q\,\|w-w_{B(2\sigma)}\|\,dx \\
&\le 2\int_\Omega (a+b\|W(Dw+Dg)\|^2)\,\theta^q\,\|w\|_{\infty,\Omega}\,dx \\
&= 2a\int_\Omega \theta^q\,\|w\|_{\infty,\Omega},dx + 2b\|w\|_{\infty,\Omega}\int_\Omega \|W(Dw+Dg)\|^2\,\theta^q dx.
\end{aligned}
\tag{3.10}
$$

Taking into account (3.5), (3.7), (3.9), (3.10), we get

$$
\begin{aligned}
&(\nu-2b\|w\|_{\infty,\Omega})\int_\Omega \|W(Dw+Dg)\|^2\,\theta^q dx \le c(q,M)\varepsilon\int_\Omega \|Dw\|^q\,\theta^q\,dx \\
&+c(q,M,\|u-g\|_{\infty,\Omega})\varepsilon\int_{B(2\sigma)} \theta^q\,dx + c(q,M,\varepsilon)\int_{B(2\sigma)} (1+\|Dg\|)^q\,dx \\
&+c(q,M,\varepsilon)\sigma^{-q}\int_{B(2\sigma)} \|w-w_{B(2\sigma)}\|^q\,dx.
\end{aligned}
\tag{3.11}
$$

Since, in virtue of (2.8),

$$
\int_\Omega \theta^q\,\|Dw\|^q\,dx \le c(q)\int_\Omega \|W(Dw+Dg)\|^2\,\theta^q dx + c(q)\int_\Omega (1+\|Dg\|)^q\,\theta^q dx,
$$

and

$$
\int_{B(2\sigma)} \theta^q\,dx \le \int_{B(2\sigma)} dx \le \int_{B(2\sigma)} (1+\|Dg\|)^q\,dx
$$

for ε sufficiently small, from (3.11) and (1.7) we obtain

$$\begin{aligned}&\int_{B(\sigma)} \|Dw)\|^q \, dx \\ &\leq c\sigma^{-q} \int_{B(2\sigma)} \left\| w - w_{B(2\sigma)} \right\|^q dx + c_1 \int_{B(2\sigma)} (1 + \|Dg\|)^q \, dx,\end{aligned} \tag{3.12}$$

that is our thesis.

We are in position to derive the interior higher summability of the gradient.

Theorem 3.2 *Assume that conditions (1.4), (1.5), (1.6) and (1.7) are fulfilled and* $g \in H^{1,s}(\Omega) \cap L^\infty(\Omega)$, *with* $s > q \geq n$. *If* $w \in H_0^{1,q}(\Omega) \cap L^\infty(\Omega)$ *is a solution of the strongly elliptic system*

$$\begin{aligned}&\int_\Omega \sum_{i=1}^n \left(a^i(x, w+g, Dw+Dg) \,|\, D_i\varphi \right) dx \\ &= \int_\Omega \left(B(x, w+g, Dw+Dg) \,|\, \varphi \right) dx \qquad \forall \varphi \in H_0^{1,q}(\Omega) \cap L^\infty(\Omega),\end{aligned} \tag{3.13}$$

then there exists a number $\tilde{r} > 1$ *such that* $Du \in L_{loc}^{q\tilde{r}}(\Omega)$ *and* $\forall B(2\sigma) \subset \Omega$ *it results*

$$\left(-\int_{B(\sigma)} \|Dw\|^{q\tilde{r}} \, dx \right)^{\frac{1}{\tilde{r}}} \leq K - \int_{B(2\sigma)} \|Dw\|^q \, dx + K \left(-\int_{B(2\sigma)} \left(1 + \|Dg\|\right)^{q\tilde{r}} dx \right)^{\frac{1}{\tilde{r}}} \tag{3.14}$$

where the constant K *does not depend on* σ.

Proof. By Poincaré inequality it follows

$$\sigma^{-q} \int_{B(2\sigma)} \left\| w - w_{B(2\sigma)} \right\|^q dx \leq c\sigma^n \left(-\int_{B(2\sigma)} \|Dw\|^{\frac{nq}{n+q}} dx \right)^{\frac{n+q}{n}}.$$

Hence if we set

$$U = \|Dw\|^{\frac{qn}{n+q}}$$

$$G = \left(1 + \|Dg\|\right)^{\frac{qn}{n+q}}$$

from "Caccioppoli's inequality" (3.2) it follows

$$-\int_{B(\sigma)} U^{\frac{n+q}{n}} dx \le c\left(-\int_{B(2\sigma)} U dx\right)^{\frac{n+q}{n}} + c_1 - \int_{B(2\sigma)} G^{\frac{n+q}{n}} dx.$$

Then, in virtue of Gehring-Giaquinta-Modica Lemma 2.1, the assert follows.

In the second step of the proof of the global higher summability, we have to prove the higher summability up to the boundary for the gradient of a solution to Dirichlet Problem.

Theorem 3.3 *Assume that conditions (1.4), (1.5), (1.6) and (1.7) are fulfilled and* $g \in H^{1,s}(B^+(1)) \cap L^\infty(B^+(1))$, *with* $s > q \ge n$. *If* $w \in H^{1,q}(B^+(1)) \cap L^\infty(B^+(1))$ *is a solution of the strongly elliptic problem*

$$\begin{cases} \int_{B^+(1)} \sum_{i=1}^{n} \left(a^i(x, w+g, Dw+Dg) \mid D_i\varphi\right) dx = \\ \int_{B^+(1)} \left(B(x, w+g, Dw+Dg) \mid \varphi\right) dx, \quad \forall \varphi \in H_0^{1,q}(B^+(1)) \cap L^\infty(B^+(1)) \\ w(x) = 0 \quad \text{on } \Gamma, \end{cases} \tag{3.15}$$

then there exists a number $r' > 1$ *such that* $Dw \in L^{qr'}_{loc}(B^+(1))$ *and for all* $B^+(2\sigma) \subset B^+(1)$ *it results*

$$-\int_{B^+(\sigma)} \|Dw\|^{qr'} dx \le K - \int_{B^+(2\sigma)} \|Dw\|^q dx + K\left(-\int_{B^+(2\sigma)} \left(1 + \|Dg\|\right)^{qr'} dx\right)^{\frac{1}{r'}} \tag{3.16}$$

where K *is a positive constant which does not depend on* σ.

Proof. Let us choose $\sigma < \frac{1}{2}$ and a function $\theta \in C_0^\infty(\mathbb{R}^n)$ having the following properties:

$$0 \le \theta \le 1, \quad \theta = 1 \text{ in } B(\sigma), \quad \theta = 0 \text{ in } \mathbb{R}^n \setminus B(2\sigma), \quad \|D\theta\| \le C\sigma^{-1} \tag{3.17}$$

with C numerical constant. Taking into account that $w = 0$ on Γ, in (3.15) we can assume $\varphi = \theta^q w$ and, arguing as in the proof of Theorem 3.1, we get the "Caccioppoli's type estimate"

$$\int_{B^+(\sigma)} \|Dw\|^q \, dx \le c\sigma^{-q} \int_{B^+(2\sigma)} \|w\|^q \, dx + c_1 \int_{B^+(2\sigma)} \left(1 + \|Dg\|\right)^q dx. \quad (3.18)$$

Now, taking into account that

$$w(x) = 0 \text{ on } \Gamma$$

we can apply the Poincaré inequality

$$\sigma^{-q} \int_{B^+(2\sigma)} \|w\|^q \, dx \le c\sigma^n \left(-\int_{B^+(2\sigma)} \|Dw\|^{\frac{nq}{n+q}} \, dx \right)^{\frac{n+q}{n}}$$

and hence, in order to obtain (3.16) we can repeat the same arguments of Theorem 3.2.

Now we may derive the global higher summability of the gradient.

Theorem 3.4 *Let conditions (1.4), (1.5), (1.6), (1.7) be fulfilled, let* $\partial\Omega$ *be of class* C^2 *and* $g \in H^{1,s}(\Omega) \cap L^\infty(\Omega)$, *with* $s > q \ge n$. *If* $w \in H_0^{1,q}(\Omega) \cap L^\infty(\Omega)$ *is a solution to the Dirichlet problem*

$$\begin{cases} \sum_{i=1}^{n} D_i a^i(x, w+g, Dw+Dg) = -B(x, w+g, Dw+Dg) \\ w = 0 \text{ on } \partial\Omega \end{cases}$$

there there exists $r > 1$ *such that* $Dw \in L^{qr}(\Omega)$.

Proof. Taking into account that $\partial\Omega$ is of class C^2, it is enough to use the usual covering procedure (see [4] Lemma 2.V, 2.VI, 2.VII and Section n.8 for details).

REFERENCES

[1] S. Campanato, *Sistemi ellittici in forma di divergenza. Regolarità all'interno*, Quaderni S.N.S. di Pisa (1980).

[2] S. Campanato, *Nonlinear elliptic systems with quadratic growth*, Seminario Matematica Bari n.208 (1986).

[3] S. Campanato, *A bound for the solutions of a basic elliptic system with non-linearity* $q \ge 2$, Atti Acc. Naz. Lincei (8) 80 (1986), 81-88.

[4] S. Campanato, *Elliptic systems with nonlinearity q greater or equal to two. Regularity of the solution of the Dirichlet Problem*, Ann. Mat. Pura e Appl. 147 (1987), 117-150.

[5] F. Colombini, *Un teorema di regolarità di sistemi ellittici quasi lineari*, Ann. Sc. Norm. Sup. Pisa 25 (1971), 115-161.

[6] E. De Giorgi, *Un esempio di estremali discontinue per un problema variazionale di tipo ellittico*, Boll. Un. Mat. Ital. (4) 1 (1968), 135-137.

[7] L. Fattorusso and G. Idone, *Partial Hölder continuity results for solutions to nonlinear nonvariational elliptic systems with limit controlled growth*, Boll. U.M.I. 8 (2002), 747-754.

[8] J. Frehse, *On the boundedness of weak solutions of higher order nonlinear elliptic partial differential equations*, Boll. Un. Mat. Ital. 4 (1970), 607-627.

[9] M. Giaquinta, *A counter-example to the boundary regularity of solutions to elliptic quasilinear systems*, Manus. Math. 24 (1978), 217-220.

[10] M. Giaquinta and E. Giusti, *Nonlinear elliptic systems with quadratic growth*, Manus. Math 24 (1978), 323-349.

[11] E. Giusti and M. Miranda, *Sulla regolarità delle soluzioni di una classe di sistemi ellittici quasi lineari*, Arch. Rational Mech. Anal. 31 (1968), 173-184.

[12] J.F. Grotowski, *Boundary regularity for quasilinear elliptic systems*, Comm. Partial Diff. Eq. 27 (2002), 2491-2512.

[13] G. Idone, *Elliptic Systems with Nonlinearity q Greater or Equal than Two with Controlled Growth. Global Hölder Continuity of Solutions to the Dirichlet Problem*, J. Math. Anal. Appl. 290 (2004), 147-170.

[14] M. Marino and A. Maugeri, *Boundary regularity results for nonvariational basic elliptic systems*, Le Matematiche 55 (2000), 109-123.

[15] G. Mingione, *The singular set of solutions to non-differentiable elliptic systems*, Arch. Rational Mech. Anal. 166 (4) (2003), 287-301.

[16] J. Nečas and J. Stará, *Principio di massimo per i sistemi ellittici quasi lineari non diagonali*, Boll. U.M.I. 6 (1972), 1-10.

[17] K.O. Widman, *Hölder continuity of solutions of elliptic systems,* Manus. Math. 5 (1971), 299-308.

[18] J. Wolf, *Partial regularity of weak solutions to nonlinear elliptic systems satisfying a Dini condition*, Zeit. Anal. und Appl., 19 (2) (2001), 315-330.

OPTIMALITY CONDITIONS FOR GENERALIZED COMPLEMENTARITY PROBLEMS

S. Giuffré[1], G. Idone[1] and A. Maugeri[2]
D.I.M.E.T., Faculty of Engineering, University of Reggio Calabria, Reggio Calabria, Italy;[1] Dept. of Mathematics and Computer Science, University of Catania, Catania, Italy[2]

Abstract: In this paper Generalized Complementarity Problems are expressed in terms of suitable optimization problems and some optimality conditions are given. The infinite dimensional Lagrangean and Duality Theories play an important role in order to achieve the main result.

Key words: Generalized Complementarity Problem, Lagrangean Function, Dual Problem, Quasirelative interior, saddle point.

1. INTRODUCTION

Let S be a nonempty subset of a real linear space X. Let Y be a partially ordered real normed space with the ordering cone C. Let Z be the set of nonnegative measurable functions and let

$$\mathcal{L}: S \to Z$$

$$\mathcal{B}: S \to Z$$

be two operators. Let $g: S \to Y$ be a given constraint mapping and let us set

$$\mathbb{K} = \{v \in S : g(v) \in -C\}. \tag{1}$$

Let us suppose that

$$\mathcal{L}(v) \geq 0 \quad \mathcal{B}(v) \geq 0 \quad \forall v \in S$$

and let us observe that the Generalized Complementarity Problem

$$\begin{cases} \mathcal{B}(u)\mathcal{L}(u) = 0 \\ u \in \mathbb{K} \end{cases} \tag{2}$$

expresses many economic and physical equilibrium problems. In fact, starting from the classical Signorini problem, it has been observed that the Obstacle problem, the Elastic–Plastic Torsion problem, the Traffic Equilibrium problem both in the discrete and continuous cases, the Spatial Price Equilibrium problem, the Financial Equilibrium problem and many others (see [7], [8], [9], [14]) satisfy the Generalized Complementarity Problem (2). For example, the continuous traffic equilibrium problem fits very well with the above scheme assuming $X = L^2_{\text{div}}(\Omega) = \{u \in L^2(\Omega, \mathbb{R}^2) : \text{div } u \in L^2(\Omega)\}$, $Y = L^2(\Omega)$, $S = L^2_{\text{div}}(\Omega)$, $\mathcal{B}v = v$, $\mathcal{L}v = \left(c_i(x, v(x)) - \dfrac{\partial \mu}{\partial x_i} \right)_{i=1,2}$. Problem (2) becomes

$$\begin{cases} \left(c_i(x, u(x)) - \dfrac{\partial \mu(x)}{\partial x_i} \right) u_i(x) = 0 \\ \qquad\qquad\qquad\qquad\qquad\qquad i = 1,2, \text{ a.e. in } \Omega \\ u \in \mathbb{K} \end{cases}$$

where $\mu \in H^1(\Omega)$ is a given function (potential) and $\mathbb{K}$ is given by

$$\mathbb{K} = \{u \in L^2_{\text{div}}(\Omega) : u_i(x) \geq 0, u_i(x)|_{\partial\Omega} = \varphi_i(x), \text{ div } u + t(x) = 0\},$$

with Ω a simply connected bounded domain in $\mathbb{R}^2$ with Lipschitz boundary $\partial\Omega$.

In this model, $u_i(x)$ $i = 1,2$ represent the traffic density through a neighbourhood of x in the direction of the increasing axis x_i. $u_i(x)$ has nonnegative fixed trace $\varphi_i(x)$ on $\partial\Omega$ which represents the entering flow. If

we associate to each point $x \in \Omega$ a scalar field $t(x) \in L^2(\Omega)$, which measures the density of the flow originating or terminating at x, the flow $u(x)$ satisfies the conservation law

$$\operatorname{div} u(x) + t(x) = 0 \quad \text{a.e. in } \Omega.$$

The function $c_i(x, v(x))$ represents the travel cost along the axis x_i $(i = 1,2)$.

The equilibrium condition is the following one:

Definition 1 *$u(x) \in \mathbb{K}$ is an equilibrium distribution flow if there exists a potential $\mu \in H^1(\Omega)$ such that*

$$\left(c_i(x,u(x)) - \frac{\partial \mu(x)}{\partial x_i} \right) u_i(x) = 0$$

$$i = 1,2 \quad a.e.\ in\ \Omega$$

$$c_i(x,u(x)) - \frac{\partial \mu(x)}{\partial x_i} \geq 0$$

The potential μ measures the cost occurred when a user travels from the point x to the boundary $\partial\Omega$ using the cheapest possible path (see [4], [5], [10]).

The same happens for the Elastic–Plastic Torsion Problem. In this case we have $X = H^2(\Omega)$ (or $H^1(\Omega)$ if we consider the weak formulation); $Y = L^2(\Omega)$; $\mathcal{B}v = 1 - \sum_{i=1}^{n} \left(\frac{\partial v}{\partial x_i} \right)^2$; $\mathcal{L}v$ an elliptic operator. The equilibrium condition is

$$\left[1 - \sum_{i=1}^{n} \left(\frac{\partial u}{\partial x_i} \right)^2 \right] \mathcal{L}u = 0$$

and

$$\mathbb{K} = \left\{ v \in H_0^1(\Omega) : v \geq 0, \quad 1 - \sum_{i=1}^{n} \left(\frac{\partial v}{\partial x_i} \right)^2 \leq 0 \right\}$$

(see [11], [12]).

The Evolutionary Financial Time Equilibrium has been considered very recently and also it perfectly agrees with the above scheme. In this case a vector of sector assets, liabilities and instrument prices $(x^*(t), y^*(t), r^*(t)) \in \prod_{i=1}^{n} P_i \times L^2(0,T,\mathbb{R}^n_+)$, where

$$P_i = \{(x_i(t), y_i(t)) \in L^2(0,T,\mathbb{R}^{2n}) : \sum_{j=1}^{n} x_{ij}(t) = s_i(t), \sum_{j=1}^{n} y_{ij}(t) = s_i(t),$$

$$x_{ij}(t), y_{ij}(t) \geq 0 \text{ a.e. in } [0,T]\},$$

with $s_i(t)$ the total financial volume held by sector i at the time t, is an equilibrium of the evolutionary financial model if and only if it satisfies the system of equalities:

$$x^*_{ij}\left[2[Q^i_{11}(t)]^T_j x^*_i(t) + 2[Q^i_{21}(t)]^T_j y^*_i(t) - r^*_j(t) - \mu^{(1)}_i(t)\right] = 0$$

$$y^*_{ij}\left[2[Q^i_{12}(t)]^T_j x^*_i(t) + 2[Q^i_{22}(t)]^T_j y^*_i(t) + r^*_j(t) - \mu^{(2)}_i(t)\right] = 0$$

$$\sum_{i=1}^{m}\left(x^*_{ij}(t) - y^*_{ij}(t)\right) r^*_j(t) = 0$$

with all the functions $x^*_{ij}(t)$, $y^*_{ij}(t)$, $2[Q^i_{11}(t)]^T_j x^*_i(t) + 2[Q^i_{21}(t)]^T_j y^*_i(t) - r^*_j(t) - \mu^{(1)}_i(t)$, $2[Q^i_{12}(t)]^T_j x^*_i(t) + 2[Q^i_{22}(t)]^T_j y^*_i(t) + r^*_j(t) - \mu^{(2)}_i(t)$, $\sum_{i=1}^{m}\left(x^*_{ij}(t) - y^*_{ij}(t)\right)$ non negative.

The meaning of this definition is the following one:

To each financial volume $s_i(t)$ invested by sector i there are associated two functions $\mu^{(1)}_i(t)$ and $\mu^{(2)}_i(t)$ related to the assets and to the liabilities which represent the "Equilibrium Utilities" per unit of the sector i, respectively; $2[Q^i_{11}(t)]^T_j x^*_i(t) + 2[Q^i_{21}(t)]_j y^*_i(t) - r^*_j(t)$ is the personal utility of the investor in the instrument j as an asset. Then if this personal utility equals the equilibrium utility $\mu^{(1)}_i(t)$, it results $x^*_{ij}(t) \geq 0$, whereas if the personal utility is greater than the equilibrium utility $\mu^{(1)}_i(t)$, it results $x^*_{ij}(t) = 0$. The meaning of the second condition is analogous, whereas the

third one $\sum_{i=1}^{n}(x_{ij}^*(t)-y_{ij}^*(t))r_j^*(t)=0$ states that if the price r_j^* of the instrument j is positive, then the amount of the assets is equal to the amount of liabilities, on the contrary if there is an excess supply of an instrument in the economy:

$$\sum_{i=1}^{m}x_{ij}^*(t)>\sum_{i=1}^{m}y_{ij}^*(t)$$

then $r_j^*(t)=0$ (see [6]).

In this paper we observe that the Generalized Complementarity Problem (2) can be written as the Optimization Problem:

$$\begin{cases}\min \mathcal{B}(v)\ \mathcal{L}(v)=0\\ v\in\mathbb{K}\end{cases} \tag{3}$$

and we investigate how we can associate to Problem (3), by means of Lagrangean and Duality Theories, some optimality conditions. For the sake of simplicity, we confine ourselves to a less general case. Let us suppose that X and Y are real Hilbert spaces with the usual inclusion $X\subseteq Y\subseteq X^*$; let it be C the ordering convex cone of Y and let L, B, g three functions defined on X with values in Y. Let us suppose that the set $\mathbb{K}=\{v\in X: g(v)\in -C\}$ is nonempty and let us assume that the Generalized Complementarity Problem (3) holds in the sense of the scalar product on Y and that $\langle \mathcal{L}v,\mathcal{B}v\rangle \geq 0$, $\forall v\in X$. Then Problem (3) becomes

$$\min_{v\in\mathbb{K}}\langle \mathcal{L}v,\mathcal{B}v\rangle=0. \tag{4}$$

The main result of this paper is the following:

Theorem 1. *Let the function* $(\langle \mathcal{L}(v),\mathcal{B}(v)\rangle, g(v))$ *be convex-like. Let us assume that qri* $(g(X)+C)\neq\varnothing$ *and* $\overline{cone(qri(g(X)+C))}$ *is not a linear subspace of* Y. *In addition suppose that C is closed,* $\overline{C-C}=Y$ *and there exists* $\overline{v}\in X$ *such that* $g(\overline{v})\in -qri\ C$. *Then, if the functions* $\mathcal{L}$, $\mathcal{B}$, g *are Fréchet differentiable and Problem (4) admits a solution* $u\in\mathbb{K}$, *then there exists an element* $\overline{l}\in C^*$ *such that*

$$\langle \mathcal{L}_u(u)v,\mathcal{B}(u)\rangle+\langle \mathcal{L}(u),\mathcal{B}_u(u)v\rangle+\langle \overline{l},g_u(u)v\rangle=0,\quad \forall v\in X$$

and

$$\langle l, g(u)\rangle \le 0, \quad \forall l \in C^*$$
$$\langle \overline{l}, g(u)\rangle = 0$$

Note that, in virtue of Proposition 6 in Section 2, qri $C \neq \emptyset$. Further, it is worth remarking that, taking into account that $\langle \mathcal{L}_u(u)v, \mathcal{B}(u)\rangle$, $\langle \mathcal{L}(u), \mathcal{B}_u(u)v\rangle$, $\langle \overline{l}, g_u(u)v\rangle$ define three continuous linear mappings on the Hilbert space X, there exist three elements of X^*, that we denote by $\mathcal{B}(u)\mathcal{L}_u(u)$, $\mathcal{L}(u)\mathcal{B}_u(u)$, $\overline{l}\, g_u(u)$, such that

$$\langle \mathcal{L}_u(u)v, \mathcal{B}(u)\rangle + \langle \mathcal{L}(u), \mathcal{B}_u(u)v\rangle + \langle \overline{l}, g_u(u)v\rangle =$$

$$= \langle \mathcal{B}(u)\mathcal{L}_u(u) + \mathcal{L}(u)\mathcal{B}_u(u) + \overline{l}\, g_u(u), v\rangle = 0, \quad \forall v \in X.$$

Hence we derive the equivalent condition:

$$\mathcal{B}(u)\mathcal{L}(u)\mathcal{B}_u(u) + \overline{l}\, g_u(u) = 0.$$

Finally we observe that the above technique is complementary to the study of the Generalized Complementarity Problems by means of Variational Inequalities.

2. THE LAGRANGEAN AND DUALITY THEORY

Let us introduce the dual cone C^*

$$C^* = \{l \in Y^* : \langle l, v\rangle \ge 0, \quad \forall v \in C\}$$

that, in virtue of the usual identification $Y = Y^*$, can be rewritten

$$C^* = \{l \in Y : \langle l, v\rangle \ge 0, \forall v \in C\}.$$

Then, using the same technique used by J. Jhan in [13], it is possible to show the following result:

Theorem 2 *Let the ordering cone be closed. Then u is a minimal solution of (4) if and only if u is a solution of the problem*

$$\min_{v\in X}\sup_{l\in C^*}\{\langle \mathcal{L}v, \mathcal{B}v\rangle + \langle l, g(v)\rangle\} \tag{5}$$

and the extremal values of the two problems are equal.

Now let us introduce the Dual Problem

$$\max_{l\in C^*}\inf_{v\in X}\{\langle \mathcal{L}v, \mathcal{B}v\rangle + \langle l, g(v)\rangle\}. \tag{6}$$

It is known (see Theorem 6.7 of [13]) that if int C is nonempty, if Problem (4) (or 5) is solvable and the generalized Slater condition is satisfied, namely there exists $\overline{v}\in X$ with $g(\overline{v})\in -$ int C, then problem (6) is also solvable and the extremal values of the two problems are equal:

$$\min_{v\in X}\sup_{l\in C^*}\{\langle \mathcal{L}v, \mathcal{B}v\rangle + \langle l, g(v)\rangle\} = \max_{l\in C^*}\inf_{v\in X}\{\langle \mathcal{L}v, \mathcal{B}v\rangle + \langle l, g(v)\rangle\}. \tag{7}$$

However in many concrete situations the request that int C is non-empty is not verified: for example if X, Y are Lebesgue spaces. For this reason in [1], the authors develops the notation of quasi-relative interior of a convex set that is an extension of the relative interior in finite dimension. Let us recall the definition and some properties of quasi-relative interior of a convex subset C of a real Hilbert space Y..

Definition 2 *Let C be a convex subset of Y. The quasi-relative interior of C, denoted by qri C, is the set of those $x\in C$ for which*

$$\overline{\mathrm{Cone}(C-x)} = \overline{\{\lambda y : \lambda \geq 0,\ y\in C-x\}}$$

is a subspace.

Proposition 1 *Let C be a convex subset of Y and $\overline{x}\in C$. Then $\overline{x}\in$ qri C if and only if the normal cone to C at $\overline{x}$ $N_C(\overline{x}) = \{l\in Y : \langle l, x-\overline{x}\rangle \leq 0,\ \forall x\in C\}$ is a subspace.*

Proposition 2 *Let C be a convex subset of Y. If qri $C\neq\varnothing$, then*

$\overline{\text{qri } C} = \overline{C}$ and $\text{qri } C = \text{qri } (\text{qri } C)$.

Proposition 3 *Let C be a convex subset of Y and suppose $x_1 \in \text{qri } C$ and $x_2 \in C$. Then $\lambda x_1 + (1-\lambda)x_2 \in \text{qri } C$ for all $0 < \lambda \leq 1$.*

Proposition 4 *Let C and D be two convex subsets of X such that qri $X \neq \varnothing$, qri $Y \neq \varnothing$ and let $\lambda \in \mathbb{R}$. Then*

$$\text{qri } C + \text{qri } D \subset \text{qri } (C + D), \quad \lambda \text{qri } C = \text{qri } (\lambda C), \quad \text{qri } (C \times D) = \text{qri } C \times \text{qri } D.$$

Proposition 5 *Let C be a convex subset of Y such that qri $C \neq \varnothing$ and $l \in Y$. If int $\langle l, C \rangle \neq \varnothing$, then $\langle l, \text{qri } C \rangle = \text{int } \langle l, C \rangle$.*

Proposition 6 *Let C be a convex closed subset of a separable Banach space. Then qri $C \neq \varnothing$.*

Proposition 7 *Let C be a nontrivial convex cone. If C is in addition acute, namely $\overline{C} \cap (-\overline{C}) = \{0_Y\}$, then $0_Y \notin$ qri C.*

The proofs of these propositions can be found in [1] and [2].

Using this concept of quasi-relative interior more general separation theorems can be proved (see [3]). In fact the following statements hold.

Lemma 1 *Let A be a convex subset of Y such that qri $A \neq \varnothing$ and $0_Y \notin \text{qri } A$. Then there exists $g \in Y - \{0_Y\}$ such that $\langle g, v \rangle \leq 0$ for all $v \in A$.*

Theorem 3 *Let S and T be two convex subsets of Y such that qri $S \neq \varnothing$ and qri $T \neq \varnothing$ and such that $\text{cone}(\text{qri } S - \text{qri } T)$ is not a linear subspace of Y or, alternatively, cone $(\text{qri } S - \text{qri } T)$ is acute. Then there exists $l \in Y - \{0_Y\}$* such that $\langle l, s \rangle \leq \langle l, t \rangle$, *for all $s \in S$, $t \in T$.*

Theorem 4 *Let S and T be two nonempty convex subsets of Y such that qri $S \neq \varnothing$ and qri $T \neq \varnothing$. Suppose that there exists a convex set $V \subseteq Y$ such that $V - V = Y$, $0_Y \in \text{qri } V$ and $\text{cone}(\text{qri}(S - T) - \text{qri } V)$ is not a linear subspace of Y or, alternatively, cone $(\text{qri}(S - T) - \text{qri } V)$ is acute. Then there exists $l \in Y - \{0_Y\}$ and $\gamma \in \mathbb{R}$ such that*

$$\langle l, s \rangle < \gamma < \langle l, t \rangle, \quad \forall s \in S, t \in T.$$

Using the new separation theorems, we can show that problem (6) is solvable and that the extremal values are equal.

At first we recall the definition of a convex–like function.

Definition 3 *Let X be a real linear space and let Y be a real linear space partially ordered by a convex cone C. A function $f: X \to Y$ is called convex–like if the set $f(X)+C$ is convex.*

Theorem 5 *Let the function $\varphi(v) = \big(\langle \mathcal{L}(v), \mathcal{B}(v)\rangle, g(v)\big)$ be convex-like with respect to the product cone $\mathbb{R}^+ \times C$ in $\mathbb{R} \times Y$. Let us assume that qri $[g(X)+C] \neq \varnothing$ and $\operatorname{cone}(\operatorname{qri}(g(X)+C)$ is not a linear subspace of Y or, alternatively, cone $[\operatorname{qri}(g(X)+C)]$ is acute. In addition suppose that qri $C \neq \varnothing$ and $C - C = Y$. If Problem (4) is solvable and there exists $\overline{v} \in X$ with $g(\overline{v}) \in -$ qri C, then also Problem (6) is solvable and the extremal values of the two problems are equal. Moreover, if u is a solution to Problem (4) and $\overline{l} \in C^*$ of (6), it turns out to be $\langle \overline{l}, g(u)\rangle = 0$.*

3. PROOF OF THE MAIN RESULT

Let us consider the Lagrangean functional $L: X \times C^* \to \mathbb{R}$

$$L(v,l) = \langle \mathcal{L}(v), \mathcal{B}(v)\rangle + \langle l, g(v)\rangle.$$

Using the preceding theorems we are able to state the following

Theorem 6 *Let the assumptions of Theorem 5 be fulfilled, with C closed. Then a point $(u,\overline{l}) \in X \times C^*$ is a saddle point of L, namely*

$$L(u,l) \leq L(u,\overline{l}) \leq L(v,\overline{l}), \quad \forall v \in X, \forall l \in C^* \tag{8}$$

if and only if u is a solution of problem (4) (or (5)), $\overline{l}$ is a solution of Problem (6) and (7) holds, namely

$$\min_{v \in X} \sup_{l \in C^*} \{\langle \mathcal{L}v, \mathcal{B}v\rangle + \langle l, g(v)\rangle\} = \max_{l \in C^*} \inf_{v \in X} \{\langle \mathcal{L}v, \mathcal{B}v\rangle + \langle l, g(v)\rangle\} =$$
$$= \langle \mathcal{L}u, \mathcal{B}u\rangle + \langle \overline{l}, g(u)\rangle = 0.$$

(See for the proof [7]).

From (8) we can derive a lot of consequences. First let us take into account the right hand side inequality

$$L(v,\overline{l}) \geq L(u,\overline{l}) = 0, \quad \forall v \in X, \tag{9}$$

namely

$$\left\langle \mathcal{L}(v), \mathcal{B}(v) \right\rangle + \left\langle \overline{l}, g(v) \right\rangle \geq 0, \quad \forall v \in X. \tag{10}$$

Now let us take into account that $\mathcal{L}: X \to Y$, $\mathcal{B}: X \to Y$, $g: X \to Y$ are Frèchet differentiable functions. Then from (9) we derive

$$\left\langle \mathcal{L}_u(u)v, \mathcal{B}(u) \right\rangle + \left\langle \mathcal{L}(u), \mathcal{B}_u(u)v \right\rangle + \left\langle \overline{l}, g_u(u)v \right\rangle = 0, \quad \forall v \in X. \tag{11}$$

Taking into account that the three terms of the left hand side of (11) define three continuous linear mappings on X and that our setting is the Hilbert one, there exist three elements of X^* that for the sake of simplicity we denote by $\mathcal{B}(u)\mathcal{L}_u(u)$, $\mathcal{L}(u)\mathcal{B}_u(u)$, $\overline{l}g_u(u)$ such that

$$\left\langle \mathcal{L}_u(u)v, \mathcal{B}(u) \right\rangle = \left\langle \mathcal{B}(u)\mathcal{L}_u(u), v \right\rangle \quad \forall v \in X$$

$$\left\langle \mathcal{L}(u), \mathcal{B}_u(u)v \right\rangle = \left\langle \mathcal{L}(u)\mathcal{B}_u(u), v \right\rangle \quad \forall v \in X$$

$$\left\langle \overline{l}, g_u(u)v \right\rangle = \left\langle \overline{l}g_u(u), v \right\rangle \quad \forall v \in X.$$

Then we get

$$\left\langle \mathcal{B}(u)\mathcal{L}_u(u) + \mathcal{L}(u)\mathcal{B}_u(u) + \overline{l}g_u(u), v \right\rangle = 0 \quad \forall v \in X$$

and hence

$$\mathcal{B}(u)\mathcal{L}_u(u) + \mathcal{L}(u)\mathcal{B}_u(u) + \overline{l}g_u(u) = 0 \quad \forall v \in X. \tag{12}$$

Now let us consider the inequality at the left hand side of (8). We get

$$\left\langle \mathcal{L}(u), \mathcal{B}(u) \right\rangle + \left\langle l, g(u) \right\rangle \leq 0, \quad \forall l \in C^*,$$

namely

$$\left\langle l, g(u) \right\rangle \leq 0, \quad \forall l \in C^*.$$

So we find that if u is a solution of Problem (4), there exists $\overline{l} \in C^*$ such that

$$\left\langle \overline{l}, g(u) \right\rangle = 0 \quad \text{and} \quad \left\langle l, g(u) \right\rangle \leq 0, \quad \forall l \in C^*,$$

namely u and $\overline{l}$ satisfy the Variational Inequality

$$\begin{cases} \left\langle g(u), l - \overline{l} \right\rangle \geq 0 \\ \forall l \in C^*. \end{cases} \tag{13}$$

REFERENCES

[1] J. M. Borwein, A.S. Lewis, Practical Conditions for Fenchel Duality in Infinite Dimensions, Pitman Research Notes in Mathematic Series 252, M.A. Therà - J.B. Baillon Editors, 1989, 83–89.

[2] J.M. Borwein, R. Goebel, Notions of Relative Interior in Banach Space, 2001.

[3] F. Cammaroto, B. Di Bella, A Separation Theorem based on the Quasi–Relative Interior. An Application to the Theory of Duality for Infinite Dimensional extremum problems, to appear.

[4] S. Dafermos, Continuum Modeling of Transportation Networks, Transportation Res. 14 B 1980, 295–301.

[5] P. Daniele, G. Idone, A. Maugeri, Variational Inequalities and the Continuum Model of Transportation Problems, Int. Journal of Nonlinear Sciences and Numerical Simulation, 4, 2003, 11–16.

[6] P. Daniele, Variational Inequalities for Evolutionary Financial Equilibrium, Innovations in Financial and Economic Networks, 2003, 84–109.

[7] P. Daniele, F. Giannessi, A. Maugeri Editors, Equilibrium Problems and Variational Models, Kluwer Academic Publishers, 2002.

[8] F. Giannessi, A. Maugeri, P. Pardalos Editors, Equilibrium Problems: Nonsmooth Optimization and Variational Inequality Models, Kluwer Academic Publishers, 2001.

[9] F. Giannessi, A. Maugeri Editors, Variational Inequalities and Network Equilibrium Problems, Plenum Press, New York, 1995.

[10] G. Idone, Variational Inequalities and applications to a Continuum Model of Transportation Network with Capacity Constraints, Journal of Global Optimization, 2002.

[11] G. Idone, A. Maugeri, C. Vitanza, Variational Inequalities and the Elastic–Plastic Torsion Problem, J. Optim. Theory Appl., 117, 2003.

[12] G. Idone, A. Maugeri, C. Vitanza, Topics on variational Analysis and applications to Equilibrium Problems, Journal of Global Optimization, to appear.

[13] J. Jahn, Introduction to the theory of Nonlinear Optimization, Springer, 1996.

[14] A. Nagurney, Network Economics: A Variational Inequality Approach, Kluwer Academic Publishers, Dordrecht 1993.

VARIATIONAL INEQUALITIES FOR TIME DEPENDENT FINANCIAL EQUILIBRIUM WITH PRICE CONSTRAINTS

S. Giuffrè[1] and S. Pia[1]
D.I.M.E.T., Faculty of Engineering, University of Reggio Calabria, Reggio Calabria, Italy[1]

Abstract: We study a financial evolutionary problem, when variance-covariance matrices, sector financial holding volumes, instrument prices are time-dependent. As in P.Daniele [1], but assuming the realistic condition of a lower constraint for the price of each instrument, we give the evolutionary financial equilibrium condition, prove an equivalent variational inequality formulation and an existence result.

Key words: financial problem, equilibrium condition, variational inequality formulation, time-dependent requirements.

1. INTRODUCTION

In the paper [1] P. Daniele studied an evolutionary model for a multi-sector, multi-instrument financial equilibrium problem, extending the important results on stationary financial equilibrium by J. Dong, M. Hughes, K. Ke, A. Nagurney, S. Siokos and D. Zhang (see [3] and [7]-[14]).

In the above paper [1] the variance-covariance matrices associated with risk perception, the financial volumes held by the sectors, the optimal portfolio compositions, as well as the instrument prices, all are time-dependent.

Although equilibrium excludes time, time is, nevertheless, central in both the physical-technological world as well as in the socio-economic world. For

example P. Daniele emphasize the fact that in presence of uncertainty and of risky perspectives, the volume held by each sector cannot be considered stable and may decrease or increase depending on unfavorable or favorable economic conditions.

In the above paper P. Daniele provides the evolutionary financial equilibrium conditions, gives an equivalent variational inequality formulation, establishes an existence result and proposes a computational procedure.

The instrument price equilibrium condition introduced by P. Daniele are obtained in the following way. Let, at time t, $x_{ij}(t)$ be the amount of instrument j held as an asset in sector i's portfolio, $y_{ij}(t)$ be the amount of instrument j held as a liability in sector i's portfolio, then the equilibrium condition for price $r_j(t)$ of instrument j is the following:

$$\begin{cases} \sum_{i=1}^{m}(x_{ij}(t)-y_{ij}(t)) \geq 0 & \text{a.e. in } [0,T] \\ \sum_{i=1}^{m}(x_{ij}(t)-y_{ij}(t))r_j(t)=0 & r(t)\in L^2([0,T],\mathbb{R}^n_+). \end{cases} \tag{1.1}$$

From these conditions it derives that if $\sum_{i=1}^{m} x_{ij}(t) > \sum_{i=1}^{m} y_{ij}(t)$ a.e.in $[0,T]$, then $r_j(t)=0$. Namely, if there is an excess supply of an instrument in the economy, then its price must be zero. It seems to be reasonable to generalize the assumption of zero prices in presence of excess supply and make it able to cover a larger range of financial behaviours; then we can suppose that, as a result of policy interventions, a price floor $\underline{r}_j(t) \geq 0$, for each instrument j, is guaranteed. As a consequence, the equilibrium condition must be replaced by

$$\begin{cases} \sum_{i=1}^{m}(x_{ij}(t)-y_{ij}(t)) \geq 0 & \text{a.e. in } [0,T] \\ \sum_{i=1}^{m}(x_{ij}(t)-y_{ij}(t))(r_j(t)-\underline{r}_j(t))=0 & r(t),\underline{r}_j(t)\in L^2([0,T],\mathbb{R}^n_+) \end{cases} \tag{1.2}$$

The meaning of this condition is that, if there is an excess supply of an instrument in the economy, then its price must be the floor.

In this paper we intend to study this financial evolutionary problem, proving that the equilibrium conditions are equivalent to a variational

inequality and giving an existence result. It is worth mentioning that the presence of the prices floor does not alter the optimal assets and liabilities obtained in absence of prices floor, because the corresponding varational inequalities are equivalent to the same variational inequality (2.9).

2. STATEMENT OF THE PROBLEM AND MAIN RESULTS

Consider a financial economy consisting of m sectors, with a typical sector denoted by i, and of n instruments, with a typical financial instrument denoted by j, in the period $[0,T]$.

Let $s_i(t)$ be the total financial volume held by sector i at the time t.

The assets x_{ij} in sector i's portfolio are grouped into the column vector $x_i(t)=\left[x_{i1}(t),x_{i2}(t),...,x_{ij}(t),...,x_{in}(t)\right]^T$, the liabilities y_{ij} in sector i's portfolio into the column vector $y_i(t)=\left[y_{i1}(t),y_{i2}(t),...,y_{ij}(t),...,y_{in}(t)\right]^T$, the instrument prices $r_j(t)$ into the column vector $r(t)=\left[r_1(t),r_2(t),...,r_i(t),...,r_n(t)\right]^T$, the minimal instrument prices $\underline{r}_j(t)$ into the column vector $\underline{r}(t)=\left[\underline{r}_1(t),\underline{r}_2(t),...,\underline{r}_i(t),...,\underline{r}_n(t)\right]^T$. Moreover, we group the sector asset vectors into the matrix

$$x(t)=\begin{bmatrix} x_1^T(t) \\ ... \\ x_i^T(t) \\ ... \\ x_n^T(t) \end{bmatrix}=\begin{bmatrix} x_{11}(t) & ... & x_{1j}(t) & ... & x_{1n}(t) \\ ... & & & & \\ x_{i1}(t) & ... & x_{ij}(t) & ... & x_{in}(t) \\ ... & & & & \\ x_{m1}(t) & ... & x_{mj}(t) & ... & x_{mn}(t) \end{bmatrix}$$

and the sector liability vectors into the matrix

$$y(t)=\begin{bmatrix} y_1^T(t) \\ ... \\ y_i^T(t) \\ ... \\ y_n^T(t) \end{bmatrix}=\begin{bmatrix} y_{11}(t) & ... & y_{1j}(t) & ... & y_{1n}(t) \\ ... & & & & \\ y_{i1}(t) & ... & y_{ij}(t) & ... & y_{in}(t) \\ ... & & & & \\ y_{m1}(t) & ... & y_{mj}(t) & ... & y_{mn}(t) \end{bmatrix}.$$

Assuming as the functional setting the Lebesgue space $L^2([0,T],\mathbb{R}^p)$, the set of feasible assets and liabilities becomes:

$$P_i = \left\{ \begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix} \in L^2([0,T],\mathbb{R}^{2n}) : \sum_{j=1}^{n} x_{ij}(t) = s_i(t), \sum_{j=1}^{n} y_{ij}(t) = s_i(t) \text{ a.e. in } [0,T], \right.$$

$$\left. x_{ij}(t) \geq 0, y_{ij}(t) \geq 0 \text{ a.e. in } [0,T] \right\}$$

and the set of feasible instrument price is:

$$\mathcal{R} = \left\{ r(t) \in L^2([0,T],\mathbb{R}^n) : r_j(t) \geq \underline{r}_j(t),\ j = 1,...,n, \text{ a.e. in } [0,T] \right\}.$$

Moreover, let $Q^i(t)$ be the $2n \times 2n$ variance-covariance matrix $Q^i(t) = \begin{bmatrix} Q^i_{11}(t) & Q^i_{12}(t) \\ Q^i_{21}(t) & Q^i_{22}(t) \end{bmatrix}$ associated with sector i's assets and liabilities. We assume $Q^i(t)$ to be symmetric and positive definite with $L^\infty([0,T])$ entries. Further, we denote by $\left[Q^i_{\alpha,\beta}(t) \right]_j$ the j-th column of $Q^i_{\alpha,\beta}(t)$ with $\alpha = 1,2$ and $\beta = 1,2$. Then (see [5], [6]) the aversion to the risk at time $t \in [0,T]$ is given by:

$$\begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix}^T Q^i(t) \begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix}.$$

We can provide the following definition of an evolutionary financial equilibrium.

Definition 2.1 *A vector of sector assets, liabilities, and instrument prices* $(x^*(t), y^*(t), r^*(t)) \in \prod_{i=1}^{m} P_i \times \mathcal{R}$ *is an equilibrium of the evolutionary financial model if and only if it satisfies the system of inequalities and equalities*

$$2[Q^i_{11}(t)]^T_j x^*_i(t) + 2[Q^i_{21}(t)]^T_j y^*_i(t) - (r^*_j(t) - \underline{r}_j(t)) - \mu^{(1)}_i(t) \geq 0, \qquad (2.3)$$

$$2[Q^i_{12}(t)]^T_j x^*_i(t) + 2[Q^i_{22}(t)]^T_j y^*_i(t) + (r^*_j(t) - \underline{r}_j(t)) - \mu^{(2)}_i(t) \geq 0, \qquad (2.4)$$

$$x_{ij}^*(t)\Big[2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t) - (r_j^*(t) - \underline{r}_j(t)) - \mu_i^{(1)}(t)\Big] = 0, \tag{2.5}$$

$$y_{ij}^*(t)\Big[2[Q_{12}^i(t)]_j^T x_i^*(t) + 2[Q_{22}^i(t)]_j^T y_i^*(t) + (r_j^*(t) - \underline{r}_j(t)) - \mu_i^{(2)}(t)\Big] = 0, \tag{2.6}$$

where $\mu_i^{(1)}(t)$, $\mu_i^{(2)}(t) \in L^2([0,T])$ *are Lagrangean functions, for all sectors* i: $i = 1,2,...,m$, *and for all instruments* j, $j = 1,2,...,n$ *and verifies condition (1.2) a.e., that is*

$$\begin{cases} \sum_{i=1}^{m} (x_{ij}^*(t) - y_{ij}^*(t)) \geq 0 & \text{a.e. in } [0,T] \\ \sum_{i=1}^{m} (x_{ij}^*(t) - y_{ij}^*(t))(r_j^*(t) - \underline{r}_j(t)) = 0. \end{cases} \tag{2.7}$$

As referred in [1], the meaning of Definition 2.1 is the following: to each financial volume $s_i(t)$ held by the sector i, we associate the functions $\mu_i^{(1)}(t)$, $\mu_i^{(2)}(t)$, related, respectively, to the assets and to the liabilities and which represent the "equilibrium utilities" for unit of the sector i. The financial volume held in the instrument j as assets $x_{ij}^*(t)$ is greater or equal than zero if the j-th component

$$2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t) - (r_j^*(t) - \underline{r}_j(t))$$

of the utility is equal to $\mu_i^{(1)}(t)$, whereas if

$$2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t) - (r_j^*(t) - \underline{r}_j(t)) > \mu_i^{(1)}(t)$$

then $x_{ij}^*(t) = 0$. The same occurs for the liabilities.

The functions $\mu_i^{(1)}(t)$, $\mu_i^{(2)}(t)$ are Lagrangean functions associated, respectively, with the constraints $\sum_{j=1}^{n} (x_{ij}(t) - s_i(t)) = 0$ and $\sum_{j=1}^{n} (y_{ij}(t) - s_i(t)) = 0$. They are not known a priori, but this has not influence, since we will prove later that Definition 2.1 is equivalent to a variational inequality in which $\mu_i^{(1)}(t)$, $\mu_i^{(2)}(t)$ do not appear.

Conditions (2.7), that, as we said in the Introduction, represent the equilibrium condition for the prices, express the equilibration of the total assets and the total liabilities of each instrument.

We can give the following equivalent variational inequality formulation.

Theorem 2.1 *A vector* $(x^*(t), y^*(t), r^*(t)) \in \prod_{i=1}^{m} P_i \times \mathcal{R}$ *is an evolutionary financial equilibrium if and only if it satisfies the following variational inequality: Find* $(x^*(t), y^*(t), r^*(t)) \in \prod_{i=1}^{m} P_i \times \mathcal{R}$*:*

$$\int_0^T \{\sum_{i=1}^{m} \Big[2[Q_{11}^i(t)]^T x_i^*(t) + 2[Q_{21}^i(t)]^T y_i^*(t) - (r^*(t) - \underline{r}(t))\Big] \times [x_i(t) - x_i^*(t)]$$

$$+\sum_{i=1}^{m} \Big[2[Q_{12}^i(t)]^T x_i^*(t) + 2[Q_{22}^i(t)]^T y_i^*(t) + (r^*(t) - \underline{r}(t))\Big] \times [y_i(t) - y_i^*(t)]$$

$$+\sum_{i=1}^{m} (x_i^*(t) - y_i^*(t)) \times [r(t) - r^*(t)]\} dt \geq 0, \qquad \forall (x(t), y(t), r(t)) \in \prod_{i=1}^{m} P_i \times \mathcal{R}. \tag{2.8}$$

For (2.8) it is possible to establish the following equivalence result, from which an existence result will follow.

Theorem 2.2 *If* $(x^*(t), y^*(t), r^*(t)) \in \prod_{i=1}^{m} P_i \times \mathcal{R}$ *is a financial equilibrium, then the equilibrium asset and liability vector* $(x^*(t), y^*(t))$ *is a solution to the variational inequality:*

$$\sum_{i=1}^{m} \int_0^T \{\sum_{j=1}^{n} \Big[2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t)\Big] \times [x_{ij}(t) - x_{ij}^*(t)]$$

$$+\sum_{j=1}^{n} \Big[2[Q_{12}^i(t)]_j^T x_i^*(t) + 2[Q_{22}^i(t)]_j^T y_i^*(t)\Big] \times [y_{ij}(t) - y_{ij}^*(t)]\} dt \geq 0, \tag{2.9}$$

$$\forall (x(t), y(t)) \in S$$

where

$$S = \left\{(x(t), y(t)) \in \prod_{i=1}^{m} P_i; \sum_{i=1}^{m} (x_{ij}(t) - y_{ij}(t)) \geq 0,\ j = 1, 2, \ldots, n\right\}.$$

Conversely, if $(x^*(t), y^*(t))$ *is a solution to (2.9), then there exists an* $r^*(t) \in \mathcal{R}$ *such that* $(x^*(t), y^*(t), r^*(t))$ *is a financial equilibrium.*

3. PROOF OF THEOREM 2.1

The proof of the variational inequality formulation of the governing equilibrium conditions is obtained in the following way. In a first step we prove the equivalence between a first variational inequality and the following problem:

$$\min_{P_i} \int_0^T \left\{ \begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix}^T Q^i(t) \begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix} - (r^*(t) - \underline{r}(t)) \times [x_i(t) - y_i(t)] \right\} dt, \forall \begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix} \in P_i \tag{3.10}$$

for a fixed $r^*(t) \in \mathcal{R}$. Then we obtain the equivalence between conditions (2.3)-(2.6) and (3.10), (let us remark that at least one solution to (3.10) exists, since P_i is a bounded, convex and closed set of an Hilbert space, then also weakly compact, and the functional

$$U_i(x_i(t), y_i(t)) = \int_0^T \left\{ \begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix}^T Q^i(t) \begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix} - (r^*(t) - \underline{r}(t)) \times [x_i(t) - y_i(t)] \right\} dt$$

is weakly lower semicontinuous (see [4], Lemma 2.11, Theorem 2.3)).

In a second step we prove a variational formulation of the equilibrium condition related to the instrument prices (2.7).

From these two variational inequalities, we then derive variational inequality (2.8).

Let us start with the equivalence between problem (3.10) and a first variational inequality.

Theorem 3.1 $\begin{bmatrix} x_i^*(t) \\ y_i^*(t) \end{bmatrix}$ *is a solution to (3.10) if and only if it is a solution to the variational inequality*

$$\int_0^T \sum_{i=1}^m \left[2[Q_{11}^i(t)]^T x_i^*(t) + 2[Q_{21}^i(t)]^T y_i^*(t) - (r^*(t) - \underline{r}(t))\right] \times [x_i(t) - x_i^*(t)]\, dt$$

$$+ \int_0^T \sum_{i=1}^m \left[2[Q_{12}^i(t)]^T x_i^*(t) + 2[Q_{22}^i(t)]^T y_i^*(t) + (r^*(t) - \underline{r}(t))\right] \times [y_i(t) - y_i^*(t)] dt \geq 0,$$

$$\forall \begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix} \in P_i, \tag{3.11}$$

for a given $r^*(t) \in \mathcal{R}$.

Proof. Let us prove the necessary condition. Assume that $\begin{bmatrix} x_i^*(t) \\ y_i^*(t) \end{bmatrix}$ is a solution to problem (3.10) and consider, $\forall \begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix} \in P_i$, the function

$$F(\lambda) = \int_0^T \sum_{i=1}^m \{[\lambda x_i^*(t) + (1-\lambda)x_i(t)]^T Q_{11}^i(t)[\lambda x_i^*(t) + (1-\lambda)x_i(t)]$$

$$+[\lambda y_i^*(t) + (1-\lambda)y_i(t)]^T Q_{21}^i(t)[\lambda x_i^*(t) + (1-\lambda)x_i(t)]$$

$$+[\lambda x_i^*(t) + (1-\lambda)x_i(t)]^T Q_{12}^i(t)[\lambda y_i^*(t) + (1-\lambda)y_i(t)]$$

$$+[\lambda y_i^*(t) + (1-\lambda)y_i(t)]^T Q_{22}^i(t)[\lambda y_i^*(t) + (1-\lambda)y_i(t)]$$

$$-(r^*(t) - \underline{r}(t)) \times [\lambda x_i^*(t) + (1-\lambda)x_i(t) - \lambda y_i^*(t) - (1-\lambda)y_i(t)]\}\, dt, \quad \forall \lambda \in [0,1].$$

$\lambda = 1$ is a minimum point for $F(\lambda)$ and then $F'(1) \leq 0$.

After differentiating, we obtain

$$F'(\lambda) = \int_0^T \sum_{i=1}^m [x_i^*(t) - x_i(t)]^T Q_{11}^i(t)[\lambda x_i^*(t) + (1-\lambda)x_i(t)]\,dt$$

$$+ \int_0^T \sum_{i=1}^m [\lambda x_i^*(t) + (1-\lambda)x_i(t)]^T Q_{11}^i(t)[x_i^*(t) - x_i(t)]\,dt$$

$$+ \int_0^T \sum_{i=1}^m [y_i^*(t) - y_i(t)]^T Q_{21}^i(t)[\lambda x_i^*(t) + (1-\lambda)x_i(t)]\,dt$$

$$+ \int_0^T \sum_{i=1}^m [\lambda y_i^*(t) + (1-\lambda)y_i(t)]^T Q_{21}^i(t)[x_i^*(t) - x_i(t)]\,dt$$

$$+ \int_0^T \sum_{i=1}^m [x_i^*(t) - x_i(t)]^T Q_{12}^i(t)[\lambda y_i^*(t) + (1-\lambda)y_i(t)]\,dt$$

$$+ \int_0^T \sum_{i=1}^m [\lambda x_i^*(t) + (1-\lambda)x_i(t)]^T Q_{12}^i(t)[y_i^*(t) - y_i(t)]\,dt$$

$$+ \int_0^T \sum_{i=1}^m [y_i^*(t) - y_i(t)]^T Q_{22}^i(t)[\lambda y_i^*(t) + (1-\lambda)y_i(t)]\,dt$$

$$+ \int_0^T \sum_{i=1}^m [\lambda y_i^*(t) + (1-\lambda)y_i(t)]^T Q_{22}^i(t)[y_i^*(t) - y_i(t)]\,dt$$

$$- \int_0^T \sum_{i=1}^m (r^*(t) - \underline{r}(t)) \times [x_i^*(t) - x_i(t) - y_i^*(t) + y_i(t)]\,dt.$$

Then

$$F'(1)=\int_0^T \sum_{i=1}^m [x_i^*(t)-x_i(t)]^T Q_{11}^i(t)\,x_i^*(t)\,dt+\int_0^T \sum_{i=1}^m [x_i^*(t)]^T Q_{11}^i(t)[x_i^*(t)-x_i(t)]\,dt$$

$$+\int_0^T \sum_{i=1}^m [y_i^*(t)-y_i(t)]^T Q_{21}^i(t)\,x_i^*(t)\,dt+\int_0^T \sum_{i=1}^m [y_i^*(t)]^T Q_{21}^i(t)[x_i^*(t)-x_i(t)]\,dt$$

$$+\int_0^T \sum_{i=1}^m [x_i^*(t)-x_i(t)]^T Q_{12}^i(t)\,y_i^*(t)\,dt+\int_0^T \sum_{i=1}^m [x_i^*(t)]^T Q_{12}^i(t)[y_i^*(t)-y_i(t)]\,dt$$

$$+\int_0^T \sum_{i=1}^m [y_i^*(t)-y_i(t)]^T Q_{22}^i(t)\,y_i^*(t)\,dt+\int_0^T \sum_{i=1}^m [y_i^*(t)]^T Q_{22}^i(t)[y_i^*(t)-y_i(t)]\,dt$$

$$-\int_0^T \sum_{i=1}^m (r^*(t)-\underline{r}(t))\times[x_i^*(t)-x_i(t)-y_i^*(t)+y_i(t)]\,dt,$$

and finally, taking into account the symmetry of $Q^i(t)$,

$$F'(1)=\int_0^T \sum_{i=1}^m \left[2[Q_{11}^i(t)]^T x_i^*(t)+2[Q_{21}^i(t)]^T y_i^*(t)-(r^*(t)-\underline{r}(t))\right]\times[x_i^*(t)-x_i(t)]\,dt$$

$$+\int_0^T \sum_{i=1}^m \left[2[Q_{12}^i(t)]^T x_i^*(t)+2[Q_{22}^i(t)]^T y_i^*(t)+(r^*(t)-\underline{r}(t))\right]\times[y_i^*(t)-y_i(t)]dt\le 0,$$

that is, the variational inequality (3.11).

The sufficient condition follows as in [1].

Now we may prove equivalence between problem (3.10) or problem (3.11) and the equilibrium conditions (2.3), (2.4), (2.5), (2.6).

Theorem 3.2 $\begin{bmatrix} x_i^*(t) \\ y_i^*(t) \end{bmatrix}$ *is a solution to (3.10) or to (3.11) if and only if it satisfies, a.e. in* $[0,T]$, *conditions (2.3), (2.4), (2.5), (2.6), where* $\mu_i^{(1)}(t)$, $\mu_i^{(2)}(t) \in L^2([0,T])$ *are Lagrangean functions.*

Proof. Let $\begin{bmatrix} x_i^*(t) \\ y_i^*(t) \end{bmatrix}$ be a solution to (3.10). In order to obtain conditions

(2.3), (2.4), (2.5), (2.6) we use the infinite-dimensional Lagrangean theory (see [2], [4]).

Let us consider the function

$$\mathcal{L}(x_i(t), y_i(t), \lambda_i^{(1)}(t), \lambda_i^{(2)}(t), \mu_i^{(1)}(t), \mu_i^{(2)}(t))$$

$$= \Psi(x_i(t), y_i(t)) - \int_0^T \sum_{j=1}^n \lambda_{ij}^{(1)}(t) x_{ij}(t) dt - \int_0^T \sum_{j=1}^n \lambda_{ij}^{(2)}(t) y_{ij}(t) dt$$

$$- \int_0^T \mu_i^{(1)}(t) (\sum_{j=1}^n x_{ij}(t) - s_i(t)) dt - \int_0^T \mu_i^{(2)}(t) (\sum_{j=1}^n y_{ij}(t) - s_i(t)) dt,$$

where

$$\Psi(x_i(t), y_i(t))$$

$$= \int_0^T \sum_{j=1}^n \left[2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t) - (r_j^*(t) - \underline{r}_j(t)) \right] \times [x_{ij}(t) - x_{ij}^*(t)] \, dt$$

$$+ \int_0^T \sum_{j=1}^n \left[2[Q_{12}^i(t)]_j^T x_i^*(t) + 2[Q_{22}^i(t)]_j^T y_i^*(t) + (r_j^*(t) - \underline{r}_j(t)) \right] \times [y_{ij}(t) - y_{ij}^*(t)] dt,$$

$$\begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix} \in L^2([0,T], \mathbb{R}^{2n}) \quad \text{and} \quad (\lambda_i^{(1)}(t), \lambda_i^{(2)}(t), \mu_i^{(1)}(t), \mu_i^{(2)}(t)) \in C =$$

$$= \{ \lambda_i^{(1)}(t), \lambda_i^{(2)}(t) \in L^2([0,T], \mathbb{R}^n), \lambda_i^{(1)}(t), \lambda_i^{(2)}(t) \geq 0, \mu_i^{(1)}(t), \mu_i^{(2)}(t) \in L^2([0,T]);$$

$$i = 1, \ldots, m \}$$

Applying Lagrange Multiplier Theorem [2], it is possible to prove that there exist $\lambda_i^{(1)}(t), \lambda_i^{(2)}(t), \mu_i^{(1)}(t), \mu_i^{(2)}(t) \in C$ such that

$$\int_0^T \sum_{j=1}^n \lambda_{ij}^{(1)}(t) x_{ij}^*(t) \, dt = 0; \quad \int_0^T \sum_{j=1}^n \lambda_{ij}^{(2)}(t) y_{ij}^*(t) \, dt = 0,$$

from which it follows

$$\lambda_{ij}^{(1)}(t) x_{ij}^*(t) = 0, \quad \lambda_{ij}^{(2)}(t) y_{ij}^*(t) = 0 \quad \text{a.e. in } [0,T]. \tag{3.12}$$

Moreover, using the characterization of the solution by means of a saddle

point (see [2]), we obtain

$$\mathcal{L}(x_i(t), y_i(t), \lambda_i^{(1)}(t), \lambda_i^{(2)}(t), \mu_i^{(1)}(t), \mu_i^{(2)}(t))$$

$$= \int_0^T \sum_{j=1}^{n} \left[2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t) - (r_j^*(t) - \underline{r}_j(t)) - \lambda_{ij}^{(1)}(t) - \mu_i^{(1)}(t)\right]$$

$$\times [x_{ij}(t) - x_{ij}^*(t)]\, dt$$

$$+ \int_0^T \sum_{j=1}^{n} \left[2[Q_{12}^i(t)]_j^T x_i^*(t) + 2[Q_{22}^i(t)]_j^T y_i^*(t) + (r_j^*(t) - \underline{r}_j(t)) - \lambda_{ij}^{(2)}(t) - \mu_i^{(2)}(t)\right]$$

$$\times [y_{ij}(t) - y_{ij}^*(t)] dt \geq 0 \quad \forall \begin{bmatrix} x_i(t) \\ y_i(t) \end{bmatrix} \in L^2([0,T], \mathbb{R}^{2n}). \tag{3.13}$$

Choosing

$$x_i(t) = x_i^*(t) + \varepsilon_1(t), \quad y_i(t) = y_i^*(t) + \varepsilon_2(t),$$

with

$$\varepsilon_1(t) = 2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t) - (r_j^*(t) - \underline{r}_j(t)) - \lambda_{ij}^{(1)}(t) - \mu_i^{(1)}(t),$$

$$\varepsilon_2(t) = 2[Q_{12}^i(t)]_j^T x_i^*(t) + 2[Q_{22}^i(t)]_j^T y_i^*(t) + (r_j^*(t) - \underline{r}_j(t)) - \lambda_{ij}^{(2)}(t) - \mu_i^{(2)}(t),$$

(3.13) becomes

$$\int_0^T \sum_{j=1}^{n} \left[2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t) - (r_j^*(t) - \underline{r}_j(t)) - \lambda_{ij}^{(1)}(t) - \mu_i^{(1)}(t)\right]^2 dt$$

$$+ \int_0^T \sum_{j=1}^{n} \left[2[Q_{12}^i(t)]_j^T x_i^*(t) + 2[Q_{22}^i(t)]_j^T y_i^*(t) + (r_j^*(t) - \underline{r}_j(t)) - \lambda_{ij}^{(2)}(t) - \mu_i^{(2)}(t)\right]^2 dt \geq 0.$$

Similarly, choosing

$$x_i(t) = x_i^*(t) - \varepsilon_1(t), \quad y_i(t) = y_i^*(t) - \varepsilon_2(t),$$

we obtain

$$-\int_0^T \sum_{j=1}^{n} \left[2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t) - (r_j^*(t) - \underline{r}_j(t)) - \lambda_{ij}^{(1)}(t) - \mu_i^{(1)}(t)\right]^2 dt$$

$$-\int_0^T \sum_{j=1}^{n} \left[2[Q_{12}^i(t)]_j^T x_i^*(t) + 2[Q_{22}^i(t)]_j^T y_i^*(t) + (r_j^*(t) - \underline{r}_j(t)) - \lambda_{ij}^{(2)}(t) - \mu_i^{(2)}(t)\right]^2 dt \geq 0.$$

Then we may conclude

$$2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t) - (r_j^*(t) - \underline{r}_j(t)) - \mu_i^{(1)}(t) = \lambda_{ij}^{(1)}(t) \geq 0,$$

$$2[Q_{12}^i(t)]_j^T x_i^*(t) + 2[Q_{22}^i(t)]_j^T y_i^*(t) + (r_j^*(t) - \underline{r}_j(t)) - \mu_i^{(2)}(t) = \lambda_{ij}^{(2)}(t) \geq 0, \tag{3.14}$$

that are (2.3), (2.4).

Moreover from (3.12) and (3.14), we get (2.5), (2.6).

Conversely, supposing that (2.3)-(2.6) are fulfilled, let us show that (3.11) holds. From (2.3), (2.5) we have

$$\sum_{j=1}^{n} \left[2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t) - (r_j^*(t) - \underline{r}_j(t)) - \mu_i^{(1)}(t)\right] \times [x_{ij}(t) - x_{ij}^*(t)]\, dt \geq 0.$$

Since $\sum_{j=1}^{n} x_{ij}(t) = s_i(t)$, $\sum_{j=1}^{n} x_{ij}^*(t) = s_i(t)$ a.e. in $[0,T]$, after integrating, we derive

$$\int_0^T \sum_{j=1}^{n} \left[2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t) - (r_j^*(t) - \underline{r}_j(t))\right] \times [x_{ij}(t) - x_{ij}^*(t)]\, dt \geq 0. \tag{3.15}$$

In a similar way we get

$$\int_0^T \sum_{j=1}^{n} \left[2[Q_{12}^i(t)]_j^T x_i^*(t) + 2[Q_{22}^i(t)]_j^T y_i^*(t) + (r_j^*(t) - \underline{r}_j(t))\right] \times [y_{ij}(t) - y_{ij}^*(t)]\, dt \geq 0. \tag{3.16}$$

Summing (3.15), (3.16) for all $i = 1,...,m$, we obtain (3.11).

Now we can show the following characterization of the equilibrium condition related to the instrument prices.

Theorem 3.3 *$r^*(t) \in R$ is a solution of (2.7) if and only if it satisfies*

$$\int_0^T \sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] \times [r_j(t) - r_j^*(t)]\, dt \geq 0, \quad \forall r(t) \in R. \tag{3.17}$$

Proof. Suppose $r^*(t)$ satisfies (2.7) and define

$$E_+ = \{t \in [0,T] : r_j^*(t) > \underline{r}_j(t)\} \quad \text{and} \quad E_0 = \{t \in [0,T] : r_j^*(t) = \underline{r}_j(t)\}$$

From condition (2.7) it follows that, in E_+, $\sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] = 0$ and, in E_0, $\sum_{i=1}^m [x_{ij}^*(t) - y_{ij}^*(t)] \geq 0$, then

$$\int_0^T \sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] \times [r_j(t) - r_j^*(t)]\, dt = \int_{E_0} \sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] \times [r_j(t) - \underline{r}_j(t)]\, dt$$
$$+ \int_{E_+} \sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] \times [r_j(t) - r_j^*(t)]\, dt \geq 0,$$

that is (3.17).

Conversely, suppose (3.17) holds. We may rewrite (3.17) as

$$\int_0^T \sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] \times [r_j(t) - r_j^*(t)]\, dt = \int_{E_0} \sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] \times [r_j(t) - \underline{r}_j(t)]\, dt$$
$$+ \int_{E_+} \sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] \times [r_j(t) - r_j^*(t)]\, dt \geq 0.$$

If $\sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] > 0$ in E_+ (or in a subset of E_+ with positive measure), choosing

$$r_j(t) = \begin{cases} \underline{r}_j(t) & \text{in } E_0 \\ r_j^*(t) - \varepsilon(t) & \text{in } E_+, \end{cases}$$

where $\underline{r}_j(t) < \varepsilon(t) < r_j^*(t)$, we get

$$\int_0^T \sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] \times [r_j(t) - r_j^*(t)]\, dt = \int_{E_1} \sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] \times [-\varepsilon(t)]\, dt < 0,$$

which is an absurdity.

On the other hand, if $\sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] < 0$ in E_+ (or in a subset of E_+ with positive measure), choosing

$$r_j(t) = \begin{cases} \underline{r}_j(t) & \text{in } E_0 \\ r_j^*(t) + \varepsilon(t) & \text{in } E_+, \end{cases}$$

where $\varepsilon(t) > 0$, we reach

$$\int_0^T \sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] \times [r_j(t) - r_j^*(t)]\, dt = \int_{E_+} \sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] \times [\varepsilon(t)]\, dt < 0,$$

which is also an absurdity.

Moreover if $\sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] < 0$ in E_0 (or in a subset of E_0 with positive measure), choosing

$$r_j(t) = \begin{cases} r_j^*(t) & \text{in } E_+ \\ \underline{r}_j(t) + \varepsilon(t) & \text{in } E_0, \end{cases}$$

where $\varepsilon(t) > 0$, we get

$$\int_0^T \sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] \times [r_j(t) - r_j^*(t)]\, dt = \int_{E_0} \sum_{i=1}^m \left[x_{ij}^*(t) - y_{ij}^*(t)\right] \times [\varepsilon(t)]\, dt < 0,$$

which is an absurdity.

Then we may conclude that (2.7) holds.

From Theorems 3.1, 3.2, 3.3, it immediately follows that if $(x^*(t), y^*(t), r^*(t)) \in \prod_{i=1}^{m} P_i \times R$ is a financial equilibrium, then it satisfies variational inequalities (3.11), (3.17) and hence variational inequality (2.8) and vice versa. Thus Theorem 2.1 is completely proved.

4. PROOF OF THEOREM 2.2

Assume that $(x^*(t), y^*(t), r^*(t))$ is a financial equilibrium. Then, choosing in the variational inequality formulation (2.8) $x_i(t) = x_i^*(t)$, $y_i(t) = y_i^*(t)$, $r(t) = \underline{r}(t)$ a.e. in $[0,T]$ and $\forall i = 1,...,m$, we derive

$$\sum_{i=1}^{m} \int_0^T \sum_{j=1}^{n} \left[x_{ij}^*(t) - y_{ij}^*(t)\right] \times (\underline{r}_j(t) - r_j^*(t))\, dt \geq 0. \tag{4.18}$$

If now we set in (2.8) $(x(t), y(t)) \in \mathcal{S}$, $r(t) = r^*(t)$, it follows

$$\begin{aligned} &\sum_{i=1}^{m} \int_0^T \Big\{ \sum_{j=1}^{n} \left[2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t)\right] \times [x_{ij}(t) - x_{ij}^*(t)] \\ &\quad + \sum_{j=1}^{n} \left[2[Q_{12}^i(t)]_j^T x_i^*(t) + 2[Q_{22}^i(t)]_j^T y_i^*(t)\right] \times [y_{ij}(t) - y_{ij}^*(t)] \Big\}\, dt \qquad (4.19) \\ &\geq \int_0^T \sum_{j=1}^{n} (r_j^*(t) - \underline{r}_j(t)) \left| \sum_{i=1}^{m} (x_{ij}(t) - y_{ij}(t)) - \sum_{i=1}^{m} (x_{ij}^*(t) - y_{ij}^*(t)) \right| dt. \end{aligned}$$

The right hand side of (4.19) is nonnegative, because the constraint set $\mathcal{S}$ and (4.18), and then we derive (2.9).

Conversely, if $(x^*(t), y^*(t))$ is a solution to (2.9), let us prove that there exists $r^*(t) \in R$ such that $(x^*(t), y^*(t), r^*(t))$ is a financial equilibrium.

Let us apply the Lagrange Multiplier Theorem to the function

$$\mathcal{L}(x(t),y(t),\lambda^{(1)}(t),\lambda^{(2)}(t),\mu^{(1)}(t),\mu^{(2)}(t),r(t))$$

$$=\varphi(x(t),y(t))-\sum_{i=1}^{m}\int_0^T\sum_{j=1}^{n}\lambda_{ij}^{(1)}(t)x_{ij}(t)dt-\sum_{i=1}^{m}\int_0^T\sum_{j=1}^{n}\lambda_{ij}^{(2)}(t)y_{ij}(t)dt$$

$$-\sum_{i=1}^{m}\int_0^T\mu_i^{(1)}(t)(\sum_{j=1}^{n}x_{ij}(t)-s_i(t))dt-\sum_{i=1}^{m}\int_0^T\mu_i^{(2)}(t)(\sum_{j=1}^{n}y_{ij}(t)-s_i(t))dt$$

$$-\int_0^T\sum_{j=1}^{n}(r_j(t)-\underline{r}_j(t))\left[\sum_{i=1}^{m}(x_{ij}(t)-y_{ij}(t))\right]dt,$$

where

$$\varphi(x(t),y(t))$$

$$=\sum_{i=1}^{m}\int_0^T\sum_{j=1}^{n}\left[2[Q_{11}^i(t)]_j^T x_i^*(t)+2[Q_{21}^i(t)]_j^T y_i^*(t)\right]\times[x_{ij}(t)-x_{ij}^*(t)]\,dt$$

$$+\sum_{i=1}^{m}\int_0^T\sum_{j=1}^{n}\left[2[Q_{12}^i(t)]_j^T x_i^*(t)+2[Q_{22}^i(t)]_j^T y_i^*(t)\right]\times[y_{ij}(t)-y_{ij}^*(t)]dt,$$

$$\begin{bmatrix}x(t)\\y(t)\end{bmatrix}\in\prod_{i=1}^{m}L^2([0,T],\mathbb{R}^{2n})\text{ and }(\lambda^{(1)}(t),\lambda^{(2)}(t),\mu^{(1)}(t),\mu^{(2)}(t),r(t))$$

$$\overline{C}=\Bigg\{\lambda^{(1)}(t),\lambda^{(2)}(t)\in\prod_{i=1}^{m}L^2([0,T],\mathbb{R}^n),\ \lambda^{(1)}(t),\lambda^{(2)}(t)\geq 0,$$

$$\mu^{(1)}(t),\mu^{(2)}(t)\in L^2([0,T],\mathbb{R}^m);\quad r(t)\in L^2([0,T],\mathbb{R}^n),$$

$$r(t)-\underline{r}(t)\geq 0\ \text{ a.e. in }[0,T]\Bigg\}.$$

Note that for the Lagrangean multiplier associated to the constraint $\sum_{i=1}^{m}(x_{ij}(t)-y_{ij}(t))$ we have used the form $r_j(t)-\underline{r}_j(t)$.

By means of the Lagrange Multiplier Theorem [2], it is possible to prove that there exist $\lambda^{(1)}(t),\lambda^{(2)}(t),\mu^{(1)}(t),\mu^{(2)}(t),r^*(t)\in\overline{C}$ such that

$$2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t) - (r_j^*(t) - \underline{r}_j(t)) - \mu_i^{(1)}(t) = \lambda_{ij}^{(1)}(t) \geq 0, \tag{2.3}$$

$$2[Q_{12}^i(t)]_j^T x_i^*(t) + 2[Q_{22}^i(t)]_j^T y_i^*(t) + (r_j^*(t) - \underline{r}_j(t)) - \mu_i^{(2)}(t) = \lambda_{ij}^{(2)}(t) \geq 0. \tag{2.4}$$

$$x_{ij}^*(t)\Big[2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t) - (r_j^*(t) - \underline{r}_j(t)) - \mu_i^{(1)}(t)\Big] = 0, \tag{2.5}$$

$$y_{ij}^*(t)\Big[2[Q_{12}^i(t)]_j^T x_i^*(t) + 2[Q_{22}^i(t)]_j^T y_i^*(t) + (r_j^*(t) - \underline{r}_j(t)) - \mu_i^{(2)}(t)\Big] = 0, \tag{2.6}$$

and

$$\begin{cases} \sum_{i=1}^m (x_{ij}^*(t) - y_{ij}^*(t)) \geq 0 & \text{a.e. in } [0,T] \\ \sum_{i=1}^m (x_{ij}^*(t) - y_{ij}^*(t))(r_j^*(t) - \underline{r}_j(t)) = 0. \end{cases} \tag{2.7}$$

Using the same arguments as in the proof of Theorem 2.1, from conditions (2.3)-(2.6) it follows

$$\sum_{i=1}^m \sum_{j=1}^n \Big[2[Q_{11}^i(t)]_j^T x_i^*(t) + 2[Q_{21}^i(t)]_j^T y_i^*(t) - (r_j^*(t) - \underline{r}_j(t))\Big] \times [x_{ij}(t) - x_{ij}^*(t)]$$

$$+ \sum_{i=1}^m \sum_{j=1}^n \Big[2[Q_{12}^i(t)]_j^T x_i^*(t) + 2[Q_{22}^i(t)]_j^T y_i^*(t) + (r_j^*(t) - \underline{r}_j(t))\Big] \times [y_{ij}(t) - y_{ij}^*(t)] \geq 0$$

and from condition (2.7) we derive

$$\sum_{i=1}^m \sum_{j=1}^n [x_{ij}^*(t) - y_{ij}^*(t)] \times [r_j(t) - r_j^*(t)] \geq 0.$$

After summing and integrating, it results

$$\sum_{i=1}^{m}\int_0^T \{\sum_{j=1}^{n}\left[2[Q_{11}^i(t)]_j^T x_i^*(t)+2[Q_{21}^i(t)]_j^T y_i^*(t)-(r_j^*(t)-\underline{r}_j(t))\right]\times[x_{ij}(t)-x_{ij}^*(t)]$$

$$+\sum_{j=1}^{n}\left[2[Q_{12}^i(t)]_j^T x_i^*(t)+2[Q_{22}^i(t)]_j^T y_i^*(t)+(r_j^*(t)-\underline{r}_j(t))\right]\times[y_{ij}(t)-y_{ij}^*(t)]$$

$$+\sum_{j=1}^{n}[x_{ij}^*(t)-y_{ij}^*(t)]\times[r_j(t)-r_j^*(t)]\}\,dt\geq 0,$$

that is our equivalence result.

Finally the existence of solution is ensured since S is weakly compact and for each $(u(t),v(t))\in S$ the operator

$$(x(t),y(t))\rightarrow$$

$$\sum_{i=1}^{m}\int_0^T \{\sum_{j=1}^{n}\left[2[Q_{11}^i(t)]_j^T x_i(t)+2[Q_{21}^i(t)]_j^T y_i(t)\right]\times[u_{ij}(t)-x_{ij}(t)]$$

$$+\sum_{j=1}^{n}\left[2[Q_{12}^i(t)]_j^T x_i(t)+2[Q_{22}^i(t)]_j^T y_i(t)\right]\times[v_{ij}(t)-y_{ij}(t)]\}\,dt$$

is weakly upper semicontinuous (see [1]).

REFERENCES

[1] P. Daniele, *Variational Inequalities for Evolutionary Financial Equilibrium*, Advances in Economic and Financial Networks, A. Nagurney Ed. (2003), 84-108.

[2] P. Daniele, *Lagrangean Funtion for Dynamic Variational Inequalities*, Rendiconti del Circolo Matematico di Palermo, Serie II, Suppl. 58, 101-119.

[3] J. Dong, D. Zhang and A. Nagurney, *A projected Dynamical Systems Model of General Financial Equilibrium with Stability Analysis*, Mathematical and Computer Modeling 24 (1996), 35-44.

[4] J. Jahn, Introduction to the Theory of Nonlinear Optimization, Springer-Verlag, Berlin, 1996.

[5] H.M. Markowitz, *Portfolio Selection*, Journal of Finance 7 (1952), 77-91.

[6] H.M. Markowitz, Portfolio Selection: Efficient Diversification of Investments, Wiley & Sons, New York, 1959.

[7] A. Nagurney, *Variational Inequalities in the Analysis and Computation of Multi-Sector Multi-Instrument Financial Equilibria*, Journal of Economic Dynamics and Control 18 (1994), 161-184.

[8] A. Nagurney, Network Economics - A Variational Inequality Approach, second and revised version, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.

[9] A. Nagurney, *Financial and Variational Inequalities*, Quantitative Finance 1 (2001), 309-317.

[10] A. Nagurney, J. Dong and M. Hughes, *Formulation and Computation of General Financial Equilibrium*, Optimization 26 (1992), 339-354.

[11] A. Nagurney and K. Ke, *Financial Networks with Intermediation*, Quantitative Finance 1 (2001), 441-451.

[12] A. Nagurney and S. Siokos, *Variational Inequalities for International General Financial Equilibrium Modeling and Computation*, Mathematical and Computer Modelling 25 (1997), 31-49.

[13] A. Nagurney and S. Siokos, Financial Networks: Statics and Dynamics, Springer-Verlag, Heidelberg, Germany, 1997.

[14] A. Nagurney and D. Zhang, Projected Dynamical Systems and Variational Inequalities with Applications, Kluwer Academic Publishers, Boston, Massachussets, 1996.

REMARKS ABOUT DIFFUSION MEDIATED TRANSPORT: THINKING ABOUT MOTION IN SMALL SYSTEMS

S. Hastings[1] and D. Kinderlehrer[2*]
Dept. of Mathematics, University of Pittsburgh, Pittsburgh, PA, USA;[1] Center for Nonlinear Analysis and Department of Mathematical Sciences Carnegie Mellon University, Pittsburgh, PA, USA[2]

Abstract: We describe a dissipation principle/variational principle which may be useful in modeling motion in small viscous systems and provide brief illustrations to brownian motor or molecular rachet situations which are found in intracellular transport. Monge-Kantorovich mass transport and Wasserstein metric play an interesting role in these developments. Some properties of the system that ensure the presence of transport are discussed.

INTRODUCTION

Here we describe a dissipation principle that describes transport in a typical molecular motor system, like conventional kinesin, [20], [22]. As background to this application, we recount that intracellular transport in eukarya is attributed to motor proteins that transduce chemical energy into directed mechanical motion. Muscle mysosin has been known since the mid-nineteenth century and its role in muscle contraction demonstrated by A.F. Huxley and H.E. Huxley in the 1950's. Kinesins and their role in intracellular transport were discovered around 1985. These nanoscale motors

* Partially supported by the National Science Foundation Grants DMS 0072194 and DMS 0305794.

tow organelles and other cargo on microtubules. They function in a highly viscous setting with overdamped dynamics; the Reynolds' number is about 5×10^{-2}. The dissipation principle begins a chain of events. It suggests, in a natural way, a variational principle and an implicit scheme in the sense of Otto [14], [15] and Jordan, Kinderlehrer and Otto [9]. This determines, in turn, a system of equations analogous to that proposed by Adjari and Prost [1] or Peskin, Ermentrout, and Oster [18]. Viewed as an ensemble, this system occupies configurations that are distant from conventional notions of equilibrium. Thismeans that to understand the stability properties of the process we must discover an appropriate environment for its kinetics. The novelty in our development is that the dynamical process is set in a weak topology as described by a Kantorovich-Wasserstein metric. This owes in part to a result of Brenier and Benamou, [3]. It illustrates the feasibility of mesoscale modeling for these systems.

The flashing rachet, a different type of Brownian motor, was discussed in [10]. One explanation of this was given in [2] and it has been suggested as a description of processivity in the KIF-1A family of kinesins, [12], [13].There is a discussion in [6] as well as the Parrondo Paradox, a coin toss game somethimes thought to mimic molecular motor behavior, in [7].

With a thermodynamically consistent system of differential equations in hand, we inquire of conditions that ensure transport. In the example we describe, a model for conventional kinesin, diffusion and conformational change collaborate with transport in periodic potentials. This model is highly over simplified. Asymmetry of the potentials within their period intervals is critical for transport, and a particular such condition based on this property is explained.

This is a description of joint work with Michal Kowalczyk, Michel Chipot, and Jean Dolbeault, to whom we are grateful for their collaboration.

1. A VARIATIONAL PRINCIPLE

Consider an ensemble of statistically homogeneous non-interacting particles in a highly viscous medium, thought of simply as spring-mass-dashpots. For our setup, suppose we have probability densities $f^*(x)$ and $f(x), x \in \Omega = (0,1)$, and interpolating densities $f(x,t), x \in \Omega, 0 \leq t \leq \tau$ with $f^*(x) = f(x,0)$ and $f(x) = f(x,\tau)$. For this 'Eulerian' description, there is a 'Lagrangian' description in terms of a family of measure preserving mappings, transfer functions, homeomorphisms of the interval into itself, $\phi(x,t), x \in \Omega, 0 \leq t \leq \tau$ related by

$$\int_\Omega \zeta(y)f(y,t)dy = \int_\Omega \zeta(\phi(x,t))f^*(x)dx.$$

The velocities in the two descriptions satisfy $\phi_t(x,t) = v(\phi(x,t),t)$. For $f(x,t)$ there is

$$f_t + (vf)_x = 0 \text{ in } \Omega, 0 < t < \tau \quad (\textit{continuity equation}) \tag{1}$$

and likewise in the 'Lagrangian' version

$$f(\phi(x,t),t)\phi_x = f^*(x). \tag{2}$$

This is actually the Monge-Ampere Equation. For example, if v is given and we wish to solve (1), (2) corresponds to a characteristic equation.

For the ensemble of spring-mass-dashpots, the viscous dissipation moving from f^* to f via $f(x,t)$ is simply

$$\gamma \int_0^\tau \int_\Omega v^2 f dx dt$$

for a parameter γ. When the system moves in response to a potential ψ, its free energy at a density φ is

$$F(\varphi) = \int_\Omega (\psi\varphi + \sigma\varphi \ log \ \varphi)dx$$

In this way, we arrive at a simple mesoscopic dissipation principle. The state f is admissible from f^* provided

$$\gamma \int_0^\tau \int_\Omega v^2 f dx dt + F(f) \leq F(f^*) \tag{3}$$

for some interpolating density $f(x,t)$ with $f^*(x) = f(x,0)$ and $f(x) = f(x,\tau)$. We regard τ as a relaxation time. To connect this to a variational principle, we observe that [3]

$$\frac{1}{2\tau} d(f,f^*)^2 = inf_A \frac{1}{2} \int_0^\tau \int_\Omega v^2 f dx dt \tag{4}$$

where A is the family of interpolating densities and d is the Kantorovich-Wasserstein metric defined by

$$d(f,f^*)^2 = inf_P \int_{\Omega\times\Omega} |\, x-y\,|^2 \; dp(x,y)$$

$$P = \textit{joint distributions with marginals } f, f^*.$$

The optimality condition for f, v in (4) is

$$v_t + vv_x = 0 \textit{ in } \Omega, 0 < t < \tau \quad (\textit{Burgers' Equation})$$

Its 'Lagrangian' form is the geodesic equation, [3], [16],

$$\frac{d^2}{dt^2} d(\phi(x,t),\phi(x,\tau))^2 = 0$$

which implies

$$\phi(x,t) = x + \frac{t}{\tau}(\phi(x,\tau) - x), x \in \Omega, 0 < t < \tau$$

The metric d delivers the weak* topology on measures, i.e., its topology as the dual space of $C(\Omega)$, and the 'Lagrangian' form suggests that the optimality condition describes a geodesic path in this space.

For convenience we set $\gamma = \frac{1}{2}$. Our variational principle is now: given f^*, determine f such that

$$\frac{1}{2\tau} d(f,f^*)^2 + F(f) = min \tag{5}$$

The variational principle (5) provides an implicit scheme: Given $f^{(k-1)}$, set $f^* = f^{(k-1)}$ and determine f^k from the minimum principle. Then define $f^{(\tau)}$

$$f^{(\tau)}(x,t) = f^k(x) \quad k\tau < t \leq (k+1)\tau$$

The great merit of the Wasserstein metric is that it may be, in essence, differentiated. Thus, in the limit as $\tau \to 0, f^{(\tau)}$ tends to the solution f of the ordinary Fokker-Planck Equation, [9], [14], [15],

$$\frac{\partial f}{\partial t} = \sigma \frac{\partial^2 f}{\partial x^2} + \frac{\partial}{\partial x}(\psi' f) \qquad in \quad \Omega, t > 0 \tag{6}$$

$$\sigma \frac{\partial}{\partial x} f + \psi' f = 0 \qquad on \quad \partial\Omega, \; t > 0 \tag{7}$$

Variational principles such as (5) above may be considered without discussing natural systems, of course, and there is now a significant literature in this topic, and even traditional problems have unexpected interpretations, [21]. (5) establishes that the coarse graining of the microscopic system gives rise to weak topology dynamics at the mesoscale. For situations, like the one below, where equilibrium is never achieved, this may provide additional insight into their metastable nature.

From the analysis point of view, one observes that the basic variational principle is convex and superlinear, so existence of the iterates in the implicit scheme is not usually a difficulty. Convergence as $\tau \to 0$ could be, especially for nonlinear problems.

2. A LOOK AT CONVENTIONAL KINESIN

Conventional kinesin has two identical head domains (heavy chains) which walk in a hand over hand fashion along a rigid microtubule. This is an intricate process with a complicated transformation path comprising both the ATP hydrolysis (chemical states)and the motion (mechanical states), [8], [22]. For a crude reckoning, at a gross combinatorial level, each head is attached or in motion and is nucleotide bound or not. Assuming that a given motor has one head bound and one free at any instant leads to eight possible pathways for each cycle. We shall give a simplified description by considering the nucleotide binding and then the subsequent motion. Our dissipation/variational principle is flexible enough to accomodate this process.

The ensemble of motor heads may be divided into two sets, set 1 and set 2; for example, the set 1 motors bind to odd labled sites on microtubules and the set 2 motors bind to even labeled sites at a given time t.This permits distance along the microtubule to be used as a process variable. Regard the conformational change and nucleotide binding to be the result of first order chemistry and the motion to be the result of interaction with potentials, diffusion, and dissipation. Let ρ_1 and ρ_2 denote the relative densities of the set 1 and set 2 motors in the powerstroke state. Introduce potentials and coefficients for conformational change,

$\sigma > 0$ *constant*

$\psi_i \geq 0$ *and* $\nu_i \geq 0, i = 1,2,$ *smooth and periodic of period* $\frac{1}{N}$

with $supp\,\nu_1 = supp\,\nu_2$ and $\nu_1 + \nu_2 \leq 1$. Let

$$P = 1 + \tau \begin{pmatrix} -\nu_1 & \nu_2 \\ \nu_1 & -\nu_2 \end{pmatrix}$$

where τ is a relaxation time. Denote the free energy of this system by

$$F(\rho) = \sum_{i=1}^{2} \int_\Omega (\psi_i \rho_i + \sigma \rho_i \ log \ \rho_i) dx \tag{8}$$

We may envision a cycle starting with density $\rho^* = (\rho_1^*, \rho_2^*)$ and proceeding by

$$\rho^* \to \rho^* P \to \rho$$

subject to the dissipation principle: given ρ^* with

$$\int_\Omega (\rho_1^* + \rho_2^*) dx = 1 \quad and \quad \rho_i^* \geq 0 \ in \ \Omega, \tag{9}$$

determine ρ by

$$\sum_{i=1}^{2} \frac{1}{2\tau} d(\rho_i, (\rho^* P)_i)^2 + F(\rho) = min \tag{10}$$

$$\int_\Omega \rho_i dx = \int_\Omega (\rho^* P)_i dx \tag{11}$$

The variational principle (10) separates the roles of the dissipation, conformational change, and free energy in the system. It gives the incremental state of the system in terms of a step in a Markov chain from its prior state. Although there are some subtleties, (10) admits an Euler Equation which is the system [5]

$$\frac{\partial \rho_1}{\partial t} = \frac{\partial}{\partial x}(\sigma \frac{\partial \rho_1}{\partial x} + \psi'_1 \rho_1) - \nu_1 \rho_1 + \nu_2 \rho_2 \ in \ \Omega, t > 0 \tag{12}$$

$$\frac{\partial \rho_2}{\partial t} = \frac{\partial}{\partial x}(\sigma \frac{\partial \rho_2}{\partial x} + \psi'_2 \rho_2) + \nu_1 \rho_1 - \nu_2 \rho_2 \ in \ \Omega, t > 0 \tag{13}$$

$$\sigma \frac{\partial \rho_1}{\partial x} + \psi'_1 \rho_1 = 0 \text{ on } \partial\Omega, t > 0$$

$$\sigma \frac{\partial \rho_2}{\partial x} + \psi'_2 \rho_2 = 0 \text{ on } \partial\Omega, t > 0$$

$$\rho_i(x,0) = \rho_i^0 \geq 0, \qquad \text{in } \Omega, \quad i = 1,2$$

$$\int_\Omega (\rho_1 + \rho_2) dx = 1$$

and moreover this system has a solution for all time. The general program to obtain (12), (13) from the variational principle (10) consists of two parts. First there is some type of estimate of iterates and second an approximate Euler Equation. When estimating the left hand side of (5), we choose f^* as a test function, which gives

$$\frac{1}{2\tau} d(f, f^*)^2 + F(f) \leq F(f^*)$$

When applied to the sequence of iterates (f^k), this provides the basic estimate

$$\frac{1}{2\tau} \sum_{k=1}^{\infty} d(f^{k-1}, f^k)^2 \leq F(f^0) \quad \text{and}$$
$$F(f^k) \leq F(f^0), \quad k = 1,2,3,...$$

In our variational principle (10), $\rho^* P$ is an admissible competitor but ρ^* is not. Hence

$$\sum_{i=1}^{2} \frac{1}{2\tau} d(\rho_i, (\rho^* P)_i)^2 + F(\rho) \leq F(\rho^* P) \tag{14}$$

To replace $\rho^* P$ by ρ^* in (14), we use the simple property of Markov chains that relative entropy of successive states decreases. Namely, for a probablity matrix P with stationary state $\mu^\sharp$, given a vector ofnon-negative components μ,

$$\sum_{j=1}^{n} (\mu P)_j log \frac{(\mu P)_j}{\mu_j^\sharp} \leq \sum_{j=1}^{n} \mu_j log \frac{\mu_j}{\mu_j^\sharp}$$

For the 2 matrix P, the (x-dependent) stationary state is just proportional to (ν_2, ν_1) so we obtain

$$\sum_{i=1}^{2} \frac{1}{2\tau} d(\rho_i, (\rho^* P)_i)^2 + F(\rho) \leq F(\rho^*) + const.\tau \tag{15}$$

This estimate is sufficient to establish the approximate Euler equation

$$\begin{aligned} &| \sum_{i=1,2} \int_\Omega \{(\frac{1}{\tau}(\rho_i - \rho_i^*) - (\rho^* \nu)_i)\zeta_i - \sigma \rho_i \zeta''_i + \psi' \rho_i \zeta'_i\} dx | \\ &\leqq \frac{1}{2} max\ sup\ | \zeta''_i | (F(\rho^*) - F(\rho) + C\tau), \quad \zeta \in C_0^\infty(\Omega) \end{aligned} \tag{16}$$

and to prove that the sequence $\rho^{(\tau)}$,

$$\rho^{(\tau)}(x,t) = \rho^k(x) \quad k\tau < t \leq (k+1)\tau,$$

converges as $\tau \to 0$ to a solution of (12), (13). Along the way, we are assisted by a novel maximum principle. Suppose that ρ is the solution of the (10) for ρ^*. If

$$\frac{\rho_i^*}{e^{-\psi_i/\sigma}} \leq M_i$$

then

$$\frac{\rho_i}{e^{-\psi_i/\sigma}} \leq M_i(1 + \alpha\tau) \quad i = 1,2$$

for a suitable $\alpha > 0$. The interesting feature is that the proof is a truncation argument involving joint distributions. The first use of the idea was by Otto, [15], and new ingredients have been added to it by Petrelli and Tudorascu, [19]. There is a similar minimum principle. These estimates do not permit us to deduce the behaviour of the system as $t \to \infty$, which will be dealt with elsewhere.

The foregoing may be generalized easily to n species with potentials ψ_i and a matrix $\nu = (\nu_{ij})$, with $\nu_{ij} > 0$ for $i \neq j$ and $\sum_{j=1}^{n} \nu_{ij} = 0$. i.e., P as defined above a probability matrix. Allowing more complex interactions

among the species requires more thought about the form of the interactions and their statistical properties.

3. THE STATIONARY SOLUTION

There is, in addition, a unique stationary solution $\rho^\sharp$ of (12), (13) provided

$$\nu_1 \geq 0 \quad \textit{and} \quad \nu_2 \geq 0$$

and neither are identically zero. Namely, $\rho^\sharp$ is the solution of the system of ordinary differential equations [4]

$$\frac{d}{dx}(\sigma \frac{d\rho_1^\sharp}{dx} + \psi'_1 \rho_1^\sharp) - \nu_1 \rho_1^\sharp + \nu_2 \rho_2^\sharp = 0 \ \textit{in} \ \Omega \tag{17}$$

$$\frac{d}{dx}(\sigma \frac{d\rho_2^\sharp}{dx} + \psi'_2 \rho_2^\sharp) + \nu_1 \rho_1^\sharp - \nu_2 \rho_2^\sharp = 0 \ \textit{in} \ \Omega \tag{18}$$

$$\sigma \frac{d\rho_1^\sharp}{dx} + \psi'_1 \rho_1^\sharp = 0 \ \textit{on} \ \partial\Omega$$

$$\sigma \frac{d\rho_2^\sharp}{dx} + \psi'_2 \rho_2^\sharp = 0 \ \textit{on} \ \partial\Omega$$

$$\int_\Omega (\rho_1^\sharp + \rho_2^\sharp) dx = 1$$

Note that in general $\rho^\sharp$ does not minimize (8). There are two ways to attack this, one starting with the Schauder Fixed Point Theorem and one by a shooting method, based on writing (17),(18) as a first order system, [4].

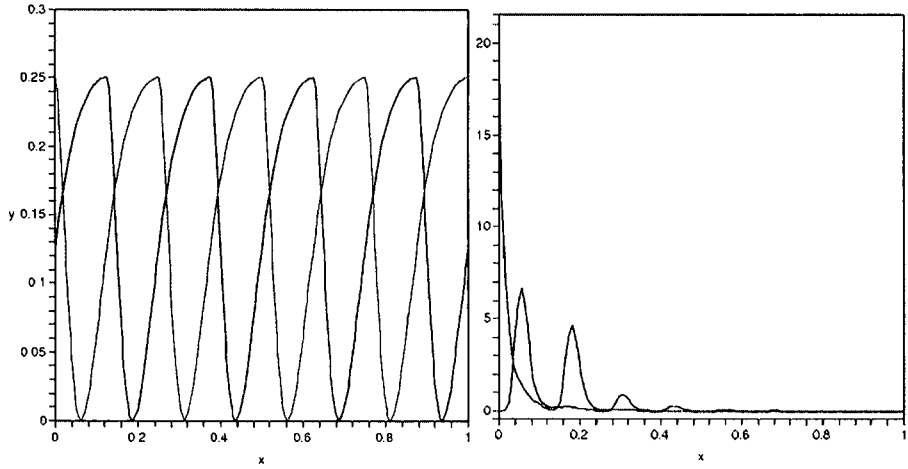


Figure 1. Interdigitated asymmetric potentials ψ_1 and ψ_2 (left) and stationary state $\rho^{\natural}$ demonstrating about 0,9 of its mass on the left half of the interval.

We would like to briefly discuss the origins of transport and the role of the asymmetry of the potentials. Assume that ψ_1 and ψ_2 are periodic of period $1/N$, in fact, for purposes of discussion,

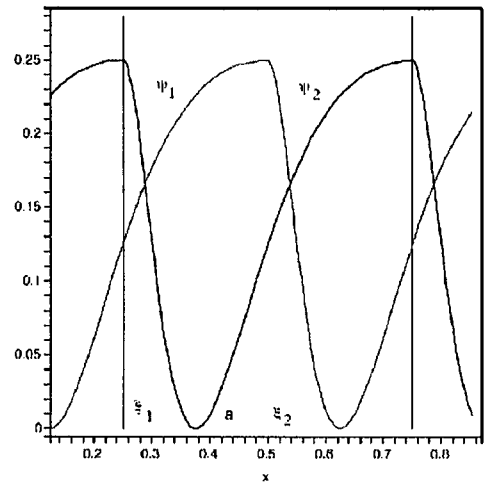


Figure 2. For this pair of ψ_1 and ψ_2, there is no interval where both are decreasing and transport to the left is anticipated

let us take

$$\psi_2(x) = \psi_1(x - \frac{1}{2N})$$

so that they interdigitate each other. Assume that ψ_i decreases monotonely from its maximum to its minimum and then increases monotonely to its maximum in each period interval. Choose a period interval, max to max, for ψ_1, say $[\xi_1, \xi_1 + 1/N]$ and suppose we are in the situation where

$$\xi_1 < a < \xi_2 < b < \xi_1 + 1/N \text{ and}$$
$$\psi_1(\xi_1) = \psi_1(\xi_1 + 1/N) = max\,\psi_1, \quad \psi_1(a) = min\,\psi_1 = 0$$
$$\psi_2(\xi_2) = max\,\psi_2, \quad \psi_2(b) = min\,\psi_2 = 0$$

Think of σ as very small. Now we have that

1. in (a, ξ_2), $\psi_1 > 0$ and $\psi_2 > 0$, so $\rho_1^\sharp$ and $\rho_2^\sharp$ are both exponentially decreasing regardless of ν_i.
2. in (ξ_1, a), there is a large population of $\rho_1^\sharp$, and, because of the equations (17), (18), some is passed to $\rho_2^\sharp$ because $\nu_2 > 0$. Little is passed from $\rho_2^\sharp$ to $\rho_1^\sharp$ because we are not close to the minimum of ψ_2.
3. the net effect is movement to the left

The condition for the balance in 2, and for similar behavior near the minima of ψ_2, is that

- ψ_1 is increasing where ψ_2 is decreasing and ψ_2 is increasing where ψ_1 is decreasing.

This means, in particular, that the minima of the $\psi_i s$ are located asymmetrically in their period intervals. Unfortunately, the above reads like just one of many plausible scenarios and so does not serve well for intuition, but it is the correct one. The result may be loosely formulated in this way:

Suppose there is no interval where ψ_1 and ψ_2 are both decreasing, and

$$\nu_1 > 0 \text{ and } \nu_2 > 0 \text{ in } \Omega$$

then

$$\rho_1^\sharp(x + \frac{1}{N}) + \rho_2^\sharp(x + \frac{1}{N}) \leq Ke^{-\frac{c}{\sigma}}(\rho_1^\sharp(x) + \rho_2^\sharp(x)), \qquad x \geq 1 + \frac{2}{N} \quad (19)$$

To prove this result, we rewrite (12), (13) as the first order system (dropping the $^\sharp$ superscript), with

$$\phi = \sigma\rho_1' + \psi_1'\rho_1,$$

$$\sigma\rho_1 = \phi - \psi_1'\rho_1 \tag{20}$$

$$\sigma\rho_2 = -\phi - \psi_2'\rho_2 \tag{21}$$

$$\phi' = \nu_1\rho_1 - \nu_2\rho_2 \tag{22}$$

$$\phi(0) = \phi(1) = 0$$

or

$$\rho' = A\rho, \text{ with } \rho = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \phi \end{pmatrix} \text{ and } A = \frac{1}{\sigma}\begin{pmatrix} -\psi_1' & 0 & 1 \\ 0 & -\psi_2' & -1 \\ \sigma\nu_1 & -\sigma\nu_2 & 0 \end{pmatrix} \tag{23}$$

Let $R(\xi, x)$ be a fundamental solution to this system with $R(\xi, \xi) = 1$, say. Write

$$R = \begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{21} & \rho_{22} & \rho_{23} \\ \phi_1 & \phi_2 & \phi_3 \end{pmatrix} \tag{24}$$

Thus, in particular,

$$\rho(a) = R(\xi_1, a)\rho(\xi_1) \quad \text{and} \quad \rho(\xi_2) = R(a, \xi_2)\rho(a)$$

Since $\rho_i > 0$, the additional function ϕ can be eliminated from the equation in favor of an inequality. Indeed,

$$0 < \rho_1(x) = \rho_{11}\rho_1(\xi) + \rho_{12}\rho_2(\xi) + \rho_{13}\phi(\xi), \quad x < \xi \tag{25}$$

$$0 < \rho_2(x) = \rho_{21}\rho_1(\xi) + \rho_{22}\rho_2(\xi) + \rho_{23}\phi(\xi), \quad x < \xi \tag{26}$$

where the ρ_{ij} are evaluated at x. Hence,

$$\phi(\xi) < -\frac{\rho_{11}}{\rho_{13}}\rho_1(\xi) - \frac{\rho_{12}}{\rho_{13}}\rho_2(\xi) \text{ and}$$
$$\phi(\xi) < -\frac{\rho_{21}}{\rho_{23}}\rho_1(\xi) - \frac{\rho_{22}}{\rho_{23}}\rho_2(\xi)$$

Combining this with (25), (26) and reconfiguring gives that

$$\rho_1(x) < \frac{\rho_{13}\rho_{21} - \rho_{11}\rho_{23}}{-\rho_{23}}\rho_1(\xi) + \frac{\rho_{22}\rho_{13} - \rho_{12}\rho_{23}}{-\rho_{23}}\rho_2(\xi)$$
$$\rho_2(x) < \frac{\rho_{13}\rho_{21} - \rho_{11}\rho_{23}}{\rho_{13}}\rho_1(\xi) + \frac{\rho_{22}\rho_{13} - \rho_{12}\rho_{23}}{\rho_{13}}\rho_2(\xi)$$

A first thought is that when a typical ρ_{ij} varies with $exp(c/\sigma)$, the fraction varies like $exp(c/\sigma)^2/exp(c/\sigma) = exp(c/\sigma)$, that is, exponential in $1/\sigma$. Interesting here is that the numerators in the fractions are the terms $(adjR)_{23}$ and $(adjR)_{13}$ and the adjugate itself satisfies an equation (variation of Abel's formula)

$$\frac{d}{dx}adjR = adjRM, \quad M = (traceA)1 - A$$

which means that the numerator and the denominator are typically of the same order. This is the starting point of the proof. The details require careful analysis of R and $adjR$ in the appropriate intervals.

With Bryce McLeod, we are preparing a second approach which would extend to an arbitrary number of components ρ_i weakly coupled by matrix $N = (\nu_{ij})$.

At this writing, the relationship of the supports of the conformational change coefficients ν_i and the potentials ψ_i is still not clear. One obvious situation where no transport can be expected is when the system (12), (13) decouples. This happens when

$$\nu \propto (e^{-\frac{\psi_2}{\sigma}}, e^{-\frac{\psi_1}{\sigma}}) \tag{27}$$

This is sometimes referred to as detailed balance, but it only concerns the balance in part of the equations. However, even in this case, retaining the $\sigma = \sigma_0$ above in (27) but diminishing sufficiently the diffusion coefficient σ in (12), (13) will result in transport according to our theorem provided the ν_i are positive.

A more amazing result is given in the last figure. Here the potentials are the same as before, although there are eight periods instead of four, but the support of the ν_i are where one ψ_j is decreasing and the other increasing. The result istransport in the reverse, that is the "wrong" direction. Much remains to be studied in these problems.

REFERENCES

[1] Adjari, A. and Prost, J. (1992) Mouvement induit par un potentiel périodique de basse symétrie: dielectrophorese pulse, C. R. Acad. Sci. Paris t. **315**, Série II, 1653.

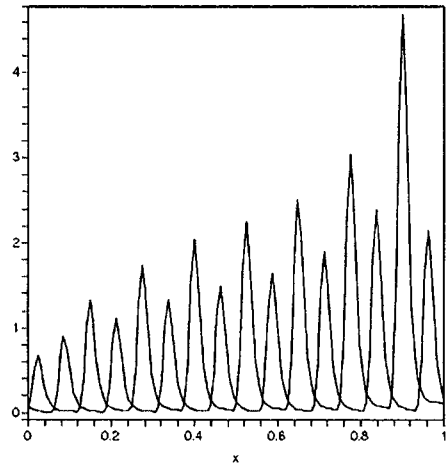


Figure 3. Reverse transport achieved by selecting the support of the ν_i in a region where one ψ_j is increasing and the other decreasing

[2] Astumian, R.D. (1997) Thermodynamics and kinetics of a Brownian motor, Science **276** (1997), 917–922.

[3] Benamou, J.-D. and Brenier, Y. (2000) A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, Numer. Math. **84**, 375–393.

[4] Chipot, M., Hastings, S., and Kinderlehrer, D., to appear

[5] Chipot, M., D. Kinderlehrer, D. and Kowalczyk, M. (2003) A variational principle for molecular motors, Meccanica, **38**, 505–518

[6] Dolbeault, J., Kinderlehrer, D., and Kowalczyk, M. Remarks about the flashing rachet, to appear Proc. PASI 2003

[7] Heath, D., Kinderlehrer, D. and Kowalczyk, M. (2002) Discrete and continuous ratchets: from coin toss to molecular motor, Discrete and continuous dynamical systems Ser.B **2** no. 2 , 153–167.

[8] Howard, J. (2001) Mechanics of Motor Proteins and the Cytoskeleton, Sinauer Associates, Inc., 2001.

[9] Jordan, R., Kinderlehrer, D. and Otto, F. (1998) The variational formulation of the Fokker-Planck equation, SIAM J. Math. Anal. Vol. **29** no. 1, 1–17.

[10] Kinderlehrer, D. and Kowalczyk, M (2002) Diffusion-mediated transport and the flashing ratchet, Arch. Rat. Mech. Anal. **161**, 149–179.

[11] Kinderlehrer, D. and Walkington, N. (1999) Approximation of parabolic equations based upon Wasserstein's variational principle, Math. Model. Numer. Anal. (M2AN) **33** no. 4, 837–852.

[12] Okada, Y. and Hirokawa, N. (1999) A processive single-headed motor: kinesin superfamily protein KIF1A, Science Vol. **283**, 19

[13] Okada, Y. and Hirokawa, N. (2000) Mechanism of the single headed processivity: diffusional anchoring between the K-loop of kinesin and the C terminus of tubulin, Proc. Nat. Acad. Sciences **7** no. 2, 640–645.

[14] Otto, F. (1998) Dynamics of labyrinthine pattern formation: a mean field theory, Arch. Rat. Mech. Anal. **141**, 63-103

[15] Otto, F. (2001) The geometry of dissipative evolution equations: the porous medium equation, Comm. PDE **26**, 101-174

[16] Otto, F. and Villani, C. (2000) Generalization of an inequality by Talagrand and links with the logarithmic Sobolev Inequality, J. Funct. Anal. **173**, 361–400

[17] Parmeggiani,A., Jülicher, F., Adjari, A. and Prost, J. (1999) Energy transduction of isothermal ratchets: generic aspects and specific examples close and far from equilibrium, Phys. Rev. E, **60** no. 2, 2127–2140.

[18] Peskin, C.S.. Ermentrout, G.B. and Oster, G.F. (1995) *The correlation ratchet: a novel mechanism for generating directed motion by ATP hydrolysis*, in Cell Mechanics and Cellular Engineering (V.C Mow et.al eds.), Springer, New York

[19] Petrelli, L. and Tudorascu, A. Variational principle for general Fokker-Planck equations , to appear

[20] Reimann, P. (2002) Brownian motors: noisy transport far from equilibrium, Phys. Rep. **361** nos. 2–4, 57–265.

[21] Tudorascu, A. A one phase Stefan problem via Monge-Kantorovich theory,(CNA Report 03-CNA-007)

[22] Vale, R.D. and Milligan, R.A. (2000) The way things move: looking under the hood of motor proteins, Science **288**, 88–95.

[23] C. Villani (2003) Topics in optimal transportation, AMS Graduate Studies in Mathematics vol. 58, Providence

AUGMENTED LAGRANGIAN AND NONLINEAR SEMIDEFINITE PROGRAMS[1]

X. X. Huang, X. Q. Yang and K. L. Teo
Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, P.R. of China

Abstract: In this paper, we introduce an augmented Lagrangian for nonlinear semidefinite programs. Some basic properties of the augmented Lagrangian such as differentiabilty, monotonicity and convexity, are discussed. Necessary and sufficient conditions for a strong duality property and an exact penalty representation in the framework of augmented Lagrangian are derived. Under certain conditions, it is shown that any limit point of a sequence of stationary points of augmented Lagrangian problems is a Karuh, Kuhn-Tucker (for short, KKT) point of the original semidefinite program.

Key words: Semidefinite programming, augmented Lagrangian, duality, exact penalization, convergence, stationary point.

1. INTRODUCTION

It is well-known that semidefinite programming has wide applications in engineering, economics and combinatorial optimization and has received considerable attention in the optimization community (see, e.g., [23,11] and the references therein). Linear semidefinite programs are mainly solved by interior-point algorithms (see, e.g., [23,25,24,2,15] and the references therein). Nonlinear semidefinite programming arises in optimal structural

[1] This work is supported by a Postdoctoral Fellowship of The Hong Kong Polytechnic University.

design (see [18]), optimal robust control (see [11]) and robust feedback control design (see [10]). Various applications of nonlinear (nonconvex) semidefinite optimization were recently summarized in [1,14].

It is worth noting that the study of nonlinear semidefinite programming, in particular, nonconvex semidefinite programming is very limited (see [17,21,4,9,1,14]). Recently, a class of penalty/barrier multiplier methods was proposed for the solution of convex semidefinite programming with a linear matrix inequality constraint (see [16]). More recently, a class of semidefinite programs have been solved by converting them into nonlinear programs (see [5,6]). Barrier methods were developed for nonlinear semidefinite programs (see [18,1,14]). However, barrier methods require a strict (interior) feasible solution as the starting feasible point, which is not easy to be found even if it exists. Augmented Lagrangian method is popular and effective in constrained nonlinear programming (see, e.g., [3,19,20]). An advantage of augmented Lagrangian method is that it is robust and need not a starting feasible point.

In this paper, we propose an augmented Lagrangian approach to a nonlinear semidefinite program. Strong duality and exact penalization results are established. Furthermore, it is shown that, under certain assumptions, any limit point of first order stationary points of augmented Lagrangian problems is a KKT (stationary) point of the original semidefinite program.

The outline of the paper is as follows. In Sect. 2, we introduce augmented Lagrangian for nonlinear semidefinite programming problems. We also investigate some basic properties of the augmented Lagrangian. In Sect. 3, we study necessary and sufficient conditions for a zero duality gap property between augmented Lagrangian dual problem and the original semidefinite program. In Sect. 4, conditions for the exact penalty representation in the framework of augmented Lagrangian are established. In Sect. 5, we show that any limit point of a sequence of stationary points of the augmented Lagrangian problems satisfies the KKT condition of the original semidefinite program.

2. AUGMENTED LAGRANGIAN

Consider the following nonlinear semidefinite program:

$$\text{(SDP)} \quad \min f(x), \quad \text{s.t.} x \in R^n, \qquad g(x) \preceq 0,$$

where $f : R^n \to R$, and $g : R^n \to S^m$ are continuously differentiable, S^m is the set of $m \times m$ real symmetric matrices, and for $A \in S^m$, the notation $A \preceq 0$ means that A is negative semidefinite.

Let $A, B \in S^m$. By $A \succeq 0$ we mean that A is positive semidefinite. We write $A \succeq (\preceq) B$ if and only if $A - B \succeq (\preceq) 0$. Let $A \succeq 0$. Denote by $A^{1/2}$ or $\sqrt{A}$ the unique (positive semidefinite) square root of A. For $A \in S^m$, define $|A| = (A^2)^{1/2}$. If A is nondegenerate, denote by A^{-1} or $1/A$ the inverse of A. Denote $A \prec 0$ ($A \succ 0$) iff A is negative (positive) definite.

Suppose that X and Y are two normed spaces. Let $h: X \to Y$ be a (Fréchet) differentiable operator. Let $x \in X$. We use $Dh(x)$ to denote the (Fréchet) derivative of h at x. Let $d \in X$. We use $Dh(x)(d)$ to denote the directional derivative of h at x in the direction d.

The matrix-valued function $g(x)$ is said to be convex on R^n iff for any $x_1, x_2 \in R^n$ and any $\theta \in [0,1]$, there holds

$$g(\theta x_1 + (1-\theta)x_2) \preceq \theta g(x_1) + (1-\theta) g(x_2).$$

If both $f(x)$ and $g(x)$ are convex on R^n, we say that (SDP) is a convex semidefinite program.

Denote by X_0 the feasible set of (SDP), i.e., $X_0 = \{x \in R^n : g(x) \succeq 0\}$. Throughout the paper, we assume that $X_0 \neq \varnothing$.

Consider the following *augmented Lagrangian*:

$$L(x, \Omega, r) = f(x) + \frac{1}{2r}\left[\left\|\frac{\Omega + rg(x) + |\Omega + rg(x)|}{2}\right\|^2 - \|\Omega\|^2\right], \qquad (1)$$

where $x \in R^n, \Omega \in S^m, r \geq 1$ and the norm $\|\cdot\|$ is the Frobenius norm of an $m \times m$ matrix, i.e., $\|A\| = \sqrt{\mathrm{tr}[A^T A]}$, for any $m \times m$ matrix A.

The *augmented Lagrangian dual function* is defined as

$$q(\Omega, r) = \inf_{x \in R^n} L(x, \Omega, r), \quad \Omega \in S^m, r \geq 1.$$

The problem of evaluating $q(\Omega, r)$ is called an *augmented Lagrangian problem.*

The augmented Lagrangian dual problem is defined as

$$(SDD) \quad \sup_{\Omega \in S^m, r \geq 1} q(\Omega, r).$$

Denote by M_{SDD} the optimal value of (SDD).

Next, we discuss some basic properties of the augmented Lagrangian $L(x,\Omega,r)$ such as monotonicity and convexity. Next lemma follows immediately from Proposition 4.3 of [7].

Lemma 2.1. Consider the matrix valued function $A: S^m \to S^m$ defined by $A(X) = X \mid X \mid$. Then for any $Y \in S^m$, there holds

$$DA(X)(Y) = U^T \left(B \circ (UYU^T)\right)U,$$

where

$$B = (b_{ij})_{m\times m},$$

$$b_{ij} = \begin{cases} \frac{\lambda_i|\lambda_i|-\lambda_j|\lambda_j|}{\lambda_i-\lambda_j}, \lambda_i \neq \lambda_j \\ 0, \ \lambda_i=\lambda_j=0 \\ 2\lambda_i, \lambda_i = \lambda_j > 0 \\ -2\lambda_j, \lambda_i = \lambda_j < 0, \end{cases}$$

$\circ$ is the Hadamard product of two matrices:

$$C = (c_{ij})_{m\times m}, D = (d_{ij})_{m\times m}, C \circ D = (c_{ij}d_{ij})_{m\times m},$$

$$X = U^T \mathrm{diag}\left[\lambda_1,\cdots,\lambda_m\right]U,$$

$$U^TU = I, \lambda_1 \geq \cdots \lambda_s \geq 0 > \lambda_{s+1} \geq \cdots \geq \lambda_m.$$

Lemma 2.2. Let $h(X) = \mathrm{tr}\left[(\mid X \mid + X)^2\right]: S^m \to R$. Then
(i) h is convex on S^m.
(ii) For any $Y \in S^m$, there holds

$$Dh(X)(Y) = 2\mathrm{tr}\left[(X + \mid X \mid)Y\right].$$

Proof. (i) The conclusion follows from Theorem 2.3.14 of [13].
(ii) By Lemma 2.1, we need only to show that

$$D\left(\mathrm{tr}[X \mid X \mid]\right)(Y) = \mathrm{tr}\left[U^T \left(B \circ (UYU^T)\right)U\right] = 2\mathrm{tr}\left[\mid X \mid (Y)\right], \qquad (2)$$

where X, B and U are as in Lemma 2.1. Indeed, we have

$$\mathrm{tr}\left[U^T\left(A\circ(UYU^T)\right)U\right]=\mathrm{tr}\left[\left(U^T\left(A\circ(UYU^T)\right)U\right)I\right].$$

By Lemma 5.1.4 of [12], we have

$$\begin{aligned}&\mathrm{tr}\left[\left(U^T\left(A\circ(UYU^T)\right)U\right)I\right]\\&=\mathrm{tr}\left[(A\circ I)(U(Y)^TU^T)\right]\\&=\mathrm{tr}\left[A_1(UYU^T)\right],\end{aligned}$$

where

$$A_1=2\mathrm{diag}\left[\lambda_1,\cdots,\lambda_s,-\lambda_{s+1},\cdots,-\lambda_m\right]=2U\mid X\mid U^T.$$

Thus, (2) follows. □

Lemma 2.3. Let h be defined as in Lemma 2.2. Then h is nondecreasing, i.e., for any A_1, $A_2\in S^m$ satisfying $A_2\succeq A_1$, there holds $h(A_2)\geq h(A_1)$.

Proof. By Lemma 2.2, we have

$$h(A_2)-h(A_1)\geq Dh(A_1)(A_2-A_1)=2\mathrm{trace}\left[(\mid A_1\mid+A_1)(A_2-A_1)\right]\geq 0$$

because $\mid A_1\mid+A_1\succeq 0$ and $A_2-A_1\succeq 0$. □

Now, we show that the augmented Lagrangian L and the augmented Lagrangian dual function q are concave in $(\Omega,r)\in S^m\times(0,+\infty)$. It is also nondecreasing in r.

Let $A\in S^m$. Define

$$F(x,A)=\begin{cases}f(x),if\mathrm{g}(\mathrm{x})\preceq A,\\+\infty,else,.\end{cases}$$

$$p(A)=\inf_{x\in R^n}F(x,A) \tag{3}$$

It is clear that $p(0)$ is the optimal value of (SDP). We need the following lemma.

Lemma 2.4. Let $x \in R^n$, $\Omega \in S^m$ and $r \geq 1$. Then, we have
(i)

$$L(x,\Omega,r) = \inf_{A\in S^m}\{F(x,A) + \mathrm{tr}[\Omega A] + r/2\|A\|^2\}; \tag{4}$$

(ii)

$$q(\Omega,r) = \inf_{A\in S^m}\{p(A) + \mathrm{tr}[\Omega A] + r/2\|A\|^2\}. \tag{5}$$

Proof. (i) Let $a = \inf_{A\in S^m}\{F(x,A) + \mathrm{tr}[\Omega A] + r/2\|A\|^2\}$. First we prove that $L(x,\Omega,r) \leq a$. Otherwise, there exist $\delta > 0$ and $A \in S^m$ such that

$$L(x,\Omega,r) \geq F(x,A) + \mathrm{tr}[\Omega A] + r/2\|A\|^2 + \delta. \tag{6}$$

By the definition of $F(x,A)$, we see that

$$g(x) \preceq A \tag{7}$$

and

$$f(x) = F(x,A).$$

Thus, from (6), we deduce

$$1/(8r)\mathrm{tr}\left[\left(|\Omega + rg(x)| + \Omega + rg(x)\right)^2\right] - 1/(2r)\mathrm{tr}[\Omega^2] \geq \mathrm{tr}[\Omega A] + r/2\|A\|^2 + \delta. \tag{8}$$

By Lemma 2.3, (7) and (8), we have

$$\begin{aligned}
&\mathrm{tr}[\Omega A] + r/2\|A\|^2 = 1/(8r)\mathrm{tr}\left[(2|\Omega + rA)^2\right] - 1/(2r)\mathrm{tr}[\Omega^2] \\
&\geq 1/(8r)\mathrm{tr}\left[\left(|\Omega + rA| + \Omega + rA\right)^2\right] - 1/(2r)\mathrm{tr}[\Omega^2] \\
&\geq \mathrm{tr}[\Omega A] + r/2\|A\|^2 + \delta,
\end{aligned}$$

which is impossible. So $L(x,\Omega,r) \le a$. Now we prove that $L(x,\Omega,r) \ge a$. Otherwise, there exists $\delta > 0$ such that

$$\begin{aligned} & f(x) + r/8\mathrm{tr}\Big[\big(|\,\Omega/r + g(x)\,|_+ \Omega/r + g(x)\big)^2\Big] - 1/(2r)\mathrm{tr}[\Omega^2] + \delta \\ & = L(x,\Omega,r) + \delta \\ & \le F\big(x,A\big) + \mathrm{tr}[\Omega A] + r/2\|A\|^2 + \delta. \end{aligned} \tag{9}$$

Assume that

$$g(x) + \Omega/r = U^T \mathrm{diag}\big[\lambda_1,\cdots,\lambda_s,\lambda_{s+1},\cdots,\lambda_m\big]U,$$

where $\lambda_1 \ge \cdots \ge \lambda_s > 0 \ge \lambda_{s+1} \ge \cdots \ge \lambda_m$ and $U^T U = I$. Let

$$B = U^T \mathrm{diag}\big[\lambda_1,\cdots,\lambda_s,0,\cdots,0\big]U.$$

Then $g(x) \preceq B - \Omega/r$. Let $A = B - \Omega/r$. Then, $g(x) \preceq A$. Thus,

$$F(x,A) + \mathrm{tr}[\Omega A] + r/2\|A\|^2 = f(x) - 1/(2r)\|\Omega\|^2 + r/2\|B\|^2, \tag{10}$$

$$\begin{aligned} & f(x) + r/8\mathrm{tr}\Big[\big(|\,\Omega/r + g(x)\,|_+ \Omega/r + g(x)\big)^2\Big] - 1/(2r)\mathrm{tr}[\Omega^2] + \delta \\ & = f(x) - 1/(2r)\|\Omega\|^2 + r/2\|B\|^2. \end{aligned} \tag{11}$$

The combination of (9)-(11) leads to a contradiction.
(ii) By the definition of $q(\Omega,r)$ and (4), we have

$$\begin{aligned} q(\Omega,r) &= \inf_{x\in R^n} L(x,\Omega,r) \\ &= \inf_{x\in R^n}\ \inf_{A\in S^m}\{F(x,A) + \mathrm{tr}[\Omega A] + r/2\|A\|^2\}. \end{aligned} \tag{12}$$

Moreover, it is elementary to show that

$$\begin{aligned} & \inf_{x\in R^n}\ \inf_{A\in S^m}\{F(x,A) + \mathrm{tr}[\Omega A] + r/2\|A\|^2\} \\ & = \inf_{A\in S^m}\ \inf_{x\in R^n}\{F(x,A) + \mathrm{tr}[\Omega A] + r/2\|A\|^2\} \\ & = \inf_{A\in S^m}\{p(A) + \mathrm{tr}[\Omega A] + r/2\|A\|^2\}. \end{aligned} \tag{13}$$

The combination of (12) and (13) gives (5). □

The following proposition is a direct consequence of Lemmas 2.1-2.4.

Proposition 2.1. (i) The augmented Lagrangian $L(x,\Omega,r)$ and the augmented Lagrangian dual function $q(\Omega,r)$ are both concave in $(\Omega,r) \in S^m \times [1,+\infty)$;
(ii) The augmented Lagrangian $L(x,\Omega,r)$ is nondecreasing in r on $[1,+\infty)$.
(iii) The augmented Lagrangian L is continuously differentiable in x;
(iv) If f and g are convex on R^n, then the augmented Lagrangian $L(x,\Omega,r)$ is also convex in x on R^n.

3. DUALITY

In this section, we establish duality results based on augmented Lagrangian. Throughout this section, $p(A)$ is defined by (3). The following proposition, whose proof is elementary, establishes a weak duality property between (SDP) and (SDD).

Proposition 3.1 (weak duality). The following relation holds

$$M_{SDD} \leq p(0). \tag{14}$$

The next theorem presents conditions for the strong duality property between (SDP) and (SDD).

Theorem 3.1 (strong duality). Assume that there exist $\bar{\Omega} \in S^m$ and $\bar{r} \geq 1$ such that

$$q(\bar{\Omega},\bar{r}) \geq m_0 \tag{15}$$

for some $m_0 \in R$.
(i) If

$$\liminf_{A \to 0} p(A) = p(0) \tag{16}$$

holds, then, there holds

$$p(0) = M_{SDD}. \tag{17}$$

(ii) If (17) holds, then (16) holds.

Proof. (i) First, we prove that if (16) holds, then (17) holds. Suppose (by Proposition 3.1) to the contrary that there exists $\delta > 0$ such that

$$q(\Omega, r) \le p(0) - \delta, \quad \forall \Omega \in S^m, r > 0.$$

Let $1 \le r_k \to +\infty$. Then

$$q(\bar{\Omega}, r_k) \le p(0) - \delta, \quad \forall k.$$

Consequently, by (5), there exists $A_k \in S^m$ such that

$$p(A_k) + \mathrm{tr}[\bar{\Omega} A_k] + r/2 \|A_k\|^2 \le p(0) - \delta/2, \quad \forall k. \tag{18}$$

It follows from (15) and (18) that

$$\|A_k\|^2 \le \frac{2(p(0) - \delta/2 - m_0)}{r_k - \bar{r}} \quad Ak.$$

Thus,

$$\lim_{k \to +\infty} \|A_k\| = 0.$$

Taking the lower limit in (18), we obtain

$$\liminf_{A \to 0} p(A) \le \liminf_{A_k \to 0} p(A_k) \le p(0) - \delta/2,$$

contradicting (16). Therefore, (17) holds.

(ii) From $M_{SDD} = p(0)$, we see that, for any $\delta > 0$, there exist $\Omega^* \in S^m, r^* \ge 1$ such that

$$q(\Omega^*, r^*) \ge p(0) - \delta.$$

Thus, it follows from (5) that we have

$$p(A)+\mathrm{tr}[\Omega^* A]+r^*/2\|A\|^2 \geq p(0)-\delta.$$

Passing to the lower limit as $A \to 0$, we have

$$\liminf_{A\to 0} p(A) \geq p(0)-\delta.$$

By the arbitrariness of $\delta > 0$, we see that

$$\liminf_{A\to 0} p(A) = p(0).$$

(16) follows. □

Remark 3.1. Previously, *classical Lagrangian* is used to deal with (SDP):

$$L'(x,\Omega) = f(x)+\mathrm{tr}[\Omega g(x)], \quad \Omega \succeq 0, x \in R^n.$$

Accordingly, *classical Lagrangian dual problem* is formulated as follows:

$$(SDD') \quad \sup_{\Omega\succeq 0}\inf_{x\in R^n} L'(x,\Omega).$$

Denote by $M_{SDD'}$ the optimal value of (SDD'). It can be shown that If the zero duality gap property between (SDP) and its classical Lagrangian dual problem (SDD'), $p(0) = M_{SDD'}$ holds, then the zero duality gap property holds by means of augmented Lagrangian, i.e., $p(0) = M_{SDD}$. In particular, if (SDP) is convex and the Slater constraint qualification holds: there exists $x_0 \in R^n$ such that $g(x_0) \prec 0$, then $p(0) = M_{SDD'}$ (see [21]). This further implies that $p(0) = M_{SDD}$.

4. EXACT PENALIZATION

The next result is concerned with an exact penalty representation property via augmented Lagrangian.

Theorem 4.1. Let $\overline{\Omega}' \in S^m$. Assume that there exist $\overline{\Omega} \in S^m$ and $\overline{r} \geq 1$ such that (15) holds.

(i) Suppose that there exist a $\bar{r}' \geq 1$ and a neighbourhood W of $0 \in S^m$ such that

$$p(A) + \mathrm{tr}[\bar{\Omega}'A] + \bar{r}'\|A\|^2 \geq p(0), \quad \forall A \in W. \tag{19}$$

Then there exists $r^* \geq 1$ such that

$$p(0) = q(\bar{\Omega}', r), \quad \forall r \geq r^*. \tag{20}$$

(ii) If (20) holds, then there exist a $\bar{r}' \geq 1$ and a neighbourhood W of $0 \in S^m$ such that (19) holds.

Proof. (i) First, we show that (19) is equivalent to that there exists $\bar{r}'' \geq 1$ such that

$$p(A) + \mathrm{tr}[\bar{\Omega}'A] + \bar{r}''\|A\|^2 \geq p(0), \quad \forall A \in S^m. \tag{21}$$

It is obvious that if (21) holds, then (19) holds (by taking $\bar{r}' = \bar{r}''$). Now we show that (19) implies (21). Without loss of generality, suppose to the contrary that there exists $\{A_k\} \subset S^m$ and $1 \leq r_k \to +\infty$ such that

$$\|A_k\| \geq a > 0, \quad \forall k \tag{22}$$

for some constant $a > 0$ and

$$p(A_k) + \mathrm{tr}[\bar{\Omega}'A_k] + \bar{r}''\|A_k\|^2 \geq p(0), \quad \forall A \in S^m. \tag{23}$$

On the other hand, by (15) and statement (ii) of Lemma 2.4, we have

$$p(A_k) + \mathrm{tr}[\bar{\Omega}A_k] + \bar{r}\|A_k\|^2 \geq m_0. \tag{24}$$

(23), together with (24), gives us

$$p(0) > m_0 + \mathrm{tr}[(\bar{\Omega}' - \bar{\Omega}))A_k] + (r_k - \bar{r})\|A_k\|^2.$$

As a result,

$$\|A_k\| \le \frac{p(0) - m_0 - \mathrm{tr}[(\bar{\Omega}' - \bar{\Omega}))A_k / \|A_k\|]}{r_k - \bar{r}},$$

which contradicts (22) as $k \to +\infty$. So (21) holds. By statement (ii) of Lemma 2.4, there exists $r^* \ge 1$ such that (20) holds.

(ii) If (20) holds, then, by statement (ii) of Lemma 2.4, we have (21), hence (19). The proof is complete. □

5. CONVERGENCE ANALYSIS

In this section, we consider convergence of first-order stationary points of the augmented Lagrangian problems when the penalty parameter tends to $+\infty$.

Denote by $(P(\Omega, r))$ the augmented Lagrangian problem:

$$\inf_{x \in R^n} L(x, \Omega, r).$$

Definition 5.1 [21]. Let $x_0 \in R^n$ be feasible to (SDP). We say that the *Mangasarian-Fromovitz constraint qualification* (MFCQ in short) holds at x_0 iff there exists $d \in R^n$ such that $g(x_0) + Dg(x_0)(d) \prec 0$.

Definition 5.2. Let $\bar{x}$ be feasible to (SDP). We say that $\bar{x}$ satisfies the *KKT optimality condition* of (SDP) iff there exists $\Lambda \in S^m$ with $\Lambda \succeq 0$ such that

$$\frac{\partial f(\bar{x})}{\partial x_i} + \mathrm{tr}\left[\Lambda \frac{\partial g(\bar{x})}{\partial x_i}\right] = 0 \tag{25}$$

and

$$\Lambda g(\bar{x}) = 0. \tag{26}$$

It was established in [21] that if $\bar{x}$ is a local solution of (SDP) and the Mangasarian-Fromovitz constraint qualification holds at $\bar{x}$, then $\bar{x}$ satisfies the KKT optimality condition of (SDP).

The next result follows immediately from Lemma 2.2.

Theorem 5.1. Suppose that $\bar{x}(\Omega, r)$ is a local minimum of $(P(\Omega, r))$. Then

$$\frac{\partial f(\bar{x}(\Omega,r))}{x_i} + 1/2\mathrm{tr}\left[\left(\Omega + rg(\bar{x}(\Omega,r))+ \mid \Omega + rg(\bar{x}(\Omega,r))\mid\right)\frac{\partial g(\bar{x}(\Omega,r))}{\partial x_i}\right]$$

$$= 0, \quad i = 1,\cdots,n. \tag{27}$$

The next lemma is useful for convergence analysis.

Lemma 5.1. Let $\{\Omega_k\} \subset S^m$ be bounded and $0 < r_k \to +\infty$. Let $\bar{x}_k \in R^n, \forall k$. Suppose that there exists $M \in R$ such that

$$L(\bar{x}_k, \Omega_k, r_k) \le M. \tag{28}$$

Then any limit point of $\{\bar{x}_k\}$ is feasible to (SDP).

Proof. Suppose that $\bar{x}$ is a limit point of $\{\bar{x}_k\}$ and assume without loss of generality that $\bar{x}_k \to \bar{x}$. Then from (28), we deduce that

$$\left\|\Omega_k/r_k + g(x_k)+ \mid \Omega_k/r_k + g(x_k)\mid\right\|^2 \le \frac{8}{r_k}\left[f(x_0) - f(x_k) + \frac{1}{2r_k}\mathrm{tr}[\Omega_k^2]\right].$$

We have following convergence results for augmented Lagrangian.

Theorem 5.2. Let $\{\Omega_k\} \subset S^m$ be bounded, $1 \le r_k \to +\infty$. Consider the problems (SDP) and $(P(\Omega_k, r_k))$. Let $\bar{x}_k$ be a local minimum of $(P(\Omega_k, r_k))$. Suppose that there exists a constant M such that

$$L(\bar{x}_k, \Omega_k, r_k) \le M, \quad \forall k$$

holds. Then each limit point of $\{\bar{x}_k\}$ is feasible for (SDP). Furthermore, if $\bar{x}$ is a limit point of $\{\bar{x}_k\}$ and MFCQ holds at $\bar{x}$, then $\bar{x}$ satisfies the KKT optimality condition of (SDP).

Proof. By Lemma 5.1, each limit point of $\{\bar{x}_k\}$ is feasible for (SDP). Assume without loss of generality that $\bar{x}_k \to \bar{x}$ as $k \to +\infty$. Let

$$\Lambda_k = \left(\Omega_k + r_k g(\bar{x}_k)+ \mid \Omega_k + r_k g(\bar{x}_k)\mid\right) \succeq 0. \tag{29}$$

Then (27) (with r and $\bar{x}(\Omega,r)$ replaced by r_k and $\bar{x}_k$, respectively) becomes

$$\frac{\partial f(\bar{x}_k)}{\partial x_i} + \mathrm{tr}\left[\Lambda_k \frac{\partial g(\bar{x}_k)}{\partial x_i}\right] = 0, \quad i = 1, \cdots, n. \tag{30}$$

We assert that $\{\Lambda_k\}$ is bounded. Otherwise, assume without loss of generality that $\|\Lambda_k\| \to +\infty$ and

$$\lim_{k \to +\infty} \Lambda_k / \|\Lambda_k\| = \Lambda' \succeq 0.$$

Dividing (30) by $\|\Lambda_k\|$ and passing to the limit as $k \to +\infty$, we get

$$\mathrm{tr}\left[\Lambda' \frac{\partial g(\bar{x})}{\partial x_i}\right] = 0, \quad i = 1, \cdots, n. \tag{31}$$

Note that

$$\begin{aligned}
\mathrm{tr}[\Lambda' g(\bar{x})] &= \lim_{k \to +\infty} \mathrm{tr}\left[\frac{\Lambda_k}{\|\Lambda_k\|} g(\bar{x}_k)\right] \\
&= \tfrac{1}{2} \lim_{k \to +\infty} \mathrm{tr}\left[\frac{\Omega_k + r_k g(\bar{x}_k) + |\Omega_k + r_k g(\bar{x}_k)|}{\|\Lambda_k\|} g(\bar{x}_k)\right] \\
&\geq \tfrac{1}{2} \liminf_{k \to +\infty} \mathrm{tr}\left[\frac{r_k g(\bar{x}_k) + |r_k g(\bar{x}_k)|}{\|\Lambda_k\|} g(\bar{x}_k)\right] \\
&= \tfrac{1}{2} \liminf_{k \to +\infty} \mathrm{tr}\left[r_k \frac{g(\bar{x}_k) + |g(\bar{x}_k)|}{\|\Lambda_k\|} g(\bar{x}_k)\right]
\end{aligned}$$

because $\{\Omega_k\}$ is bounded and $\|\Lambda_k\| \to +\infty$. Furthermore,

$$\begin{aligned}
&\tfrac{1}{2} \liminf_{k \to +\infty} \mathrm{tr}\left[r_k \frac{g(\bar{x}_k) + |g(\bar{x}_k)|}{\|\Lambda_k\|} g(\bar{x}_k)\right] \\
&= \tfrac{1}{2} \liminf_{k \to +\infty} \mathrm{tr}\left[r_k \frac{g(\bar{x}_k) + |g(\bar{x}_k)|}{\|\Lambda_k\|} |g(\bar{x}_k)|\right] \\
&\geq 0
\end{aligned}$$

because $|g(\bar{x}_k)| \succeq 0$ and $g(\bar{x}_k)+|g(\bar{x}_k)| \succeq 0$. On the other hand, from $\Lambda' \succeq 0$ and $g(\bar{x}) \preceq 0$, we deduce that

$$\mathrm{tr}[\Lambda' g(\bar{x})] \leq 0.$$

Hence, we have

$$\mathrm{tr}[\Lambda' g(\bar{x})] = 0. \tag{32}$$

By MFCQ at $\bar{x}$, there exists $d \in R^n$ such that $g(\bar{x}) + Dg(\bar{x})(d) \prec 0$. It is obvious that $\Lambda' \neq 0$. It follows that

$$\mathrm{tr}\left[\Lambda'(g(\bar{x}) + Dg(\bar{x})(d)\right] < 0.$$

This combined with (32) yield

$$\mathrm{tr}\left[\Lambda' Dg(\bar{x})(d)\right] < 0,$$

contradicting (31). So we assume without loss of generality that $\Lambda_k \to \Lambda \succeq 0$. Taking the limit in (30) as $k \to +\infty$, we obtain (25). Moreover, if $\Lambda = 0$, then (26) holds automatically. Now assume that $\Lambda \neq 0$. We have

$$\mathrm{tr}[\Lambda g(\bar{x})] = \lim_{k\to+\infty} \mathrm{tr}\left[\Lambda_k g(\bar{x}_k)\right]$$

$$= \tfrac{1}{2} \lim_{k\to+\infty} \|\Lambda_k\| \mathrm{tr}\left[\frac{\Omega_k + r_k g(\bar{x}_k)+|\Omega_k + r_k g(\bar{x}_k)|}{\|\Omega_k + r_k g(\bar{x}_k)+|\Omega_k + r_k g(\bar{x}_k)|\|} g(\bar{x}_k)\right]$$

$$= \tfrac{1}{2} \lim_{k\to+\infty} \|\Lambda_k\| \mathrm{tr}\left[\frac{\Omega_k/r_k + g(\bar{x}_k)+|\Omega_k/r_k + g(\bar{x}_k)|}{\|\Omega_k/r_k + g(\bar{x}_k)+|\Omega_k/r_k + g(\bar{x}_k)|\|} (\Omega_k/r_k + g(\bar{x}_k))\right] \tag{33}$$

because $\{\Omega_k\}$ is bounded. Let $A_k = \Omega_k/r_k + g(\bar{x}_k), \quad \forall k$. Assume that

$$g(\bar{x}) = U^T \mathrm{diag}\left[\lambda_1, \cdots, \lambda_m\right] U,$$

where $\lambda_m \leq \cdots, \leq \lambda_{s+1} < 0 = \lambda_s = \cdots, \lambda_1$ and $U^T U = I$, and

$$A_k = U_k^T \operatorname{diag}\left[\lambda_{1,k}, \cdots, \lambda_{m,k}\right] U_k,$$

where $\lambda_{m,k} \leq \cdots \leq \lambda_{1,k}$ and $U_k^T U_k = I$. Since $A_k \to g(\bar{x})$, we see that $\lambda_{i,k} \to \lambda_i$, $i = 1, \cdots, m$. As a result,

$$\lim_{k \to +\infty} \left| \frac{\lambda_{i,k} + |\lambda_{i,k}|}{\| A_k + |A_k| \|} \lambda_{i,k} \right| \leq \lim_{k \to +\infty} |\lambda_{i,k}| = 0, \quad i = 1, \cdots, s \tag{34}$$

and

$$\lim_{k \to +\infty} \frac{\lambda_{i,k} + |\lambda_{i,k}|}{\| A_k + |A_k| \|} \lambda_{i,k} = \lim_{k \to +\infty} 0 \times \lambda_{i,k} = 0, \quad i = s+1, \cdots, m \tag{35}$$

The combination of (33)-(35) yields (26). This completes the proof. □

REFERENCES

[1] Bel-Tal, A., Jarre, F., Kocvara, M., Nemirovski and Zowe, J., "Optimal design of trusses under a nonconvex global buckling constraints". Optimization and Engineering, Vol. 1, 2000, pp. 189-213.

[2] Benson, H. Y. and Vanderbei, R. J., "Solving problems with semidefinite and related constraints using interior-point methods for nonlinear programming". Mathematical Programming, Ser. B, Vol. 93, 2002.

[3] Bertsekas, D. P., "Constrained Optimization and Lagrangian Multiplier Methods". Academic Press, New York, 1982.

[4] Bonnans, J. F., Cominetti, R. and Shapiro, A., "Second order optimality conditions based on second order tangent sets". SIAM Jou. Optimization, Vol. 9, 1999, pp. 466-492.

[5] Burer, S., Monteiro, R. D. C. and Zhang, Y., "Solving a class of semidefinite programs via nonlinear programming". Mathematical Programming, Ser. A., Vol. 93, 2002, pp. 97-122.

[6] Burer, S., Monteiro, R. D. C. and Zhang, Y., "Interior-point algorithms for semidefinite programming based on a nonlinear formulation". Computational Optimization and Applications, Vol. 22, 2002, pp. 49-79.

[7] Chen, X., Qi, H. D. and Tseng, P., "Analysis of nonsmooth symmetric-matrix-valued functions with applications to semidefinite complementarity constraints". SIAM J. Optimization. To appear.

[8] Fan, K., "On a theorem of Wely concerning eigenvalues of linear transformations". I., Proc. Nat. Acad. Sci. U. S. A., Vol. 35, 1949, pp. 652-655.

[9] Forsgren, A., "Optimality conditions for nonconvex semidefinite programming". Mathematical Programming, Ser. A., Vol. 88, 2000, pp. 105-128.

[10] Fares, B., Noll, D. and Apkarian, P., "Robust control via sequential semidefinite programming". SIAM J. Control and Optimi., Vol. 40, 2002, pp. 1791-1820.

[11] Ghaoui, L. E. and Niculescu, S. I., "Advances in Linear Matrix Inequality Methods in Control". Advances in Design Control, SIAM, Philadelphia, 2000.

[12] Horn, R. A. and Johnson, C. R., "Topics in Matrix Analysis". Cambridge University Press, Cambridge, 1991.

[13] Jarre, F., "Convex analysis on symmetric matrices". In "Handbook of Semidefinite Programming, Theory, Algorithms and Applications", H. Wolkowicz, R. Saigal and Vandenberghe (eds), Kluwer Academic Publishers, 2000.

[14] Jarre, F., "An interior point method for semidefinite programs". Optimization and Engineering, Vol. 1, 2000, pp. 347-372.

[15] Kanzow, C. and Nagel, C., "Semidefinite programs: new search directions, smoothing-type methods, and numerical results". SIAM Jou. Optimization, Vol. 13, 2002, pp. 1-23.

[16] Mosheyev, L. and Zibulevsky, M., "Penalty/barrier multiplier algorithm for semidefinite programming". Optimization Methods and Software, Vol. 13, 2000, pp. 235-261.

[17] Overton, M. L. and Womersley, R. S., "Second derivatives for optimizing eigenvalues of symmetric matrices". SIAM J. Matrix Analysis and Applications, Vol. 16, 1995, pp. 697-718.

[18] Ringertz, U. T., "Eigenvalues in optimal structural design". In: Biegler, L. T., Coleman, T. F. Conn, A. R. and Santosa, F. N. (eds), "Large Scale Optimization and Applications, Part I: Optimization in Inverse Problems and Design", Vol. 92 of the IMA Volumes in Mathematics and its Applications, pp. 135-149, Springer, New York, 1997. of the IMA Volumes in Mathematics and its Applications, Springer, New York, 1997, pp. 135-149.

[19] Rockafellar, R. T., "Augmented Lagrange multiplier functions and duality in nonconvex programming", SIAM Jou. on Control and Optimization, Vol. 12, 1974, pp. 268-285.

[20] Rockafellar, R. T., "Lagrange multipliers and optimality". SIAM Review, Vol. 35, 1993, pp. 183-238.

[21] Shapiro, A., "First and second order analysis of nonlinear semidefinite programs". Mathematical Programming, Ser. B., Vol. 77, 1997, pp. 301-320. National University of Singapore, Singapore, 2002.

[22] Todd, M., "Semidefinite Optimization". Acta Numerica, Vol. 10, 2001, pp. 515-560.

[23] Vandenberghe, L. and Boyd, S., "Semidefinite programming". SIAM Review, Vol. 38, 1996, pp. 49-95.

[24] Wolkowicz, H., Saigal, R. and Vandenberghe, L. (eds), "Handbook of semidefinite programming, theory, algorithms and applications". International Series in Operations Research and Management Science, Vol. 27, Kluwer Academic Publishers, Boston,MA, 2000.

[25] Ye, Y., "Interior Point Algorithms: Theory and Analysis". John Wiley & Son, New York, 1997.

OPTIMALITY ALTERNATIVE: A NON-VARIATIONAL APPROACH TO NECESSARY CONDITIONS

A.D. Ioffe[1]
Dept. of Mathematics, Technion, Haifa, Israel

1. INTRODUCTION. THE OPTIMALITY ALTERNATIVE

In this paper we shall discuss an approach to necessary optimality conditions that can be qualified as "non-variational". The essence of the approach can be summarized in the following theorem.

Consider an abstract minimization problem

$$\text{minimize} \quad f(x), \text{s.t.} \; x \in M \subset X,$$

where X is the domain space, M is the constraint set and f is the cost function.

Theorem 1 (Optimality alternative). *Suppose that* (X,d) *is a metric space,* $\overline{x} \in M$ *and* f *satisfies the Lipschitz condition near* $\overline{x}$. *Let further* $\varphi(x)$ *be a nonnegative extended-real-valued function equal to zero at* $\overline{x}$. *Suppose finally that* $\overline{x}$ *is a local solution of the problem. Then the following alternative holds:*

[1] The research was supported by the USA-Israel Binational Science Foundation under the grant 2000157

– either there is a $\lambda > 0$ such that the function $\lambda f + \varphi$ attains an unconditional minimum at $\overline{x}$;

– or there is a sequence (x_n) converging to $\overline{x}$ such that $\varphi(x_n) < n^{-1} d(x_n, M)$. In particular, if X is complete, M is closed and φ is lower semicontinuous, then there is a sequence (z_n) converging to $\overline{x}$ such that $z_n \notin M$ and each of the function $\varphi(x) + n^{-1} d(x, z_n)$ attains an absolute minimum at z_n.

We shall speak about the regular case if the first possibility takes place and about singular case otherwise.

Proof. Indeed, either there are a neighborhood of $\overline{x}$ and $R > 0$ such that $R\varphi(x) \geq d(x, M)$ for all x of the neighborhood, or there is a sequence (x_n) converging to $\overline{x}$ with $2n\varphi(x_n) < d(, M)$. In the first case, as f is Lipschitz (e.g. with constant L), we can choose for any x close to $\overline{x}$ a $u \in M$ such that, say $d(x,u) \leq 2d(x, M)$. Then

$$f(x) \geq f(u) - Ld(x,u) \geq f(\overline{x}) - 2LR\varphi(x).$$

In the second case, if X is complete, M is closed and φ l.s.c., we apply Ekeland's variational principle to the function $\varphi(x)$. As it is nonnegative, we have $\varphi(x_n) \leq \inf \varphi + (2n)^{-1} d(x_n, M)$, so Ekeland's principle guarantees the existence of (z_n) such that $d(x_n, z_n) \leq d(x_n.M)/2$ and $\varphi(x) + n^{-1} d(x, z_n)$ attains an absolute minimum at z_n. □

The theorem reduces the problem to one or a sequence of unconstrained problems. This means that a necessary condition to the original problem can be obtained by way of analysis of necessary conditions in unconstrained problems.

We observe further that in case when $\varphi(x) = 0$ on M, the regular case conclusion that $\overline{x}$ is an unconditional local minimum of $f + K\varphi$ is also sufficient for $\overline{x}$ to be a solution. Therefore the optimality alternative allows in principle to get all kind of necessary conditions, not just the first order conditions.

To get a better look at the alternative and its relationship to the method of variations, let us consider a very simple example of an equality constrained problem in $\mathbb{R}^n$:

$$\text{minimize} \quad f(x), \ \text{s.t.} \ F(x) = 0,$$

where F is defined and continuous in a neighborhood of $\overline{x}$ and takes values in $\mathbb{R}^m$.

Suppose $\bar{x}$ is a local solution. We first apply the optimality alternative (in which case f should be assumed Lipschitz continuous near $\bar{x}$). Take $\varphi(x) = \|F(x)\|$. Then by the theorem

– either there is a $\lambda > 0$ such that $\lambda f(x) + \|F(x)\|$ attains an unconditional local minimum at $\bar{x}$,

– or there is a sequence (x_n) converging to such that $F(x_n) \neq 0$ and $\|F(x)\| + n^{-1}\|x - x_n\|$ has an absolute minimum at $\bar{x}$.

Assuming that f is Gâteaux differentiable at x and F is Gâteaux differentiable in a neighborhood of $\bar{x}$, we conclude that in the regular case there is a $y \in \mathbb{R}^m$ with $\|y\| \leq K$ such that $f'(\bar{x}) + y \circ F'(\bar{x}) = 0$, and in the singular case for any n there is a y_n with $\|y_n\| = 1$ (since $F(x_n) \neq 0$) such that $\|y_n \circ F'(x_n)\| \leq n^{-1}$. If finally we assume that F' is continuous at $\bar{x}$ (hence F is Fréchet differentiable at $\bar{x}$), then for any limit point y of (y_n) we would get $\|y\| = 1$ and $y \circ F'(\bar{x}) = 0$.

Summarizing, we arrive at the standard Lagrange multiplier rule: there is a nontrivial pair (λ, y) with $\lambda \geq 0$ (actually $\lambda \in \{0,1\}$) such that $\lambda f'(\bar{x}) + y \circ F'(\bar{x}) = 0$.

Applying the standard variational argument (that is to say, based on variations of the cost function and the constraint map along curves $x(t) = \bar{x} + th + o(t)$) to the same problem, we shall arrive at the same result with the help of either implicit function theorem or Brouwer fixed point theorem, in the latter case under a weaker assumption on F, just that F is Fréchet differentiable at $\bar{x}$. Here also we can distinguish between the regular and singular cases, the first corresponding to Im $F'(\bar{x}) = \mathbb{R}^m$ in which case $\lambda > 0$.

If we compare both results and proofs, we have to conclude that in the two lines of arguments we

(a) consider different types of minimum;

(b) impose different assumptions on the components of the problem;

(c) use different analytic apparatus in proofs.

Indeed, regarding (a) we observe that in the proof based on the optimality alternative we deal with the real local minimum. It is clear from the analysis of the regular case in the proof of the optimality alternative that the reduction to unconstrained minimization cannot be implemented if $\bar{x}$ is not a point of a local minimum in the problem.

On the contrary, in the proof based on the method of variations, we actually deal with weaker concepts, namely local minimum on each variation curve $x(t) = \bar{x} + th + o(t)$ if we use the implicit function theorem or the property that the image of a neighborhood of $\bar{x}$ under mapping (f, F) does not cover a neighborhood of $(f(\bar{x}), 0)$ if the Brouwer theorem is used.

In connection with (b) we notice that the method of variations may work under weaker assumptions (which is not surprising in view of (a)) both on

f (no need to require f to be Lipschitz near $\overline{x}$) and F (e.g. in the case of the proof based on the Brouwer theorem).

Finally, as far as (c) is concerned, we notice that the proof based on optimality alternative does not use any powerful tools of analysis (such as the implicit function theorem or the Brouwer fixed point theorem or, more generally, any kind of an open mapping theorem) which are central in any application of the method of variations.

This is probably the main advantage of the optimality alternative. Indeed, in many cases, for instance in optimal control, there is no need in any kind of existence theorems for solutions of equations or inclusions involved in the statement of the problem when necessary optimality conditions for the problem are obtained by means of the optimality alternative.

Of course, the optimality alternative subsumes many essential elements of the techniques developed in nonsmooth optimization starting with Clarke's studies of mid-70s and related to subdifferential-oriented theory of necessary conditions in nonsmooth optimization problems (as say decoupling and reduction to unconstrained minimization). One can also trace a (more distant) relationship with the penalty function method which was also occasionally used to prove necessary conditions, although the question about existence of solutions in approximating unconstrained minimizations, the most painful for applications of penalty function methods, does not even appear when the optimality alternative is concerned. And typically, all so far available proofs do relay on certain powerful theorems of analysis.

The above observations suggest the possibility that the method of variations and the optimality alternative may produce different necessary conditions in certain situations. A stronger necessary condition can be expected of the optimality alternative, which may be not necessary for the type of a minimum analyzed by means of variational methods, and which therefore cannot be obtained by the latter. On the other hand, it can be expected that variational methods can be applied under weaker assumptions on the components of the problem.

It seems that this is precisely what we observe in certain branches of optimal control theory, namely optimal control of systems governed by differential inclusions. There are two parallel theories (e.g. [1,4,12] on the one hand and [11] on the other) in which similar problems are considered under different sets of assumptions and different necessary optimality conditions are obtained which do not coincide even in the intersection settings.

Before we pass to more detailed discussions concerning demonstration of the work of the optimality alternative, we need to make an excursion into nonsmooth analysis in the next section since meaningful applications of the optimality alternative always require nonsmooth function φ, so that non-

differentiable functionals appear in accompanying unconstrained minimization problems. The information collected in the next section is sufficient mainly to understanding the statements of results and some simple steps in proofs. This is more or less sufficient to present complete proofs in abstract setting but many technical details concerning specific functionals, that appear e.g. in connection with optimal control problems, require more elaborate techniques which goes beyond the scope and possible size of this paper (which is basically a transcript of the talk at the conference) and will be published elsewhere. As more specific applications, we consider two problems, one is rather abstract and the other more specific: optimal control of a system governed by a parabolic differential inclusion.

2. DIGRESSION INTO NONSMOOTH ANALYSIS

Let X be a Banach space. All definitions and results below are formulated under the additional assumption that X is weakly compactly generated (that is to say, it contains a weakly compact set whose linear hull is dense in X). This does not mean that certain definitions or propositions do not have counterparts valid for arbitrary Banach spaces but, on the one hand, some formulations are much simpler in the WCG setting and, on the other hand, the class of WCG spaces is sufficient for the vast majority of applications of nonsmooth analysis, in particular those considered later in this paper. (Perhaps this is appropriate to mention that all reflexive spaces and all separable spaces are weakly compactly generated.)

So let X be a WCG space, and let f be a function on X taking values in $(-\infty,\infty]$. If f is finite at x, then the function (of h)

$$f^-(x;h)=\liminf_{\substack{t\to+0\\ u\to h}}\frac{f(x+tu)-f(x)}{t}$$

is called the *Dini-Hadamard directional derivative* of f at x. The *Dini-Hadamard subdifferential* of f at x is defined by

$$\partial^- f(x)=\{x^*\in X^*:\langle x^*,h\rangle\le f^-(x;h),\ \forall h\in X\}$$

(where X^* as usual stands for the dual space).

Suppose now that f satisfies the Lipschitz condition near x. Then the approximate G-*subdifferential* (to be called in what follows simply *approximate subdifferential* as no other types of approximate subdifferential is considered) is the upper limit of the Dini-Hadamard subdifferentials:

$$\partial f(x) = \limsup_{u \to x} \partial^- f(u).$$

The upper limit is considered with respect to the norm topology in X and the weak*-topology in X^*. As far as the latter is concerned, the limit can be understood either topologically or sequentially: recall that in the dual to a WCG space, every bounded sequence contains a weak*-converging subsequence Each approach has its advantages but the theories, including proofs and formulas are very similar. In case when the sequential limit is considered, the subdifferential is often called *limiting*.

To define approximate subdifferential for non-Lipschitz functions, we need the notion of approximate normal. Let $S \subset X$ and $x \in S$. The set

$$N(S,x) = \bigcup_{\lambda>0} \lambda \partial d(\cdot, S)(x)$$

is called the *approximate normal* cone to S at x. Observe that this cone is not necessarily closed even if the subdifferential is defined through the topological weak*-limit. Now, given a function f which is finite at x, we define its approximate subdifferential at x as follows:

$$\partial f(x) = \{x^* : (x^*, -1) \in N(\operatorname{epi} f, (x, f(x))\}.$$

In case of a Lipschitz function, this definition is equivalent to the original definition. For Lipschitz functions the generalized gradient of Clarke coincides with the convex closure (convex hull in the finite dimensional case) of the approximate subdifferential:

$$\partial_c f(x) = \operatorname{cl}(\operatorname{conv} \partial f(x)).$$

Suppose now that F is a set-valued mapping from X into Y. The empty set is allowed to be a possible value for F, so we can always assume that F is defined on the whole of X. Let $y \in F(x)$. Then the set-valued mapping from Y^* into X^* defined by

$$D^*F(x,y)(y^*) = \{x^* \in X^* : (x^*, -y^*) \in N(\operatorname{Gr} F, (x,y))\}$$

is called the *(approximate) coderivative* of F at (x,y). If F is single-valued, we usually write $D^*F(x)(y^*)$. If F is single-valued and Lipschitz near x, then

$$D^*F(x)(y^*) = \partial(y^* \circ F)(x).$$

Subdifferential, normal cone and coderivative are the three main classes of objects studied in local nonsmooth analysis. Below we list some properties of theirs which will be used in the subsequent discussions. As only approximate subdifferentials, normal cones and coderivatives will appear in the sequel, we do not speak here about objects of other types.

The sum rule: Let $f_1,...,f_k$ be lower semicontinuous functions and all of them, save at most one, satisfy the Lipschitz condition in a neighborhood of $\overline{x}$. Then

$$\partial(f_1+...+f_k)(\overline{x}) \subset \partial f_1(\overline{x})+...+\partial f_k(\overline{x}).$$

Chain rule: (a) Let G be a continuous mapping from a neighborhood of $\overline{x} \in X$ into Y, and let f be a function on Y defined and Lipschitz in a neighborhood of $\overline{y} = G(\overline{x})$. Then

$$\partial(f \circ G)(\overline{x}) \subset \bigcup_{y^* \in \partial f(\overline{y})} D^*G(\overline{x})(y^*).$$

(b) Consider the *Nemytzki operator* $G: x(\cdot) \to g(t, x(t))$ e.g. from $\mathcal{L}^p$ into $\mathcal{L}^s$ over some measure space ($p, s < \infty$), where g is a Carathéodory function satisfying the conditions which guarantee continuity of G (see e.g. [7]). Then

$$[D^*G(x(\cdot)\xi(\cdot))](t) = \xi(t)\partial(\operatorname{sign}\xi(t)(t, x(t)).$$

Fermat rule: If f attains a local minimum at $\overline{x}$, then $0 \in \partial f(\overline{x})$.

Connection with convexity: If f is a convex function, then $\partial f(x)$ coincides with the subdifferential of f at x in the sense of convex analysis, that is $\partial f(x) = \{x^* : f(x+h) - f(x) \geq \langle x^*, h \rangle, \forall h\}$.

Connection with differentiability: If f is continuously (or strictly) Fréchet differentiable at $\overline{x}$, then $\partial f(\overline{x}) = \{f'(\overline{x})\}$; if $F: X \to Y$ is continuously (or strictly) Fréchet differentiable at $\overline{x}$, then $D^*F(\overline{x}) = (F'(\overline{x}))^*$.

Lipschitz properties: (a) If f satisfies the Lipschitz condition near x with constant K and $x^* \in \partial f(x)$, then $\|x^*\| \leq K$;
(b) A set-valued mapping $F: X \rightrightarrows Y$ is called *pseudo-Lipschitz* (or having the Aubin property near $(\overline{x}, \overline{y}) \in \operatorname{Gr} F$) if there are $\varepsilon > 0$ and $K > 0$ such that

$$F(x') \bigcap B(\overline{u}, \varepsilon) \subset F(x) + K\|x - x'\|B,$$

provided $\|x-\bar{x}\|,\|x'-\bar{x}\| \le \varepsilon$. If F is pseudo-Lipschitz near $(\bar{x},\bar{y})$ and $x^* \in D^*F(\bar{x},\bar{y})(y^*)$, then $\|x^*\| \le K\|y^*\|$.

Robustness of the limiting subdifferential. The limiting subdifferential of a function on $\mathbb{R}^n$ is upper semicontinuous set-valued mapping: if $x_n \to x$ and $y_n \in \partial f(x_n)$ converge to some y, then $y \in \partial f(x)$.

Subdifferential of the distance function. Let $Q \subset \mathbb{R}^n$ and $f(x) = d(x,Q)$ Then the norm of any element of $\partial f(x)$ is not greater than one; it is exactly one if $f(x) > 0$. In the latter case if $y \in \partial f(x)$ and $v \in Q$ is such that $\|x-v\| = f(x)$, then $y \in N(Q,v)$.

We refer to [3,9,10] for further information and proofs.

3. AN ABSTRACT INCLUSION PROBLEM.

Consider the following problem:

$$\text{minimize} \quad l(x), \ \text{s.t.} \ Ax \in F(x), G(x) \in S. \qquad \textbf{(P)}$$

Here X,Y,Z are WCG Banach spaces, A is a linear operator from X into Y, $F: X \rightrightarrows Y$ is a set-valued mapping, $G: X \to Z$ and $S \subset Z$. Let $\bar{x} \in X$ be an admissible element in the problem. Set $\bar{y} = A(\bar{x}), \bar{z} = G(\bar{x})$. With every problem of this type we can associate two concepts of a local minimum:

- *strong* minimum which is a local minimum in the usual sense: $l(x) \ge l(\bar{x})$ for all admissible x satisfying $\|x-\bar{x}\| < \varepsilon$;
- $A-$*weak* minimum: $l(x) \ge l(\bar{x})$ for all admissible x satisfying $\|x-\bar{x}\| + \|A(x-\bar{x})\| < \varepsilon$.

Concerning the components of the problem, the functional l, the operator A, the constraint set S and the set-valued mapping F, we shall assume the following:

- $(\mathbf{A}_1)$ l satisfies the Lipschitz condition near $\bar{x}$;
- $(\mathbf{A}_2)$ A is densely defined closed linear operator with $A = Y$ and a compact inverse $\Lambda: Y \to X$;
- $(\mathbf{A}_3)$ G is Lipschitz in a neighborhood of $\bar{x}$ and S is closed and normally compact near $\bar{z}$, that is

$$z_n \to \bar{z}, z_n^* \in N(S,x_n), z_n^* \to 0 (\text{weak}^*) \Rightarrow \|z_n^*\| \to 0.$$

This holds in particular when S is a convex set with a nonempty interior.

The assumptions on F depend on the type of the minimum to be considered:

$(\mathbf{A}_4)_w$ the graph of F is closed and F is pseudo-Lipschitz near $(\overline{x},\overline{y})$,
$(\mathbf{A}_4)_s$ the graph of F is closed and for any $N>0$ there are there are $\varepsilon>0, K>0$ such that

$$F(x')\bigcap B(\overline{y},N)\subset F(x)+K\|x-x'\|B$$

if $\|x-\overline{x}\|<\varepsilon, \|x'-\overline{x}\|<\varepsilon$;

Both conditions can be reformulated in terms of Lipschitz properties of the distance function $d(y,F(x))$. Namely, $(\mathbf{A}_4)_w$ means that this function satisfies the Lipschitz condition in a neighborhood of $(\overline{x},\overline{y})$, while $(\mathbf{A}_4)_s$ implies that for any N there is a ε such that the function is locally Lipschitz on the product $B(\overline{x},\varepsilon)\times B(\overline{y},N)$.

The following theorem gives a necessary optimality condition for the A-weak minimum in (**P**).

Theorem 2. *Assume (A_1)–(A_3), $(A_4)_w$. If $\overline{x}$ is an A-weak local solution to (P), then there are $\lambda\geq 0$, $y^*\in Y^*$, $z^*\in N(S,\overline{z})$ such that the following two relations are satisfied*

(a) $\lambda+\|y^*\|+\|z^*\|>0$ *(non-triviality);*
(b) $A^*y^*\in\lambda\partial l(\overline{x})+D^*F(\overline{x},\overline{y})(y^*)+D^*G(\overline{x})(z^*)$ *(adjoint inclusion).*

Proof. We can reformulate our problem as follows:

$$\text{minimize } l(x) \quad \text{s.t. } Ax-y=0,\ y\in F(x),\ G(x)\in S. \qquad (\mathbf{P}')$$

Then $(\overline{x},\overline{y})$ is a local solution in (**P'**). This is obvious. Set

$$\varphi(x,y)=d(y,F(x))+d((x,y),\operatorname{Gr}A)+d(G(x),S). \qquad (1)$$

This functions satisfies the requirement of the optimality alternative, so we can apply it to(**P'**) with this φ.

Regular case. Assume that there is a $\lambda>0$ such that the function $g_\lambda(x,y)=\lambda l(x)+\varphi(x,y)$ attains a local minimum at $(\overline{x},\overline{y})$. Then

$$(0,0)\in\partial g_\lambda(\overline{x},\overline{y}). \qquad (2)$$

As was explained prior to the statement of the theorem, the function $d(y,F(x))$ satisfies the Lipschitz condition in a neighborhood of $(\overline{x},\overline{y})$. It is possible to show that in this case the normal cone to the graph of F is generated by the subdifferential of $d(\cdot,F(\cdot))$. Furthermore, the other two items in the right-hand part of (1) are also Lipschitz functions. Therefore by

the sum rule, (2) implies that zero belongs to the sum of the (approximate) subdifferentials of the three functions at $(\bar{x},\bar{y})$. We have

- if (x^*,y^*) belongs to the subdifferential of $d(\cdot,F(\cdot))$ at $(\bar{x},\bar{y})$, then $\|x^*\|\le K, \|y^*\|\le 1, x^*,\ x^*\in D^*F(\bar{x},\bar{y})(-y^*)$;

- if (x^*,y^*) belongs to the subdifferential of $d(\cdot,\mathrm{Gr}\,A)$ at $(\bar{x},\bar{y})$, then $\|x^*\|\le 1, \|y^*\|\le 1$ (if we take the l_1-norm in $X\times Y$), and $x^*+A^*y^*=0$;

- if x^* belongs to the subdifferential of $d(G(\cdot),S)$ at $\bar{x}$, then by the chain rule there is a $z^*\in N(S,\bar{x})$ such that $\|z^*\|\le 1$ and $x^*\in D^*G(\bar{x})(z^*)$.

Combining this with (2), we conclude, that there are y^* and $z^*\in N(S,\bar{z})$ such that

$$0\in\lambda\partial l(\bar{x})+D^*F(\bar{x},\bar{y})(y^*)+D^*G(\bar{x})(z^*)-A^*y^*$$

as claimed.

Singular case. If g_λ does not attain a local minimum at $(\bar{x},\bar{y})$ for any positive λ, the by Theorem 1 there is a sequence (x_n,y_n) converging to $(\bar{x},\bar{y})$ and consisting of inadmissible vectors for (**P'**) such that for each n the function

$$\psi_n(x,y)=\varphi(x,y)+\frac{1}{n}(\|x-x_n\|+\|y-y_n\|)$$

attains global minimum at (x_n,y_n). Then $(0,0)\in\partial\psi_n(x_n,y_n)$. The latter means that there are $(x_n^*,-y_n^*)$ in the subdifferential of $d(\cdot,F(\cdot))$, (u^*,v^*) in the subdifferential of $d(\cdot,\mathrm{Gr}\,A)$ at (x_n,y_n) and a w^* in the subdifferential of $d(G(\cdot),S)$ at x_n^* such that

$$\|x_n^*+u_n^*+w_n^*\|\le 1/n,\quad \|y_n^*-v^*\|\le 1/n. \tag{3}$$

Furthermore, by the chain rule, $w_n^*\in D^*G(x_n)(z_n^*)$ for some z_n^* in the subdifferential of $d(\cdot,S)$ at $G(x_n)$ and as follows from $(\mathbf{A}_3)$ and $(\mathbf{A}_4)_w$ there is a $K>0$ such that

$$\|x_n^*\|\le K\|y_n^*\|,\ \|w_n^*\|\le K\|z_n^*\|,\forall n. \tag{4}$$

The norms of functionals $(x_n^*),(y_n^*),\ldots$ are uniformly bounded as the corresponding functions are Lipschitz, so without any loss of generality we may assume (as both spaces are WCG) that each of the sequences weak*-converges to certain vectors in the corresponding spaces: $x_n^* \to x^*$, $y_n^* \to y^*, u_n^* \to u^*, v_n^* \to v^*, w_n^* \to w^*, z_n^* \to z^*$. It follows from (3) that

$$x^* + u^* + w^* = 0, \quad y^* = v^*, \quad w^* \in D^*G(x)z^* \tag{5}$$

Furthermore, as in the regular case, we have $u_n^* + A^*v_n^* = 0$ which by $(\mathbf{A}_2)$ means that $v_n^* = -\Lambda^* u_n^*$ norm converge to v^*, hence by (3),(5) y_n^* norm converge to y^*.

As neither of (x_n, y_n) is admissible in the problem, it must violate one of the three constraints in $(\mathbf{P}')$. This means that at least one of the three constraints must be violated for infinitely many n. Again, without any loss of generality, we assume that a certain condition is not satisfied by all (x_n, y_n). Thus:

- either $y_n \notin F(x_n)$ for all n,
- or $y_n \neq Ax_n$ for all n,
- or $z_n = G(x_n) \notin S$ for all n.

We shall show that in either case either $y^* \neq 0$ or $z^* \neq 0$. Indeed, if $y_n \notin F(x_n)$ for all n, then $\|y_n^*\| = 1$, hence $y^* = 1$. If $y_n \neq Ax_n$ for all n, then $\max\{\|u_n^*\|, \|v_n^*\|\} = 1$ for all n. If we admit that $\|v_n^*\| \to 0$, then $\|u_n^*\| = 1$ for all n, hence $\|x_n^* + w_n^*\| \to 1$. But (3) and (4) imply together that $\|x_n^*\| \to 0$, hence $\|w_n^*\| \to 1$. By (4) the latter implies that the norms of z_n^* are bounded away from zero and (A_3) implies that $z^* \neq 0$.

Finally, if $z_n \notin S$ for all n, then $\|z_n^*\| = 1$ for all n, and as above, we conclude that $z^* \neq 0$. Note also that in case $w^* \neq 0$ we have $x^* + u^* \neq 0$ which would imply that $y^* = v^* \neq 0$ for otherwise we would have $x^* = u^* = 0$.

Applying (5) along with the fact that $u^* = -A^*v^*$, we conclude that $A^*y^* \in D^*F(\bar{x},\bar{y})(y^*) + D^*G(\bar{x})(z^*)$ which is the desired inclusion with $\lambda = 0$. □

4. RELAXABILITY AND THE MAXIMUM PRINCIPLE

Property $(\mathbf{A}_2)$ is typical for differential operators. Many of them, actually many of the most important differential operators, have an additional property that allows to strengthen the necessary optimality condition of Theorem 2 up to a necessary condition for strong minimum in the form that can be naturally called an *abstract maximum principle*. Let us agree to say

that a function $f(x,y)$ attains a strong local minimum at $(\overline{x},\overline{y})$ subject to $Ax=y$ if ($A\overline{x}=\overline{y}$ and) there is an $\varepsilon>0$ such that $f(x,Ax)\geq f(\overline{x},\overline{y})$), provided $\|x-\overline{x}\|\leq\varepsilon$.

Definition. Let f be a function on $X\times Y$. We shall say that f is *relaxable with respect to* A at $(\overline{x},\overline{y})$ if there is an $\varepsilon>0$ such that for any (x,y) with $\|x-\overline{x}\|<\varepsilon$, any finite collection $\{y_1,...,y_k\}\subset Y$ and any $\alpha_1\geq 0,...,\alpha_k\geq 0$ with $\sum\alpha_i\leq\varepsilon$ there is a sequence (x_n,y_n) such that $Ax_n=y_n, y_n$ weakly converge to $y+\sum\alpha_i(y_i-y)$ and $f(x_n,y_n)$ converge to

$$g(u,y,\alpha_1,...,\alpha_k)=f(u,y)+\sum\alpha_i(f(u,y_i)-f(u,y)),$$

where $u=\lim x_n$.

Proposition 3. *Suppose that* f *has the following property*:

($\mathbf{A_5}$) for any $N>0$ there is an $\varepsilon>0$ such that $f(\cdot,y)$ satisfies the Lipschitz condition in the ε-ball around $\overline{x}$ if $\|y-\overline{y}\|\leq N$.

If A *satisfies* ($\mathbf{A}_2$), f *is relaxable with respect to* A *at* $(\overline{x},\overline{y})$ *and attains at* $(\overline{x},\overline{y})$ *a strong local minimum subject to* $Ax=y$, *then*

(a) for any finite collection $y_1,...,y_k$ *of elements of* Y *the function*

$$h(x,y,\alpha_1,...,\alpha_k)=g(\overline{x}+x,\overline{y}+y,\alpha_1,...,\alpha_k)$$

attains at $(0,0,0,...,0)$ *a local minimum subject to*

$$Ax=y+\sum_{i=1}^{k}\alpha_i(y_i-y);\quad \alpha_i\geq 0,\ i=1,...,k;$$

(b) there is a y^* *such that*

$$(-A^*y^*,y^*)\in\partial f(\overline{x},\overline{y});\quad f(\overline{x},y)-f(\overline{x},\overline{y})\geq\langle y^*,y-\overline{y}\rangle,\ \forall\ y\in Y.$$

Proof. Fix some $y_1,...,y_k\in Y$ and let T be a linear operator from $X\times Y\times\mathbb{R}^k$ into X defined by

$$T(x,y,\alpha_1,...,\alpha_k)=x-\Lambda(y+\sum_{i=1}^{k}\alpha_i(y_i-y)).$$

(Recall that Λ is the inverse of A.) Clearly T is a bounded (even Fredholm) operator with $T = X$. Moreover, the image under T of the convex closed cone $K = X \times Y \times \mathbb{R}^n_+$ (the latter being the positive orthant in $\mathbb{R}^n$) is the whole of X. By the theorem of Robinson-Ursesku (see e.g. [5]) there is an $M > 0$ such that

$$d((x, y, \alpha_1, ..., \alpha_k), \ker T \bigcap K) \le M(\| x - \Lambda(y + \sum_{i=1}^{k} \alpha_i (y_i - y)) \| + \sum_{i=1}^{k} \alpha_i^+), \quad (6)$$

where $\alpha^+ = \max\{\alpha, 0\}$.

Observe that a common M can be chosen for all possible choices of k and $y_1, ..., y_k$.

Since f is relaxable with respect to A at $(\overline{x}, \overline{y})$, the function $h(x, y, \alpha_1, ..., \alpha_k)$ attains at $(0,0,0,...,0)$ a local minimum subject to $(x, y, \alpha_1, ..., \alpha_k) \in \ker T \bigcap K$. Indeed, if $Ax = y$, the norm of x, y and $\sum \alpha_i$ are sufficiently small and $Au = y + \sum \alpha_i (y_i - y)$ is close to x and hence to zero . By definition there is a sequence (x_n, y_n) such that y_n converge to $y + \sum \alpha_i (y_i - y)$ and $f(x_n, y_n)$ converge to $g(\overline{x} + u, \overline{y} + y, \alpha_1, ..., \alpha_k)$. But x_n converge to u by (A_2) and therefore $f(\overline{x} + x_n, \overline{y} + y_n) \ge f(\overline{x}, \overline{y})$. Therefore $g(\overline{x} + u, \overline{y} + y, \alpha_1, ..., \alpha_k) \ge f(\overline{x}, \overline{y}) = g(\overline{x}, \overline{y}, 0, ...0)$. This proves (a).

Thanks to $(\mathbf{A}_5)$, we can apply the optimality alternative to the problem of minimizing $h(x, y, \alpha_1, ..., \alpha_k)$ subject to $Ax = y + \sum \alpha_i (y_i - y), \alpha_i \ge 0$. Moreover, it follows from the proof of Theorem 1 and (6) that we have the regular situation in this case. Thus there is an N, not depending on $y_1, ..., y_k$ such that $(0,0,0,...,0)$ is an unconditional minimum of

$$h(x, y, \alpha_1, ..., \alpha_k) + N\left(\|x - \Lambda(y + \sum_{i=1}^{k} \alpha_i (y_i - y))\| + \sum_{i=1}^{k} \alpha_i^+\right).$$

It follows that zero belongs to the subdifferential of the function at $(0,0,0,...,0)$ and standard arguments show the existence of $(x^*, y^*) \in \partial f(\overline{x}, \overline{y})$, z^* in the unit ball of X^* and $\xi_i \in [0, N], i = 1, ..., k$, such that $x^* + Nz^* = 0, y^* - N\Lambda^* z^* = 0$ and $f(\overline{x}, \overline{y} + y_i) - f(\overline{x}, \overline{y}) - N\langle z^*, \Lambda(y_i)\rangle - \xi_i = 0$ for all i, that is to say $f(\overline{x}, \overline{y} + y_i) - f(\overline{x}, \overline{y}) \ge \langle y^*, y_i \rangle$ for $i = 1, ..., k$.

The proof is now completed by the following standard argument. Denote by $\mathcal{Y}^*(y_1, ..., y_k)$ the collection of $y^* \in Y^*$ such that $\|y^*\| \le N$ and

$$(-A^*y^*, y^*) \in \partial f(\bar{x}, \bar{y}); \quad f(\bar{x}, y_i) - f(\bar{x}, \bar{y}) \geq \langle y^*, y_i - \bar{y} \rangle, i = 1, ..., k.$$

Clearly, each such set is weak* compact and the sets decrease as the collections of y_i increase. Therefore there is an element belonging to all $\mathcal{Y}^*(y_1, ..., y_k)$. □

Returning back to our problem (**P**), we can get the following *maximum principle* as a necessary condition for a strong minimum in the problem.

Theorem 4. *Assume* $(\mathbf{A}_1)$ *-*$(\mathbf{A}_3)$ *,* $(\mathbf{A}_4)_s$. *Suppose further that the function* $f(x, y) = d(y, F(x)) + \varepsilon \|y - z\|$ *is relaxable with respect to* A *for any* ε *and* z. *If under these conditions,* $\bar{x}$ *is a strong local solution to* (**P**), *then there are* $\lambda \geq 0$, $y^* \in Y^*$, $z^* \in N(S, G(\bar{x}))$ *such that the following three relations are satisfied*

(a) $\lambda + \|y^*\| + \|z^*\| > 0$ *(non – triviality);*

(b) $A^*y^* \in \lambda \partial l(\bar{x}) + D^*F(\bar{x}, \bar{y})(y^*) + D^*G(\bar{x})(z^*)$ *(adjoint inclusion);*

(c) $\langle y^*, A \rangle = \max_{y \in F(\bar{x})} \langle y^*, y \rangle$ *(maximum principle).*

The proof of the theorem is an easy adaptation of the proof of Theorem 2 with Proposition 3 taken into account and in view of the fact that $l(x)$ and $d(G(x), S)$ both satisfy the Lipschitz condition.

5. A CONTROL PROBLEM WITH PARABOLIC INCLUSION AND STATE CONSTRAINTS

The proofs in the previous section offer basic guidelines to the alternative-based approach to necessary optimality conditions in more specialized problems which usually require more careful analysis to get desired results. As an example, we shall consider in this section the following optimal control problem for systems governed by "parabolic inclusions".

Let Ω be a bounded domain in $\mathbb{R}^n$ with a regular boundary (e.g. C^2-smooth). Set $Q = (0, T) \times \Omega$ and let $\Gamma = (0 \times \Omega) \bigcup (0, T) \times \mathrm{bd}\Omega$. We shall be interested in necessary optimality conditions for a strong minimum in the following problem

$$\begin{aligned} \text{minimize} \quad & J(u(\cdot)) = \int_{\Omega} L(x,u(T,x))dx \\ \text{s.t.} \quad & \frac{\partial}{\partial t}u - \Delta_x u \in \Phi(t,x,u); \\ & u|_{\Gamma} = 0 \\ & g(\cdot,u(\cdot)) \in C. \end{aligned} \tag{C}$$

Our purpose is to demonstrate the work of the mechanism developed in this paper not obscured by technical complications usually accompanying problems involving partial differential operators in sufficiently general situations. The statement of the problem reflects this intention and so do the assumptions on the components of the problem which allow to avoid main technical complications and to concentrate on the optimization content of the subsequent arguments. (I am sure that the assumptions can be substantially weakened without affecting the result.) The calculation of subdifferentials, not directly connected with the work of the optimality alternative will also be omitted.

Before stating the assumption, it seems appropriate to mention that optimal control problems in standard form with control equations rather than differential inclusions and terminal functionals as in the statement can be easily reduced to this form. Certain complications may appear in case when the cost function is defined by the area functional with integrand also depending on control. But in this case too, a reduction to a problem with differential inclusion is possible although the inclusion will be a bit different from the considered above. We shall not consider such problems here.

- **(A$_6$)** L is nonnegative, measurable with respect to (t,x) and there is a $k_L \in \mathbb{R}$ such that $|L(t,x,u) - L(t,x,v)| \le k_L |u-v|$;
- **(A$_7$)** Φ is measurable with respect to (t,x) for any u, the graph of $\Phi_{t,x}(\cdot) = \Phi(t,x,\cdot)$ is closed for almost every (t,x) and there is a $k_\Phi \in \mathbb{R}$ such that the Hausdorff distance between $\Phi(t,x,u)$ and $\Phi(t,x,v)$ is not greater than $k_\Phi |u-v|$; moreover, we shall assume that there is an $r_\Phi(t,x) \in \mathcal{L}^2$ such that $|y| \le r_\Phi(t,x)$ if $y \in \Phi(t,x,u)$;
- **(A$_8$)** C is a closed convex set in a Banach space $\mathcal{L}^1$ with a nonempty interior, $g(t,x,u)$ is measurable with respect to (t,x) and there are $k_g \in \mathbb{R}$ and $r_g(\cdot) \in \mathcal{L}^2$ such that

$$|g(t,x,u)| \le r_g(t,x); \quad |g(t,x,u) - g(t,x,v)| \le k_g |u-v| \tag{7}$$

for all t,x,u,v.

The problem is naturally interpreted as (**P**) if we set $X = Y = Z = Z^1$, $A = (\partial/\partial t) - \Delta_x$ and weak solutions of the heat equation $Au = y$ are considered, F is the set-valued mapping which with every $u(t,x)$ associates the collection of all measurable selections of $\Phi(t,x,u(t,x))$ and G is the Nemytzki operator defined by $G(u(\cdot))(t,x) = g(t,x,u(t,x))$. With this choice of X, $\overline{u}(\cdot)$ is a strong minimum in the problem if there is an $\varepsilon > 0$ such that $J(u(\cdot)) \geq J(\overline{u}(\cdot))$ for any admissible $u(\cdot)$ satisfying

$$\int_Q |u(t,x) - \overline{u}(t,x)| \, dxdt \leq \varepsilon.$$

The choice of spaces is dictated by the desire to use relaxability at an appropriate stage of the proof. However it leads to some difficulties. The first is that the cost function is in general not well defined for a $u \in Z^1$. This difficulty can be overcome by interpreting the integral in the cost function as

$$\liminf_{\varepsilon \to 0} \varepsilon^{-1} \int_{T-\varepsilon}^{T} \int_\Omega L(x,u(t,x)) dxdt. \tag{8}$$

The second and more fundamental problem is that the choice of $\mathcal{L}^1$ as the space from which right-hand parts of the heat equation can be taken may negatively affect properties similar to those established by ($\mathbf{A}_2$). We bypass it by imposing (in ($\mathbf{A}_6$)-($\mathbf{A}_8$)) conditions which effectively make the right-hand parts belong to $\mathcal{L}^2$ when necessary.

Theorem 5. *Assume* ($\mathbf{A}_6$) –($\mathbf{A}_8$). *Suppose that* $\overline{u}(\cdot)$ *is a strong local minimum in the problem. Then there are* $\lambda \geq 0$, $p(\cdot) \in W^{1,2}$ *and* $\xi(\cdot) \in \mathcal{L}^2$ *such that*

(a) $\lambda + \|p(\cdot)\| + \|\xi(\cdot)\| > 0$;

(b) $\int_Q \xi(t,x)(g(t,x,\overline{u}(t,x)) - v(t,x)) dxdt \geq 0, \forall\, v(\cdot) \in C$;

(c) $-p(T,x) \in \partial_c L(x,\overline{u}(T,x))$ a.e. on Ω;

(d) $\dfrac{\partial p(t,x)}{dt} + \Delta_x p(t,x) \in \operatorname{conv}[D^* \Phi_{t,x}(\overline{u}(t,x),\overline{y}(t,x))(p(t,x))]$
$+\partial(\xi(t,x) g(t,x,\cdot)(\overline{u}(t,x))$ a.e. on Q;

(e) $p(t,x)\overline{y}(t,x) = \max\limits_{y \in \Phi(t,x,\overline{u}(t,x))} p(t)y$ a.e. on Q.

(Here of course L is subdifferentiated w.r.t the second variable.)

Proof. 1) We start by reformulating the problem. Let $\mathcal{M}^p$ denote the closure in $\mathcal{L}^p \times \mathcal{L}^p$ of the collection of all pairs $(u(\cdot), y(\cdot))$ such that $u(\cdot) \in C^\infty$, $((\partial/\partial t) - \Delta_x)u = y$ and u vanishes in a neighborhood of Γ. Then

$\mathcal{M}^p$ is a linear subspace of $\mathcal{L}^p \times \mathcal{L}^p$ which can be considered a graph of a linear operator A from $\mathcal{L}^p(Q)$ into itself.

We shall be interested here only in $p=1$ and $p=2$. Observe that $\mathcal{M}^2 \subset \mathcal{M}^1$ and for any $y \in \mathcal{L}^2$ there is a unique u such that $(u,y) \in \mathcal{M}^2$. This u is a solution of the equation $((\partial/\partial t) - \Delta_x)u = y$ subject to $u|_\Gamma = 0$ and actually belongs to $\overset{\circ}{W}{}^{1,2}$. Moreover the mapping $t \to u(t,\cdot)$ from $[0,T]$ into $\mathcal{L}^2$ is continuous as well as the mapping $\mathcal{L}^2 \to \overset{\circ}{W}{}^{1,2}$ which associates with $y \in \mathcal{L}^2$ the corresponding u; hence it is a compact mapping from $\mathcal{L}^2$ into itself (see e.g. [8]).

The problem can now be reformulated as follows

$$\begin{array}{ll} \text{minimize} & J(u) \\ \text{s.t.} & y(t,x) \in \Phi(t,x,u(t,x)),\ \text{a.e.},\ G(u) \in C;\quad (u,y) \in \mathcal{M}^1, \end{array} \tag{$\mathbf{C}_1$}$$

where G is a mapping (into $\mathcal{L}^2$ - see $(\mathbf{A}_8)$) defined by $G(u(\cdot))(t,x) = g(t,x,u(t,x))$.

As $\overline{u}$ is a strong local minimum in $(\mathbf{C})$, $(\overline{u},\overline{y})$ is a local minimum in $(\mathbf{C}_1)$ in the following sense: $J(u(\cdot)) \geq J(\overline{u}(\cdot))$ for any admissible pair $(u,y) \in \mathcal{M}^1$ such that u is $\mathcal{L}^1$-close to $\overline{u}$. By $(\mathbf{A}_6)$ J satisfies the Lipschitz condition on $\mathcal{L}^1$, hence we can apply the optimality alternative with

$$\varphi(u,y) = \int_Q d(y(t,x),\Phi(t,x,u(t,x))dxdt + d(g(\cdot,u(\cdot)),C).$$

It follows that

- either there is a $\lambda > 0$ such that $\lambda J + \varphi$ attains an unconditional local minimum on $\mathcal{M}^1$ at $(\overline{u},\overline{y})$ (in the above defined sense);
- or there is a sequence $(u_n y_n,) \in \mathcal{M}^1$ converging to $(\overline{u},\overline{y})$ such that each $(u_n y_n,)$ is not admissible in $(\mathbf{C}_1)$ and each function

$$\psi_n(u,y) = \varphi(u,y) + n^{-1}(\|u-u_n\|_1 + \|y-\overline{y}\|_1)$$

attains absolute minimum at (u_n, y_n). (The subscript refers to the fact that we consider here the $\mathcal{L}^1$-norms).

It follows from $(\mathbf{A}_7)$ that $\overline{y}$ belongs to $\mathcal{L}^2$. Therefore we conclude that

- in the regular case $\lambda J + \varphi$ attains an unconditional local minimum on $\mathcal{M}^2$ at $(\overline{u},\overline{y})$;
- in the singular case ψ_n attains a (finite) absolute minimum on $\mathcal{M}^1$ at (u_n, y_n).

2) Next we show that in either case the functions to be minimized are relaxable with respect to A. This will follow from the lemma below.

Lemma 6. *Consider the functional*

$$I(u(\cdot),y(\cdot))=\int_Q m(t,x,u(t,x),y(t,x))dxdt$$

assuming that m *is a Carathéodory function (measurable with respect to* (t,x) *and continuous with respect to* (u,y)*). Suppose that for given pair* (u_0,y_0) *for which* $I(u_0,y_0)$ *well defined and finite*

$$|\,m(t,x,u,y)-m(t,x,u_0,y_0)\,|\le \rho(t,x)(|\,u-u_0\,|+|\,y-y_0\,|)$$

with some $\rho\in\mathcal{L}^2$.

Then for any $y_1,\dots y_k,\in\mathcal{L}^2$ *and any* $\alpha_1\ge 0,\dots,\alpha_k\ge 0$ *with* $\sum\alpha_i\le 1$ *there is a sequence* $(w_n,z_n)\in\mathcal{M}^2$ *such that* z_n *converge weakly to* $y_0+\sum\alpha_i\,(y_i-y_0)$ *and*

$$I(u_0+w_0,y_0+z_n)\to I(u_0+\sum_{i=1}^{k}\alpha_i u_i,y_0)+\sum_{i=1}^{k}\alpha_i(I(u_0+\sum\alpha_i\,u_i,y_i)-I(u_0+\sum\alpha_i\,u_i,y_0)),$$

where $Au_i=y_i$.

Proof. Let $\alpha_{in}(t,x),(i=1,\dots,k,n=1,2,\dots)$ be functions on Q with the following properties: $\alpha_{in}\in\{0,1\}$, $\sum_i\alpha_{in}\le 1$ (which means that for any n α_{in} are characteristic functions of a certain collection of disjoint subsets of Q) and for each i the sequence (α_{in}) weakly converge (e.g. in $\mathcal{L}^2$) to the function identically equal to α_i. Set $z_n=y_0+\sum_i\alpha_{in}(y_i-y_0)$, and let w_n be the corresponding solution of $Aw=z_n$ (which as $z_n\in\mathcal{L}^2$ is well defined). Then

$$I(w_n,z_n)=\int_Q[m(t,x,w_n(t,x),y_0(t,x)$$
$$+\sum\nolimits_{i=1}^{k}\alpha_{in}(t,x)(m(t,x,w_n(t,x),y_i(t,x))-m(t,x,w_n(t,x)y_0(t,x)))]dxdt,$$

and the result follows since w_n norm converge to $\sum\alpha_i\,u_i$ (as all $z_n\in\mathcal{L}^2$). □

3) Set

$m_0(t,x,u,y)) = d(\overline{y}(t,x)+y, \Phi(t,x,\overline{u},(t,x)+u))$;

$m_n(t,x,u,y) = d(y_n(t,x)+y, \Phi(t,x,u_n(t,x)+u)) + n^{-1}\,|\,y\,|$;

$g_0(t,x,u) = g(t,x,\overline{u}(t,x)+u); \quad g_n(t,x,u) = g(t,x,u_n(t,x)+u)$,

and let for $n = 0,1,...$

$$M_n(u(\cdot),y(\cdot)) = \int_Q m_n(t,x,u(t,x),y(t,x))dxdt.$$

Then each m_i satisfies the condition of Lemma 6, thanks to $(\mathbf{A}_7)$, with $\rho = k_\Phi$. Set further for given $y_1,...,y_k$ in $\mathcal{L}^2$

$$J_0(u(\cdot),y(\cdot),\alpha_i,...,\alpha_k) = \lambda J(\overline{u}+u) + d(G_0(\cdot,u(\cdot)),C) + M_0(u(\cdot),y(\cdot)) + \sum_{i=1}^{k} \alpha_i (M_0(u(\cdot),y_i(\cdot)) - M_0(u(\cdot),y(\cdot)))$$

and for $n \geq 1$

$$J_n(u(\cdot),y(\cdot),\alpha_i,...,\alpha_k) = d(G_n(\cdot,u_n+u(\cdot)),C) + M_n(u(\cdot),y(\cdot)) + \sum_{i=1}^{k} \alpha_i (M_n(u(\cdot),y_i(\cdot)) - M_n(u(\cdot),y(\cdot))) + n^{-1}\|u(\cdot)\|_1,$$

where $G_n, n = 0,1,...$ is the Nemytzki operator $u(t,x) \mapsto g_n(t,x,u(t,x))$, and consider the following problems for $n = 0,1,2,...$.

$$\begin{array}{ll} \text{minimize} & J_n(u,y,\alpha_i,...\alpha_k), \\ \text{s.t} & \alpha_i \geq 0,...,\alpha_k \geq 0; \quad (u, y+\sum \alpha_i(y_i - y)) \in \mathcal{M}^2, \end{array} \qquad (\mathbf{P}_n)$$

Lemma 6 and the first part of Proposition 3 together with the result obtained at the first step of the proof imply that

– either $(0,0,0,...,0)$ is a local minimum in $(\mathbf{P}_0)$;

– or $(0,0,0,...,0)$ is an absolute minimum in each of $(\mathbf{P}_n)$ ($n = 1,2,...$).

On the other hand, the second part of Proposition 3 guarantees that whenever $(0,0,0,...,0)$ is a a local minimum in a certain $(\mathbf{P}_n)$, there is a $p_n \in \mathcal{L}^2$ such that

$$(-A^* p_n(\cdot), p_n(\cdot)) \in \partial M_n(0,0) + \lambda_n \partial J(0) + \partial d(G(\cdot),C)(0)) + \delta_n B^\infty \qquad (9)$$

where by B^{∞} we have denoted the unit ball in $\mathcal{L}^{\infty}$, and

$$M_n(0,y(\cdot))-M_n(0,0)\geq\int_Q p_n(t,x)y(t,x)dxdt,\quad \forall y(\cdot)\in\mathcal{L}^2. \tag{10}$$

Here we have set $\lambda_0=\lambda,\lambda_n=0,n=1,2,\ldots$ and $\delta_0=0,\delta_n=n^{-1},n=1,2,\ldots$.
4) We need to decipher (9), (10). The standard application of measurable selection arguments allows to conclude that (10) implies that

$$m_n(t,x,0,y)-m_n(t,x,0,0)\geq p_n(t,x)y,\quad \forall y, \text{ a.e. on } Q. \tag{11}$$

To decipher (9) we have to understand the general form of of each of the four suddifferentials in the right hand part of (9) and the action of A^*p. The latter is straightforward: action of A^*p on any smooth $u(\cdot)$ satisfying $u|_{\Gamma}=0$ is defined by

$$\left\langle A^*p,u\right\rangle=-\int_Q[(\frac{\partial}{\partial t}+\Delta_x)p]udxdt+\int_{\Omega}p(T,x)u(T,x)dx. \tag{12}$$

By the chain rule the approximate subdifferential of $d(G_n(\cdot),C)$ at zero consists of vectors belonging to $D^*G_n(0)(\xi_n)$, where ξ_n belongs to the subdifferential of the distance to C (in the $\mathcal{L}^1$-metric) at $G_n(0)$, that is $|\xi_n(t,x)|\leq 1$ a.e. on Q and

$$\sup\{\int_Q\xi_n(t,x)v(t,x))dxdt:v\in C\}=\int_Q\xi_n(t,x)g_n(t,x,0)dxdt-d(G_n(0),C) \tag{13}$$

On the other hand, as G is a Nemytzki operator satisfying the Lipschitz condition by $(\mathbf{A}_8)$, $\eta\in D^*G_n(0)(\xi_n)$ means that

$$\eta(t,x)\in\partial(\xi_n(t,x)g_n(t,x,\cdot))(0)\quad \text{a.e. on } Q, \tag{14}$$

This is a complete description of $D^*G_n(0)$. The analysis of the structures of first two subdifferentials in (9) is much mure involved and goes beyond the scope of this paper for it actually uses the entire power of the modern subdifferential calculus. We can refer to [2] and [6] for the guidelines of the techniques which is in the basis of calculation of subdifferential of integral functionals. The important point that should be taken into account is that $(\mathbf{A}_7)$, together with the fact that η_n,ξ_n and ρ_n are bounded, implies that $r_n=((\partial/\partial t)+\Delta_x)p_n\in\mathcal{L}^{\infty}$ (and moreover that the $\mathcal{L}^{\infty}$-norms of r_n are

uniformly bounded) and consequently, that all p_n are continuous on the closure of Q.

The results of the calculations based on these techniques can be summarized as follows:

$$-p_n(T,x(t)) \in \partial_c L(T,u_n(T,x)), \quad \text{a.e. on } \Omega; \tag{15}$$

$$r_n(t,x) \in \text{conv } \{q : (-q, p_n) \quad \in \partial m_n(t,x,0,0)\} + \eta_n(t,x)\xi_n(t,x) + \rho_n(t,x), \tag{16}$$

where $\rho_n(t,x) \le \delta_n$ for almost all (t,x) and subdifferentiation is considered with respect to u in (15) inclusion and with respect to (u,y) in (16).

We shall further consider separately the regular and the singular case.

The **regular case** corresponds to $n=0$ and $\lambda > 0$. Then the part (a) of the theorem is automatic, (b) follows from (12), (c) is established in (15) and (d) is a consequence of (16) in view of the definition of m_0. Finally, (e) follows from (11). Indeed, for $n=0$ (11) means that

$$d(\overline{y}(t,x) + y, \Phi(t,x,\overline{u}(t,x)) \ge p(t,x)y$$

for all y a.e. on Q (recall that $\overline{y}(t,x) \in \Phi(t,x,\overline{u}(t,x))$. Fix such (t,x). Then for any $y \in \Phi(t,x,u(t,x)) - \overline{y}(t,x)$ we have $p(t,x)y \le 0$ which implies (e).

The **singular case**, when zero is a solution to each ($\mathbf{P}_n$), requires a bit harder work. First we note that as $\lambda = 0$, the terminal term disappears and we get $p(T,x) \equiv 0$ by (16). Hence (c). Furthermore, as all r_n are uniformly bounded, the sequence (p_n) is relatively compact in the space of continuous functions and we may assume that p_n converge uniformly to some p which is a continuous function on the closure of Q. As all u_n are not admissible in the problem then for infinitely many n either $(Au_n)(t,x) \notin \Phi(t,x))$ on a set of positive measure or $G(u_n) \notin C$. In the latter case $\|\xi_n\|_\infty = 1$ for infinitely many n and, as C is normally compact any weak*-limiting point ξ of the sequence is distinct from zero.

If $\xi_n = 0$ for all but finitely many n, then $p_n(t,x) = 1$ at least at one point. As Q is a bounded domain, it follows that p must be equal to one at least at one point. Thus either the limiting ξ or the limiting p is different from zero. This completes the proof of (a). The second statement (b) follows from (12) and (7) as u_n converge to $\overline{u}$ in $\mathcal{L}^1$, hence almost everywhere, and $d(G_n(0),C) \to 0$.

The adjoint equation (d) as in the regular case as a consequence of (15). Indeed, as r_n are uniformly bounded, we may assume that they converge weakly in $\mathcal{L}^2$ and the limit function r must be $((\partial/\partial t) + \Delta_x)p$. On the other

hand, u_n and y_n converge almost everywhere to $\overline{u}$ and $\overline{y}$ and (e) follows from Lemma 3 of [6]. Finally,it follows from (11) and the definition of m_n that for every n

$$d(y_n(t,x)+y,\Phi(t,x,u_n(t,x))-d(y_n(t,x),\Phi(t,x,u_n(t,x))\geq p_n(t,x)y,\ \forall y$$

for almost every $(t,x)\in Q$. Using again the fact that u_n, y_n and p_n converge almost everywhere to $\overline{u},\overline{y}$, and p respectively, we get (e). □

REFERENCES

[1] F.H. Clarke, Necessary Conditions in Dynamic Optimization, to appear.

[2] A.D. Ioffe, Absolutely continuous subgradients of nonconvex integral functionals, Nonlinear Analysis, TMA 11 (1987).

[3] A.D. Ioffe, Approximate subdifferentials and applications 4. The metric theory, Mathematika,36 (1989),1-38.

[4] A.D. Ioffe, Euler-Lagrange and Hamiltonian formalisms in dynamic optimization, Trans. AMS 349 (1997), 2871-2900.

[5] A.D. Ioffe, Metric regularity and subdifferential calculus Russian Mathematical Surveys, 55:3 (2000), 501-558.

[6] A.D. Ioffe and R.T. Rockafellar, The Euler and Weierstrass conditions for nonsmooth variational problems, Calculus of variations and PDEs, 4 (1996), 59-87.

[7] Krasnoselskii M.A., Zabreiko P.P., Pustylnik E.I., Sobolevskii P.E., Integral Operators in Spaces of Summable Functions, Noordhoff International Publishing, Leyden, 1976.

[8] O.A. Ladyzhenskaya, Boundary Value Problems of Mathematical Physics, Springer-Verlag, 1985.

[9] B.S. Mordukhovich and Y. Shao, Nonsmooth sequential analysis in Asplund spaces, Trans. Amer. Math. Soc., 348 (1996), 1235-1280.

[10] R.T. Rockafellar and R.J.B. Wets, Variational Analysis, Springer-Verlag 1998

[11] H. Sussmann, New Theories of set-valued differentials and new versions of the maximum principle in the optimal control theory, in Nonlinear Control in the Year 2000, A. Isidori, F. Lamnabhi-Lagarrigue and W. Respondek Eds., Springer-Verlag, 2000; pp 487-526.

[12] R. Vinter, Optimal Control, Birkhäuser, 2000.

A VARIATIONAL INEQUALITY SCHEME FOR DETERMINING AN ECONOMIC EQUILIBRIUM OF CLASSICAL OR EXTENDED TYPE

A. Jofre,[1*] R.T. Rockafellar[2**] and R.J.-B. Wets[3***]
Center for Mathematical Modelling and Dept. of Mathematical Engineering, University of Chile, Santiago, Chile;[1] Dept. of Mathematics, University of Washington, Seattle, USA;[2] Dept. of Mathematics, University of California, Davis, USA[3]

Abstract: The existence of an equilibrium in an extended Walrasian economic model of exchange is confirmed constructively by an iterative scheme. In this scheme, truncated variational inequality problems are solved in which the agents' budget constraints are relaxed by a penalty representation. Epi-convergence arguments are employed to show that, in the limit, a virtual equilibrium is obtained, if not actually a classical equilibrium. A number of technical hurdles are, in this way, surmounted.

Key words: variational inequalities, Walras exchange equilibrium, virtual equilibrium, epi-convergence, penalization, equilibrium computations

1. INTRODUCTION

Mathematical models of equilibrium in economics attempt to capture the effects of competing interests among different "agents" in face of the limited availability of goods and other resources. They typically revolve around the

* Reserach supported by MI Nucleus Complex Engineering Systems.
** Reserach supported by the U.S. National Science Foundation under grant DMS-104055.
*** Reserach supported by the U.S. National Science Foundation under grant DMS-0205699 and Office of Naval Research under grant MURI N00014-00-1-0637.

existence of prices for the goods under which the optimization carried out by these agents, individually, leads collectively to a balance between supply and demand.

Although the fundamental ideas go back to Walras and others, the work of Arrow and Debreu [1], [3], initiated the solidly mathematical form of the subject, still continuing in its development. Notions from game theory, such as Nash equilibrium and its counterpart for generalized games (where each agent's strategy set can depend on the other agents' actions), have entered strongly too. Nowadays, influences are also coming from applications beyond the academic, for instance to traffic equilibrium and the practical consequences of deregulation of markets in electrical power.

In the economics literature, fixed-point theory has long provided the environment for establishing whether an equilibrium exists. Fixed-point approaches to calculation were promoted by Scarf [15], [16]. The emphasis on the theory side, though, has largely been on broadening the models so as to encompass preference relations expressed by set-valued mappings that satisfy weakened semicontinuity assumptions and the like. The question of how agents might discover an equilibrium through a Walras-type procedure of tatonnement has been of interest as well, but economists have not devoted much effort to achieving a structured format conducive to large-scale numerical computation. General fixed-point algorithms are notoriously slow and unpromising in anything but simple, low-dimensional situations.

Alternative approaches have been opening up, however, in the optimization literature in connection with variational inequality formulations, including "complementarity" models; see [2], the 1990 survey of Harker and Pang [9], and the 2003 book of Facchinei and Pang [6] for background. Such approaches offer ways of tying the computation of equilibrium into the major advances that have been made in numerical optimization, although this kind of computation is nevertheless much more difficult than mere minimization or maximization.

The task of setting up a variational inequality model for equilibrium involves not only challenges but compromises for the sake of tractibility. Some levels of generality have to be abandoned, at least within present capabilities. For example, the expression of preferences by abstract relations has to be dropped in favor of expression by utility functions, which moreover may need to satisfy assumptions like differentiability. Certain constraints need to be handled with Lagrange multipliers. Such maneuvers run into some serious technical issues, however, two of the main ones being the existence of an equilibrium and the existence of a solution to the proposed variational inequality.

The question of whether an equilibrium exists can be very subtle, even in a purely economic framework. The Arrow-Debreu model [1], as applied to

pure exchange, for instance, effectively requires that each agent start out with a tradable quantity of every possible good. Much effort has successfully gone into weakening that sort of provision, but the techniques appear, at least on the surface, to conflict with the features desired for a readily computible representation. The constraint qualifications ordinarily invoked to ensure access to Lagrange multipliers can fail, in particular. On the other side, the variational inequality models achieved by introducing Lagrange multipliers have the drawback of leading to problems in which the underlying convex sets are unbounded and adequate coercivity is absent. They tend then to fall outside the domain of the standard criteria for confirming that a solution exists.

Our aim in this paper is to demonstrate how these difficulties can be overcome in the fundamental case of a Walras equilibrium, which we take for simplicity (rather than technical necessity) to be a pure exchange equilibrium among consumers, with no producers. We carefully introduce assumptions that enable us to prove the existence, at least, of a "virtual" exchange equilibrium, which might have some agents just barely surviving without optimizing, but can be approximated arbitrarily closely by an exchange equilibrium in the classical sense. Moreover, we show that a *virtual equilibrium* can be computed in principle by solving a sequence of variational inequality problems in which the underlying convex sets are actually compact.

A key contribution lies in showing how the iterative truncations needed technically in order to achieve compactness in the variational inequality, for existence of solutions, can be interpreted as corresponding to penalty representations of the agents' budget constraints, which surprisingly, however, furnish classical equilibrium relative to nearby endowments in place of the original ones. In verifying that the equilibrium sequence from the truncated problems yields, in the limit, a virtual equilibrium, we develop detailed progress estimates and break new ground in utilizing arguments about epi-convergence.

We do not try to answer, here, the question of how the truncated variational inequalities can, themselves, be solved. Some guidance toward the future prospects is available, though, in the recent papers [10], [11], which deal with generalized games, and of course in the book [6], which addresses variational inequalities more generally.

Beyond computation, it should be noted that variational inequality representations of equilibrium are able also to take advantage of the extensive theory on how solutions to variational inequality problems respond to data perturbations, as for instance in [14], [5]. Our work can be viewed as contributing also in that direction.

2. EQUILIBRIUM MODEL

The space of goods is $\mathbb{R}_+^l$; the goods are indexed by $j = 1,...,l$. Each agent $a \in \mathcal{A}$ has an endowment $e_a \in \mathbb{R}_+^l$ and a utility function u_a to be applied to consumption choices. The consumption vector x_a must belong to a certain subset $X_a \subset \mathbb{R}_+^l$. The condition $x_a \in X_a$ is the *survival constraint*, and X_a is the *survival set*. In elementary models, $X_a = \mathbb{R}_+^l$.

Subject to survival and the feasibility of exchanging the goods j at appropriate prices p_j, which are not given but have to be determined from the data elements e_a, X_a and u_a, the agents seek individually to arrange their consumption so as to maximize their utility. The focus is on *relative* price vectors, i.e., vectors p that belong to the price simplex

$$P = \{p = (p_1,..., p_l) \in \mathbb{R}^l \mid p_j \geq 0, p_1 + ... p_l = 1\}. \tag{1}$$

Definition 1 (exchange equilibrium). *A* classical *exchange equilibrium consists of a price vector* $\overline{p}$ *and consumption vectors,* $\overline{x}_a$, *such that*

(a) $\Sigma_{a\in\mathcal{A}} \overline{x}_{aj} \leq \Sigma_{a\in\mathcal{A}} e_{aj}$ *for all goods* j, *with equality holding if* $\overline{p}_j > 0$,

(b) $\overline{x}_a \in \operatorname{argmax}\{u_a(x_a) \mid x_a \in X_a, \overline{p} \cdot x_a \leq \overline{p} \cdot e_a)\}$, *and* $\overline{p} \cdot \overline{x}_a = \overline{p} \cdot e_a$.

A two-tier exchange equilibrium is the same, except that some of the agents a *may satisfy as a substitute for (b) the condition*

(b$^-$) $\overline{x}_a \in \operatorname{argmin}\{\overline{p} \cdot x_a \mid x_a \in X_a\}$, *and* $\overline{p} \cdot \overline{x}_a = \overline{p} \cdot e_a$.

An agent satisfying (b) *will be called an optimizing agent , whereas an agent satisfying* (b$^-$) *will be called a barely surviving agent.*

The requirement that $\overline{p} \cdot x_a \leq \overline{p} \cdot e_a$ is the *budget* constraint for agent a. In a two-tier equilibrium, the barely surviving agents have their budgets so tight that they can only choose cheapest possible consumption vectors from their survival sets, and that uses up all their wealth.

If the argmin in (b$^-$) consists of a unique vector, that is what must be chosen. In that case, (b$^-$) trivially entails (b), so the situation special interest in (b$^-$) is mainly the one where the argmin isn't just a singleton. It's conceivable then that a small amount of freedom may be left for utility optimization while keeping to lowest cost. No such secondary optimization is claimed in the definition, but we don't exclude the possibility that an optimizing agent might also be a barely surviving agent. However, we will really be concerned with a sharpened form of two-tier equilibrium, defined next, in which the barely surviving agents, if any, are "arbitrarily close" to

being optimizing agents and fall short only because of a slightest lack of resources.

Definition 2 (virtual exchange equilibrium). *A two-tier exchange equilibrium, with price vector* $\overline{p}$ *and consumption vectors* $\overline{x}_a$, *is a virtual exchange equilibrium if (when not itself actually a classical equilibrium) it includes at least one optimizing agent and can be approximated arbitrarily closely by a classical equilibrium in the following sense. There are price vectors* p^ν *and consumption vectors* x_a^ν, $\nu = 1,2,...,$ *with*

$$\lim_{\nu\to\infty} p^\nu = \overline{p}, \qquad \lim_{\nu\to\infty} x_a^\nu = \overline{x}_a,$$

which for each ν *furnish a classical exchange equilibrium with respect to the same sets* X_a *and functions* u_a *but possibly different endowments* e_a^ν *satisfying*

$$e_a^\nu \geq e_a, \qquad \lim_{\nu\to\infty} e_a^\nu = e_a.$$

Although any classical equilibrium is a virtual equilibrium in particular (and fits the sequence prescription with $p^\nu = \overline{p}$, $x_a^\nu = \overline{x}_a$, $e_a^\nu = e_a$), the converse is false. Likewise, not every two-tier equilibrium is a virtual equilibrium. Examples of these differences will be provided in the final section of this paper.

In the economic literature, what we are calling a classical exchange equilibrium in Definition 1 is a special case of a Walras equilibrium, namely one in which preferences are expressed by utilities, free disposal is assumed, and "production" has not been introduced. Production is omitted here mainly for the sake of simplicity. The results that will be described can be extended in that way, but we wish to avoid the notational complications in order to focus here on the newer features more clearly.

What we call a two-tier exchange equilibrium in Definition 1 corresponds, under the same specializations, to a model first developed by Debreu [4] as a *quasi*-equilibrium. We prefer to speak of a two-tier equilibrium because the term quasi-equilibrium has shifted over the years to mean something different from what Debreu originally indicated. It regularly refers now, in a utility context like ours, to substituting for (b) the condition that $\overline{x}_a \in X_a$ with $\overline{p}\cdot\overline{x}_a = \overline{p}\cdot e_a$, but there is no $x_a \in X_a$ satisfying both $\overline{p}\cdot x_a < \overline{p}\cdot e_a$ and $u_a(x_a) > u_a(\overline{x}_a)$. This property is not as sharp as (b$^-$); it is implied by (b$^-$) but is insufficient to yield (b$^-$) in return.

The notion of a virtual exchange equilibrium in Definition 2 does not seem to have been introduced or explored previously in economics. Beyond its potential in the theoretical understanding of equilibrium, it has natural significance for numerical work, where limits of computed sequences of approximations to a desired equilibrium may inevitably need to be contemplated anyway.

In our variational approach to equilibrium, each agent's utility maximization problem will be translated into optimality conditions involving a Lagrange multiplier. It is partly for the extra benefit accruing from such conditions, but also for enhancing the computational possibilities when given specific data, that we concentrate on utility functions (instead of abstract preference relations) and furthermore make the following restrictions. Although these restrictions could be relaxed in several ways, they will assist us here in getting some basic ideas across without too many technical complications.

Ongoing Assumptions (utility and constraint structure).

(A1) X_a *is convex and closed, with nonempty interior.*

(A2) u_a *is concave and continuously differentiable on* X_a.

(A3) u_a *does not attain a maximum on* X_a.

Because we are operating in an environment of free disposal, there is no real loss of generality in stipulating in (A1) that $X_a \neq \emptyset$; we could harmlessly replace X_a by $\hat{X}_a = X_a + \mathbb{R}^l_+$ while extending u_a to the nondecreasing utility $\hat{u}_a$ defined by $\hat{u}_a(x_a) = \sup\{u_a(\hat{x}_a) | \hat{x}_a \leq x_a\}$. The continuous differentiability in (A2) can be interpreted merely as continuous differentiability on $\operatorname{int} X_a$ with the mapping ∇u_a having a continuous extension from $\operatorname{int} X_a$ to the boundary of X_a.

Definition 3 (utility scaling). *By an equilibrium with* utility scaling *will be meant an equilibrium in the sense of Definition 1 or Definition 2 in which condition* (b) *is replaced by the existence of a coefficient* $\overline{\lambda}_a$, *called a* utility scale factor *for agent* a, *such that*

$(\mathrm{b}^*)\ \overline{x}_a \in \operatorname{argmax}\{u_a(x_a) - \overline{\lambda}_a \overline{p} \cdot (x_a - e_a) \mid x_a \in X_a\}$ *with*

$$\overline{\lambda}_a \in [0,\infty) \quad \textit{and} \quad \overline{p} \cdot (\overline{x}_a - e_a) \begin{cases} \leq 0 \ \textit{if} \ \overline{\lambda}_a = 0, \\ = 0 \ \textit{if} \ \overline{\lambda}_a > 0, \end{cases}$$

and, in Definition 2, this also to the sequence of approximate equilibria.

Proposition 1 (status of utility scale factors). *Condition* (b^*) *implies condition* (b) *always. Thus, an exchange equilibrium with utility scaling (whether classical or two-tier) in the sense of Definition 3 always entails the*

corresponding equilibrium in Definition 1 or Definition 2. Conversely, (b) *implies* (b^*) *in particular when there exists* $x_a \in X_a$ *such that* $\overline{p} \cdot x_a > \overline{p} \cdot e_a$.

Proof. In fact, (b^*) gives the Kuhn-Tucker conditions for the maximization problem in (b), inasmuch as X_a is convex by (A1) and u_a is concave by (A2). These conditions are always sufficient for optimality, and they are necessary under a Slater assumption, which by virtue of (A3) comes out here as the existence of an $x_a \in X_a$ satisfying the budget constraint strictly. □

The point is that (b^*) is, in general, an enhancement of (b), so that in establishing the existence of an equilibrium with utility scaling, we will be accomplishing more than just proving the existence of a equilibrium by itself.

Proposition 2 (positivity of utility scale factors). *Because of* (A3), *condition* (b^*) *can only hold with* $\overline{\lambda}_a > 0$ *and* $\overline{p} \cdot (\overline{x}_a - e_a) = 0$.

Proof. If we had $\lambda_a = 0$ in (b^*), the maximum of u_a over X_a would be attained at $\overline{x}_a$, in contradiction to (A3). □

The reason for calling $\overline{\lambda}_a$ a utility scale factor is that it acts as a coefficient for converting the price $\overline{p}_j$ for a good j into to a price $\overline{\lambda}_a \overline{p}_j$ measured in the utility units of agent a. According to (b^*), once such utility prices are available they can be brought into play by maximizing $u_a(x_a) - \overline{\lambda}_a \overline{p}_a \cdot (x_a - e_a)$ instead of $u_a(x_a)$, with the original budget constraint pushed into the background. This alternative maximization converts the cost $\overline{p} \cdot (x_a - e_a)$ of passing from e_a to x_a into an adjustment of the utility associated with x_a, as compared to e_a.

If u_a were strictly concave, the maximization in (b^*) would by itself determine $\overline{x}_a$ uniquely, and the budget constraint would therefore turn out to be satisfied automatically. Even when the maximization in (b^*) doesn't determine $\overline{x}_a$ uniquely, however, the budget constraint is not invoked directly in this maximization and is only needed, if at all, in the aftermath, for the purpose of eliminating some of the vectors in the argmin set.

Theorem 1 (existence of virtual equilibrium). *A two-tier exchange equilibrium that is a virtual exchange equilibrium with utility scaling is sure to exist under the following assumptions on the initial endowments:*

(S1) *for every agent* a *there is a vector* $x_a \in X_a$ *such that* $x_a \leq e_a$,

(S2) *there are vectors* $x_a \in X_a$ *such that* $\sum_{a\in\mathcal{A}} x_a < \sum_{a\in\mathcal{A}} e_a$.

The proof of Theorem 1 will come later and, in a major respect, it will be "constructive" (as elaborated in Theorem 3). In contrast to Theorem 1, the existence result of Debreu [4] for this sort of model, although posed in a somewhat broader setting, was not constructive and didn't provide utility scaling. It didn't confirm the presence of at least one optimizing agent or yield the approximation property that distinguishes a virtual equilibrium.

Of course, any agent a for which there exists $x_a \in X_a$ such that $x_a < e_a$ must in particular be an optimizing agent, since this strict vector inequality precludes (b^-). Other, more subtle criteria for an agent to be optimizing are known as well; cf. [7], [8], and their references. In combination with Theorem 1, such criteria immediately lead to conclusions about the existence of a *classical* equilibrium in our setting. We omit the details, because our interest centers on the proof of Theorem 1 by way of a variational inequality formulation having computational potential.

Nonetheless, it's worth noting that both of our survival assumptions (S1) and (S2) automatically do hold when every agent a has some $x_a \in X_a$ with $x_a < e_a$ (which amounts to the main case treated in [1] by Arrow and Debreu).

3. VARIATIONAL REPRESENTATION

The variational inequality representation of an equilibrium with utility scaling will now be set up. In general in a space $\mathbb{R}^L$ of vectors v, the variational inequality problem VI(C,F) associated with a nonempty, convex set $C \subset \mathbb{R}^L$ and a mapping $F: C \to \mathbb{R}^L$ consists of finding

$$\overline{v} \in C \text{ such that } -F(\overline{v}) \in N_C(\overline{v}),$$

where $N_C(\overline{v})$ is the normal cone to C at $\overline{v}$:

$$w \in N_C(\overline{v}) \Longleftrightarrow w \cdot (v - \overline{v}) \le 0 \text{ for all } v \in C.$$

It's well known that if C is compact and F is continuous, a solution $\overline{v}$ to problem VI(C,F) exists.

In our formulation of equilibrium, the variational inequality we set up will have C closed and F continuous, but C unbounded, so this criterion for the existence of a solution to VI(C,F) will not be applicable directly. That will oblige us to introduce truncations to create compactness. Such truncations will be construed as corresponding to penalty formulations of the

budget constraints in the agent's maximization problems. In obtaining an equilibrium through an iterative process of truncation, we will employ an argument crucially based on epi-convergence, which is a concept of variational analysis associated with the convergence of solutions when optimization problems are approximated.

Theorem 2 (variational inequality format for classical equilibrium). *A classical exchange equilibrium with utility scaling is furnished by* $\overline{p}$, $\{\overline{x}_a\}_{a\in\mathcal{A}}$ *and* $\{\overline{\lambda}_a\}_{a\in\mathcal{A}}$, *if and only if the variational inequality* VI(C,F) *in the form*

$$-F(\overline{p};...,\overline{x}_a,...;...,\overline{\lambda}_a,...)\in N_C(\overline{p};...,\overline{x}_a,...;...,\overline{\lambda}_a,...) \tag{2}$$

holds for the nonempty, closed, convex set $C\subset\mathbb{R}^l\times[\Pi_{a\in\mathcal{A}}\mathbb{R}^l]\times[\Pi_{a\in\mathcal{A}}\mathbb{R}]$ *defined by*

$$C=P\times[\Pi_{a\in\mathcal{A}}X_a]\times[\Pi_{a\in\mathcal{A}}[0,\infty)] \tag{3}$$

and the continuous mapping $F:C\to\mathbb{R}^l\times[\Pi_{a\in\mathcal{A}}\mathbb{R}^l]\times[\Pi_{a\in\mathcal{A}}\mathbb{R}]$ *defined by*

$$\begin{aligned}&F(p;...,x_a,...;...,\lambda_a,...)\\&=\left(\Sigma_{a\in\mathcal{A}}[e_a-x_a];...,\lambda_a p-\nabla u_a(x_a),...;...,p\cdot[e_a-x_a],...\right)\end{aligned} \tag{4}$$

Proof. The closedness and convexity claimed for C and the continuity claimed for F are evident from (A1) and (A2). The variational inequality in question decomposes into the conditions

$$\begin{aligned}&\Sigma_{a\in\mathcal{A}}[\overline{x}_a-e_a]\in N_P(\overline{p}),\\&\nabla u_a(\overline{x}_a)-\overline{\lambda}_a\overline{p}\in N_{X_a}(\overline{x}_a)\text{ for all }a\in\mathcal{A},\\&\overline{p}\cdot(\overline{x}_a-e_a)\in N_{[0,\infty)}(\overline{\lambda}_a)\text{ for all }a\in\mathcal{A}.\end{aligned} \tag{5}$$

The second condition means that the function $x_a\mapsto u_a(x_a)-\overline{\lambda}_a\overline{p}\cdot(x_a-e_a)$, which is concave, has its maximum over X_a at $\overline{x}_a$, whereas the third condition refers to the complementarity relations

$$\overline{p}\cdot(\overline{x}_a-e_a)\le 0,\qquad \overline{\lambda}_a\ge 0,\qquad \overline{\lambda}_a\overline{p}\cdot(\overline{x}_a-e_a)=0. \tag{6}$$

Those two conditions together, therefore, are equivalent to (b*) holding for every $a\in A$.

The first condition in (5) is not, on the surface, the same as the market condition (a) in Definition 1, which in principle would be stronger. We will see, however, that in the presence of the other conditions, the first condition in (5) implies (a). In terms of

$$\zeta = \max_{j=1,\ldots,J} \left\{ \Sigma_{a \in \mathcal{A}} [\overline{x}_{aj} - e_{aj}] \right\}, \tag{7}$$

the first condition in (5) says that

$$\overline{p}_j = 0 \text{ unless } \Sigma_{a \in \mathcal{A}} [\overline{x}_{aj} - e_{aj}] = \zeta, \tag{8}$$

so that in particular

$$\overline{p} \cdot \Sigma_{a \in \mathcal{A}} [\overline{x}_a - e_a] = \zeta. \tag{9}$$

Now we bring in Proposition 2: we actually must have $\overline{p} \cdot [\overline{x}_a - e_a] = 0$ for all $a \in \mathcal{A}$. Then (9) implies $\zeta = 0$, hence $\Sigma_{a \in \mathcal{A}} [\overline{x}_{aj} - e_{aj}] \leq 0$ for every j in (7), and we are able to conclude through (8) that (a) holds. □

Although the unboundedness of the set C in our variational inequality representation of classical equilibrium is an unavoidable consequence of the multiplier conditions we have introduced, a partial kind of boundedness, at least, can be achieved under our assumptions by a trouble-free truncation of the survival sets X_a.

Proposition 3 (underlying boundedness of consumption). *Under* (S1) *and* (S2), *there exist bounded subsets* $X_a^b \subset X_a$ *still satisfying these assumptions and such that a classical exchange equilibrium with utility scaling is furnished for* $\{X_a^b\}_{a \in \mathcal{A}}$ *by* $\overline{p}$, $\{\overline{x}_a\}_{a \in \mathcal{A}}$ *and* $\{\lambda_a\}_{a \in \mathcal{A}}$, *if and only if these elements give such an equilibrium for* $\{X_a\}_{a \in \mathcal{A}}$. *Specifically, this is true when*

$$X_a^b = \{x_a \in X_a \mid x_a \leq b\} \text{ for any } b > \sum_{a \in \mathcal{A}} e_a, \tag{10}$$

in which case there definitely exist elements $x_a \in X_a^b$ *satisfying* $x_a < b$, *whereas any elements* $x_a \in X_a^b$ *satisfying* $\sum_{a \in \mathcal{A}} x_a \leq \sum_{a \in \mathcal{A}} e_a$ *must satisfy* $x_a < b$.

Proof. Take b and X_a^b as in (10). Clearly X_a^b is still convex and closed, but also bounded, and (S2) is preserved. Since $e_a \geq 0$ for every $a \in \mathcal{A}$, the strict inequality in (10) implies that $e_a < b$ for every $a \in \mathcal{A}$. The condition that $e_a \in X_a$, from (S1), thus carries over to having $e_a \in X_a^b$ and in particular informs us, by taking $x_a = e_a$, that there exists $x_a \in X_a^b$ satisfying $x_a < b$. Indeed, in the background of e_a belonging to $\{x_a | x_a < b\} \cap X_a$, we get from (A1) that $\operatorname{int} X_a^b = \{x_a | x_a < b\} \cap \operatorname{int} X_a \neq \emptyset$ (cf. [12,Theorem 6.5]) and can conclude that (A1) holds for X_a^b. Trivially, (A2) persists when X_a is replaced by the truncation X_a^b.

Because $X_a \subset \mathbb{R}_+^l$, the conditions $x_a \in X_a$ and $\sum_{a \in \mathcal{A}} x_a \leq \sum_{a \in \mathcal{A}} e_a$ in the definition of an equilibrium imply $x_a < b$. Hence any equilibrium with respect to the sets X_a is an equilibrium with respect to the sets X_a^b, and conversely as well, the constraints $x_a \leq b$ necessarily being inactive in either case. □

According to this observation, we can replace the sets X_a by bounded sets X_a^b in the formulation of the variational inequality in Theorem 2 without undermining the equivalence with the desired equilibrium. This still leaves the unboundedness caused by the multiplier conditions, however. To handle that, our approach is to truncate the interval $[0,\infty)$ to $[0,r]$ for a value $r > 0$, which will turn out to act as a penalty parameter.

Proposition 4 (truncated variational inequality). *Consider the variational inequality* $\mathrm{VI}(C_r^b, F)$ *for the same mapping* F *as in Theorem 2 but with the set* C *there replaced for* $r > 0$ *by*

$$C_r^b = P \times [\Pi_{a \in \mathcal{A}} X_a^b] \times [\Pi_{a \in \mathcal{A}} [0,r]],$$

the sets X_a^b *being defined as in Proposition 3. Then* C_r^b *is nonempty, closed and convex, but also bounded, and a solution to* $\mathrm{VI}(C_r^b, F)$ *therefore exists. A solution to* $\mathrm{VI}(C_r^b, F)$ *is comprised of a relative price vector* $\overline{p}$ *along with* $\{\overline{x}_a\}_{a \in \mathcal{A}}$ *and* $\{\overline{\lambda}_a\}_{a \in \mathcal{A}}$ *for which there is a value* $\zeta \in \mathbb{R}$ *such that*

(a_r) $\Sigma_{a \in \mathcal{A}} \overline{x}_{aj} \leq \Sigma_{a \in \mathcal{A}} e_{aj} + \zeta$ *for all goods* j, *with equality when* $\overline{p}_j > 0$,

(b_r^+) $\overline{x}_a \in \operatorname{argmax}\{u_a(x_a) - \overline{\lambda}_a \overline{p} \cdot (x_a - e_a) | x_a \in X_a^b\}$, *with*

$$\overline{\lambda}_a \in [0,r] \text{ and } \overline{p}\cdot(\overline{x}_a - e_a)\begin{cases} \le 0 \text{ if } \overline{\lambda}_a = 0, \\ = 0 \text{ if } 0 < \overline{\lambda}_a < r, \\ \ge 0 \text{ if } \overline{\lambda}_a = r. \end{cases}$$

Proof. The standard existence criterion for variational inequalities, invoked for the compact set C_r^b, produces $\overline{p}$, $\{\overline{x}_a\}_{a\in\mathcal{A}}$ and $\{\overline{\lambda}_a\}_{a\in\mathcal{A}}$ for which the corresponding $\overline{v} = (\overline{p};\ldots,\overline{x}_a,\ldots;\ldots,\overline{\lambda}_a,\ldots)$ solves $\mathrm{VI}(C_r^b, F)$, i.e., has $-F(\overline{v}) \in N_{C_r^b}(\overline{v})$. Adopting the pattern in the proof of Theorem 2, we decompose this variational inequality into the conditions

$$\begin{aligned} &\Sigma_{a\in\mathcal{A}}[\overline{x}_a - e_a] \in N_P(\overline{p}), \\ &\nabla u_a(\overline{x}_a) - \overline{\lambda}_a\overline{p} \in N_{X_a^b}(\overline{x}_a) \text{ for all } a \in \mathcal{A}, \\ &\overline{p}\cdot[\overline{x}_a - e_a] \in N_{[0,r]}(\overline{\lambda}_a) \quad \text{for all } a \in \mathcal{A}. \end{aligned}$$

The fact that the first of these conditions is equivalent to (a_r) was effectively argued already in the proof of Theorem 2. The second and third of these conditions is (b_r^*). □

In working with the truncated variational inequality and understanding its meaning, it will be helpful to have the notation

$$[t]_+ = \max\{0,t\} \text{ for } t \in \mathbb{R}.$$

We use it to set up a linear penalty approximation to the budget constraint $p\cdot(x_a - e_a) \le 0$ in terms of the expression

$$r[p\cdot(x_a - e_a)]_+ = \begin{cases} 0 & \text{when } p\cdot(x_a - e_a) \le 0, \\ rp\cdot(x_a - e_a) & \text{when } p\cdot(x_a - e_a) > 0. \end{cases}$$

Proposition 5 (penalty interpretation). *Condition (b_r^*) of Proposition 4 holds with respect to $\overline{p}$ for $\overline{x}_a$ and some $\overline{\lambda}_a$ if and only if $\overline{x}_a$ satisfies*

(b_r) $\overline{x}_a \in \operatorname{argmax}\{u_a(x_a) - r[\overline{p}\cdot(x_a - e_a)]_+ \,|\, x_a \in X_a^b\}$.

Proof. The equivalence can be seen by thinking of (b_r) as referring to the minimization of $\varphi_a + \psi_a$ over $\mathbb{R}^l$, where

$$\varphi_a(x_a)=\begin{cases}-u_a(x_a) & \text{when } x_a\in X_a^b,\\ \infty & \text{when } x_a\notin X_a^b,\end{cases}\qquad \psi_a(x_a)=r[\overline{p}\cdot(x_a-e_a)]_-.$$

Here φ_a is a lower semicontinuous, proper, convex function, while ψ_a is a finite convex function on $\mathbb{R}^l$. The subgradient condition both necessary and sufficient for the minimum of $\varphi_a+\psi_a$ to occur at $\overline{x}_a$, namely $0\in\partial(\varphi_a+\psi_a)(x_a)$, comes out therefore as the existence of a subgradient $z_a\in\partial\psi_a(\overline{x}_a)$ such that $-z_a\in\partial\varphi_a(\overline{x}_a)$, where moreover (cf. [12, Theorem 23.8):

$$\partial\varphi_a(\overline{x}_a)=-\nabla u_a(\overline{a})+N_{X_a^b}(\overline{x}_a)\ .$$

The necessary and sufficient condition thus refers to the existence of $\nabla u_a(\overline{x}_a)-z_a\in N_{X_a^b}(\overline{x}_a)$.

By a basic chain rule in convex analysis (cf. [12, Theorem 23.9]), we have $z_a\in\partial\psi_a(\overline{x}_a)$ if and only if $z_a=\overline{\lambda}_a\overline{p}$ for some $\overline{\lambda}_a$ satisfying the conditions in (b_r^*). In this manner, we have $\nabla u_a(\overline{x}_a)-z_a\in N_{X_a^b}(\overline{x}_a)$ if and only if $\nabla u_a(\overline{x}_a)-\overline{\lambda}_a\overline{p}\in N_{X_a^b}(\overline{x}_a)$ for some such $\overline{\lambda}_a$, and this can be recognized as the necessary and sufficient condition for optimality in the maximization in condition (b_r). □

4. ITERATIVE SCHEME

The existence result in Theorem 1 will be derived by an iterative scheme based on the variational inequality representations of equilibrium we have been developed above. In this scheme, we replace the survival sets X_a to the bounded sets X_a^b specified in 10 and consider for $\nu=1,2,\ldots$, a sequence of penalty parameter values $r^\nu\nearrow\infty$, denoting by C^ν the set C_b^r of Proposition 4 in the case of $r=r^\nu$. For each ν we solve the variational inequality $\mathrm{VI}(C^\nu,F)$, which is possible by Proposition 4 in principle (and moreover should be approachable numerically by methods developed along the lines of those in [6], [10], [11], as mentioned in the introduction).

This way, we generate a sequence of price vectors $p^\nu\in P$ together with sequences of consumption vectors $x_a^\nu\in X_a^b$, multipliers λ_a^ν and values ζ^ν satisfying

(a_{r^ν}) $\Sigma_{a\in\mathcal{A}} x_{aj}^\nu \le \Sigma_{a\in\mathcal{A}} e_{aj} + \zeta^\nu$ for all goods j, with equality when $p_j^\nu > 0$,

$(b_{r^\nu}^*)$ $x_a^\nu \in \operatorname{argmax}\{u_a(x_a) - \lambda_a^\nu p^\nu \cdot (x_a - e_a) | x_a \in X_a^b\}$ with

$$\lambda_a^\nu \in [0, r^\nu] \text{ and } p^\nu \cdot (x_a^\nu - e_a) \begin{cases} \le 0 & \text{if } \lambda_a^\nu = 0, \\ = 0 & \text{if } 0 < \lambda_a^\nu < r^\nu, \\ \ge 0 & \text{if } \lambda_a^\nu = r^\nu. \end{cases}$$

Note that because the components p_j^ν of p^ν are nonnegative, but not all zero, condition (a_{r^ν}) means that

$$\zeta^\nu = \max_{j=1,\ldots,l} \sum_{a\in\mathcal{A}} [x_{aj}^\nu - e_{aj}]. \tag{11}$$

Condition $(b_{r^\nu}^*)$, on the other hand, can be interpreted through Proposition 5 as the condition

(b_{r^ν}) $x_a^\nu \in \operatorname{argmax}\{u_a(x_a) - r^\nu [p^\nu \cdot (x_a - e_a)], | x_a \in X_a^b\}$,

which relaxes the budget constraint $p^\nu \cdot (x_a - e_a) \le 0$ by allowing it to be exceeded at a penalty rate which is increased in each iteration.

Theorem 3 (limits in the iterative scheme). *Once r^ν is higher than a certain threshold value, p^ν and x_a^ν furnish a classical equilibrium, with utility scaling, respect to the same sets X_a and functions u_a but possibily different endowment vectors $e_a^\nu \ge e_a$ with $e_a^\nu \to e_a$. The sequence of these nearby classical equilibria $(p^\nu, \{x_a^\nu\}_{a\in\mathcal{A}})$ is bounded, and every cluster point $(\overline{p}, \{\overline{x}_a\}_{a\in\mathcal{A}})$ furnishes a virtual equilibrium for the original data. Hence if only one virtual equilibrium exists, the entire sequence must converge to it.*

Obviously, in proving Theorem 3 we will have proved Theorem 1, so we can concentrate on Theorem 3. Since the sets P and X_a^b containing p^ν and x_a^ν are closed and bounded, the sequences of vectors p^ν and x_a^ν are bounded, as claimed in Theorem 3, and cluster points do exist. Note, however, that no such claim is made about the sequences of multipliers λ_a^ν. The possible unboundedness of such a sequence is exactly what can lead to

an agent a being only a barely surviving agent. The following facts will be crucial, in view of the definition of a virtual equilibrium,

Proposition 6 (convergence estimates). *For each* $a \in \mathcal{A}$, *choose any* $\hat{x}_a \in X_a$ *with* $\hat{x}_a \leq e_a$, *as exists by (S1), and let* $\mu_a^b = \max\{u_a(x_a) | x_a \in X_a^b\}$. *Then* $\mu_a^b \geq u_a(\hat{x}_a)$, *and one has*

$$p^\nu \cdot (x_a^\nu - e_a) \leq \frac{\mu_a^b - u_a(\hat{x}_a)}{r^\nu} \text{ for all } a \in \mathcal{A}, \tag{12}$$

This implies that

$$x_a^\nu < b \text{ for all } a \in \mathcal{A} \text{ when } r^\nu \text{ is sufficiently large,} \tag{13}$$

as is true specifically when

$$r^\nu \geq \hat{r} \quad \text{for} \quad \hat{r} = \frac{N}{\beta} \max_{a \in \mathcal{A}} \{ \mu_a^b - u_a(\hat{x}_a) \}, \tag{14}$$

where N *is the number of agents* $a \in \mathcal{A}$ *and* β *is any positive number small enough that* $\sum_{a \in \mathcal{A}} e_{aj} \leq b_j - \beta$ *for every good* j. *Thereafter, one will have*

$$p^\nu \cdot (x_a^\nu - e_a) \geq 0 \text{ for all } a \in \mathcal{A}, \tag{15}$$

and the vectors p^ν *and* x_a^ν *will furnish a classical equilibrium with respect to the sets* X_a, *functions* u_a, *and the endowment vectors* e_a^ν *defined by*

$$e_{aj}^\nu = e_{aj} + \zeta_a^\nu \text{ with } \zeta_a^\nu = p^\nu \cdot (x_a^\nu - e_a), \tag{16}$$

in which the multipliers λ_a^ν *are positive and serve as utility scale factors. Thus, one will have*

(a^ν) $\Sigma_{a \in \mathcal{A}} x_{aj}^\nu \leq \Sigma_{a \in \mathcal{A}} e_{aj}^\nu$ *for all* j, *with equality holding if* $p_j^\nu > 0$,

($b^{\nu+}$) $x_a^\nu \in \operatorname{argmax}\{u_a(x_a) - \lambda_a^\nu p^\nu \cdot (x_a - e_a^\nu) \mid x_a \in X_a\}$ *with*

$$\lambda_a^\nu > 0, \qquad p^\nu \cdot (x_a^\nu - e_a^\nu) = 0.$$

Proof. Because $\hat{x}_a \le e_a$, we have $\hat{x}_a \in X_a^b$ with $p^\nu \cdot (\hat{x}_a - e_a) \le 0$. Hence $\mu_a^b \ge 0$ and

$$\begin{aligned} u_a(\hat{x}_a) &= u_a(\hat{x}_a) - r^\nu [p^\nu \cdot (\hat{x}_a - e_a]_+ \\ &\le u_a(x_a^\nu) - r^\nu [p^\nu \cdot (x_a^\nu - e_a]_+ \le \mu_a^b - r^\nu [p^\nu \cdot (x_a^\nu - e_a]_+, \end{aligned}$$

so that $r^\nu [p^\nu \cdot (x_a^\nu - e_a]_+ \le \mu_a^b - u_a(\hat{x}_a)$. This inequality guarantees (12). From (a_{r^ν}) we have $\sum_{a\in\mathcal{A}} p^\nu \cdot (x_a^\nu - e_a) = p^\nu \cdot \sum_{a\in\mathcal{A}} (x_a^\nu - e_a) = \zeta^\nu$ with ζ^ν expressed by (11), and therefore $\sum_{a\in\mathcal{A}} [x_{aj}^\nu - e_{aj}] \le \zeta^\nu$. It follows that $\sum_{a\in\mathcal{A}} x_{aj}^\nu \le \sum_{a\in\mathcal{A}} e_{aj} + \beta$ when $\zeta^\nu \le \beta$, and in particular

$$x_{aj}^\nu \le b_j - \beta \text{ when } \zeta^\nu \le \beta \text{ with } \beta \le b_j - \sum_{a\in\mathcal{A}} e_{aj}. \tag{17}$$

Thus, $x_a^\nu < b$ as claimed in (13) when r^ν is beyond the value $\hat{r}$ in (14).

Once we have $x_a^\nu < b$, the maximum over X_a^b in condition $(\mathrm{b}^+_{r^\nu})$ is the same as the maximum over X_a, due to convexity. Then necessarily $\lambda_a^\nu > 0$, since otherwise our nonsatiation assumption (A3) would be violated. In $(\mathrm{b}^+_{r^\nu})$ we then have $p^\nu \cdot (x_a^\nu - e_a) \ge 0$ for all $a \in \mathcal{A}$. In that case, by taking $\zeta^\nu = p^\nu \cdot (x_a^\nu - e_a)$ and defining e_a^ν as indicated, we get $e_a^\nu \ge e_a$ and $p^\nu \cdot (x_a^\nu - e_a^\nu) = 0$, so that conditions (a_{r^ν}) and $(\mathrm{b}^+_{r^\nu})$ have been converted to (a^ν) and $(\mathrm{b}^{\nu+})$. That implies by Proposition 1 that the elements p^ν, x_a^ν and λ_a^ν furnish a classical equilbrium with respect to the endowments e_a^ν. □

These estimates immediately reveal key properties of our iterative scheme. As $r^\nu \to \infty$, we eventually have (14) and, in the augmentation rule in (16),

$$0 \le \zeta_a^\nu \le \frac{\mu_a^b - u_a(\hat{x}_a)}{r^\nu} \to 0, \text{ so that } e_a^\nu \to e_a.$$

By taking limits of in (a^ν) and $(\mathrm{b}^{\nu+})$, we see then that cluster points $\bar{p}$ and $\bar{x}_a$ must satisfy the market clearing condition (a) and the budget condition $\bar{p} \cdot (\bar{x}_a - e_a) = 0$. The extent to which they satisfy (b^+) or (b^-), however, remains to be established.

The key to further analysis lies in the utility scale factors λ_a^ν. For simplicity of notation in this analysis, we can suppose we have passed to subsequences so that actually $p^\nu \to \overline{p}$ and $x_a^\nu \to \overline{x}_a$ for every agent $a \in \mathcal{A}$, and that ($b^{\nu+}$) and (15) hold for all ν, furthermore with $\overline{x}_a < b$, as comes out of the uniformity of the bound derived in (17). We look at

$$\mathcal{A}_+ = \{\text{agents } a \in \mathcal{A} \text{ such that } \{\lambda_a^\nu\}_{\nu=1}^\infty \text{ is bounded}\},$$
$$\mathcal{A}_- = \{\text{agents } a \in \mathcal{A} \text{ such that } \{\lambda_a^\nu\}_{\nu=1}^\infty \text{ is unbounded}\},$$

By a further reduction to subsequences if necessary, we can arrange that

$$\begin{cases} \text{for each } a \in \mathcal{A}_+, \text{actually } \lambda_a^\nu \to \overline{\lambda}_a \geq 0, \\ \text{for each } a \in \mathcal{A}_-, \text{actually} \lambda_a^\nu \to \infty. \end{cases}$$

Consider now an agent $a \in \mathcal{A}_+$. Define the functions φ_a^ν and φ_a on the entire space $\mathbb{R}^l$ by

$$\varphi_a^\nu(x_a) = \begin{cases} -u_a(x_a) + \lambda_a^\nu p^\nu \cdot (x_a - e_a^\nu) & \text{if } x_a \in X_a^b, \\ \infty & \text{if } x_a \notin X_a^b, \end{cases}$$
$$\varphi_a(x_a) = \begin{cases} -u_a(x_a) + \overline{\lambda}_a \overline{p} \cdot (x_a - e_a) & \text{if } x_a \in X_a^b, \\ \infty & \text{if } x_a \notin X_a^b, \end{cases}$$

these functions being convex and lower semicontinuous by virtue of (A1) and (A2). Conditions ($b^{\nu+}$) and (b^+) correspond respectively to

$$x_a^\nu \in \operatorname*{argmin}_{x_a \in R^l} \varphi_a^\nu(x_a), \qquad \overline{x}_a \in \operatorname*{argmin}_{x_a \in R^l} \varphi_a(x_a), \tag{18}$$

inasmuch as $x_a^\nu < b$ and $\overline{x}_a < b$, along with $\lambda_a^\nu > 0$ and $p^\nu \cdot (x_a^\nu - e_a^\nu) = 0$, as well as $\overline{\lambda} \geq 0$ and $\overline{p} \cdot (\overline{x}_a - e_a) = 0$. Therefore, if we can show that the second condition in (18) follows in the limit from the first condition as $\nu \to \infty$, we will be able to conclude that $\overline{p}$, $\overline{x}_a$ and $\overline{\lambda}_a$ satisfy (b^+) and thus that agent a is an optimizing agent.

This is an issue addressed, in general, by the theory of "epi-convergence" of sequences of functions and its role in minimization, as expounded for instance in [14, Chapter 7]. Here, the circumstances are especially simple because the functions are convex and all have the same effective domain,

namely X_a^b, which moreover has nonempty interior. As $\nu \to \infty$, we have $\varphi_a^\nu(x_a) \to \varphi_a(x_a)$ for each $x_a \in X_a^b$, and that guarantees the epi-convergence of φ_a^ν to φ_a by [14, Theorem 7.17]. Then by [14, Theorem 7.33], because these functions are lower semicontinuous with their effective domains uniformly bounded, the first condition in (18) yields the second, as required.

Next, consider instead an agent $x \in \mathcal{A}_-$. Define the functions ψ_a^ν and ψ_a on the entire space $\mathbb{R}^l$ by

$$\psi_a^\nu(x_a) = \begin{cases} -(1/\lambda_a^\nu)u_a(x_a) + p^\nu \cdot (x_a - e_a^\nu) & \text{if } x_a \in X_a^b, \\ \infty & \text{if } x_a \notin X_a^b, \end{cases}$$

$$\psi_a(x_a) = \begin{cases} \overline{p} \cdot (x_a - e_a) & \text{if } x_a \in X_a^b, \\ \infty & \text{if } x_a \notin X_a^b. \end{cases}$$

Again, these functions are convex and lower semicontinuous by virtue of (A1) and (A2), so ψ_a^ν epi-converges to ψ_a for the reasons already mentioned, coming from [14, Theorem 7.17]. On the basis of ($b^{\nu+}$), we have $x_a^\nu \in \operatorname{argmin} \psi_a^\nu$, and can conclude through [14, Theorem 7.33] that $\overline{x}_a \in \operatorname{argmin} \psi_a$. That tells us that $\overline{x}_a$ minimizes $\overline{p} \cdot x_a$ subject to $x_a \in X_a^b$, and since $\overline{x}_a < b$, it establishes that (b^-) holds. Thus, agent a is a barely surviving agent.

Finally, we confirm that the agents can't all be just barely surviving. If indeed $\mathcal{A}_- = \mathcal{A}$, we would have

$$\inf\{\overline{p} \cdot \textstyle\sum_{a \in \mathcal{A}} x_a \mid x_a \in X_a\} = \overline{p} \cdot \textstyle\sum_{a \in \mathcal{A}} e_a.$$

But that's incompatible with our assumption (S2), inasmuch as $\overline{p} \neq 0$.

In summary, we have demonstrated that $\overline{p}$ and $\{\overline{x}_a\}_{a \in \mathcal{A}}$ provide a virtual equilibrium as in Definition 2, in which moreover the agents $a \in \mathcal{A}_-$ are barely surviving, whereas the agents $a \in \mathcal{A}_+$ are optimizing and have the limits $\overline{\lambda}_a$ as utility scale factors. □

5. EXAMPLES

Illustrations will now be provided of the distinctions between the various equilibrium concepts in Definitions 1 and 2 and how they relate to the

existence result in Theorem 1 and the iterative scheme addressed in Theorem 3.

In these examples, we have just two goods and two agents: here $l=2$ and $\mathcal{A}=\{1,2\}$. Price vectors have the form $p=(p_1,p_2)$ with $p_1\geq 0$, $p_2\geq 0$ and $p_1+p_2=1$. Agent $a=1$ has an endowment vector $e_1=(e_{11},e_{12})$ and chooses a consumption vector $x_1=(x_{11},x_{12})$ with utility $u_1(x_{11},x_{12})$ from a survival set $X_1\subset\mathbb{R}^2_+$, whereas agent $a=2$ has an endowment vector $e_2=(e_{21},e_{22})$ and chooses a consumption vector $x_2=(x_{21},x_{22})$ with utility $u_2(x_{21},x_{22})$ from a survival set $X_2\subset\mathbb{R}^2_+$.

Example 1 (a classical equilibrium without strict feasibility). *Let* $X_1=\mathbb{R}^2_+$ *and* $X_2=\mathbb{R}^2_+$, *and take*

$$\begin{cases} e_1=(1,1),\ u_1(x_{11},x_{12})=x_{11}, \\ e_2=(1,0),\ u_2(x_{21},x_{22})=x_{21}+x_{22}. \end{cases}$$

In this case there is an $x_1\in X_1$ *with* $x_1<e_1$, *but no* $x_2\in X_2$ *with* $x_2<e_2$. *Nonetheless, a classical equilibrium exists, given by*

$$\overline{p}=(1/2,1/2),\quad \overline{x}_1=(2,0),\quad \overline{x}_2=(0,1).$$

There is no other equilibrium, even two-tier. The iterative scheme, applied to this data, would necessarily converge to the unique classical equilibrium.

Detail. Here $p_2=1-p_1$, so $p=(p_1,1-p_1)$ with $0\leq p_1\leq 1$. For agent $a=1$ the utility maximizing set is

$$\begin{aligned} M_1 &= \operatorname{argmax}\{u_1(x_1)\,|\,x_1\in X_1, p\cdot x_1\leq p\cdot e_1\} \\ &= \operatorname{argmax}\{x_{11}|x_{11}\geq 0, x_{12}\geq 0, \\ &\qquad\qquad p_1x_{11}+(1-p_1)x_{12}\leq 1\} \\ &= \begin{cases} \varnothing & \text{if } p_1=0, \\ \{(p_1^{-1},0)\} & \text{if } p_1>0, \end{cases} \end{aligned}$$

whereas for agent $a=2$ the utility maximizing set is

$$
\begin{aligned}
M_2 &= \operatorname{argmax}\{u_2(x_2) \mid x_2 \in X_2, p \cdot x_2 \le p \cdot e_2\} \\
&= \operatorname{argmax}\{x_{21} + x_{22} \mid x_{21} \ge 0, x_{22} \ge 0 \\
&\qquad\qquad p_1 x_{21} + (1 - p_1) x_{22} \le p_1\} \\
&= \begin{cases} \varnothing & \text{if } p_1 = 0, \\ \{(1,0)\} & \text{if } 0 < p_1 < 1/2, \\ \{(\tau, 1-\tau) \mid 0 \le \tau \le 1\} & \text{if } p = 1/2, \\ \{(0,(1-p_1)^{-1})\} & \text{if } 1/2 < p_1 < 1, \\ \varnothing & \text{if } p_1 = 1. \end{cases}
\end{aligned}
$$

The total endowment $e_1 + e_2$ is $(2,1)$, so the condition for market clearing is

$$
\begin{cases} x_{11} + x_{21} \le 2, & \text{with equality if } p_1 > 0, \\ x_{12} + x_{22} \le 1, & \text{with equality if } p_1 < 1. \end{cases} \tag{19}
$$

Having $p_1 = 0$ or $p_1 = 1$ in a classical equilibrium is excluded by the emptiness then of M_2, so any candidates would have to have $0 < p_1 < 1$ and obey both of the inequalities in (19) as equations. In choosing (x_{11}, x_{12}) from M_1 and (x_{21}, x_{22}) from M_2, it's impossible to get the second of these equations satisfied when $0 < p_1 < 1/2$, or to get the first satisfied when $1/2 < p_1 < 1$. Hence the only available candidate is $p_1 = 1/2$. And indeed, for $\overline{p} = (1/2, 1/2)$ we can take $\overline{x}_1 = (2,0)$ from M_1 and and $\overline{x}_2 = (0,1)$ from M_2 and have $\overline{x}_1 + \overline{x}_2 = (2,1)$, as required for a classical equilibrium.

This is the only possibility for a classical equilibrium, but what about a two-tier equilibrium more generally? The investigation of that requires us to look at the set of cheapest consumption vectors, which here happens to be the same for both agents:

$$
\begin{aligned}
M_{\_} &= \operatorname{argmin}\{p \cdot x_1 \mid x_1 \in X_1\} = \operatorname{argmin}\{p \cdot x_2 \mid x_2 \in X_2\} \\
&= \begin{cases} \{(\tau, 0) \mid \tau \ge 0\} & \text{if } p_1 = 0, \\ \{(0,0)\} & \text{if } 0 < p_1 < 1, \\ \{(0,\tau) \mid \tau \ge 0\} & \text{if } p_1 = 1. \end{cases}
\end{aligned}
$$

In a two-tier equilibrium with both agents barely surviving, both (x_{11}, x_{12}) and (x_{21}, x_{22}) would be selected from $M_{\_}$. Thus, both would have 0 in the first component if $p_1 > 0$, or both would have 0 in the second

component if $p_1 < 1$, which would be inconsistent with (19), no matter how p_1 is selected.

For a two-tier equilibrium with agent $a = 1$ barely surviving and agent $a = 2$ optimizing, we would need to satisfy (19) with a choice of $(x_{11}, x_{12}) \in M_-$ and $(x_{21}, x_{22}) \in M_2$. Again, the cases $p_1 = 0$ and $p_1 = 1$ are excluded by the emptiness of M_2 for those values, but on the other hand, when $0 < p_1 < 1$ we are forced to take $(x_{11}, x_{12}) = (0,0)$, and yet both of the conditions in (19) are required to be fulfilled as equations. But there is no way to choose p_1 to get $(x_{21}, x_{22}) \in M_2$ with $(x_{21}, x_{22}) = (2,1)$.

For a two-tier equilibrium with agent $a = 1$ optimizing and agent $a = 2$ barely surviving, we would need (19) to hold for some $(x_{11}, x_{12}) \in M_1$ and $(x_{21}, x_{22}) \in M_-$. Because $M_1 = \varnothing$ when $p_1 = 0$, we are limited to $0 < p_1 \le 1$ and $(x_{11}, x_{12}) = (p_1^{-1}, 0)$, with at least the first condition in (19) holding as an equation. Since x_{21} has to be 0 when $p_1 > 0$, we can only get this equation with $p_1 = 1/2$, but then the second condition in (19) must hold as an equation too, even though x_{22} has to be 0. Thus, this mode of equilibrium is impossible as well. □

Example 2 (a nonclassical virtual equilibrium along with other equilibria). *Let* $X_1 = \mathbb{R}^2_+$ *and* $X_2 = \mathbb{R}^2_+$ *and take*

$$\begin{cases} e_1 = (1,1), & u_1(x_{11}, x_{12}) = x_{11}, \\ e_2 = (0,1), & u_2(x_{21}, x_{22}) = x_{21} + x_{22}. \end{cases}$$

In this case there is no classical equilibrium, but two-tier equilibria in which agent $a = 1$ *is optimizing and agent* $a=2$ *is barely surviving are furnished by*

$$\overline{p} = (1,0), \quad \overline{x}_1 = (1,0), \quad \overline{x}_2 = (0,\theta), \quad \textit{for any } \theta \in [0,2]. \tag{20}$$

These are the only two-tier equilibria, and among them, only the one for $\theta = 2$ *is a virtual equilibrium. That unique virtual equilibrium, with utility scaling, must be the limit of any sequence of vectors* p^ν *and* x_a^ν *generated by the iterative scheme.*

Detail. This is close in many respects to Example 1, having the same sets M_1 and M_- and only a coordinate-switched version of M_2, namely

$$
\begin{aligned}
M'_2 &= \operatorname{argmax}\{u_2(x_2) \mid x_2 \in X_2, p \cdot x_2 \le p \cdot e_2\} \\
&= \operatorname{argmax}\{x_{21} + x_{22} \mid x_{21} \ge 0, x_{22} \ge 0, \\
&\qquad\qquad p_1 x_{21} + (1-p_1)x_{22} \le 1 - p_1\} \\
&= \begin{cases}
\varnothing & \text{if } p_1 = 0, \\
\{((p_1^{-1} - 1, 0)\} & \text{if } 0 < p_1 < 1/2, \\
\{(\tau, 1-\tau) \mid 0 \le \tau \le 1\} & \text{if } p = 1/2, \\
\{(0,1)\} & \text{if } 1/2 < p_1 < 1, \\
\varnothing & \text{if } p_1 = 1.
\end{cases}
\end{aligned}
$$

The total endowment $e_1 + e_2$ is $(1,2)$, so now the condition for market clearing takes the form

$$
\begin{cases}
x_{11} + x_{21} \le 1, & \text{with equality if } p_1 > 0, \\
x_{11} + x_{21} \le 1, & \text{with equality if } p_1 < 1.
\end{cases} \tag{21}
$$

A classical equilibrium requires $0 < p_1 < 1$ because of the emptiness otherwise of M'_2, and therefore two equations in (21). That can't be met; no choice of $(x_{11}, x_{12}) \in M_1$ and $(x_{21}, x_{22}) \in M'_2$ can yield $x_{12} + x_{22} \ge 1$.

A two-tier equilibrium with both agents barely surviving is impossible for the reasons already explained in Example 1. A two-tier equilibrium with agent $a = 1$ barely surviving and agent $a = 2$ optimizing is likewise impossible for the reasons seen earlier.

A two-tier equilibrium with agent $a = 1$ optimizing and agent $a = 2$ barely surviving does turn out to be possible, however. For this, we need $0 < p_1 \le 1$ in order to avoid M_1 being empty. But $0 < p_1 < 1$ would make the choice of $(x_{21}, x_{22}) \in M_-$ reduce to $(0,0)$ while requiring two equations in (21), which doesn't work. In taking $p_1 = 1$, we merely have to satisfy the first condition in (21) with equality. The only vector in M_1 is $(1,0)$, whereas M_- consists of the vectors $(0,\tau)$ with $\tau \ge 0$. We have (21) fulfilled when $0 \le \tau \le 2$.

In view of Theorem 2 (and Theorem 3), at least one of these two-tier equilibria must be a virtual equilibrium, but which? To sort that out, we have to inspect the possibilities for having a classical equilibrium when the endowment vectors e_1 and e_2 are perturbed to

$$
e_1^\varepsilon = (1 + \varepsilon_{11}, 1 + \varepsilon_{12}), \qquad e_2^\varepsilon = (\varepsilon_{21}, 1 + \varepsilon_{22}),
$$

where the increments are all ≥ 0. The calculations focus then on the set

$$\begin{aligned} M_1^\varepsilon &= \operatorname{argmax}\{u_1(x_1) \mid x_1 \in X_1, p \cdot x_1 \leq p \cdot e_1^\varepsilon\} \\ &= \operatorname{argmax}\{x_{11} \mid x_{11} \geq 0, x_{12} \geq 0, \\ &\qquad p_1 x_{11} + (1-p_1)x_{12} \leq 1 + p_1\varepsilon_{11} + (1-p_1)\varepsilon_{21}\} \\ &= \begin{cases} \varnothing & \text{if } p_1 = 0, \\ \{(p_1^{-1}(1 + p_1\varepsilon_{11} + (1-p_1)\varepsilon_{21}), 0))\} & \text{if } p_1 > 0, \end{cases} \end{aligned}$$

for agent $a = 1$ and the set

$$\begin{aligned} M_2^\varepsilon &= \operatorname{argmax}\{u_2(x_2) \mid x_2 \in X_2, p \cdot x_2 \leq p \cdot e_2^\varepsilon\} \\ &= \operatorname{argmax}\{x_{21} + x_{22} \mid x_{21} \geq 0, x_{22} \geq 0, \\ &\qquad p_1 x_{21} + (1-p_1)x_{22} \leq p_1\varepsilon_{21} + (1-p_1)(1+\varepsilon_{22})\} \\ &= \begin{cases} \varnothing & \text{if } p_1 = 0, \\ \{(\varepsilon_{21} + p_1^{-1}(1-p_1)(1+\varepsilon_{22}), 0)\} & \text{if } 0 < p_1 < 1/2, \\ \{(\tau, 1-\tau) \mid 0 \leq \tau \leq 1\} & \text{if } p = 1/2, \\ \{(0, 1 + p_1(1-p_1)^{-1}\varepsilon_{21} + \varepsilon_{22}))\} & \text{if } 1/2 < p_1 < 1, \\ \varnothing & \text{if } p_1 = 1, \end{cases} \end{aligned}$$

for agent $a = 2$. The perturbed total endowment is

$$e_1^\varepsilon + e_2^\varepsilon = (1 + \varepsilon_{11} + \varepsilon_{21}, 2 + \varepsilon_{12} + \varepsilon_{22}),$$

and the market clearing conditions come out therefore as

$$\begin{cases} x_{11} + x_{21} \leq 1 + \varepsilon_{11} + \varepsilon_{21}, & \text{with equality if } p_1 > 0, \\ x_{12} + x_{22} \leq 2 + \varepsilon_{12} + \varepsilon_{22}, & \text{with equality if } p_1 < 1. \end{cases} \tag{22}$$

Once more the cases where $p_1 = 0$ or $p_1 = 1$ can be eliminated because of emptiness in M_2^ε, so we must have $0 < p_1 < 1$ along with equality in both of the conditions in (22). This can be achieved with

$$p^\varepsilon = (1+\varepsilon_{12}+\varepsilon_{21})^{-1}(1+\varepsilon_{12},\varepsilon_{21}),$$
$$x_1^\varepsilon = (1+\varepsilon_{11}+\varepsilon_{12},0),\ x_2^\varepsilon = (0,2+\varepsilon_{21}+\varepsilon_{22})$$

when $\varepsilon_{21} > 0$, but not otherwise. As the increments tend to 0, the only possible limit of such classical equilibria is the two-tier equilibrium in (20) for $\theta = 2$. Hence that is the unique virtual equilibrium, and the iterative scheme must converge to it. □

Other insights can be gleaned from Example 2 as well. The utility scale factors associated with agent $a = 2$ in the perturbed equilibria must tend to ∞ as the increments go to 0, specifically as $\varepsilon_{21} \to 0$. If that were not the case, the virtual equilibrium obtained in the limit would actually be a classical equilibrium. This feature of the iterative scheme came out in the proof of Theorem 3. The interpretation is that, as the amount of good 1 available to agent 2 shrinks to nothing, the interest of agent 2 in acquiring some of good 1 increases without bound. The scaling between utility for agent 2 and the relative prices at equilibrium blows up. This emerges as the essential reason why agent 2 ends up barely surviving without optimizing, even though, with an infinitesimal amount of good 1, optimization would be possible.

REFERENCES

[1] K.J. Arrow, G. Debreu, Existence of an equilibrium for a competitive economy. *Econometrica* ,Vol.22 (1954), 265–290.

[2] S. Dafermos, Exchange price equilibria and variational inequalities. *Mathematical Programming*, Vol.46 (1990), 391–402.

[3] G. Debreu, *Theory of Valu,* (Wiley, 1959).

[4] G. Debreu, New concepts and techniques for equilibrium analysis. Chapter 10 of *Mathematical Economics* (Econometric Soc. Monographs, No. 4), 1983.

[5] A.D. Dontchev, R.T. Rockafellar, Ample parameterization of variational inclusions. *SIAM J. Optimization,* Vol.12 (2002), 170–187.

[6] F. Facchinei, J.-S. PANG, *Finite-dimensional variational inequalities and complementarity problems.* Vols. I and II, Springer Series in Operations Research. Springer-Verlag, New York, 2003.

[7] M. Florig, On irreducible economies. *Annales d'Économie et de Statistique*, Vol.61 (2001), 184–199.

[8] M. Florig, Hierarchic competitive equilibria. *Journal of Mathematical Economics*, Vol.35 (2001), 515–546.

[9] P.T. Harker, J.-S. Pang, Finite-dimensional variational inequalities and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Mathematical Programming, Ser. B,* Vol.48 (1990), 161–220.

[10] J.-S. Pang, Computing generalized Nash equilibria. Preprint, 2002.

[11] J.-S. Pang, M. Fukushima, Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games. Preprint, 2002.
[12] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.
[13] Oiko Nomia, Existence of Walras equilibria: the excess demand approach, Cahiers MSE, University of Paris I, 1996.
[14] R.T. Rockafellar, R. J-B Wets, *Variational Analysis*, Springer-Verlag, Berlin, 1997.
[15] H.E. Scarf, The approximate fixed points of a continuous mapping. SIAM J. Applied Math.,Vol.15 (1967), 1328–1343.
[16] H.E. Scarf, *The Computation of Economic Equilibria*, Yale University Press, 1973.

ON TIME DEPENDENT VECTOR EQUILIBRIUM PROBLEMS

A. Khan[1] and F. Raciti[2]
Department of Mathematical Sciences, Fisher Hall, Michigan Technological University, Houghton MI, USA,[1]; Dipartimento di Matematica e Informatica, Università di Catania, and Facoltà di Ingegneria dell' Università di Catania, Catania, Italy[2]

Abstract: We consider the time dependent traffic equilibrium problem in the case of a vector valued cost operator. The motivation for this approach is that users can decide to choose a path according to several criteria. In fact, they may want to choose a minimum delay path as well as a minimum tax path. Other criteria can be introcuced in the model, depending on the particular problem under consideration. Thus, we are led to a multicriteria equilibrium problem which can be related to vector variational inequalities. The functional setting is the space $L^2([0,T],R^n)$. The extension of the definition of weak equilibria in such a space is not straightforward due to the fact that the cone made up of the non-negative functions has empty interior. We overcome this problem by using the notion of quasi interior of a closed convex set of a Hilbertspace and give sufficient conditions for the existence of weak equilibria.

Key words: Time Dependent Traffic Networks,Vector Variational Inequalities, Pareto optimization, multicriteria equilibrium problems, quasi interior.

1. INTRODUCTION

Many problems of physics, economics and applied mathematics can be formulated as equilibrium problems [8]. Depending on the particular structure of the problem, the search for equilibria can be performed by using optimization techniques, variational inequalities, projected dynamical

systems and other methods. In this note, by using the paradigmatic example of the traffic equilibrium problem, we present a formulation of time dependent vector equilibrium problems. In particular, we extend the concept of *weak* equilibrium (see below) and overcome the difficulty due to the fact that the cone of nonnegative functions has empty interior by using the notion of quasi interior of a closed convex subset of a Hilbert space. We also connect our vector equilibrium problem to the theory of vector variational inequalities. The concept of vector variational inequality has been introduced by Giannessi [6] in finite dimension and successively extended in infinite dimension by several authors (see for instance [7]. By using the concept of quasi interior we can also formulate a general variational inequality in the weak case which, in a special case, can be used to give sufficent condition for the existence of weak equilibria. The general analysis of this new variational inequality is object of future research.

The plan of the paper is the following: we complete this introduction by establishing some notations and definitions. In section 2. we give some existence results for the vector variational inequality in the weak case. Then we consider in detail the time dependent vector traffic problem and give sufficient conditions for the existence of weak equilibria.

Let us introduce some ordering relations between vectors of $\mathbb{R}^r$.

- $\xi \leq \eta \Longleftrightarrow \eta - \xi \in \mathbb{R}^r_+$
- $\xi \lneqq \eta \Longleftrightarrow \eta - \xi \in \mathbb{R}^r_+ \setminus 0$
- $\xi \not\leq \eta \Longleftrightarrow \eta - \xi \notin \mathbb{R}^r_+ \setminus 0$
- $\xi < \eta \Longleftrightarrow \eta - \xi \in int\mathbb{R}^r_+$
- $\xi \not< \eta \Longleftrightarrow \eta - \xi \notin int\mathbb{R}^r_+$

where $\mathbb{R}^r_+$ is the non negative orthant.

In view of our time dependent extension we generalize these relations to the Hilbert space $\mathcal{L} := L^2([0,T],\mathbb{R}^r)$:

- $\xi(t) \leq \eta(t) \Longleftrightarrow \eta(t) - \xi(t) \in \mathcal{L}_+$
- $\xi(t) \lneqq \eta(t) \Longleftrightarrow \eta(t) - \xi(t) \in \mathcal{L}_+ \setminus 0$
- $\xi(t) \not\leq \eta(t) \Longleftrightarrow \eta(t) - \xi(t) \notin \mathcal{L}_+ \setminus 0$
- $\xi(t) < \eta(t) \Longleftrightarrow \eta(t) - \xi(t) \in qi\mathcal{L}_+$
- $\xi(t) \not< \eta(t) \Longleftrightarrow \eta(t) - \xi(t) \notin qi\mathcal{L}_+$

where $\mathcal{L}_+$ denotes the subset of $\mathcal{L}$ made up of the (vector) functions whose components are non negative almost everywhere on $[0,T]$, $\{0\}$ represents here the null vector function (almost everywhere on $[0,T]$), and $\mathrm{Qi}\mathcal{L}_+ := \mathcal{L}_{++}$ denotes the subset of $\mathcal{L}$ made up of the (vector) functions whose components are positive almost everywhere on $[0,T]$. The notion of quasi

interior enables us to overcome the problem that the topological interior of $\mathcal{L}_+$ is empty. Thus, let us then recall some further definitions of convex analysis: if $Z \subset \mathcal{L}$ is convex and closed, the *tangent cone* to Z at point $x(\tau)$ is defined as:

$$S_Z(x(\tau)) := \mathrm{Cl}\left\{\bigcup_{\lambda>0} \lambda(Z - x(\tau))\right\}$$

Following Borwein and Lewis [15], let us introduce the *quasi relative interior* of Z, qriZ as the set of those $x(\tau) \in Z$ for which $S_Z(x(\tau))$ is a subspace. In the particular case when $S_Z(x(\tau)) = \mathcal{L}$ we shall denote the same set as the *quasi interior* of Z, qi$Z$. The set $Z \setminus \mathrm{qi}Z$ will be denoted as the *quasiboundary* of Z, qbdry$Z$. These concepts have been used quite recently by Gwinner [12] to extend the notion of a projected dynamical system [10] to an abstract Hilbert space and by the author [13] in order to establish the connection between projected dynamical systems and the time dependent variational inequalities presented in [1].

Definition 1.1 A (finite dimensional) vector variational inequality represents the following problem:

$$\text{find } x \in C : F(x)(y - x) \not\leq 0, \forall y \in C \tag{1}$$

where C is a closed convex subset of $\mathbb{R}^n$ and $F : C \mapsto \mathbb{R}^{r \times n}$ is a matrix valued function.

Definition 1.2 A (finite dimensional) *weak* Variational Inequality represents the following problem:

$$\text{find } x \in C : F(x)(y - x) \not< 0, \forall y \in C \tag{2}$$

Definition 1.3 A *weak* variational inequality in L represents the following problem:

$$\text{find } x \in K : (T(x), (y - x)) \not< 0, \forall y \in K \tag{3}$$

where $T : \mathcal{L} \mapsto L(\mathcal{L}, \mathcal{L})$, the natural ordering cone is $\mathcal{L}_+$, and the symbol $\not<$ is the one defined above. In view of the application to the traffic equilibrium problem it will be useful to consider the particular case where $T : \mathcal{L} \mapsto L(\mathcal{L}, \mathbb{R}^n)$.

2. EXISTENCE RESULTS

In this section we recall and generalize some definitions and results due to [14].

Definition 2.1 Let X be a real Banach space and (Y,P) an ordered Banach space, where P is a convex ordering cone. The mapping $T: X \mapsto L(X,Y)$ is called *monotone* if:

$$(T(x)-T(y),x-y) \geq 0, \forall x,y \in X$$

Definition 2.2 Let X,Y be normed spaces. $T: X \mapsto L(X,Y)$ is called v-hemicontinuous if $\forall x,y \in X$ the map $t \mapsto (T(x+ty),y)$ is continuous at 0^+.

Lemma 2.1 (see [14]) Let (X,C) and (Y,P) be ordered Banach spaces with ordering cones C and P, respectively. Let T be monotone and v-hemicontinuous. Then, the following two problems are equivalent for each convex subset K of X:

$$(I)\ x \in K\ (T(x),y-x) \not< 0, \forall y \in K$$

$$(II)\ x \in K\ (T(y),y-x) \not< 0, \forall y \in K$$

Remark 2.1 In this Lemma we understand that $a \not< 0$ means that $-a \in intP$ if $intP$ is non empty, $-a \in qiY$, otherwise. Thus, the equivalence between the original vector variational inequality (I) and the so called Minty variational inequality (II) can be generalized to spaces whose ordering cones have empty interiors.

In view of the application to the traffic problem it will be convenient to specialize the existence theorem in [14] to our functional setting.

Theorem 2.1 Let K a nonempty closed, convex and bounded subset of $\mathcal{L}$. Let $T: K \mapsto L(\mathcal{L},\mathbb{R}^n)$ be a monotone and v-hemicontinuous map on $\mathcal{L}$. Then, the vector variational inequality (I) has a solution.

3. THE SCALAR AND VECTOR TRAFFIC EQUILIBRIUM

The traffic assignment problem has a relatively recent history. For a variational inequality formulation of equilibrium conditions we refer to the influential paper by Smith [5]. For an interesting survey on models and

methods we refer to [8] and to the reach bibliography therein. In the last years some authors have tried to enlarge the classical traffic assignment problem. For instance, in [1] a time dependent formulation has been proposed, while in [2] a vector model has been put forth (see also [3]). In the first part of this note we combine these two approaches and consider time dependent vector equilibria. While the extension of the strong vector equilibrium definition is quite straightforward, in order to extend the definition of weak vector equilibrium given in [2] we shall use the notion of quasi interior of a convex closed set. Then we formulate two variational inequalities which imply each type of equilibrium, respectively. Let us first introduce the notation commonly used to state the standard traffic equilibrium problem.

A traffic network consists of a set W of origin-destination pairs and a set $\mathcal{R}$ of routes. The set of all $r \in \mathcal{R}$ which link a given $w \in W$ is denoted by $\mathcal{R}(w)$. In our analysis we are not interested on the link structure of the routes. A route-flow vector is an element $F \in \mathbb{R}^R$. Feasible flows are flows which satisfy the capacity constraints and demands, i.e., which belongs to the set:

$$K := \{F \in \mathbb{R}^{\mathcal{R}} \mid \lambda \leq F \leq \mu) \ \phi F = \rho \ \}$$

where $\lambda \leq \mu$ and ρ are given, and ϕ is the well known pair-route incidence matrix whose elements $(\phi)_{w,r}$ are set equal 1 if route r connects the pair w, 0 else. A Mapping $C := K \mapsto \mathbb{R}^{\mathcal{R}}$ is then given which assigns to each flow $F \in K$ its cost $C(F) \in \mathbb{R}^{\mathcal{R}}$.

Definition 3.1 A flow is called an equilibrium flow (or Wardrop Equilibrium) iff: $H \in K$ and

$\forall w \in W, \forall q, s \in \mathcal{R}(w)$, there holds :

$$C_q(H) < C_s(H) \Rightarrow H_q = \mu_q \ \text{ or } \ H_s = \lambda_s$$

that is equivalent to say that:

$$H \in K \text{ and } \langle C(H), F - H \rangle \geq 0 \ \ \forall F \in K$$

Let us notice that we are considering capacity constrained network as done by some authors [1]. Roughly speaking, the meaning of Wardrop Equilibrium is that the road users choose minimum cost paths, and the

meaning of the cost is usually that of traversal time. However, in many situations users behave according to more than one criteria, and this can be expressed by introducing, for each path, a vector cost whose components represent the various criteria in the choice of the path. In [2] the scalar weights $C_k[F]$ are generalized to vector weights. In order to keep notation as compact as possible we shall use the same notation of the scalar case. Now, for each k, $C_k[F] \in R^r$, and a matrix $C[F]$ is built, having as columns the vectors $C_k[F]$. In the spirit of Pareto optimization [2] proposed the following strong and weak vector equilibrium principle:

Definition 3.2 A flow $H \in K$ is a strong vector equilibrium if

$\forall w \in W, \forall q,s \in \mathcal{R}(w)$, there holds :

$$C_q(H) - C_s(H) \gneqq 0 \Rightarrow H_q = 0$$

Definition 3.3 A flow $H \in K$ is a weak vector equilibrium if

$\forall w \in W, \forall q,s \in \mathcal{R}(w)$, there holds :

$$C_q(H) - C_s(H) > 0 \Rightarrow H_q = 0$$

The following strong and weak variational inequalities are sufficient condition for H to be a strong and weak equilibrium, respectively.

$$H \in K : C(H)(F - H) \lneqq 0, \forall F \in K \tag{4}$$

$$H \in K : C(H)(F - H) \lneq 0, \forall F \in K \tag{5}$$

4. THE TIME DEPENDENT VECTOR MODEL

The traffic network is now considered at all times $t \in \tau$, where $\tau = [0,T]$. For each time $t \in \tau$ there is a route-flow vector $F(t) \in \mathbb{R}^{\mathcal{R}}$. Feasible flows are flows which satisfy the time dependent capacity constraints and demands, i.e., which belongs to the set:

$$K := \{F \in \mathcal{L} | \lambda(t) \leq F(t) \leq \mu(t) \ \ \phi F(t) = \rho(t) \text{ a.e. on } \tau\}$$

where $\lambda(t) \leq \mu(t)$ and $\rho(t)$ are given, and ϕ is the well known pair-route incidence matrix whose elements $(\phi)_{w,r}$ are set equal 1 if route r connects the pair w, 0 else. $F : \tau \mapsto \mathbb{R}^{\mathcal{R}}$ is the flow trajectory over time. Flows trajectories are supposed to be elements of $\mathcal{L} := L^2([0,T], \mathbb{R}^{\mathcal{R}})$. A matrix $C[F(t)] : K \mapsto L^2([0,T], R^{r \times R})$ is then given whose columns are the vector costs for each path in correspondence to each flow trajectory $F(t) \in K$. For the sake of simplicity we prefer to consider the simpler feasible set: $K := \{F \in \mathcal{L} | 0 \leq F(t) \ \ \phi F(t) = \rho(t) \text{ a.e. on } \tau\}$

and give the following definitions:

Definition 4.1 A flow $H \in K$ is a strong dynamical vector equilibrium if

$\forall w \in W, \forall q, s \in \mathcal{R}(w)$, there holds :

$$C_q[H(t)] - C_s[H(t)] \gneqq 0 \Rightarrow H_q(t) = 0$$

Definition 4.2 A flow $H(t) \in K$ is a weak dynamical vector equilibrium if

$\forall w \in W, \forall q, s \in \mathcal{R}(w)$, there holds :

$$C_q[H(t)] - C_s[H(t)] > 0 \Rightarrow H_q(t) = 0$$

Let us now consider the following two problems:

Problem 4.1 Find $H(t) \in K$:

$$\int C[H(t)](F(t) - H(t))dt \not\leq 0 \ \forall F(t) \in K$$

Problem 4.2 Find $H(t) \in K$:

$$\int C[H(t)](F(t)-H(t))dt \not< 0 \ \forall F(t) \in K$$

Theorem 4.1 A solution to problem 4.1 is a *strong equilibrium*
Proof Let us suppose, by contradiction that $\exists w \in W$ and $k, j \in \mathcal{R}(w)$, and $E \subset [0,T]$, $|E| > 0$:

$$C_k[H(t)] - C_j[H(t)] \gneqq 0 \text{ and } H(t) > 0 \ a.e.t \in E \tag{6}$$

an Then, if we choose a feasible flow $F(t)$ such that: $F_i(t) = H_i(t)$ if $i \neq k, j$, $F_j(t) = 0$, $F_k(t) = H_k(t) + H_j(t)$ we get:

$$\int_E C[H(t)](F(t)-H(t))dt = \int_E H_j(t)(C_k[H(t)] - C_j[H(t)])dt$$

But the first factor in the integrand is a positive scalar function, and the vector function $(C_k[H(t)] - C_j[H(t)]) \gneqq 0$. As a consequence it can not be

$$\int_E C[H(t)](F(t)-H(t))dt = \int_E H_j(t)(C_k[H(t)] - C_j[H(t)]) \not< 0 dt$$

and we get the absurd.

Theorem 4.2 A solution to problem 4.2 is a strong equilibrium.

Proof The proof is analogous to that of theorem 4.1.

As our time dependent vector model extends the approach of [2] we inherit here the same conclusion that variational inequalities are only sufficient for equilibria. For a different approach which yields equivalent conditions see [3].

REFERENCES

[1] P. Daniele, A. Maugeri and W. Oettli, *Time-Dependent Traffic Equilibria*, Jota,Vol. 103, No. 3 December 1999.

[2] X.Q. Yang, C.J. Goh, *On Vector Variational Inequalities: Application to Vector Equilibria*, Journal of Optimization theory and Applications, Vol.95,No.2, pp.431-443, November 1997.

[3] W. Oettli, *Necessary and sufficient Conditions of Wardrop Type for vectorial traffic equilibria*, in: Equilibrium Problems: Non smooth Optimization and Variational

Inequality Models, Kluwer Academic Publishers, F.Giannessi, A.Maugeri and P.Pardalos (Eds.).

[4] Dafermos S.,*Traffic Equilibrium and Variational Inequalities*, Transp. Sc., 14, 42-54, 1980.

[5] Smith M.J., *The existence, uniqueness and stability of traffic equilibria*,Transp. Res., 13B, 295-304. Smith M.J.,*A new Dinamic Traffic Model And the Existence and calculation of Dynamic User Equilibra on Congested Capacity-Constrained Road Networks*, Transportation Research, Vol. 27B, pp.49-63,1993.

[6] F. Giannessi, *Theorems of the Alternative,Quadratic Programs, and Complementary Problem, Variational Inequalities and Complementarity Problems*, Ed. by R.W. Cottle, F. Giannessi and J.L. Lions,Wiley, New York, pp.151-186, 1980.

[7] Vector Variational Inequalities and Vector Equilibria, Ed. by F.Giannessi, Kluwer Academic Publishers, 2000.

[8] Variational Inequalities and Equilibrium Models: NonSmooth Optimization , F. Giannessi, A.Maugeri and P.Pardalos Eds., Kluwer Academic Publishers 2001.

[9] F.Raciti,*Time Dependent Equilibrium in Traffic Networks with delay* , in: Variational Inequalities and Equilibrium Models: NonSmooth Optimization , F. Giannessi, A.Maugeri and P.Pardalos Eds., Kluwer Academic Publishers, 2001.

[10] A.Nagurnay and D.Zhang, *Projected Dynamical Systems and Variational Inequalities with Applications*, Kluwer, Boston, Dordrecht, 1996.

[11] D. Zhang and A. Nagurney, *On the Stability of Projected Dynamical Systems*, *Journal of Optimization Theory and Applications* (1995), pp. 97-124.

[12] Gwinner J., *Time Dependent Variational Inequalities- Some Recent Trends*, in *Equilibrium Models and Variational Models*, P.Daniele, F. Giannessi and A. Maugeri Eds., Kluwer Academic Publishers,2002.

[13] F. Raciti, *Equilibria trajectories as stationary solutions of infinite dimensional dynamical systems*, accepted for the publication in AML.

[14] Chen Guang-Ya and Yang Xiao-Qi,*The Vector Complementary Problem and Its Equivalences with the Weak Minimal Element in Ordered Spaces*, Journal of Mathematical Analysis and Applications , pp. 136-158 (1990).

[15] J.M. Borwein, A.S. Lewis, *Partially finite convex programming,part I: quasi relative interiors and duality theory*, *Math. Programming* B 57 (1992), pp. 15-48.

ON SOME NONSTANDARD DYNAMIC PROGRAMMING PROBLEMS OF CONTROL THEORY

A.B. Kurzhanski[1] and P. Varaiya[2]
Moscow State (Lomonosov) University, fac. CMC;[1] University of California at Berkeley, EECS, ERL[2]

Abstract: The present report indicates an array of nonstandard target problems of control under state constraints. The problems are solved through dynamic optimization techniques where the systems are optimalized under nonintegral costs. In the general case this leads to new classes of HJB - type variational inequalities. In the linear case these problems may be treated through duality methods of nonlinear analysis and minimax theory.

INTRODUCTION

The recent activities in advanced automation and navigation as well as in scientific computation have motivated new interest in various *target problems* of control theory, [15], [20]. A particular question is whether a certain target set or group of sets representing, for example a safety (unsafe) zone or configuration could be reached (avoided) by a controlled system despite the acting *state and control constraints*. The posed question is obviously not an optimization problem. However here we indicate *variational techniques* that give some answers to the question.

These techinques reduce to control problems under nonintegral optimality criteria. The value functions for such problems, if solved in backward time, produce level sets which are the sets of states from which

various target problems are solvable (the "backward" reach sets). In "forward" time they also define some sets of states reachable under various types of constraints on the system trajectories.

Rather than investigating reachability sets only for *given* instances of time, the interest here is also in sets reachable at *some* instances of time under state constraints true either for *some* instances or for the whole time interval. We then introduce some *variational inequalities* or generalized Hamilton-Jacobi-Bellman (HJB) equations for such value functions which grasp the required properties. These equations and inequalities allow to treat classes of problems with nonsmooth parameters and solutions. For linear systems explicit formulas for the value functions are given in terms of duality relations of nonlinear analysis. (Such explicit solutions are mostly confined to convex optimization problems, however they are also available for some types of problems with complementary convex constraints yielding nonconvex solutions).

A direct calculation of value functions and possibly nonconvex reach sets through either exact HJB equations or through duality relations is complicated. For linear systems a parametrized sequence of HJB equations may be suggested which approximates the exact ones and allows to avoid calculation of generalized (viscosity) solutions. The level sets for such approximate equations could produce ellipsoids whose intersections allow to externally approximate convex reach sets and whose unions allow to internally approximate nonconvex reach sets.

1. THE SYSTEM

Consider a controlled system described by an ordinary differential equation:

$$\dot{x} = f(t,x,u), \tag{1}$$

which in particular can be linear,

$$\dot{x} = A(t)x + B(t)u + C(t)v(t), \quad t_0 \le t \le \tau, \tag{2}$$

Here $x \in \mathbb{R}^n$ is the state , $u \in \mathbb{R}^m$ is the control, $v(t)$ - a given disturbance, while $f(t,x,u)$ is continuous in all the variables and satisfies conditions of uniqueness and extendability of solutions for all starting points and all $t \ge t_0$, whatever be the control $u(t)$ restricted by hard bounds

$$u(t) \in \mathcal{P}(t), \ t \ge t_0. \tag{3}$$

Here $\mathcal{P}(t)$ is a compact set-valued function, continuous in t in the Hausdorff metric.

We also require set $f(t,x,\mathcal{P}(t)) = F(t,x)$ to be convex and compact and differential inclusion (DI)

$$\dot{x} \in F(t,x)$$

to have a Caratheodory solution extendable within the intervals under consideration. The tube of solutions to the latter DI which start at set X^* at time τ is denoted as $X[t] = X(t;\tau,X^*)$. This is the "reach set" of system (1).

For linear systems we require the $n \times n$ matrix function $A(t)$ as well as $n \times p-$ and $n \times q-$ matrices $B(t), C(t)$ to be continuous and $\mathcal{P}(t)$ to be convex. Next are the topics discussed in this paper.

2. BACKWARD REACHABILITY AND THE TARGET PROBLEMS

In this section we present some target problems together with closely related problems of *reachability analysis*.

Denote $x[t] = x(t;\tau,x)$ to be the *system trajectory* which starts from *position* $\{\tau,x\}, x = x[\tau],\ x \in \mathbb{R}^n$, set $\mathcal{M} = \{x \in \mathbb{R}^n : \varphi_1(x) \le 1\}$ to be the *the target set* and $\mathcal{Y}(t) = \{x \in \mathbb{R}^n : \varphi(t,x) \le 1\}$ to be the *state constraint.* Functions $\varphi(t,x), \varphi_1(x)$ are assumed to satisfy the inclusions

$$\varphi(t,\cdot) \in \Phi, t \in [t_0,\vartheta],\ \varphi_1(\cdot) \in \Phi, \tag{4}$$

where $\Phi = \{\phi(\cdot)\}$ is the class of *proper closed convex functions* $\phi(x), x \in \mathbb{R}^n$, whose Fenchel conjugates ϕ^* are such that $0 \in intdom\,\phi^*$. (Here $dom\phi = \{x : \phi(x) \le \infty\}$ and $int\mathcal{P}$ is the set of interior points of set $\mathcal{P}$). Function $\varphi(t,x)$ is assumed *continuous in both variables* and inclusion (4) for this function is satisfied in the second variable for each t.

Class Φ ensures that the level sets of functions $\varphi(t,x), \varphi_1(x)$, when nonempty, are convex and compact.

Problem 2.1. *Given time interval* $[\tau,\vartheta]$ *and functions* $\varphi(t,x), \varphi_1(x)$, *find* $W_1[\tau]$ - *the set of points* x, *such that*

$$W_1[\tau] = \{x : \{\exists u(\cdot), \forall t \in [\tau,\vartheta] : x[t] \in \mathcal{Y}(t),\ x[\vartheta] \in \mathcal{M}.\}\}$$

Here $W_1[\tau] = \{x : V_1(\tau) \le 1\}$ is a level set of the *value function*

$$V_1(\tau, x) =$$
$$= \min_u \max\{\max_t \{\varphi(t, x[t]) \mid t \in [\tau, \vartheta]\}, \varphi_1(x[\vartheta])\} \mid x[\tau] = x\}.$$

$W_1[\tau]$ is the *backward reach set relative to* $\mathcal{M}$ *under state constraints* $\mathcal{Y}(t)$, namely, the set of points $\{x\}$ for each of which there exists *some* control $u(t)$ which steers the trajectory $x[t] = x(t;\tau,x)$ to $\mathcal{M}$ under state constraint $\mathcal{Y}(t)$.

If $\varphi(t,x) \equiv \varphi_1(x)$, then $\mathcal{Y}(t) \equiv \mathcal{M}$, and $W_1[\tau]$ is the set of points x each of which generates *some* controlled trajectory $x(t,\tau,x) = x[t] \in \mathcal{M}$, $\forall t \in [\tau, \vartheta]$. It is the set of so-called "viable" states relative to state constraint $\mathcal{M} \equiv \mathcal{Y}(t)$, $\forall t \in [\tau, \vartheta]$.

Problem 2.2. *Given time interval* $[\tau, \vartheta]$ *and functions* $\varphi(t,x), \varphi_1(x)$, *find* $W_2[\tau]$ *- the set of points* x, *such that*

$$W_2[\tau] = \{x \in \mathbb{R}^n : \{\forall u(\cdot), \forall t \in [\tau, \vartheta] : x[t] \in \mathcal{Y}(t), x[\vartheta] \in \mathcal{M}\}\}.$$

Here $W_2[\tau] = \{x : V_2(\tau, x) \le 1\}$ is a level set of the *value function*

$$V_2(\tau, x) =$$
$$= \max_u \max\{\max_t \{\varphi(t, x[t]) \mid t \in [\tau, \vartheta]\}, \varphi_1(x[\vartheta])\} \mid x[\tau] = x\}.$$

$W_2[\tau]$ is the set of points from which *all* the controlled trajectories reach set $\mathcal{M}$ at time ϑ and also satisfy the state constraint $\mathcal{Y}(t), \forall t \in [\tau, \vartheta]$.

If $\varphi(t,x) \equiv \varphi_1(x)$, then $\mathcal{Y}(t) \equiv \mathcal{M}$, and $W_2[\tau]$ is the set of points x for each of which the reach tube $X[t] = X(t;\tau,x)$ *without state constraint* satisfies the inclusion $X[t] \subseteq \mathcal{M}, \forall t \in [\tau, \vartheta]$.

Problems 2.1 and 2.2 respectively reflect the properties of *weak and strong invariance* of the backward reach set $W_i[\tau], i = 1,2$, relative to equation (1) and the state constraint $\mathcal{Y}(t)$. Therefore these sets $W_i[\tau]$ may also be referred to as *invariant sets*, (see [1]).

Problem 2.3. *Given time interval* $[\tau, \vartheta]$, *and functions* $\varphi(t,x), \varphi_1(x)$, *find* $W_3[\tau]$ *- the set of points* x, *such that*

$$W_3[\tau] = \{x \in \mathbb{R}^n : \{\exists u(\cdot), \exists t \in [\tau, \vartheta] : x[t] \in \mathcal{Y}(t), x[\vartheta] \in \mathcal{M}\}\}.$$

Here $W_3[\tau] = \{x : V_3(\tau, x) \le 1\}$ is a level set of the *value function*

$$V_3(\tau,x)=\min_u \max\{\min_t\{\varphi(t,x[t])\mid t\in[\tau,\vartheta]\},\varphi_1(x[\vartheta])\}\mid x[\tau]=x\}.$$

This is the set of all points x such that *some* controlled trajectory $x[t]=x(t;\tau,x)$, which starts from x at time τ, reaches set $\mathcal{M}$ at time $t=\vartheta$ and also satisfies the state constraint $\mathcal{Y}(t)$ *at some* instant $t\in[\tau,\vartheta]$. If $\varphi(t,x)\equiv\varphi_1(x)$, then $\mathcal{Y}(t)\equiv\mathcal{M}$, and $W_3[\tau]$ is the set of all points x that at time τ eject *some* controlled trajectory $x[t]=x(t;\tau,x)$ which satifies the inclusion $x[t]\in\mathcal{M}$, *for some* $t\in[\tau,\vartheta]$. This is the union $W_3[\tau]=\cup\{W(\tau;t,\mathcal{M})\mid t\in[\tau,\vartheta]\}$ of backward reach sets from set-valued position $\{t,\mathcal{M}\}$, (without state constraints), over the time interval $[\tau,\vartheta]$.

Problem 2.4. *Given time interval* $[\tau,\vartheta]$ *and functions* $\varphi(t,x),\varphi_1(x)$, *find* $W_4[\tau]$ *- the set of all points* x, *such that*

$$W_4[\tau]=\{x\in\mathbb{R}^n:\{\forall u(\cdot),\exists t\in[\tau,\vartheta]:x[t]\in y(t),x[\vartheta]\in\mathcal{M}\}\}.$$

Here $W_4[\tau]=\{x:V_4(\tau,x)\le 1\}$ is a level set of the *value function*

$$V_4(\tau,x)=\max_u \max\{\min_t\{\varphi(t,x[t])\mid t\in[\tau,\vartheta]\},\varphi_1(x[\vartheta])\}\mid x[\tau]=x\}.$$

This is the set of all points x such that *all* the controlled trajectories $x[t]=x(t;\tau,x)$ which start from x at time τ reach set $\mathcal{M}$ at time $t=\vartheta$ and also satisfy the state constraint $\mathcal{Y}(t)$ *at some* instant $t\in[\tau,\vartheta]$. If $\varphi(t,x)\equiv\varphi_1(x)$, then $\mathcal{Y}(t)\equiv\mathcal{M}$, and $W_4[\tau]$ is the set of of all points x for which *each* of the controlled trajectories $x[t]=x(t;\tau,x)$ satisfies the inclusion $x[t]\in\mathcal{M}$, *for some* $t\in[\tau,\vartheta]$.

Problems 2.3 and 2.4 respectively reflect the weak and strong possibilities of reaching the target set at some instant of time within the interval $[\tau,\vartheta]$.

The sets W_1,W_2,W_3,W_4 are the possible types of *backward reach sets* or *solvability sets* for target problems. Other options for such problems are beyond the scope of the present paper. Note that in general, for a linear system (1.1), the sets W_1,W_2 are closed convex, while W_3,W_4 are closed, but need not be convex.

Problems 2.1, 2.3 and 2.2, 2.4 are related to the description of positions from which the target set $\mathcal{M}$ is *reachable* at *given time* or *at some time* (in the strong or weak sense respectively). It may also be necessary to specify the positions from which it is possible to *avoid* the target set.

Assume $\mathcal{D}_\varepsilon(t)=\{x:\varphi(t,x)\le 1+\varepsilon\}$, $\mathcal{D}_0(t)=\mathcal{D}(t)$.

Problem 2.5. *Given time interval* $[\tau,\vartheta]$, *set* $\mathcal{D}(t)$, *and number* $\varepsilon>0$, *find*

$$W_5[\tau,\varepsilon] = \{x : \{\exists u(\cdot), \forall t \in [\tau,\vartheta] : x[t] \notin int\mathcal{D}_\varepsilon(t)\}\}.$$

$W_5[\tau,\varepsilon] = \{x : V_5(\tau,x) \geq 1+\varepsilon\}$, where

$$V_5(\tau,x) = \max_u \{\{\min_t \varphi(t,x[t]) \mid t \in [\tau,\vartheta]\} \mid x(\tau) = x\}$$

Here $W_5[\tau,\varepsilon]$ is the set of all points x for which there exists *some control* $u(t)$ which ensures the trajectory $x[t]$ to lie beyond $int\mathcal{D}_\varepsilon(t)$ *for all* $t \in [\tau,\vartheta]$.

Problem 2.6. *Given time interval* $[\tau,\vartheta]$, *set* $\mathcal{D}(t)$, *and number* $\varepsilon > 0$, *find*

$$W_6[\tau,\varepsilon] = \{x : \{\forall u(\cdot), \forall t \in [\tau,\vartheta] : x[t] \notin int\mathcal{D}_\varepsilon(t)\}\}.$$

$W_6[\tau,\varepsilon] = \{x : V_6(\tau,x) \geq 1+\varepsilon\}$, where

$$V_6(\tau,x) = \min_u \{\{\min_t \varphi(t,x[t]) \mid t \in [\tau,\vartheta]\} \mid x(\tau) = x\}.$$

$W_6[\tau]$ is the set of all points x for which *all the controls* $u(t)$ ensure the respective trajectories $x[t]$ to lie beyond $int\mathcal{D}_\varepsilon(t)$ *for all* $t \in [\tau,\vartheta]$. Such trajectories *avoid* the "tube" $\mathcal{D}(t)$. In general the backward reach sets are nonconvex.

The next section deals with forward reachability and with the so-called "reach-evasion" set.

3. FORWARD REACHABILITY AND THE REACH - EVASION SET

In this section, when dealing with forward reachability, we denote $x[t] = x(t,t_0,x^*)$, also taking $\varphi_0(x) \in \Phi$.

Problem 3.1. *Given time interval* $[t_0,\vartheta]$ *and functions* $\varphi(t,x), \varphi_0(x)$, *find the* **value function**

$$V_1(\vartheta,x) = \min_u \max\{\max_t \{\varphi(t,x[t]) \mid t \in [t_0,\vartheta]\}, \varphi_0(x^*) \mid x[\vartheta] = x\}.$$

Then

$$\mathcal{X}(\vartheta;t_0,\mathcal{X}_0)=\mathcal{X}_1[\vartheta]=\{x:\mathcal{V}_1(\vartheta,x)\le 1\}$$

is the set of all points x for each of which there exists some controlled trajectory $x[t]=x(t,t_0,x^*)$ which starts at time t_0 from a certain $x^*\in\mathcal{X}_0=\{x^*:\varphi_0(x^*)\le 1\}$ and ensures $x[t]\in\mathcal{Y}(t),\forall t\in[t_0,\vartheta]$ with $x[\vartheta]=x$. It is the union $\cup\{\mathcal{X}(\vartheta,t_0,x^*)\mid x^*\in\mathcal{X}_0\}$.

And this is the conventional *reach set under state constraints*

$$\mathcal{X}_1[\theta]=\{x:\{\exists u(\cdot),\forall t\in[t_0,\vartheta]:x(t)\in\mathcal{Y}(t),x[\vartheta]=x\}\}.$$

An analogy of this problem with $\min_u$ substituted by $\max_u$ usually leads to degenerate types of reach sets.

Problem 3.2. *Given time interval* $[t_0,\vartheta]$, *and functions* $\varphi(t,x),\varphi_0(x)$, *find the* **value function**,

$$\mathcal{V}_2(\vartheta,x)=\min_u\{\max\{\min_t\{\varphi(t,x[t])\mid t\in[\tau,\vartheta]\},\varphi_0(x^*)\}\mid x[\vartheta]=x\}.$$

Here $\mathcal{X}_2(\vartheta;t_0,\mathcal{X}_0)=\mathcal{X}_2[\vartheta]=\{x:\mathcal{V}_2(\vartheta,x)\le 1\}$ is the set of points x for which there exists a control $u(\cdot)$ and a starting point $x^*\in\mathcal{X}_0$ which ensure the respective trajectory $x[t]=x(t;t_0,x^*)$ to satisfy the inclusion $x[t]\in\mathcal{Y}(t)$ *for some* $t\in[t_0,\vartheta]$ and $x[\vartheta]=x$. Here $\mathcal{X}_2[\vartheta]=\cup\{\mathcal{X}_2(\vartheta;t_0,x^*)\mid x^*\in\mathcal{X}_0\}$.

An analogy of the last problem with $\min_u$ substituted by $\max_u$ uaully leads to degenerate situations and will not be discussed.

Note that in general, for a linear system (1.2), the sets $\mathcal{X}_1$ are closed convex, while $\mathcal{X}_2$ are closed but need not be convex.

The given array of problems may also include backward reachability for linear systems under *complementary convex constraints*. Here is an example of a *reach-evasion set.*

Problem 3.3. *Given time interval* $[\tau,\vartheta]$ *and functions* $\varphi(t,x),\varphi_1(x)$, *find* $\mathcal{W}[\tau]$ *- the set of points* x, *such that*

$$\mathcal{W}[\tau]=\{x:\{\exists u(\cdot),\forall t\in[t_0,\vartheta]:x[t]\in\mathcal{Z}(t),x[\vartheta]\in\mathcal{M}\}\}.$$

where $\mathcal{Z}(t)=\overline{\mathbb{R}^n/\mathcal{Y}(t)}$ and $\overline{Y}$ is the closure of set Y.

Here $\mathcal{W}[t]=\{x:\mathcal{V}(t,x)\ge 1\}$ is the complement of the open level set $\{x:\mathcal{V}(t,x)<1\}$ of the **value function**

$$\mathcal{V}(\tau,x)=\max_u\{\min\{\min_t\{\varphi(t,x[t])\mid t\in[\tau,\vartheta]\},-\varphi_1(x[\vartheta])+2\}\mid x[\tau]=x\}.$$

$\mathcal{W}[\tau]$ is the set of points $\{x\}$ for each of which there exists a controlled trajectory $x[t]=x(t,\tau,x)$ which ensures the inclusion $x[t]\in\mathcal{Z}(t)$, $\forall t\in[\tau,\vartheta]$ and $x[\vartheta]\in\mathcal{M}$. Therefore, it is the set of points from which it is possible to *avoid* the domain $int\mathcal{Y}(t)$ for all t while *reaching* the target set $\mathcal{M}$ (which is assumed to lie beyond $\mathcal{Y}(\vartheta):\mathcal{M}\cap Y(\vartheta)=\varnothing$). In general $\mathcal{W}[\tau]$ is a nonconvex set.

The calculation of value functions described in this section is not simple. We now indicate some approaches to its solution.

4. SOLUTION METHODS. THE HJB EQUATIONS

In the general case the respective value functions may be calculated through the generalized HJB equation. We shall indicate such equations for problems 2.1, 3.1.

Suppose $\varphi_0(x)=d^2(x,\mathcal{X}_0),\varphi_1(x)=d^2(x,\mathcal{M}),\varphi(t,x)=d^2(x,\mathcal{Y}(t))$, where $d^2(x,\mathcal{Q})=\min\{(x-q,x-q)\,|\,q\in\mathcal{Q})\}$ is the square of the *Euclid distance* of point x from compact set $\mathcal{Q}$.

Starting with problem 2.1, denote $V_1(t,x)=V_1(t,x\,|\,V_1(t_0,x^0))$, emphasizing the dependence of $V_1(t,x)$ on the boundary condition – the function $V_1(t_0,x^0)$.

Theorem 4.1. *Value function* $V_1(t,x)$ *satisfies* the **principle of optimality**, *which has the semigroup form:*

$$V_1(\vartheta,x\,|\,V_1(\tau,\cdot))=V_1(\vartheta,x\,|\,V_1(t,\cdot\,|\,V_1(\tau,\cdot))), \tag{5}$$

with $\tau\le t\le\vartheta$.

This property is established through a conventional argument [6] and its consequence is a similar property for respective reach sets. Namely, if we redenote $W_1[\tau]=W_1(\tau;\vartheta,\mathcal{M})$, then we have:

$$W_1(\tau,\vartheta\,|\,\mathcal{M})=W_1(\tau,t\,|\,W_1(t,\vartheta\,|\,\mathcal{M})).$$

Relation (5) yields for the value function $V_1(t,x)$ the next "backward" relation (the "variational inequality") of the HJB-type:
when $V_1(t,x)\ne\varphi_1(t,x)$ it is

$$V_{1t}(t,x)+\min_u(V_{1x},f(t,x,u))=0,u\in\mathcal{P}(t) \tag{6}$$

and when $V_1(t,x) = \varphi_1(t,x)$ it is

$$\mathcal{H}(t,x,V_1,u) \geq 0, u \in \mathcal{P}(t) \cap \{u : \mathcal{H}(t,x,\varphi,u) \leq 0\}, \tag{7}$$

where

$$\mathcal{H}(t,x,V,u) = V_t(t,x) + (V_x(t,x), f(t,x,u)),$$

is the total derivative of function $V(t,x)$ due to equation (1) under control u.

Here V_t, V_x stand for the partial derivatives of $V(t,x)$, if these exist. Otherwise (6), (7) is a symbolic relation for the generalized HJB - type relations which have to be described in terms of subdifferentials, Dini derivatives or their equivalents. But the typical situation is that V *is not differentiable.* The treatment of relations (6), (7) then has to be worked out within the notion of generalized "viscosity" - type solutions or their equivalents, [14], [6], [19], [2], [3]. However, for linear systems with convex constraints, as those of Section 5, the value functions are indeed differentiable.

Relation (7) further yields (when $V_1(t,x) = \varphi_1(t,x)$):

$$0 = \mathcal{H}(t,x^0,V_1,u^0) \geq \mathcal{H}(t,x^0,\varphi,u^0). \tag{8}$$

Here $u^0 = u^0(t,x)$ is the minimizer in (7), $x^0 = x^0(t)$ – the vector of the respective optimal trajectory. (In the sequel the upper index 0 in u,x will denote the respective optimalizer and the phase space vector which it generates).

Note that the boundary condition for $V_1(t,x)$ is

$$V_1(\vartheta,x) = \max\{\varphi(\vartheta,x), \varphi_1(\vartheta,x)\}. \tag{9}$$

Taking Problem 2.2 we will have

$$V_{2t}(t,x) + \max_u (V_{2x}, f(t,x,u)) = 0, u \in \mathcal{P}(t), \tag{10}$$

when $V_{2x}(t,x) \neq \varphi(t,x)$
and

$$\max\{\mathcal{H}(t,x,V_{2x},u), \mathcal{H}(t,x,\varphi,u)\} \leq 0, \forall u \in \mathcal{P}(t),$$

when $V_2(t,x) = \varphi(t,x)$. Here the last relation also yields

$$0 = \mathcal{H}(t, x^0, \mathcal{V}_{2x}, u^0) \geq \mathcal{H}(t, x^0, \varphi, u^0), \tag{11}$$

The boundary condition is

$$V_2(t_0, x) = \max\{\varphi(\vartheta, x), \varphi_1(\vartheta, x)\}.$$

Functions $V_3(t_0, x), V_4(t_0, x)$ may be described along the lines of the previous two and of the forthcoming decription of the reach-evasion set.

We now pass to reach sets of the forward type (Problem 3.1).

Theorem 4.2. *Value function* $\mathcal{V}_1(t, x)$ *satisfies the* **principle of optimality**, *which has the semigroup form:*

$$\mathcal{V}_1(\tau, x \mid \mathcal{V}_1(t_0, x^0)) = \mathcal{V}_1(\tau, x \mid \mathcal{V}_1(t, \cdot \mid \mathcal{V}_1(t_0, \cdot))), \tag{12}$$

with $t_0 \leq t \leq \tau$.

This property is established through a conventional argument [6] and its consequence is a similar property for respective reach sets. Relation (12) yields the next "forward" HJB-type relations:
when $\mathcal{V}_1(t, x) \neq \varphi(x)$ we have

$$\mathcal{V}_{1t}(t, x) + \max_u(\mathcal{V}_{1x}, f(t, x, u)) = 0, u \in \mathcal{P}(t), \tag{13}$$

and

$$\min_u\{\mathcal{H}(t, x, \mathcal{V}_1, u) \mid u \in \mathcal{P}(t) \cap \{u : \mathcal{H}(t, x, \varphi, u) \leq 0\}\} = 0,$$

when $\mathcal{V}_1(t, x) = \varphi(t, x)$. Here the last relation further yields

$$0 = \mathcal{H}(t, x^0, \mathcal{V}_{1x}, u^0) \geq \mathcal{H}(t, x^0, \varphi, u^0)\}. \tag{14}$$

The boundary condition is

$$\mathcal{V}_1(t_0, x) = \max\{\varphi(t_0, x), \varphi_0(t_0, x)\}.$$

Finally we indicate the HJB equation for Problem 3.3. Then

$$\mathcal{V}(t, x) + \min_u(\mathcal{V}_x, f(t, x, u)) = 0, u \in \mathcal{P}(t) \tag{15}$$

when $\mathcal{V}(t,x) \neq \varphi(t,x)$ and

$$\min_u \{\mathcal{H}(t,x,\mathcal{V}_x,u) \mid u \in \mathcal{P}(t) \cap \{u : \mathcal{H}(t,x,\varphi,u) \geq 0\}\} = 0, \tag{16}$$

or

$$\mathcal{H}(t,x,\mathcal{V}_x,u) \geq 0, u \in \mathcal{P}(t) \cap \{u : \mathcal{H}(t,x,\varphi,u) \geq 0\}$$

when $\mathcal{V}(t,x) = \varphi(t,x)$. Under this condition one further has

$$0 = \mathcal{H}(t,x^0,\mathcal{V}_x,u^0) \leq \mathcal{H}(t,x^0,\varphi,u^0)\}.$$

The boundary condition is

$$\mathcal{V}(\vartheta,x) = \min\{\varphi(\vartheta,x), -\varphi_1(\vartheta,x) + 2\}.$$

The HJB equations for the other problems of Section 3 are produced in a similar way. They follow from respective versions of the Principle of Optimality. The calculation of solutions to these equations in the general case is not simple and requires additional investigation. A promising approach seems to be emerging along the lines of so-called *level set methods*, [18], [16].

However, in the case of linear systems the value functions $V_1 - V_4, V_1, \mathcal{V}_2, \mathcal{V}$ may be described through duality relations of convex analysis and related branches of optimization theory.

5. SOLUTION METHODS. DUALITY TECHNIQUES OF OPTIMIZATION THEORY

In this section we indicate solution methods for *linear systems*, where the value functions could be found through techniques of *convex analysis, semidefinite programming and minimax theory*, [7], [9], [17]. We describe the approach through the formula for calculating $\mathcal{V}(\tau,x)$, which allows to find the *reach-evasion set*. This shows the type of relations encountered here.

Suppose $y = Kx$, $y \in \mathbb{R}^k$, $\varphi(t,x) = (y, N(t)y)$, $\varphi_1(x) = (x-m, M(x-m))$, $N(t) = N'(t) > 0$, $M = M' > 0$. Here $M, N(t)$ are positive definite, symmetric matrices of respective dimensions n,k, with $N(t)$ continuous; the prime stands for the transpose.

Denote set $\mathcal{E}(p(t),P(t)) = \{x - p, P^{-1}(t)(x-p)) \le 1$ to be an ellipsoid with center $p(t)$ and shape matrix $P = P' > 0$, taking the bound on control u as

$$u(t) \in \mathcal{P}(t) = \mathcal{E}(0,P(t)), \tag{17}$$

and presuming therefore the target set

$$\mathcal{M} = \mathcal{E}(m,M) = \{x : (x-m, M(x-m)) \le 1\},$$

and the state constraint

$$\mathcal{Y}(t) = \mathcal{E}(0,N(t)) = \{y : (y, N(t)y) \le 1\},$$

to be ellipsoidal as well, with $N(t)$ being continuously differentiable.

Let $\rho(l \mid \mathcal{X}) = \max\{(l,x) \mid x \in \mathcal{X}\}$ stand for *the support function of convex compact set* $\mathcal{X}$. Then $\rho^2(d \mid \mathcal{E}(0,P)) = (d, P^{-1}d)$.

In order to find $\mathcal{V}(\tau,x)$, we shall start by looking at solvability in the class of controls (17) of the system of inequalities

$$(y[t], N(t)y[t]) \ge \mu > 0, t \in [\tau,\theta], (x(\vartheta) - m, M(x(\vartheta) - m)) \le 2 - \mu, \tag{18}$$

where $y[t] = Kx[t]$, $x[t] = x(t;\tau,x)$.

The first inequality in (18) is equivalent to the following (see [7]):

$$\exists q(\cdot) \in \mathcal{Q} : (q(t), y(t)) - 1/4(q(t), N^{-1}(t)q(t)) \ge 1, \ t \in [\tau,\vartheta]. \tag{19}$$

Here $\mathcal{Q}$ is a compact set of functions $q(\cdot)$ defined on $[\tau,\vartheta]$ and taken here as

$$\mathcal{Q} = \{q(\cdot)\}, \ q(t) = 2N(t)z(t), \ t \in [\tau,\vartheta],$$

where $z[t] = Kx[t]\}$ and $x[t] = x(t;\tau,x)$ is *any trajectory* of system (17) generated by any $u(t) \in \mathcal{E}(p(t),P(t))$ and any $x : (x,x) \le r^2$ with r^2 sufficiently large. Relation (19) in its turn is equivalent to the next one:

$$\exists q(\cdot) \in \mathcal{Q} : \int_\tau^\vartheta ((q(t), y(t)) - 1/4(q(t), N^{-1}(t)q(t)))d\Lambda(t) \ge \mu \int_\tau^\vartheta d\Lambda(t), \tag{20}$$

$$\forall \Lambda(\cdot) \in Var_+[\tau,\vartheta],$$

where scalar function $\Lambda(t) \in Var_+[t_0, \theta]$ - the space of nondecreasing functions of bounded variation on $[t_0, \theta]$.

The second inequality in (18) is equivalent to the following:

$$(l, x[\vartheta]) - 1/4(l, M^{-1}l) \le 2 - \mu, \ \forall l \in \mathbb{R}^n,$$

or

$$-\alpha((l, x[\vartheta]) - 1/4(l, M^{-1}l) + 2) \ge \alpha\mu, \ \forall l \in \mathbb{R}^n, \forall \alpha > 0. \tag{21}$$

Combining (20), (21), we come to an equivalent system, observing that (18) is solvable iff there exists a function $q(\cdot) \in \mathcal{Q}$, such that

$$-\alpha((l, x[\vartheta]) - 1/4(l, M^{-1}l) + 2) + \int_\tau^\vartheta ((q(t), y(t)) - 1/4(q(t), N^{-1}(t)q(t)))d\Lambda(t) \ge$$

$$\ge \mu(\alpha + \int_\tau^\vartheta d\Lambda(t)), \ \forall l \in \mathbb{R}^n, \ \forall \alpha > 0, \ \forall \Lambda(\cdot) \in Var_+[\tau, \vartheta].$$

The last relation may be rewritten as

$$\max_{q(\cdot)} \min_{\Lambda(\cdot)} \min_{l} \min_{\alpha} \{-(s[\tau], x) + \int_\tau^\vartheta ((s[t], B(t)u(t) + v(t))dt - \tag{22}$$

$$-\alpha(1/4 \int_\tau^\vartheta (q(t), N^{-1}(t)q(t))d\Lambda(t) - 1/4(l, M^{-1}l) + 2)\} \ge \mu,$$

under condition

$$\{\alpha, \Lambda(\cdot)\} \in \mathcal{D} = \{\{\alpha, \Lambda(\cdot)\} : \alpha + \int_\tau^\vartheta \Lambda(t) = 1\}.$$

Here $s[t]$ is the row-vector solution to the adjoint equation

$$ds = -sA(t)dt - q'(t)K\alpha d\Lambda(t), \ s(\vartheta) = \alpha l'.$$

To get the value function $\mathcal{V}(t, x)$ we now have to minimize the left-hand side of (22) over $u(\cdot)$. Applying a standard minimax theorem ([5]), we finally have

Theorem 5.1. *The value function $\mathcal{V}(\vartheta, x)$ is given by the following formula:*

$$\mathcal{V}(\vartheta,x)=\max_{q(\cdot)}\min_{\Lambda(\cdot)}\min_{l}\min_{\alpha}\{(s[\tau],x)+\int_{\tau}^{\vartheta}((s[t]B(t)P(t)B'(t)s'[t])^{1/2}+v(t))dt- \tag{23}$$

$$-\alpha(1/4\int_{\tau}^{\vartheta}(q(t),N^{-1}(t)q(t))d\Lambda(t)-(l,M^{-1}l)+2)\},$$

where the maximum in $q(\cdot)$ *is to be taken over all functions* $q(\cdot)\in\mathcal{Q}$ *and the minimums over* $l\in\mathbb{R}^n$, $\{\alpha,\Lambda(\cdot)\}\in\mathcal{D}$.

The generally nonconvex level set

$$\mathcal{W}[\tau]=\{x:\mathcal{V}(\tau,x)\geq 1\}$$

is the set of points from which it is possible to avoid the interior $int\mathcal{Y}(t)$ while reaching the target set $\mathcal{M}$ at prescribed time ϑ.

6. CONCLUSION

This paper presents some basic solution schemes for nonstandard dynamic programming problems motivated by new trends in control for automation and navigation. The solutions are given in the form of generalized HJB-type relations or, in the linear case, through duality relations of convex analysis and minmax theory.

REFERENCES

[1] AUBIN J-P., *Viability Theory*, Birkhauser, Boston, 1991.

[2] Bardi M., Capuzzo-Dolcetta I., *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhauser, Boston, 1997.

[3] Clarke F., Ledyaev Yu.S., Stern R.J., Wolenski P.R., *Nonsmooth Analysis and Control Theory*, Springer-Verlag, 1993

[4] Crandall M.G., Evans L.C., Lions P-L., "Some properties of solutions of Hamilton-Jacobi equations", Trans.Amer.Math.Soc. v.282, N2, 1984.

[5] Demianov V.F., Rubinov A.M., *Foundations of Nonsmooth Analysis and Quasidifferential Calculus*, (in Russian), Nauka, Moscow, 1990

[6] Fleming W.H., Soner H.M.,*Controlled Markov Processes and Viscosity Solutions*, Springer - Verlag, 1993.

[7] Gusev M.I., Kurzhanski A.B. Optimization of controlled systems with bounds on the controls and state coordinates. Differential equations, (transl.from Russian "Differencialniye uravneniya")v.7, NN 9,10, 1971.

[8] Krasovskii N. N., *Game-Theoretic Problems on the Encounter of Motions*. Nauka, Moscow, 1970, (in Russian), English Translation : *Rendezvous Game Problems* Nat.Tech.Inf.Serv., Springfield, VA, 1971.

[9] Kurzhanski A.B. *Control and Observation Under Uncertainty*, Nauka, Moscow (in Russian), 1977.

[10] Kurzhanski A.B., Varaiya P., "Dynamic optimization for reachability problems", JOTA, vol.108, N2, pp.227-251, 2001.

[11] Kurzhanski A.B., Varaiya P. "Reachability Under State Constraints - the Ellipsoidal Technique", Proc. IFAC-2002 World Congress, Barcelona, 2002.

[12] Lee E.B., Marcus L.,*Foundations of Optimal Control Theory*, Wiley, NY, 1967.

[13] Leitmann G., "Optimality and reachability via feedback controls". In: *Dynamic systems and mycrophysics*, Blaquiere A., Leitmann G.eds., 1982.

[14] Lions P.L., Viscosity solutions and optimal control. Proc. of ICIAM 91, SIAM, Philadelphia, 1992, pp.182-195.

[15] Lygeros J., Tomlin C., Sastri S., "Controllers for reachability specifications for hybrid systems." *Automatica*, v.35(3), pp.349-370, 1999.

[16] Osher S., Fedkiw R., *Level Set Methods and Implicit Surfaces*. Springer, 2002.

[17] Rockafellar R.T., Wets R.J.B., *Variational Analysis*, Springer-Verlag, 1998.

[18] Sethian J.A., *Level Set Methods and Fast Marching Methods*, Cambridge Univ. Press, 1994.

[19] Subbotin A.I. *Generalized Solutions of First-Order PDE's. The Dynamic optimization Perspective*. Birkhauser, Boston, 1995.

[20] Varaiya P., "Reach set computation using optimal control". Proc. of KIT Workshop on Verification of Hybrid Systems, Verimag, Grenoble, 1998.

PROPERTIES OF GAP FUNCTION FOR VECTOR VARIATIONAL INEQUALITY*

S. J. Li[1] and G. Y. Chen[2]
Department of Information and Computer Sciences, College of Sciences, Chongqing University, Chongqing, China;[1] Institute of Systems Science, Chinese Academy of Sciences, Beijing, China[2]

Abstract: The purpose of this paper is to investigate differential properties of a class of set-valued maps and gap functions involving vector variational inequalities. Relationship between their contingent derivatives are discussed. A formula computing contingent derivative of the gap functions is established. Optimality conditions of solutions for vector variational inequalities are obtained.

Key words: Contingent derivative, gap function, vector variational inequalities

1. INTRODUCTION

The concept of a gap function is well-known both in the context of convex optimisation and variational inequalities. The minimization of gap functions is a viable approach for solving variational inequalities. In this section we generalize the gap function for variational inequalities to set-valued functions of vector variational inequalities (in short, VVI). The convexity of gap functions is studied.

* This research was partially supported by the National Nature Science Foundation of China

Let X and Y be Banach spaces, and let $C \subset Y$ be a closed and convex cone with a nonempty interior $intC$. Thus, Y is an ordered Banach space with the ordering cone C.

Given $\xi, \eta \in Y$, we consider the following relationships:

$$\xi \leq_C \eta \Longleftrightarrow \eta - \xi \in C; \qquad \xi \not\leq_C \eta \Longleftrightarrow \eta - \xi \notin C;$$
$$\xi \geq_C \eta \Longleftrightarrow \xi - \eta \in C; \qquad \xi \not\geq_C \eta \Longleftrightarrow \xi - \eta \notin C;$$
$$\xi \leq_{C\setminus\{0\}} \eta \Longleftrightarrow \eta - \xi \in C \setminus \{0\}; \quad \xi \not\leq_{C\setminus\{0\}} \eta \Longleftrightarrow \eta - \xi \in C \setminus \{0\}.$$

Since $intC \neq \varnothing$, the following partial ordering can also be defined:

$$\xi \leq_{intC} \eta \Longleftrightarrow \eta - \xi \in intC; \qquad \xi \not\leq_{intC} \eta \Longleftrightarrow \eta - \xi \notin intC;$$
$$\xi \geq_{intC} \eta \Longleftrightarrow \xi - \eta \in intC; \qquad \xi \not\geq \eta \Longleftrightarrow \xi - \eta \notin intC.$$

Given two subsets of Y, say A and B, the following ordering relationships on sets are defined:

$$A \leq_C B \Longleftrightarrow \eta \leq_C \xi, \qquad \forall \eta \in A, \xi \in B;$$
$$A \leq_{C\setminus\{0\}} B \Longleftrightarrow \eta \leq_{C\setminus\{0\}} \xi, \qquad \forall \eta \in A, \xi \in B;$$
$$A \not\leq_C B \Longleftrightarrow \eta \not\leq_C \xi, \qquad \forall \eta \in A, \xi \in B.$$

Since $intC \neq \varnothing$, the following partial ordering relationships on sets can also defined:

$$A \leq_{intC} B \Longleftrightarrow \eta \leq_{intC} \xi, \qquad \forall \eta \in A, \xi \in B;$$
$$A \not\leq_{intC} B \Longleftrightarrow \eta \not\leq_{intC} \xi, \qquad \forall \eta \in A, \xi \in B.$$

2. GAP FUNCTIONS OF VECTOR VARIATIONAL INEQUALITY

Consider following vector variational inequality problem consists in finding $y \in K$ such that

$$\text{(VVI)} \qquad \langle F(y), x - y \rangle \not\leq_{C\setminus\{0\}} 0, \quad \forall x \in K,$$

where K is a closed and convex subset of X and $F : X \to L(X,Y)$ is a function.

The weak vector variational inequality problem consists in finding $y \in K$ such that

$$\text{(WVV I)} \qquad \langle F(y), x - y\rangle \not\leq_{intC} 0, \quad \forall x \in K.$$

Definition 2.1 *Let C be convex and closed cone in Y with a nonempty interior, say int C, and K be a subset of X.*

(i) A set-valued function $\phi : X \rightrightarrows Y$ is said to be a gap function of VVI iff

1. $0 \in \phi(y)$ if and only if y solves VVI;

2. $0 \not\geq_{C\setminus\{0\}} \phi(x), x \in K$.

(ii) A set-valued function $\phi : X \rightrightarrows Y$ is said to be a gap function of WVVI iff

1. $0 \in \phi_W(y)$ if and only if y solves WVVI;

2. $0 \not\geq_{intC} 0 \in \phi_W(x), \forall x \in K$.

Let

$$\langle F(x), x - K\rangle = \bigcup \{\langle F(x), x - z\rangle : z \in K\}.$$

Definition 2.2 (First gap function for VVI) *Let C and K be as in Definition 2.1. Consider the set-valued function $\phi : X \rightrightarrows Y$, defined by*

$$\phi(x) := Max_C \langle F(x), x - K\rangle, \quad x \in K.$$

where $Max_C A$ is the set of minimal elements of A.

Let

$$dom(\phi) := \{x \in K : \phi(x) \neq \varnothing\}.$$

Theorem 2.1 *Let be C a convex and pointed cone in Y. The set-valued function*

$$\phi(x) = Max_C \langle F(x), x - K\rangle \text{ is a gap function for VVI.}$$

Proof. We first prove that $0 \in \phi(y)$ if and only if y solves VVI.

Suppose that y solves VVI. Then

$$\langle F(x), y - x\rangle \not\geq_{C\setminus\{0\}} 0, \quad \forall x \in K.$$

In particular, let $x = y$, then

$$\langle F(y), y - y\rangle = 0.$$

Thus $0 \in \phi(y)$, otherwise if there exits some $z \in K$ such that $\langle F(y), y - z\rangle \geq_{C\setminus\{0\}} \langle F(y), y - y\rangle = 0$, then this contradicts that y solves VVI.

Conversely, suppose $0 \in \phi(y)$. If y does not solve VVI, $\exists x \in K$, such that

$$\langle F(y), x - y\rangle \leq_{C\setminus\{0\}} 0,$$
$$\langle F(y), y - x\rangle \geq_{C\setminus\{0\}} 0 = \langle F(y), y - y\rangle.$$

Thus $0 \notin \phi(y)$.

Moreover, taking $x = y$, then

$$\langle F(y), y - y\rangle = 0,$$

thus

$$0 \not\geq_{C\setminus\{0\}} \phi(x), \quad \forall x \in K.$$

The proof is completed. □

Definition 2.3 (First gap function for WVVI) *Let C be as in Definition 2.1 and $intC \neq \varnothing$.* Define the set-valued function $\phi_w : X \rightrightarrows Y$:

$$\phi_w(y) := Max_{intC} \langle F(x), x - K\rangle, \quad \forall x \in K.$$

where $Max_{intC}A$ is the set of weakly minimal elements of A.

Theorem 2.2 *The set-valued function $\phi_w(x)$ is a gap function for WV V I.*

Proof. The proof is similar to that for Theorem 2.1, but with all occurence of ϕ. C and Max_C replaced by ϕ_w, $\mathrm{int}\, C$ and $\mathrm{Max}_{\mathrm{int}\, C}$, respectively. □

It is also possible to meaningfully interpret the meaning of "gap" function if we consider the duality of vector variational inequalities in a slightly more general context. Subsequently we shall relate the gap function to the generalized Young's inequality for Fenchel conjugate duality.

Consider the following general VVI:

Definition 2.4 *Let C be a closed and convex cone in Y. The general vector variational inequality (for short, GVVI) consists of finding $y \in X$ such that*

$$(GVVI) \qquad \langle F(x), x-y\rangle - f(y) - f(x) \not\leq_{C\setminus\{0\}} 0, \quad x \in X,$$

where $F : X \to L(X,Y)$ is assumed to be injective, and $f : X \to Y$ is assumed to be a $C-$convex function.

Definition 2.5 *Let $f : X \to Y$ be $C-$convex function. The Fenchel conjugate of f is a set-valued function $g : L(X,Y) \rightrightarrows Y$, such that*

$$g(u) := Max_C \{\langle u, x\rangle - f(x) : x \in X\}.$$

Remark 2.1 Note here that in order to be consistent with definition of Fenchel conjugate, f is assume to be a function from X into Y. Thus, in general, VVI is not a special case GVVI. However, if we adjoin an abstract "∞" to Y and C, written as $\dot{Y} = Y \cup \{\infty\}$, then GVVI includes the previous VVI as a special case where f is just the following indicator function for the set K :

$$f(x) = \begin{cases} 0 \in \dot{Y}, & if\, x \in K; \\ \infty \in \dot{Y}, & if\, x \notin K. \end{cases}$$

Since F is injective, we may also define the following function:

Definition 2.6 *Let $F : X \to L(X,Y)$ be injective. Let $G : L(X,Y) \to X$ be defined by*

$$G(u) = -F^{-1}(-u), \quad \forall u \in Dom(G) = -Range(F).$$

The dual general vector variational inequality (for short, DGVVI) consists in finding $u \in Range(F) \subset L(X,Y)$, such that

$$(DGVVI) \qquad \langle v-u, G(u)\rangle - g(u) + g(v) \not\leq_{C\setminus\{0\}} 0, \quad \forall v \in -Range(F),$$

where g is the Fenchel conjugate of f.

Thus, a generalization of Young's inequality follows immediately:

Lemma 2.1 (Generalized Young's inequality)

$f(x)+g(u)-<u,x> \not\leq_{C\setminus\{0\}} 0\ ,\ \forall x\in X$ and $\forall u\in L(X,Y)$.

We now present a result for the DGVVI. It is a generalization of Mosco's result [10].

Theorem 2.3 (Partial duality of GVVI and DGVVI) *Let C be a closed and convex cone in Y.*

(i) If y solves GVVI, then $u=-F(y)$ solves DGVVI;

(ii) If y solves GVVI and u solves DGVVI, then

$$0\in f(y)+g(u)-\langle u,y\rangle.$$

Proof. Suppose that y solves GVVI, then we have

$$\langle F(y),y\rangle - f(y) \not\leq_{C\setminus\{0\}} \langle -F(y),x\rangle - f(x),\quad \forall x\in X;$$

or form $G(y)=-F^{-1}(u)$ and $u=-F(y)$, we have

$$\begin{aligned}\langle -F(y),y\rangle - f(y) \ &\in \ \mathrm{Max}_C\left\{\langle -F(y),x\rangle - f(x) : x\in X\right\};\\ &= g(-F(y))\\ &= g\ (u)\end{aligned}\tag{1}$$

Now, if $u=-F(y)$ does not solves DGVVI, then there exists $v\in -Range(F)$, such that

$$\langle v-u,G(u)\rangle - g(u)+(gv) \not\leq_{C\setminus\{0\}} 0;$$

or

$$-\langle v,y\rangle + \langle v,y\rangle + g(v) \not\leq_{C\setminus\{0\}} g(u).$$

Since $\langle -F(y), y\rangle - f(y) = -\langle u, y\rangle - f(y) \in g(u)$, we have

$$f(y)+g(v)-\langle v,y\rangle \leq_{C\setminus\{0\}} 0,$$

which contradicts the generalized Yong's inequality in Lemma 2.1. Furthermore, from (1), since $u = -F(y)$, it follows that

$$0 \in f(y) + g(u) - \langle u, y\rangle.$$

We call the above a partial duality result, because the converse of either of (i) or (ii) may not hold. The converse will hold if we further assumed that C is connected in the sense that $C \cup (-C) = Y$ (See Y. Sawaragi, H. Nakayama and T. Tanino [11]).

We may now extend the definition of gap function to the problem GVVI as follows:

Definition 2.7 *A set-valued function $\phi' : Y \rightrightarrows X$ is said to be a gap function of the GVVI iff f is such that:*

1. $0 \in \phi'(y)$ if and only if y solves GVVI;
2. $0 \not\geq_{C\setminus\{0\}} \phi'(x), \quad \forall x \in X.$

Set:

$$\phi'(x) = \mathrm{Max}_C \{< F(x), x - y > + f(x) - f(y) : y \in X\}$$

It follows immediately that

$$\phi'(y) = f(y) + g(-(F(y)) + \langle F(y), y\rangle,$$

where g is the Fenchel conjugate of f. We now have a much simpler proof that ϕ' is a gap function for GVVI, and the meaning of "gap" is now apparent.

Theorem 2.4 *ϕ' is a gap function for problem GVVI.*

Proof. The fact that $\phi'(y) \not\leq_{C\setminus\{0\}} 0$ follows directly from Young's inequality of Lemma 2.1. Furthermore, by Theorem 2.3, y solves GVVI implies that $u = -F(y)$ solves DGVVI, together they implies that $0 \notin \phi'(y)$.

Conversely, suppose $0 \in \phi'(y)$, if y does not solves GVVI, then there exists some $x \in X$, such that

$$\langle F(y), x-y\rangle - f(x) + f(y) \geq_{C\setminus\{0\}} 0 = \langle (F(y), y-y) - f(y) + f(y),$$

then $0 \in \phi'(y)$, a contradiction. Furthermore, taking $x = y$, then

$$\langle (F(y), y-y) - f(y) + f(y) = 0,$$

thus $0 \not\leq_{C\setminus\{0\}} \phi'(x)$, $\forall x \in X$, and the proof is complete.

The above generalization of VVI is trivially extenable to WVVI.

Definition 2.8 *The general weak vector variational inequality (for short, GWVVI) consists of finding $y \in X$ such that*

$$\langle (F(y), x-y) - f(y) + f(x) \not\leq_{intC} 0, \quad \forall x \in X,$$

where $F : X \to L(X,Y)$ is assumed to be injective, and $f : X \to Y$ is assumed to be a C - convex function.

Definition 2.9 *A set-valued function $\phi'_w : X \rightrightarrows Y$: is said to be a gap function of the GWVVI if f*

1. $0 \in \phi_w'(y)$ if and only if y solves GWVVI;

2. $0 \not\geq_{intC} \phi_w'$, $\forall x \in X$.

We set

$$\phi_w'(y) := Max_{intC} \langle F(x), x-K\rangle, \quad \forall x \in K.$$

Theorem 2.5 *The set-valued function $\phi'_w : X \rightrightarrows Y$ is a gap function for GWVVI.*

Under some appropriate conditions, the gap function for both the (general) vector variational inequality and (general) weak vector variational inequality can be shown to be convex.

Definition 2.10 *Let K be a closed and convex subset of X. The function $F : K \to L(X,Y)$ is monotone on K if f*

$$\langle F(x') - F(x''), x' - x'' \rangle \geq_C 0, \quad \forall x', x'' \in K.$$

The function F *is affine if*, $\forall x', x'' \in K$, $\forall \alpha, \beta \in R$, *with* $\alpha + \beta = 1$ *we have*

$$F(\alpha x' + \beta x'') = \alpha F(x')\beta + F(x'').$$

The notion of convexity is well-defined for single-valued functions using ordering relationships. However, this definition cannot be extended in a straightforward way to set-valued functions, and inclusion type relationships must be used. It appears that the notion of convex set-valued function is slightly more complicated and more care is needed to deal with it.

Definition 2.11 *Let* C *be a closed and convex pointed cone in* Y *with the nonempty interior* $intC$, K *be a closed and convex subset of* X. *Let* $x, y \in K$ *and let* $t \in (0,1)$. *A set-valued function* $G : X \rightrightarrows Y$ *is said to be:*

(i) Type I C *- convex iff,* $G(tx + (1-t)y) \subset tG(x) + (1-t)G(y) - C$;

(ii) Type II C *- convex iff,* $tG(x) + (1-t)G(y) \subset G(tx + (1-t)y) + C$;

(iii) Type I C *- concave iff,* $tG(x) + (1-t)G(y) \subset G(tx + (1-t)y) - C$;

(iv) Type II C *- concave iff,* $G(tx + (1-t)y) \subset tG(x) + (1-t)G(y) + C$;

Remark 2.2 *Type II convexity and concavity have been used previously in the literature, where it was acknowledged that if* G *is type II* C *- convex, then* $-G$ *is not necessarily type II* C *- concave. However, it is not difficult to see that*

(i) G *is type I* C *- convex if*, $-G$ *is type II* C *- concave;*

and similarly,

(ii) G *is type II* C *- convex if*, $-G$ *is type I* C *- concave.*

If G is a single-valued function, then both type I and type II convexity (concavity, respectively) are equivalent to that usual C - convexity (usual C - concavity, respectively).

Lemma 2.2 *Let* $F : K \to L(X,Y)$ *be a function. If* F *is affine and monotone, then the function* $\langle F(\cdot), \cdot \rangle : K \to L(X,Y)$ *is type I* C *- convex.*

Proof. Given $t \in (0,1), x', x'' \in K$,

$$\begin{aligned}
&\langle F(tx'+(1-t)x''), tx'+(1-t)x''\rangle - t\langle F(x'),x'\rangle - (1-t)\langle F(x''),x''\rangle \\
&= t^2\langle F(x'),x'\rangle + (1-t)^2\langle F(x''),x''\rangle + t(1-t)(\langle F(x'),x''\rangle + \langle F(x''),x'\rangle \\
&\quad - t\langle F(x'),x'\rangle + (1-t)\langle F(x''),x''\rangle \\
&= -t(1-t)\langle F(x'-x''),x'-x''\rangle \\
&\leq_C 0.
\end{aligned}$$

It is well-known that the Fenchel conjugate of a scalar valued function is convex in the usual definition. With the above definition of convexity for set-valued functions, this notion is now affirmative for a vector-valued function.

Lemma 2.3 *Let $f: X \to Y$ be a C-convex function and let for any $u \in L(X,Y)$ the set $\{\langle u,x\rangle - f(x) : x \in X\}$ satisfy the domination property. Then Fenchel conjugate of f is type I C-convex.*

Proof. By the definition, the Fenchel conjugate of the vector-valued function f is a set-valued function $g: L(X,Y) \rightrightarrows Y$ such that

$$g(u) = \operatorname{Max}_C \{\langle u,x\rangle - f(x) : x \in X\}.$$

We have, $\forall t \in (0,1), u', u'' \in L(X,Y)$,

$$\begin{aligned}
&g(tu'+(1-t)u'') \\
&= \operatorname{Max}_C \{\langle tu'+(1-t)u'',x\rangle - f(x) : x \in X\}. \\
&\subset \{\langle tu'+(1-t)u'',x\rangle - f(x) : x \in X\} \\
&= \{t(\langle u',x\rangle - t(x)) + (1-t)(\langle u'',x\rangle - f(x)) : x \in X\} \\
&= \{t(\langle u',x\rangle - f(x)) + (1-t)(\langle u'',x\rangle - f(x)) : x \in X\} \\
&\subset t\{(\langle u',x\rangle - f(x)) : x \in X\} + (1-t)\{\langle u'',x\rangle - f(x)) : x \in X\} \\
&\subset t\operatorname{Max}_C \{(\langle u',x\rangle - f(x)) : x \in X\} - C + (1-t)\operatorname{Max}_C \{(\langle u'',x\rangle - f(x)) : x \in X\} - C \\
&= tg(u') + (1-t)g(u'') - C.
\end{aligned}$$

Then g is type I C-convex. □

Lemma 2.4 *If $g : L(X,Y) \rightrightarrows Y$ is type I C - convex, and $F : K \to L(X,Y)$ is affine, then the composite $g \circ F : X \rightrightarrows Y$ is type I C - convex.*

Proof. Given $x', x'' \in X$ and $t \in (0,1)$.

$$\begin{aligned} g \circ F(tx' + (1-t)x'') &= g(F(tx' + (1-t)x'')) \\ &= g(tF(x') + (1-t)F(x'')) \\ &\subset tg(F(x')) + (1-t)g(F(x'')) - C \\ &= tg \circ F(x') + (1-t)g \circ F(x'') - C. \end{aligned}$$

□

Theorem 2.6 *Let C be a closed and convex cone in Y and let for any $u \in L(X,Y)$ the set $\{\langle u, x\rangle - f(x) : x \in X\}$ satisfy the domination property. Consider the problem GVVI. If F is affine and monotone, and $f : X \to Y$ is C - convex, then the gap function ϕ' is type I C - convex.*

Proof. By the definition the gap function $\phi'(x)$ can be rewritten as,

$$\phi'(x) = g \circ (-F)(x) + \langle F(x), x\rangle + f(x),$$

where the Fenchel conjugate g of f is type I C - convex by Lemma 3.3. Since F is affine, so is $-F$. By Lemma 2.4, $g \circ (-F)$ is type I C - convex. By Lemma 3.2 $\langle F(\cdot), \cdot\rangle$ is type I C - convex, and hence ϕ' is type I C - convex. □

3. DIFFERENTIAL AND SENSITIVITY OF GAP FUNCTION

Let

$$\begin{aligned} N(x) &= \mathrm{Max}_C \langle F(x), x - K\rangle, \quad x \in K, \\ W(x) &= \mathrm{Max}_{intC} \langle F(x), x - K\rangle, \quad x \in K, \end{aligned}$$

respectively.

Note that (VVI) is equivalent to the following set-valued optimization problem:

$$\mathrm{Min}_C N(x), \text{subject to } x \in K, \tag{2}$$

and (WVVI) is equivalent to the following set-valued optimization problem:

$$\mathrm{Min}_{intC} W(x), \text{subject to } x \in K. \tag{3}$$

If F is a vector-valued function from X into X^*, then, (VVI) and (WVVI) become the ordinary variational inequality problem and the gap functions N and W reduce to Auslender's gap function [3]. Thus, set-valued optimization problems (2) and (3) reduce real mathematical problems:

$$\min\varphi(x) \text{ subject to } x \in K,$$

where $\varphi(x) = \max\langle F(x), x - K\rangle$. If φ is differentiable, then the above mathematical programming may be solved by a descent algorithm which possesses a global convergence property [7]. Therefore, it is a very important and valuable to discuss differential properties of gap functions N and W in vector variational inequalities. In sequel, we let X and Y be two real Banach spaces. Let θ and Θ denote the origin points of Y and $L(X,Y)$, respectively. For any $A \in L(X,Y)$, we introduce norm:

$$\|A\|_L = \sup\{\|A(x)\|_L : \| x \| \leq 1\}.$$

Since Y is a Banach space, $L(X,Y)$, is also a Banach space with the norm $\|\cdot\|_L$.

It is easy to verify the following lemma.

Lemma 3.1 *Let sequences* $\{\alpha_k\}$ *and* $\{\beta_n\} \subset R_+ \setminus \{0\}$ *such that* $\alpha_k \to 0$ *and* $\beta_n \to 0$.

Then, there exist subsequences $\{\alpha_{k_i}\}$ *and* $\{\beta_{n_i}\}$ *such that*

$$\lim_{i\to\infty} \frac{\alpha_{k_i}}{\beta_{n_i}} = 1.$$

Let $G : X \rightrightarrows Y$ be a set-valued function. We denote the contingent derivative of G at $(\overline{x}, \overline{y}) \in X \times Y$ as $DG(\overline{x}, \overline{y})$, which is a set-valued

function from X to Y, whose graph is the contingent cone (tangent cone) $T(\text{graph } G,(\overline{x},\overline{y}))$.

Proposition 3.1 *$y \in DG(\overline{x},\overline{y})(x)$ if and only if, for any $\{\alpha_k\} \subset R_+ \setminus \{0\}$ and $\alpha_k \to 0$, there exist subsequence $\{\alpha_{k_i}\} \subset \{\alpha_k\}$ and sequence $\{(x_i,y_i)\} \subset X \times Y$, such that $\alpha_{k_i} \to 0, (x_i,y_i) \to (x,y)$ and $\overline{y} + \alpha_{k_i} y_i \in G(\overline{x} + \alpha_{k_i} x_i)$ for all i.*

Proof. Obviously, we only need to prove the necessary condition. Suppose that $y \in DG(\overline{x},\overline{y})(x)$. Then, there exist $\{h_n\} \subset R_+ \setminus \{0\}$ and $\{(x_n,y_n)\} \subset X \times Y$, such that $h_n \to 0, (x_n,y_n) \to (x,n)$ and

$$\overline{y} + h_n y_n \in G(\overline{x} + h_n x_n), \ \forall_n.$$

Take any sequence $\{\alpha_k\} \subset R_+ \setminus \{0\}$ and $\alpha_k \to 0$. By Lemma 3.1, there exist subsequences $\{h_{n_i}\}$ and $\{\alpha_{k_i}\}$ such that

$$\lim_{i\to\infty} \frac{h_{n_i}}{\alpha_{k_i}} = 1.$$

Set

$$x_i = \frac{h_{n_i}}{\alpha_{k_i}} x_{n_i} \text{ and } y_i = \frac{h_{n_i}}{\alpha_{k_i}} y_{h_i}.$$

Thus

$$(x_i,y_i) \to (x,y).$$

It follows that

$$\overline{y} + \alpha_{k_i} y_i \in G(\overline{x} + \alpha_{k_i} x_i),$$

and this completes the proof. □

Now we let X be a finite dimensional space and K be a compact subset in X, $F: K \to L(X,Y)$ be continuous Fréchet differentiable and

$$G(x) = \langle F(x), x - K\rangle = \bigcup_{z\in K} \langle F(x), x - z\rangle.$$

Theorem 3.1 *Let* $\hat{x}, \overline{x} \in K, \hat{y} = \langle F(\hat{x}), \hat{x} - \overline{x}\rangle \in G(\hat{x})$ *and* $\lim_{\|x\|\to\infty} \|\langle F(\hat{x}), x\rangle\| = \infty$. *Then*

$$DG(\overline{x}, \overline{y})(x) = \langle \nabla F(\hat{x}), \hat{x} - \overline{x}\rangle + \bigcup_{x^*\in T(K\overline{x})} \langle F(\hat{x}), x - x^*\rangle.$$

Proof. Suppose $y \in DG(\hat{x}, \hat{y})(x)$. There exist sequences $\{(x_n, y_n)\} \subset X \times Y$ and $\{h_n\} \subset R_+ \setminus \{0\}$, such that $x_n, h_n \to (x, n), h_n \to 0$ and

$$\hat{y} + h_n y_n \in G(\hat{x} + h_n x_n) = \langle F(\hat{x} + h_n x_n), \hat{x} + h_n x_n, \hat{x} + h_n x_n - K\rangle$$

Therefore, there exists $\overline{x}_n \in K$ such that

$$\hat{y} + h_n y_n = \langle F(\hat{x} + h_n x_n), \hat{x} + h_n x_n - x_n\rangle,$$

and

$$\langle F(\hat{x}, \hat{x} - \overline{x})\rangle - \langle F(\hat{x} + h_n x_n), \hat{x} - \overline{x}_n\rangle = \langle F(\hat{x} + h_n x_n), h_n x_n\rangle - h_n y_n \quad (4)$$

Since F is continuously dfferentiable, by the Taylor expansion,

$$F(\hat{x} + h_n x_n) = F(\hat{x}) + h_n \nabla F(\hat{x}) x_n + o(h_n x_n). \quad (5)$$

It follows from (4) and (5) that

$$\langle F(\hat{x}), \overline{x}_n - \overline{x}\rangle = \langle h_n \nabla F(\hat{x}) x_n + o(h_n x_n), \hat{x} - \overline{x}_n\rangle + \langle F(\hat{x} + h_n x_n), h_n x_n - h_n y_n\rangle \quad (6)$$

Hence

$$\left\langle F(\hat{x}),(\overline{x}_n-\overline{x})/h_n\right\rangle=\left\langle \nabla F(\hat{x})x_n+o(h_nx_n)/h_n,\hat{x}-\overline{x}_n\right\rangle+\left\langle F(\hat{x}+h_nx_n),x_n\right\rangle-y_n \tag{7}$$

As K is compact set, we can assume, without loss of generality, that $\overline{x}_n \to x' \in K$. Obviously,

$$\begin{aligned}&\lim_{x\to\infty}\left(\left\langle \nabla F(\hat{x})x_n+o(h_nx_n)/h_n,x_n\right\rangle+\left\langle F(\hat{x}+h_nx_n),x_n\right\rangle-y_n\right)\\&=\left\langle \nabla F(\hat{x})x,\hat{x}-x'\right\rangle+\left\langle F(\hat{x}),x\right\rangle-y.\end{aligned} \tag{8}$$

Thus, it follows form (6) that sequence $\left\{\left\langle F(\hat{x}),(\hat{x}-\overline{x})/h_n\right\rangle\right\}$ is a convergent sequence. Let us consider two possible cases for the sequence $\left\{(\overline{x}_n-\overline{x})/h_n\right\}$.

Case I. There exists a subsequence $\left\{(\hat{x}-\overline{x})/h_n\right\}$, such that $\left\|(\overline{x}_{n'}-\overline{x})/h_n\right\|\to\infty$.

By the given assumption conditions, $\left\|\left\langle F(\hat{x}),(\overline{x}_{n'}-\overline{x})/h_{n'}\right\rangle\right\|\to\infty,$, which contradicts (8).

Case II. There exists $M>0$, such that

$$\left\|(\overline{x}_n-\overline{x})/h_n\right\|\le M,\quad \forall n \tag{9}$$

Since $h_n\to 0$, by (9), we have

$$x'=\overline{x}. \tag{10}$$

Since X is a finite dimensional space, we can assume that $(\overline{x}_n-x)/h_n\to\tilde{x}$. Thus, $\tilde{x}\in T(K,\overline{x})$ and it follows from (9) and (10) that

$$\left\langle F(\hat{x}),\tilde{x}\right\rangle=\left\langle \nabla F(\hat{x})x,\hat{x}-\overline{x}\right\rangle+\left\langle F(\hat{x}),x\right\rangle-y,$$

and so

$$\begin{aligned}y&=\left\langle \nabla F(\hat{x})x,\hat{x}-\overline{x}\right\rangle+\left\langle F(\hat{x}),x-\tilde{x}\right\rangle\\&\in\left\langle \nabla F(\hat{x})x,\hat{x}-\tilde{x}\right\rangle,+\bigcup_{x^*\in T(K\overline{x})}\left\langle F(\hat{x}),x'-x^*\right\rangle.\end{aligned}$$

Thus

$$DG(\hat{x},\hat{y})(x) \subset \langle \nabla F(\hat{x})x, \hat{x}-\overline{x}\rangle + \bigcup_{x^* \in T(K\overline{x})} \langle F(\hat{x}), x-x^*\rangle.$$

Conversely, we suppose that $x^* \in T(K,\overline{x})$ and

$$y = \langle \nabla F(\hat{x})x, \hat{x}-\overline{x}\rangle + \langle F(\hat{x}), x-x^*\rangle.$$

Thus, there exists $\{\overline{x}_n\} \subset K$ and $\{h_n\} \subset R_+ \setminus \{0\}$ such that

$$\overline{x}_n \to \overline{x}, h_n \to 0 \text{ and } (\overline{x}_n - \overline{x})/h_n \to x^*.$$

Take sequences $\{\overline{x}_n\} \subset X$ and $\{y_n\} \subset Y$ such that $x_n \to x$ and

$$y_n \langle (F(\hat{x}+h_n x_n) - F(\hat{x}))/h_n, \hat{x}-\overline{x}\rangle + \langle F(\hat{x}+h_n x_n), x_n - (\overline{x}_n - \overline{x})/h_n\rangle. \tag{11}$$

It follows from (11) that

$$y_n \to y$$

and

$$y + h_n y_n = \langle F(\hat{x}+h_n x_n), \hat{x}+h_n x_n - \overline{x}_n\rangle \in G(\hat{x}+h_n x_n).$$

So that

$$y \in DG(\hat{x},\hat{y})(x),$$

and this completes the proof. □

Now we discuss the relationship among contingent derivatives DW and DG.

Theorem 3.2 *Let* $\hat{x}, \overline{x} \in K$ *and* $\overline{y} = \langle F(\hat{x},)\hat{x}-\overline{x}\rangle \in W(\overline{x})$. *Suppose that*

$$\lim_{\|x\|\to\infty} \| < F(\hat{x}), x > \| = \infty.$$

Then

$$DW(\hat{x},\hat{y})(x) \subset Max_{intC} DG(\hat{x},\hat{y})(x).$$

Proof. Let $y \in DW(\hat{x},\hat{y})(x)$. Clearly, $y \in DG(\hat{x},\hat{y})(x)$. If $y \notin Max_{intC} DG(\hat{x},\hat{y})(x)$, then there exists $\overline{y} \in DG(\hat{x},\hat{y})(x)$ such that

$$\overline{y} - y \in intC. \tag{12}$$

Since $y \in DW(\hat{x},\hat{y})(x)$, there exists sequence $\{(x_n, y_n)\} \subset X \times Y$ and $\{h_n\} \subset R_+ \setminus \{0\}$, such that $(x_n, y_n) \to (x, y), h_n \to 0$ and

$$\hat{y} + h_n y_n \in W(\hat{x} + h_n x_n), \quad \forall n.$$

It follows from $\hat{y} \in DW(\hat{x},\hat{y})(x)$, and Proposition 3.1 that, for the above given sequence $\{h_n\} \subset R_+ \setminus \{0\}$, there exist a subsequence, without loss of generality, we still write as $\{h_n\}$ and sequence $\{(x_n, y_n)\} \subset X \times Y$ such that $(x_n, y_n) \to (x, y)$ and

$$\hat{y} + h_n \overline{y}_n \in G(\hat{x} + h_n \overline{x}_n), \quad \forall n.$$

Thus, there exists $x'_n \in K$ such that

$$\hat{y} + h_n \overline{y}_n = \langle F(\hat{x} + h_n \overline{x}_n), \hat{x} + h_n \overline{x} - x'_n \rangle \tag{13}$$

F being continuously differentiable at $\hat{x}$, we have

$$F(\hat{x} + h_n \overline{x}_n) = F(\hat{x}) + h_n \nabla F(\hat{x}) \overline{x}_n = o(h_n \overline{x}_n).$$
$$F(\hat{x} + h_n \overline{x}_n) = F(\hat{x}) + h_n \nabla F(\hat{x}) x_n = o(h_n x_n).$$

Since $\{\overline{x}_n\}$ and $\{x_n\}$ are two convergent sequences, $o(h_n \overline{x})/h_n \to \Theta$ and $o(h_n x)/h_n \to \Theta$. Thus, $(o(h_n \overline{x}) \to o(h_n x))h_n \to \Theta$. By $o(h_n)$, we denote $(o(h_n \overline{x}) - o(h_n x))$. Thus, we have

$$\begin{aligned}
&\left\langle F\left(\hat{x}+h_n\overline{x}_n\right),\hat{x}+h_nx_n-x'_n\right\rangle\\
&=\left\langle F\left(\hat{x}+h_nx_n\right),\hat{x}+h_n\overline{x}_n-x'_n\right\rangle+\\
&\left\langle h_n\nabla F(\hat{x})(\overline{x}_n-x_n)+o(h_n,)\hat{x}+h_n\overline{x}_n-x'_n\right\rangle\\
&=\left\langle F\left(\hat{x}+h_nx_n\right),\hat{x}+h_nx_n-x'_n\right\rangle+\\
&\left\langle F\left(\hat{x}+h_nx_n\right),h_n\left(\overline{x}_n-x_n\right)\right\rangle\left\langle h_n\nabla F\left(\overline{x}\right)\left(\overline{x}_n-x_n\right)+o\left(h_n,\right)\hat{x}+h_n\overline{x}_n-x'_n\right\rangle
\end{aligned}\tag{14}$$

Set

$$\begin{aligned}
\alpha(n)=\langle F\rangle\left\langle F\left(\hat{x}+h_nx_n\right),\overline{x}_n-x_n\right\rangle+\nabla F(\hat{x})\left(\overline{x}_n-x_n\right)+\\
\left(o\left(h_n\right)\right)/h_n,\hat{x}+h_nx_nx'_n.
\end{aligned}$$

Since $\left(\overline{x}_n - x_x\right)\to 0_X$ and $o\left(h_n\right)/h_n\to\Theta,\alpha(n)\to\Theta$. Thus by (13) and (14), we have

$$\hat{y}+h_n\overline{y}_n=\left\langle F\left(\hat{x}+h_nx_n\right),\hat{x}+h_nx_n-x'_n\right\rangle+h_n\alpha(n),$$

and so

$$\hat{y}+h_n\left(\overline{y}_n-\alpha(n)\right)\in G\left(\hat{x}+h_nx_n\right).$$

It follows from the definition of W that

$$\left(\hat{y}+h_n\left(\overline{y}_n-\alpha(n)\right)\right)-\left(\hat{y}+h_ny_n\right)\notin intC,$$

and

$$\hat{y}-\alpha(n)-y_n\notin intC.$$

Hence,

$$\hat{y}-y\notin intC.$$

which contradicts (12) and this completes the proof. □

Theorem 3.3 *Let* $\hat{x},\overline{x}\in K$ *and* $\hat{y}=\left\langle F(\hat{x}),\hat{x}-\overline{x}\right\rangle\in N(\hat{x})$. *Suppose that*

$$\lim_{\|x\|\to\infty} \|< F(\hat{x}), x >\| \to \infty.$$

Then

$$DN(\hat{x},\hat{y})(x) \subset Max_{intC} DG(\hat{x},\hat{y})(x).$$

Proof. Since $N \subset W$,

$$DN(\hat{x},\hat{y})(x) \subset DW(\hat{x},\hat{y})(x).$$

Thus, by Theorem 3.2, the conclusion follows readily.

Lemma 3.2 *Let $\hat{y}$ be a maximal point of $G(\hat{x})$ and C have a compact base. Suppose*

$$DG(\hat{x},\hat{y})(0_X) \cap C = \{0\}. \tag{15}$$

Then

$$D(G-C)(\hat{x},\hat{y})(x) = DG(\hat{x},\hat{y}) - C,$$

where $(G-C)(x) = G(x) - C$.

Proof. Let $y \in D(G-C)(\hat{x},\hat{y})(x)$. Thus, there exist sequences $\{(x_n, y_n)\} \subset X \times Y$, $\{h_n\} \subset R_+ \setminus \{0\}$ and $\{d_n\} \subset C$, such that

$$(x_n, y_n) \to (x, y), \quad h_n \to 0$$

and

$$\hat{y} + h_n y_n + \alpha_n \in G(\hat{x} + h_n x_n), \quad \forall n \tag{16}$$

Therefore, there exists $\overline{x}_n \in K$, such that

$$\hat{y} + h_n y_n + d_n = \langle (\hat{x} + h_n x_n), \hat{x} + h_n x_n - \overline{x}_n \rangle. \tag{17}$$

K being a compact set, we can assume, without loss of generality, that $\overline{x}_n \to x' \in K$. By (17), sequence $\{d_n\}$ is a convergent one. Suppose $d_n \to d$. Then, we have

$$\hat{y} + d = \langle F(\hat{x}), \hat{x} - x' \rangle.$$

If $d \neq 0$, then this contradicts the fact that $\hat{y}$ is a maximal point of $G(\hat{x})$. Therefore, $d_n \to \theta$. Let us consider two possible cases for sequence $\{d_n\}$.

Case I. There exists n_0 such that $d_n = \theta$, for $n \geq n_0$. By the definition of the contingent derivative, $y \in DG(\hat{x}, \hat{y})(x)$.

Case II. There exists a subsequence, without loss of generality, we still write as d_n such that $d_n \neq \theta$, for all n.

Now, we assert that the sequence $\{\| d_n \| / h_n\}$ is bounded. Indeed, suppose that the sequence $\{\| d_n \| / h_n\}$ is unbounded. Without loss of generality, we assume that $\| d_n \| / h_n \to \infty$. Since C has a compact base, by Lemma 3.1 in [13], we may assume that

$$d_n / \|d_n\| \to d' \in C \setminus \{\theta\}. \tag{18}$$

Thus, we have

$$\left(h_n / \|d_n\|\right) y_n + d_n / \|d_n\| \to d',$$

and

$$\left(h_n / \|d_n\|\right) x_n \to 0_X.$$

It follows from (16) that

$$\left(\hat{y} + \|d_n\| \left(h_n / \|d_n\|\right) y_n + d_n / \|d_n\|\right) \in G\left(\hat{x} + \|d_n\| \left(h_n / \|d_n\|\right) x_n\right).$$

Therefore,

$$d' \in DG(\hat{x}, \hat{y})(0_x),$$

which contradicts (15).

Thus, the sequence $\{\|d_n\| / h_n\}$ is bounded and we can assume

$$\|d_n\| / h_n \to \alpha \geq 0. \tag{19}$$

By (16), we have

$$\hat{y} + h_n(y_n(\|d_n\|/h_n) + (d_n/\|d_n\|) \in G(\hat{x} + h_n x_n), \quad \forall n.$$

By (18), (19) and the definition of contingent derivative,

$$y + \alpha d' \in DG(\hat{x}, \hat{y})(x),$$

and so

$$D(G - C)(\hat{x}, \hat{y})(x) \subset DG(\hat{x}, \hat{y}) - C,$$

Conversely, by Proposition 2.1 of Tanino [12],

$$DG(\hat{x}, \hat{y})(x) - C \subset D(G - C)(\hat{x}, \hat{y})(x).$$

Thus, the conclusion follows readily. □

Theorem 3.4 *Let $\hat{y}$ be a maximal point of $G(\hat{x})$ and let C has a compact base. Suppose*

$$DG(\hat{x}, \hat{y})(0_x) \cap C = \{\theta\}.$$

Then

$$Max_C (G - C)(\hat{x}, \hat{y})(x) = Max_C DG(\hat{x}, \hat{y})(x).$$

Proof. It follows from Lemma 3.2 that

$$D(G - C)(\hat{x}, \hat{y})(x) = DG(\hat{x}, \hat{y})(x) - C.$$

Therefore,

$$\begin{aligned} Max_C D(G - C)(\hat{x}, \hat{y})(x) &= Max_C (DG(\hat{x}, \hat{y})(x) - C) \\ &= Max_C DG(\hat{x}, \hat{y})(x), \end{aligned}$$

and this completes the proof. □

Theorem 3.5 *Suppose that the following conditions are satisfied:*

(i) $\hat{y} \in N(\hat{x})$;

(ii) C *has a compact base;*

(iii) $\lim_{\|x\| \to \infty} \| < F(\hat{x}), x > \| \to \infty$;

(iv) $DG(\hat{x}, \hat{y})(0_X) \cap C = \{\theta\}$.

Then

$$Max_C DG(\hat{x}, \hat{y})(x) \subset DN(\hat{x}, \hat{y})(x) \subset Max_{intC} DG(\hat{x}, \hat{y})(x).$$

Proof. Since $N(x) \subset G(x)$ for all $x \in K$, it follows from (iv) that

$$DN(\hat{x}, \hat{y})(0_X) \cap C = \{\theta\}.$$

K being a compact set, $G(x)$ is also a compact set for any $x \in K$. Thus, by Lemma 2.3 in Li, Chen and Lee [8], $G(x) - C = N(x) - C$. It follows from Theorem 3.4 that

$$\begin{aligned} Max_C DN(\hat{x}, \hat{y})(x) &= Max_C D(N - C)(\hat{x}, \hat{y})(x) \\ &= Max_C D(G - C)(\hat{x}, \hat{y})(x) \\ &= Max_C DG(\hat{x}, \hat{y})(x). \end{aligned}$$

Obviously,

$$Max_C DN(\hat{x}, \hat{y})(x) \subset DN(\hat{x}, \hat{y})(x).$$

On the other hand, we have

$$DN(\hat{x}, \hat{y})(x) \subset Max_{intC} DG(\hat{x}, \hat{y})(x),$$

by Theorem 3.3. Thus, the result of this theorem holds. □

Lemma 3.3 *Suppose that* C *has a base, and* $\tilde{C}$ *is a nonempty closed and convex cone with* $\tilde{C} \setminus \{\theta\} \subset C$. *Then,* $\tilde{C}$ *is also a base.*

Proof. By Lemma 3.1 in Shi [13], this result holds.

Theorem 3.6 *Suppose that the following conditions are satisfied:*

(i) $\hat{y}$ *is a maximal point of* $G(\hat{x})$;

(ii) C *has a compact base and there exists a nonempty closed convex cone* $\tilde{C}$ *such that* $\tilde{C} \setminus \{\theta\} \subset intC$;

(iii) $\lim_{\|x\|\to\infty} \|< F(\hat{x}), x >\| \to \infty$;

(iv) $DG(\hat{x},\hat{y})(0_X) \cap intC = \varnothing$.

Then

$$\begin{aligned} DN(\hat{x},\hat{y})(x) &= DG(\hat{x},\hat{y})(x) \\ &= Max_{intC} DG(\hat{x},\hat{y})(x) \\ &= \left\langle \nabla F(\hat{x})x, \hat{x} - \overline{x} \right\rangle + Max_{intC}\left(\bigcup_{x^* \in T(K\overline{x})} F(\hat{x})x - x^* \right). \end{aligned}$$

Proof. Since $N(x) \subset W(x)$, by Theorem 3.1 and Theorem 3.3 we only need prove that

$$Max_{intC} DG(\hat{x},\hat{y})(x) \subset DN(\hat{x},\hat{y})(x).$$

It follows from (ii) and (iv) that

$$DN(\hat{x},\hat{y})(0_X) \subset \tilde{C} = \{\theta\}$$

K being a compact set, $G(x)$ is also a compact set of any $x \in K$. Thus, by Lemma 2.3 in Li, Chen and Lee [8], $G(x) - \tilde{C} = N(x) - \tilde{C}$. It follow from Lemma 3.2 and Theorem 3.4 that

$$\begin{aligned} Max_{\tilde{C}} DN(\hat{x},\hat{y})(x) &= Max_{\tilde{C}} D\left(N - \tilde{C}\right)(\hat{x},\hat{y})(x) \\ &= Max_{\tilde{C}} D\left(G - \tilde{C}\right)\left(G - \tilde{C}\right)(\hat{x},\hat{y})(x) \\ &= Max_{\tilde{C}} DG(\hat{x},\hat{y})(x). \end{aligned}$$

Since $\tilde{C} \setminus \{\theta\} \subset intC$,

$$Max_{intC} DG(\hat{x},\hat{y})(x) \subset Max_{\tilde{C}} DG(\hat{x},\hat{y})(x).$$

Therefore,

$$Max_{intC} DG(\hat{x},\hat{y})(x) \subset Max_{\tilde{C}} DN(\hat{x},\hat{y})(x) \subset DN(\hat{x},\hat{y}),$$

and the result of this theorem holds. □

Corollary 3.1 *Suppose that the following conditions are satisfied:*

(i) $\hat{y}$ *is a weakly maximal point of* $G(\hat{x})$;

(ii) C *has a compact base and there exists a nonempty closed convex cone* $\tilde{C}$ *such that* $\tilde{C} \setminus \{\theta\} \subset intC$;

(iii) $\lim_{\|x\|\to\infty} \left\|\left\langle F(\hat{x}), x\right\rangle\right\| = \infty$;

(iv) $DG(\hat{x},\hat{y})(0_X) \cap intC = \varnothing$.

Then

$$DW(\hat{x},\hat{y})(x) == \left\langle \nabla F(\hat{x}), \hat{x} - \overline{x}\right\rangle + Max_{intC}\left(\bigcup_{x^*\in T(K\overline{x})} \left\langle F(\hat{x})x - x^*\right\rangle\right).$$

Proof. By Theorem 3.2, this result holds. □

4. CHARACTERIZATIONS OF SOLUTIONS FOR (VVI) AND (WVVI)

In this section we consider characterizations of solutions for vector variational inequalities and weak vector variational inequalities in terms of gap functions.

Theorem 4.1 *Let* $\hat{x} \in K$ *and* $\lim_{\|x\|\to\infty} \left\|\left\langle F(\hat{x}), x\right\rangle\right\| = \infty$. *If* $\hat{x}$ *is a solution of (WVVI), then*

$$DG(\hat{x},\theta)(0_X) \cap intC = \varnothing$$

Proof. Obviously, we have

$$\left\langle F(\hat{x}), x - \hat{x}\right\rangle \notin -intC,\ \forall x \in K \Leftrightarrow \left\langle F(\hat{x}), x\right\rangle \notin intC,$$
$$\forall x \in cl\left(\bigcup_{h>0} \frac{1}{h}(K - \hat{x})\right),$$

and

$$T(K,\hat{x}) \subset cl\left(\bigcup_{h>0} \frac{1}{h}(K - \hat{x})\right).$$

It follows from Theorem 3.1 that

$$DG(\hat{x},\theta)(0_X) = \bigcup_{x^* \in T(K\bar{x})} \langle F(\hat{x})x - x^* \rangle.$$

Thus, this conclusion follows readily. □

Remark 4.1 *If K is a compact and convex set, then, by Proposition 5 in Aubin and Ekeland [1], we have*

$$T(K,\hat{x}) = cl\left(\bigcup_{h>0} \frac{1}{h}(K - \hat{x})\right).$$

Thus, under the conditions of Theorem 4.1, $\hat{x}$ is a solution of (WVVI) if and only if

$$DG(\hat{x},\theta)(0_X) \cap intC = \varnothing.$$

Theorem 4.2 *Suppose that the following conditions are satisfied:*

(i) $\hat{x}$ is solution of (WVVI);

(ii) C has a compact base and there exists a closed and convex cone $\tilde{C}$ such that $\tilde{C}\{0\} \subset intC$;

(iii) $\lim_{\|x\| \to \infty} \|\langle F(\hat{x}), x \rangle\| = \infty$.

Then

$$DW(\hat{x},\theta)(x) = Max_{intC} DG(\hat{x},\theta)(x)$$

$$= Max_{intC}\left(\bigcup_{x^* \in T(K,\hat{x})} \langle F(\hat{x}), x - x^* \rangle\right).$$

Proof. Since $\hat{x}$ is a solution of (WVVI), θ is a weakly maximal point $G(\hat{x})$. If follows from Theorem 4.1 that

$$DG(\hat{x},\theta)(0_X) \cap intC = \varnothing$$

Thus, by Corollary 3.1, the conclusion follows readily. □

Theorem 4.3 *Suppose that the following conditions are satisfied:*

(i) $\hat{x}$ is a solution of (VVI);

(ii) C *has a compact base and there exists a closed and convex cone* $\tilde{C}$ *such that* $\tilde{C}\{\theta\} \subset intC$;

(iii) $\lim_{\|x\|\to\infty} \|\langle F(\hat{x}), x\rangle\| = \infty$.

$$\begin{aligned} DN(\hat{x},\theta)(x) &= DW(\hat{x},\theta)(x) \\ &= Max_{intC} DG(\hat{x},\theta)(x) \\ &= Max_{intC}\left(\bigcup_{x^*\in T(K,\hat{x})} \langle F(\hat{x}), x - x^*\rangle\right). \end{aligned}$$

Proof. Since $\hat{x}$ is a solution of (VVI), θ is a maximal point of $G(\hat{x})$. It follows from Theorem 4.1 that

$$DG(\hat{x},\theta)(0_X) \cap intC = \varnothing.$$

Thus, by Theorem 3.6, the conclusion follows readily. □

REFERENCES

[1] Aubin, J. P. and Ekeland, I. (1984) *Applied Nonlinear Analysis*, John Wiley and Sons, New York.

[2] Aubin, J. P. and Frankswska, H. (1990) *Set-valued Analysis*, Birkhauser, Boston.

[3] Auslender, A. (1976) *Optimization: Methods Numeriques*, Masson Paris.

[4] Chen, G. Y., Goh, C. J. and Yang, X. Q. (2000) On gap functions for vector variatonal inequalities, in Giannessi, *Vector Variational Inequalities and Vector Equilibria: Mathematical Theories*, Klumer, 55-72.

[5] Chen, G. Y. and Yang, X. Q. (1990) The vector complementary problem and its equivalences with vector minimal element in ordered spaces, *J. Math. Anal. Appl.*, 153, 136-158.

[6] Giannessi, F. (1980) Theorems of alternative, quadratic programs and complementary problems, in Cottle, R. W., Giannessi, F. and Lions, J. L., *Variational Inequality and Complementary Problems*, Wiley and Sons, New York.

[7] Harker, P. T and Pany, J. S. (1990) Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications, *Mathematical Programming*, 48, 161-220.

[8] Li, S. J., Chen, G. Y. and Lee G. M. (2000) Minimax theorems for set-valued mappings, *J. Optim. Theory Appl.*, 106, 183-200.

[9] Li, S. J., Chen, G. Y. and Teo, K. L. (2002) On the stablity of generalized vector quasi variational inequality problems, *J. Optim. Theory Appl.*, 113, 283-295.

[10] Mosco, U. (1972) Dual variational inequalities, *J. Math. Anal. Appl.*, 40, 202-206.

[11] Sawaragi, Y, Nakayama, H. and Tanino, T. (1985) *Theory of Multiobjective Optimization*, Academic Press, New York.

[12] Tanino, T. (1998) Sensitivity analysis in multiobjective optimization, *J. Optim. Theory Appl.*, 56, 479-499.

[13] Shi, D. S. (1991) Contingent derivative of the perturvation map in multiobjective optimization, *J. Optim. Theorey Appl.*, 70, 385-396.

ZERO GRAVITY CAPILLARY SURFACES AND INTEGRAL ESTIMATES

G.M. Lieberman
Dept. of Mathematics, Iowa State University, Ames, Iowa, U.S.A.

INTRODUCTION

Let Ω be a bounded domain in R^n and write γ for the unit inner normal to $\partial\Omega$. In [1], Finn showed that the capillary problem in zero gravity

$$\operatorname{div}\left(\frac{Du}{(1+|Du|^2)^{1/2}}\right) = K \text{ in } \Omega, \tag{0.1a}$$

$$\frac{Du}{(1+|Du|^2)^{1/2}} \cdot \gamma = \varphi \text{ on } \partial\Omega \tag{0.1b}$$

has a solution for constants K and φ satisfying $|\varphi|<1$ and $K|\Omega| = -\varphi|\partial\Omega|$ (otherwise there can be no solution) provided there is a vector field w such that

$$\operatorname{div} w = K \text{ in } \Omega, \tag{0.2a}$$

$$w \cdot \gamma = \varphi \text{ on } \partial\Omega, \tag{0.2b}$$

$$\sup_{\Omega} |w| < 1. \tag{0.2c}$$

Finn's goal was to simplify a geometric condition (due to Giusti [6]) which implies the existence of a solution to this problem. In [1, Section 2], Finn showed that the vector field w can be constructed explicitly when Ω is a suitable polygon (or polyhedron); however, his trapezoid example (see pages 8 and 9 of [1]), which shows that the geometric condition for existence is quite subtle, was not described in terms of this vector field. This approach was quickly replaced by one closer in spirit to Giusti's original condition: the introduction of a subsidiary variational problem (see [3, Theorem 5.1] and [2, Chapter 6]).

Here we revisit the vector field approach from a different point of view. We shall show that the vector field criterion is an easy consequence of a uniform L^1 estimate for a related family of problems (see (1.2) below), and this uniform estimate is our main concern. We show that a corresponding result holds in a more general setting (for example, with nonconstant φ and K) and we derive some useful consequences. We prove this uniform L^1 estimate in Section 1 and thus infer the existence result, which follows also from the method of [2, Chapter 7]. Some corollaries of this estimate, motivated by results in [1] are discussed in Section 2. We also examine the L^1 estimate and a corresponding L^∞ estimate for more general boundary value problems in Section 3. In particular, we improve the corresponding estimates in [11 Section 3]. The examples given in Section 4 illustrate the nature of our results, and some related existence theorems are given in Section 5.

A crucial tool is Poincaré's inequality which we shall use in the following form. (See [15, Lemma 1.65] for a proof in this general setting.) If Ω is a bounded Lipschitz domain, then there is a constant k, determined only by Ω, such that

$$\int_\Omega |u|\,dx \le k \int_\Omega |Du|\,dx$$

for any function $u \in W^{1,1}(\Omega)$ with

$$\int_\Omega u\,dx = 0. \tag{0.3}$$

In fact, this theorem is true for a more general class of domains, but we shall not be concerned with generality. In addition, there is an analogous statement if $u \in W^{1,m}$ which we shall refer to in a later section.

1. THE MAIN THEOREM

In this section, we prove the following result.

Theorem1.1 *Let Ω be a bounded domain in $\mathbb{R}^n$ with $\partial\Omega \in C^2$, let $K \in C^1(\overline{\Omega})$, and let $\varphi \in C^1(\partial\Omega)$ with $|\varphi|<1$ on $\partial\Omega$. Suppose also that*

$$\int_\Omega K\,dx + \int_{\partial\Omega} \varphi\,ds = 0. \tag{1.1}$$

If there is a vector field w satisfying (0.2), then there is a classical solution to (0.1).

Proof. Let u_ε be the solution of

$$\operatorname{div}\left(\frac{Du_\varepsilon}{(1+|Du_\varepsilon|^2)^{1/2}}\right) = \varepsilon u_\varepsilon + K \text{ in } \Omega, \tag{1.2a}$$

$$\frac{Du_\varepsilon}{(1+|Du_\varepsilon|^2)^{1/2}} \cdot \gamma = \varphi \text{ on } \partial\Omega \tag{1.2b}$$

given by [17, Theorem 1] or Theorem 7.2 and example 1 of [11]. (Note that these references assume more smoothness for $\partial\Omega$ and ϕ than we assume here. An easy approximation argument gives the result in exactly the form required here.) It suffices to show that the set $\{u_\varepsilon\}$ is uniformly bounded in $C^{1,\alpha}(\overline{\Omega})$ for some $\alpha > 0$.

We first note that Lemma 3.3, Example 1, and Lemma 5.2 of [11] reduce this problem to a uniform L^1 estimate on u_ε, which we now prove. For brevity, we define

$$A_\varepsilon = \frac{Du_\varepsilon}{(1+|Du_\varepsilon|^2)^{1/2}} - w,$$

Then $\operatorname{div} A_\varepsilon = \varepsilon u_\varepsilon$ in Ω and $A_\varepsilon \cdot \gamma = 0$ on $\partial\Omega$. We multiply the differential equation by u_ε and integrate over Ω. Applying the divergence theorem along with the boundary condition, we infer that

$$\int_\Omega A_\varepsilon \cdot Du_\varepsilon\,dx + \varepsilon \int_\Omega u_\varepsilon^2\,dx = 0,$$

so

$$\int_\Omega A_\varepsilon \cdot Du_\varepsilon \, dx \le 0,$$

Now we observe that, for $\sigma = \sup |w|$, we have $A_\varepsilon \cdot Du_\varepsilon \ge (1-\sigma)|Du_\varepsilon| - 1$, and hence

$$\int_\Omega |Du_\varepsilon| \, dx \le \frac{|\Omega|}{1-\sigma}.$$

Next, we integrate the equation $\operatorname{div} A_\varepsilon = \varepsilon u_\varepsilon$ over Ω and apply the divergence theorem and the boundary condition to see that

$$\int_\Omega u_\varepsilon \, dx = 0.$$

Poincaré's inequality then yields the L^1 bound on u_ε. □

Note that there is no restriction on the signs of K and φ other than the obvious one that they can't both be everywhere positive or everywhere negative. Moreover, if w_0 is a smooth vector and if $K_0 = \operatorname{div} w_0$ and $\varphi_0 = w \cdot \gamma$, with $W = \sup |w_0|$, then for any $\eta \in (-1/W, 1/W)$, the functions $K = \eta K_0$ and $\varphi = \eta \varphi_0$ satisfy all the hypotheses of this theorem with $w = w_0/\eta$. Because the vector field w is often difficult to determine in practice, we give an alternative sufficient condition for (0.1) to have a solution in Section 3.

2. CONSEQUENCES OF THE MAIN THEOREM

In fact, the existence of w is also a necessary condition for the existence of a capillary surface in zero gravity under the additional hypotheses that $\partial\Omega$ is smooth and $\sup|\varphi| < 1$. Such a result is already known (Theorem 2 in [1]) for constant K and φ. Here, we also obtain a regularity result for this surface.

Corollary 2.1. *Let Ω be a bounded domain in $\mathbb{R}^n$ with $\partial\Omega \in C^2$, let $K \in C^1(\Omega)$, and let $\varphi \in C^1(\partial\Omega)$ with $|\varphi| < 1$ on $\partial\Omega$. Suppose also that* (1.1) *holds. If there is a bounded weak solution u of* (0.1)*, then there is a function w satisfying* (0.3)*. Moreover, $u \in C^{1,\alpha}(\overline{\Omega})$ for any $\alpha \in (0,1)$.*

Proof. We start with a simple maximum estimate for $|u_\varepsilon|$ (see, for example, [17, Lemma] or [4, Theorem 1]). Let $v_1 = u - \inf u$ and $v_2 = u - \sup u$. Then

the comparison principle [2, Theorem 5.1] gives us $v_1 \geq u_\varepsilon \geq v_2$. From this uniform bound, we infer a uniform (for each fixed $\alpha \in (0,1)$) $C^{1,\alpha}$ estimate for u_ε, and hence there is a convergent subsequence with limit u_0, which is a solution of (0.1). The uniform estimates on u_ε imply that $|Du_0|$ is bounded, so

$$w = \frac{Du_0}{(1+|Du_0|^2)^{1/2}}$$

satisfies (0.2).

The comparison principle implies that $u - u_0$ is constant, so u has the same regularity as u_0. In particular, $u \in C^{1,\alpha}$. []

The uniform gradient estimate on u_ε can be used to correct a misstatement in [9]. In that work, the author asserted that the gradient bound in [17, Theorem 2] for a solution of the capillary equation

$$\operatorname{div}\left(\frac{Du}{(1+|Du|^2)^{1/2}}\right) = ku \text{ in } \Omega, \tag{2.1a}$$

$$\frac{Du}{(1+|Du|^2)^{1/2}} \cdot \gamma = \varphi \text{ on } \partial\Omega \tag{2.1b}$$

is independent of the positive constant k and the maximum of u. In fact, this theorem gives a bound which is independent of k but does depend on the maximum of u. However, if there is a solution of the zero gravity problem (0.1) with constant K and φ, then the uniform gradient bound from Corollary 2.1 shows that the gradients of the functions labeled z and v in [9] are bounded independent of k (because $z = u$ and $v = u_\varepsilon$ with $\varepsilon = k$, in our notation).

In addition, our results apply to various nonsmooth domains. Such domains have been studied intensively using variational methods, and we refer to [2], particularly Chapters 6 and 7, for a survey of results up to 1985. Here, we consider only a simple situation in which the method described here can be applied easily. Let $\Omega \subset \mathbb{R}^2$ be a piecewise smooth domain satisfying a uniform exterior circle condition. It is not difficult to see that the gradient estimates in [14] can be rewritten with our vector field w in place of the product $\overline{\varphi\gamma}$ there because φ is a function of x alone. Therefore Theorem 1.1 and Corollary 2.1 hold in this case as well.

3. RESULTS FOR OTHER BOUNDARY VALUE PROBLEMS

For a large class of boundary value problems, it is well known that existence questions are reduced to appropriate *a priori* estimates. Here, we consider a class of problems modeled on the capillary problem in zero gravity. Specifically, we look at the problem

$$\operatorname{div} A(x,u,Du) + B(x,u,Du) = 0 \text{ in } \Omega, \tag{3.1a}$$

$$A(X,u,Du)\cdot\gamma + \psi(x,u) = 0 \text{ on } \partial\Omega \tag{3.1b}$$

under various hypotheses on the functions A, B, and ψ. The theory developed in [17] and continued in [11] and [13] provides a large list of combinations of hypotheses that imply $C^{1,\alpha}$ estimates in terms of L^1 estimates, so (as in Theorem 1.1) our main concern will be with the L^1 estimate. Thus, we use the weak form of (3.1) with the test function u:

$$\int_\Omega Du\cdot A(x,u,Du)\,dx = \int_\Omega uB(x,u,Du)\,dx + \int_{\partial\Omega} u\psi(x,u)\,ds. \tag{3.2}$$

To state our hypotheses, we use z and p as dummy variables for $u(x)$ and $Du(x)$. We assume that there are constants $a_0 > 0$ and a_3 along with functions $a_1 \in L^\infty(\Omega)$, $a_2 \in L^1(\Omega)$, $\psi_1 \in L^\infty(\partial\Omega)$, $\psi_2 \in L^1(\Omega)$, and $f \in C(\mathbb{R})$ such that

$$p\cdot A(x,z,p) - zB(x,z,p) \geq a_0 \mid p \mid - a_1(x)f(z) - a_2(x) + a_3 \mid z \mid, \tag{3.3a}$$

$$z\psi(x,z) \leq \psi_1(x)f(z) + \psi_2(x), \tag{3.3b}$$

$$\int_\Omega a_1(x)\,dx + \int_{\partial\Omega} \psi_1(x)\,ds = 0. \tag{3.3c}$$

We also assume that f is uniformly Lipschitz on $\mathbb{R}$. (The cases $f(u) = u$ and $f(u) = \mid u \mid$ will be those of most interest.) We also set

$$A_0 = \int_\Omega a_2\,dx + \int_{\partial\Omega} \psi_2\,dx$$

Further restrictions on the structure will be made presently.

We then have the following estimate.

Theorem 3.1 *Let* $\partial\Omega \in C^{0,1}$ *and let* $u \in W^{1,1}(\Omega)$ *satisfy (3.2). Suppose (3.3) holds and that* $Du \cdot A(x,u,Du)$ *and* $uB(x,u,Du)$ *are in* $L^1(\Omega)$. *Suppose also that there is a vector field* $w_1 \in L^\infty(\Omega) \cap W^{1,1}(\Omega)$ *satisfying*

$$\operatorname{div} w_1 = a_1 \text{ in } \Omega, \tag{3.4a}$$

$$\psi_1 = w_1 \cdot \gamma \text{ on } \partial\Omega, \tag{3.4b}$$

$$a_0 - \sup_{\Omega} |w_1| \sup_{\mathbb{R}} |f'| \geq \mu \tag{3.4c}$$

for some nonnegative constant μ. *If* $\mu > -ka_3$ *and* $\int_\Omega u\,dx = 0$, *then*

$$\int_\Omega |u|\,dx \leq \frac{k}{\mu + ka_3} A_0. \tag{3.5}$$

Proof. Define

$$\overline{A}(x,z,p) = A(x,z,p) + \frac{f(z)}{z} w_1(x),$$

$$\overline{B}(x,z,p) = B(x,z,p) - a_1(x)\frac{f(z)}{z} - p \cdot w_1(x)\left(\frac{f'(z)}{z} - \frac{f(z)}{z^2}\right),$$

$$\overline{\psi}(x,z) = \psi(x,z) - \frac{f(z)}{z}\psi_1(x).$$

Then $\operatorname{div}\overline{A}(x,u,Du) + \overline{B}(x,u,Du) = 0$ in Ω and $\overline{A}(x,u,Du)\cdot\gamma + \overline{\psi}(x,u) = 0$ on $\partial\Omega$. In addition, we have

$$p \cdot \overline{A}(x,z,p) - z\overline{B}(x,z,p) \geq \mu |p| - a_2(x) + a_3 |z|,$$

$$z\overline{\psi}(x,z) \leq \psi_2(x).$$

From (3.2) (with $\overline{A}$ replacing A, $\overline{B}$ replacing B and $\overline{\psi}$ replacing ψ), we see that

$$\mu \int_\Omega |Du|\,dx + a_3 \int_\Omega |u|\,dx \leq A_0. \tag{3.6}$$

The proof is completed by using Poincaré's inequality and simple algebra. □

In general, we have no reason to expect the mean value of u to vanish. When a_3 is positive, we can obtain an L^1 estimate by only slightly varying the preceding argument.

Theorem 3.2 *Let $\partial\Omega \in C^2$ and let u satisfy* (3.2). *Suppose* (3.3) *holds with $a_3 > 0$. Suppose also that there is a vector field w_1 satisfying* (3.4) *with $\mu \geq 0$. Then*

$$\int_\Omega |u| dx \leq \frac{A_0}{a_3}. \tag{3.7}$$

Proof. Just as in Theorem 3.1, we infer (3.6), which immediately implies (3.7). □

As the main reason for deriving L^1 estimates is to infer an L^∞ estimate, we now derive an L^1 estimate under slightly stronger regularity hypotheses on u. These hypotheses will allow equations like

$$\operatorname{div}\left(\frac{Du}{(1+|Du|^2)^{1/2}}\right) + b_0(1+|Du|^2)^{1/2}\frac{u}{1+u^2} - ku = f(x)$$

for any positive constants b_0 and k. Specifically, we assume that there are positive constants $q \geq 1$, a_0, a_3, and M; a function f which is uniformly Lipschitz function on $\mathbb{R}$ with $f(0) = 0$; and functions a_1, b_1, and ψ_1 such that

$$p \cdot A(z,x,p) \geq a_0 |p| - a_1(x) f(z), \tag{3.8a}$$

$$zB(x,z,p) \leq b_0(p \cdot A(x,z,p) + a_1(x) f(z)) + b_1(x) f(z) - a_3 |u|^{2-q}, \tag{3.8b}$$

$$z\psi(x,z) \leq \psi_1(x) f(z) \tag{3.8c}$$

for $|z| \geq M$. We also assume that there is a vector field w such that

$$\operatorname{div} w = b_1 + q a_1 \text{ in } \Omega, \tag{3.9a}$$

$$\psi_1 = w \cdot \gamma \text{ on } \partial\Omega, \tag{3.9b}$$

$$a_0 - \sup_\Omega |w| \sup_R |f'| \geq b_0/q. \tag{3.9c}$$

We then have the following estimate.

Theorem 3.3 *Let $\partial\Omega \in C^{0,1}$ and let u satisfy (3.1) with $|u|^{q-1} Du \cdot A$ and $|u|^{q-1} uB$ in $L^1(\Omega)$. Suppose(3.8) holds with a_1, b_1 and ψ_1 bounded functions. If there is a vector field w such that (3.9) holds, then*

$$\int_\Omega |u| dx \leq \frac{q}{a_3} M^q \int_\Omega a_1 dx + M|\Omega|. \tag{3.10}$$

Proof. Now we use the test function $(|u|^q - M^q)_+ \operatorname{sgn} u$ and we set $\Omega(M) = \{x \in \Omega : |u(x)| \geq M\}$ to see that

$$\int_{\Omega(M)} q|u|^{q-1} Du \cdot A dx = \int_\Omega (|u|^q - M^q)_+ \operatorname{sgn} uB dx + \int_{\partial\Omega} (|u|^q - M^q)_+ \operatorname{sgn} u\psi ds.$$

On $\Omega(M)$, we have

$$a_0(q - b_0)|u|^{q-1}|Du| \leq \left(q|u|^{q-1} - b_0 \frac{|u|^q - M^q}{|u|} \right)(Du \cdot A + a_1 f(z)),$$

and then applying (3.8) gives

$$\begin{aligned} \int_{\Omega(M)} a_0(q - b_0)|u|^{q-1}|Du| dx + a_3 \int_{\Omega(M)} |u| dx \\ \leq \int_\Omega g(u)(b_1 + qa_1) dx + \int_{\partial\Omega} g(u)\psi_1 ds \\ + qM^q \int_\Omega a_1 dx + a_3 \int_{\Omega(M)} \frac{M^q}{|u|^{q-1}} dx. \end{aligned}$$

for

$$g(z) = (|z|^q - M^q)_+ \frac{f(z)}{|z|}.$$

It follows from (3.9a,b) that

$$\int_\Omega g(u)(b_1 + qa_1) dx + \int_{\partial\Omega} g(u)\psi_1 ds = -\int_\Omega g'(u) w \cdot Du dx$$

An elementary calculation shows that $|g'(z)| \leq q|z|^{q-1} \sup|f'|$ and hence, due to (3.9c), we see that

$$a_3 \int_{\Omega(M)} |u| dx \le qM^q \int_\Omega a_1 \, dx + a_3 |\Omega(M)| M.$$

The result now follows by adding to this inequality the obvious inequality

$$a_3 \int_{\Omega \setminus \Omega(M)} |u| dx \le a_3 |\Omega \setminus \Omega(M)| M.$$

□

We can also use these ideas to prove an L^∞ estimate similar to the one in [11, Lemma 3.3].

Theorem 3.4 *Let a_0 be a positive constant, let α_1, b_0, β_1, and M be nonnegative constants, and let b_1 and ψ_1 be bounded functions. Suppose that A, B, and ψ satisfy the structure conditions*

$$p \cdot A(x,z,p) \ge a_0 |Du| - \alpha_1 |z|, \tag{3.11a}$$

$$zB(x,z,p) \le b_0[p \cdot A(x,z,p) + \alpha_1 |z|] + b_1(x)|z| + \beta_1 |z|, \tag{3.11b}$$

$$z\psi(x,z) \le \psi_1(x)|z|. \tag{3.11c}$$

Suppose also that there is a vector field w such that

$$\operatorname{div} w = b_1 \text{ in } \Omega, \tag{3.12a}$$

$$w \cdot \gamma = \psi_1 \text{ on } \partial\Omega, \tag{3.12b}$$

$$\sup_\Omega |w| < a_0. \tag{3.12c}$$

If u is a bounded weak solution of (3.1), then

$$\sup_\Omega |u| \le C(a_0, \alpha_1, b_0, \beta_1, \sup_\Omega |w|, \Omega)\left(\int_\Omega |u| dx + M\right). \tag{3.13}$$

Proof. Now we define

$$\overline{A}(x,z,p) = A(x,z,p) + (\operatorname{sgn} z) w_1(x),$$

$$\overline{B}(x,z,p) = B(x,z,p) - b_1(x) \operatorname{sgn} z,$$

$$\overline{\psi}(x,z) = \psi(x,z) - (\operatorname{sgn} z)\psi_1(x).$$

Then $div\, \overline{A}(x,u,Du) + \overline{B}(x,u,Du) = 0$ in Ω and $\overline{A}(x,u,Du) \cdot \gamma + \overline{\psi}(x,u) = 0$ on $\partial\Omega$. In addition, we have

$$p \cdot \overline{A}(x,z,p) \geq \mu |p| - \alpha_1 |z|,$$
$$z\overline{B}(x,z,p) \leq b_0[p \cdot \overline{A}(x,z,p) + \alpha_1 |z|] + \beta_1 |z|,$$
$$z\overline{\psi}(x,z) \leq 0,$$

with $\mu = a_0 - \sup |w| > 0$, so the result follows from [11, Lemma 3.3]. □

Note that the requirement that u be bounded can be relaxed by an approximation argument to $|u|^{q-1} Du \cdot A$ and $|u|^{q-1} uB$ in L^1 for some sufficiently large q.

4. COMPARISON TO PREVIOUS RESULTS

We now show how our results improve previously known ones via some examples.

We first consider the case that $A(x,z,p) = p/(1+|p|^2)^{1/2}$ and $\Omega = \{x : |x| < 1\}$. Let k be a positive integer and set $K = (2k+1)\pi/2$, let $\sigma \in (0,1)$ and define

$$a_1(x) = \sigma[n \sin(K|x|^2) + K|x|^2 \cos(K|x|^2)].$$

Suppose $B(x,z,p)$ satisfies

$$zB(x,z,p) \leq a_1(x)|z| - |z|$$

and $|\psi| \leq \sigma$ on $\partial\Omega$. Then conditions (3.3) and (3.4) are satisfied with $a_0 = 1$, $a_2 \equiv 1$, $a_3 = 1$, $\mu = 0$, $\psi_1 = \sigma$, $\psi_2 = 0$, and $w_1 = \sigma \sin(K|x|^2)x$. Thus we obtain an L^1 estimate for u, independent of K, from Theorem 3.2. In addition, we can apply Theorem 3.4 with $\alpha_1 = 1$, $M = 1$, $\beta_1 = 0$, and $b_1 = a_1$ to infer an L^∞ estimate, independent of K. It is not difficult to show that the L^1 norm of a_1 is greater than $C(n)k$ and that $zB(x,z,p)$ can be positive for some $x \in \Omega$ regardless of the value of u at that point. Thus the L^1 norm of the coefficient a_1 is not the crucial factor in obtaining an L^∞ estimate for u. This situation is quite different from the usual version of this estimate in, for example, [5, Section 10.5] or [10, Section 4.6].

Next, we suppose that there are nonnegative constants α_1, b_0, c_0, and M along with a positive function b_3 defined on $\mathbb{R}_+$ such that

$$b_3(z) \to \infty \text{ as } z \to \infty,$$

and

$$p \cdot A(x,z,p) \geq |p| - \alpha_1 |z|,$$
$$zB(x,z,p) \leq b_0 (p \cdot A(x,z,p) + \alpha_1 |z|) - b_3(|z|)|z|,$$
$$z\psi(x,z) \leq c_0 |z|$$

for $|z| \geq M$. We also assume that $\partial\Omega \in C^2$ and that $c_0 < 1$. Because $\partial\Omega \in C^2$, there is a C^1 extension γ of the unit inner normal into all of Ω with $|\gamma| \leq 1$ in Ω. Thus conditions (3.8) and (3.9) hold with $a_0 = 1$, $a_1 \equiv \alpha_1$, $a_3 = 1$, $q = \max\{1, b_0/(1-c_0)\}$, $b_1 = c_0 \operatorname{div} \gamma - q a_1$, $f(z) = |z|$, $\psi_1 \equiv c_0$, $w = c_0 \gamma$, and $M \geq 1$ chosen so that

$$1 + \sup(-b_1) \leq b_3(|z|) \text{ for } |z| \geq M.$$

Hence our results include the L^1 estimate of [11, Lemma 3.4].

As a complement to this last estimate, we finally derive an L^∞ bound if zB is allowed to be positive but $z\psi$ is negative. For nonnegative constants α_1 and β_1 (to be further specified), and positive constants a_0 and c_1, consider the structure conditions

$$p \cdot A(x,z,p) \geq a_0 |p| - \alpha_1 |z|,$$
$$zB(x,z,p) \leq b_0 p \cdot A(x,z,p) + \beta_1 |z|,$$
$$z\psi(x,z) \leq -c_1 |z|.$$

For simplicity, we again assume that $\partial\Omega \in C^2$. We first observe that Theorem 3.4 provides an L^∞ bound in terms of an L^1 bound, so we only need to examine the L^1 bound. For this bound, we fix $q \geq 1$ so that $a_0 q > b_0$. Next, we write v for the solution of the boundary value problem $\Delta v = 1$ in Ω, $Dv \cdot \gamma = -|\Omega|/|\partial\Omega|$ with mean value zero. This solution exists by virtue of standard linear elliptic theory, which also guarantees that $|Dv| \leq C(\Omega)$. We now set $w_1 = (q\alpha_1 + \beta_1) Dv$ and note that $\sup |w_1| < 1$ if $(q\alpha_1 + \beta_1) C(\Omega) < 1$. If we further restrict the size of $q\alpha_1 + \beta_1$ by assuming that $c_1 > (q\alpha_1 + \beta_1)|\Omega|/|\partial\Omega|$, then (3.4) holds with $\psi_1(x) = w_1 \cdot \gamma$, so Theorem 3.2 gives an L^1 bound. When c_1 is sufficiently large, we can alter this argument somewhat. Specifically, we suppose that there is a constant a_2 such that

$$|A(x,z,p)| \leq a_2$$

for all (x,z,p) and that there is a constant M such that

$$|\psi(x,z)| > a_2$$

for all (x,z) with $|z| \geq M$. From the boundary condition, we see that $|u| \leq M$ on $\partial\Omega$ and then an L^∞ bound follows from the corresponding estimate for solutions to the Dirichlet problem. For example, [5, Theorem 10.9] gives such a bound if

$$p \cdot A(x,z,p) \geq |p| - \alpha_1 |z| - \alpha_2, \quad \operatorname{sgn} z B(x,z,p) \leq \beta_0$$

for some nonnegative constants α_1, α_2, and β_0 provided α_1 and β_0 are sufficiently small.

5. EXISTENCE RESULTS

We can also use our estimates to prove that the problem

$$\operatorname{div} A(x,Du) + B(x) = 0 \text{ in } \Omega, \quad A(x,Du) \cdot \gamma + \psi(x) = 0 \text{ on } \partial\Omega \tag{5.1}$$

under the obvious necessary condition

$$\int_\Omega B(x)\,dx + \int_{\partial\Omega} \psi(x)\,dx = 0 \tag{5.2}$$

is solvable. (Further assumptions will be made below.) When A is linear with respect to p, then this solvability is usually proved via Fredholm theory. Here we give a general result for the nonlinear problem (5.1).

Theorem 5.1 *Suppose there are constants a_0, a_2, b_0, and ψ_0 such that*

$$p \cdot A(x,p) \geq a_0 |p| - a_2, \quad |B(x)| \leq b_0, \quad |\psi(x)| \leq \psi_0. \tag{5.3}$$

Suppose also that the gradient estimate

$$|Du| \leq C(\sup|u|, A, B, \Omega, \psi) \tag{5.4}$$

is satisfied for any solution of (5.1). Finally, let w_0 be a C^1 vector field such that $w_0 \cdot \gamma \geq 1$ on $\partial\Omega$ and set $c_1 = \sup|w_0|$ and $c_2 = \sup|\operatorname{div} w_0|$. If

$$k[b_0 + c_2\psi_0] + c_1\psi_0 < a_0, \tag{5.5}$$

then (5.1) *has a solution.*

Proof. Let u_ε solve

$$\operatorname{div} A(x, Du_\varepsilon) - \varepsilon u_\varepsilon + B(x) = 0 \text{ in } \Omega, \quad A(x, Du_\varepsilon)\cdot\gamma + \psi(x) = 0 \text{ on } \partial\Omega.$$

Then u_ε has mean value zero, so we can apply Theorem 3.1 with $f(z) = |z|$, $a_2(x) \equiv a_2$, $a_1 = \psi \operatorname{div} w_0$, $q_3 = -b_0 - c_2\psi_0$, $\psi_1 \equiv \psi_0$, and $\psi_2 \equiv 0$. This gives a uniform estimate on the L^1 norm of u_ε, and then Theorem 3.4 gives a uniform estimate on the L^∞ norm. The uniform gradient estimate (5.4), along with Theorem 2.1 of [10, Chapter 10] and the Arzela-Ascoli theorem, shows that some subsequence $(u_{\varepsilon(j)})$ converges in $C^1(\overline{\Omega})$ to a function u, which is the desired solution of (5.1). □

Note that any solutions of (5.1) must differ by a constant under the simple assumption that

$$(p-q)\cdot[A(x,p) - A(x,q)] > 0 \text{ whenever } p \neq q. \tag{5.6}$$

In addition, (5.5) and the first inequality in (5.3) hold (for a suitable choice of a_1) if $p \cdot A/|p| \to \infty$ as $|p| \to \infty$, and the other conditions just quantify the assumptions that B and ψ are bounded. In fact, if we strengthen the growth condition in (5.3) with respect to p, then we can relax the hypotheses on B and ψ.

Theorem 5.2 *Suppose there are constants* $a_0 > 0$, $a_2 \geq 0$, *and* $m > 1$ *such that*

$$p \cdot A(x,p) \geq a_0 |p|^m - a_2. \tag{5.7}$$

Suppose also that

$$B \in L^{m/(m-1)}(\Omega), \psi \in L^{m/(m-1)}(\partial\Omega), \tag{5.8}$$

that A *is* C^1 *with respect to* x *and* p, *and that there is a positive constant* a_4 *such that*

$$|A(x,p)| \leq a_4[1+|p|]^{m-1}. \tag{5.9}$$

Suppose finally that A *satisfies (5.6). Then there is a solution of (3.1).*

Proof. First, for $\varepsilon \in (0,1]$, we define the operator $\Psi_\varepsilon : W^{1,m} \to W^{-1,m'}$ by

$$\langle \Psi_\varepsilon(u), w \rangle = \int_\Omega A(x, Du) \cdot Dw + (\varepsilon u - B(x))w \, dx + \int_{\partial\Omega} \psi(x) w \, ds.$$

Then Ψ_ε is monotone, coercive, and continuous on finite-dimensional subspaces of $W^{1,m}$, so the usual theory of monotone operators (for example Corollary 1.8 of Chapter 2 in [7] with $\mathbb{K} = X = W^{1,m}$) shows that, for any $\varepsilon > 0$, there is a unique solution u_ε to the problem

$$\operatorname{div} A(x, Du_\varepsilon) - \varepsilon u_\varepsilon + B(x) = 0 \text{ in } \Omega, \; A(x, Du_\varepsilon) \cdot \gamma + \psi(x) = 0 \text{ on } \partial\Omega.$$

With $\theta > 0$ to be further specified, we have

$$\begin{aligned} \int_\Omega Du_\varepsilon \cdot A(x, Du_\varepsilon) \, dx &\le \int_\Omega u_\varepsilon B \, dx + \int_{\partial\Omega} u_\varepsilon \psi_\varepsilon \, ds \\ &\le \theta \int_\Omega |\, u_\varepsilon \,|^m \, dx + \theta^{-1/(m-1)} \int_\Omega |\, B \,|^{m/(m-1)} \, dx \\ &+ \theta \int_{\partial\Omega} |\, u_\varepsilon \,|^m \, ds + \theta^{-1/(m-1)} \int_{\partial\Omega} |\, \psi \,|^{m/(m-1)} \, dx. \end{aligned}$$

With w_0, c_1, and c_2 as in Theorem 5.1, we have

$$\begin{aligned} \int_{\partial\Omega} |\, u_\varepsilon \,|^m \, ds &\le \int_{\partial\Omega} |\, u_\varepsilon \,|^m \, w_0 \cdot \gamma \, ds \\ &= \int_\Omega |\, u_\varepsilon \,|^m \operatorname{div} w_0 \, dx + m \int_\Omega D\,|\, u_\varepsilon \,| \cdot |\, w_0 \,| \cdot |\, u_\varepsilon \,|^{m-1} \, dx \\ &\le (c_2 + mc_1) \int_\Omega |\, u_\varepsilon \,|^m \, dx + mc_1 \int_\Omega |\, Du_\varepsilon \,|^m \, dx. \end{aligned}$$

Combining these two inequalities with (5.7) and setting

$$\Psi_0 = \int_\Omega |\, B \,|^{m/(m-1)} \, dx + \int_{\partial\Omega} |\, \psi \,|^{m/(m-1)} \, ds$$

yields

$$\begin{aligned} a_0 \int_\Omega |\, Du_\varepsilon \,|^m \, dx \le mc_1 \theta \int_\Omega |\, Du_\varepsilon \,|^m \, dx + \theta[1 + c_2 + mc_1] \int_\Omega |\, u_\varepsilon \,|^m \, dx \\ + \theta^{-1/(m-1)} \Psi_0 + a_1 |\, \Omega \,|. \end{aligned}$$

Again, u_ε has mean value zero, so Poincaré's inequality (in the form

$$\int_\Omega |u|^m dx \le k_m \int_\Omega |Du|^m dx$$

if (0.3) holds) implies that

$$a_0 \int_\Omega |Du_\varepsilon|^m dx \le \theta[mc_1 + k_m(1 + c_2 + mc_1)] \int_\Omega |Du_\varepsilon|^m dx + \theta^{-1/(m-1)} \Psi_0.$$

Choosing $\theta = a_0/2[mc_1 + k_m(1 + c_2 + mc_1)]$ gives

$$\int_\Omega |Du_\varepsilon|^m dx \le \frac{2}{a_0}\left(a_1 |\Omega| + \theta^{-1/(m-1)} \Psi_0\right),$$

which implies that u_ε is uniformly bounded in L^m and that a suitable subsequence converges weakly in $W^{1,m}$ to a function u, which is the desired solution of our problem. □

REFERENCES

[1] R. Finn, Existence and non existence of capillary surfaces, Manus. Math. **28** (1979), 1–11.

[2] R. Finn, "Equilibrium Capillary Surfaces", Springer-Verlag, Berlin, 1986.

[3] R. Finn, Existence criteria for capillary free surfaces without gravity, Indiana Univ. Math. J. **32** (1983), 439–460.

[4] R. Finn and A. A. Kosmodem'yanskii, jr., Some unusual comparison properties of capillary surfaces, Pac. J. Math. **205** (2002), 119–137.

[5] D. Gilbarg and N. S. Trudinger, "Elliptic Partial Differential Equations of Second Order", Springer-Verlag, Berlin, 1977; second edition, 1983.

[6] E. Giusti, boundary value problems for non-parametric surfaces of prescribed mean curvature, Ann. Scuola Norm. Sup Pisa **3** (1976), 501–548.

[7] D. Kinderlehrer and G. Stampacchia, "An Introduction to Variational Inequalities and their Applications", Academic Press, New York, 1980.

[8] N.J. Korevaar, Maximum principle gradient estimates for the capillary problem, Comm. Partial Differential Equations **13** (1988), 1–13.

[9] A.A. Kosmodem'yanskii, jr., The comparison of capillary surfaces heights in case of small gravity, Nonlinear Anal. **43** (2001), 937–942.

[10] O.A. Ladyzhenskaya and N.N. Ural'tseva, "Linear and Quasilinear Elliptic Equations", Nauka, Moscow, 1964 [Russian]; English transl. Academic Press, New York, 1968.

[11] G.M. Lieberman, The conormal derivative problem for elliptic equations of variational type, J. Differential Equations **49** (1983), 218–257.

[12] G.M. Lieberman, Hölder continuity of the gradient at a corner for the capillary problem and related results, Pac. J. Math. **133** (1988), 115–135.

[13] G.M. Lieberman, The conormal derivative problem for non-uniformly parabolic equations, Indiana Univ. Math. **37** (1988), 23–72.

[14] G.M. Lieberman, The conormal derivative problem for equations of variational type in nonsmooth domains, Trans. Amer. Math. Soc. **330** (1991), 41–67.

[15] J.Malý and W.P. Ziemer, "Fine Regularity of Solutions of Elliptic Partial Differential Equations', American Mathematical Society, Providence, R. I., 1997.

[16] D.Siegel, The behavior of a capillary surface for small Bond number, in Variational Methods for Free Surface Interfaces, Springer-Verlag, New York 1987, pp. 109– 113.

[17] N.N. Ural'tseva, Solvability of the capillary problem II, Vestnik Leningrad. Univ. Mat. Mekh. Astron. **1** (1975), 143–149 ([Russian]; English transl. in Vestnik Leningrad Univ. Math. **8** (1980), 151–158).

ASYMPTOTICALLY CRITICAL POINTS AND MULTIPLE SOLUTIONS IN THE ELASTIC BOUNCE PROBLEM

A. Marino[1] and C. Saccon[2]
Dept. of Mathematics, University of Pisa, Pisa, Italy;[1] Dept. of Applied Mathematics, University of Pisa, Pisa, Italy[2]

1. INTRODUCTION

A quite natural way to face the problem of the elastic bounce in a "billiard" with perfectly rigid walls is to pass to a sequence of approximating problems, where the walls are replaced by some repulsive force field which gets stronger and stronger outside the billiard.

This approach, however, poses some major difficulties, which are at least of two kinds. First of all one usually needs that a sequence (γ_n) of solutions of the approximating problems admits a subsequence converging to a bounce trajectory.

A second issue arises when looking for multiplicity results, for instance when one wishes to estimate the number of elastic bounce trajectories in the ideal billiard, with assigned end points. In this case indeed, multiple solutions of the approximating problems could converge to the same limit trajectory.

The fact that the approximating problems have a variational structure, since they verify the Hamilton principle, is quite helpful in both the above mentioned respects, especially in the former one. Actually if one chooses the approximating problems in a suitable way and adds the property that the

Lagrange integrals $I(\gamma_n)$ are bounded, then it is possible to prove that a subsequence of (γ_n) converges to a bounce trajectory γ.

The variational nature of the problem also suggests the first approach one can try for facing the second difficulty, that is trying to distinguish different limit solutions γ by the value $I(\gamma)$. It may happen indeed that different sequences (γ_n) of approximating solutions provide different limits of $I_n(\gamma_n)$, hence the limit solutions remain distinct, provided again that $I(\gamma_n) \to I(\gamma)$. This is not always the case, however. There are several meaningful problem where one expects multiple limit solutions possibly with the same value of the Lagrange integral. In this case it is clear that the previous approach fails.

To treat this kind of problems a very useful tool turned out to be the notion of "asymptotically critical points" for a sequence of functions f_n and a "limit" function f. In that notion γ is "critical" provided that there exists a sequence (γ_n) such that $\gamma_n \to \gamma$, $f_n(\gamma_n) \to f(\gamma)$ and grad $f_n(\gamma_n) \to 0$. Roughly speaking "asymptotically critical points" are limits of "almost critical points" for f_n. It is worth noting that the underlying metric influences both the meaning of the gradients and really determines the properties of γ. For this kind of points a multiplicity "asymptotic theorem" was proven (see [11, 12] for the case of smooth f_n and [15]), which gives an estimate of the number of asymptotically critical points in terms of the topological structure of the f_n's. Such a theorem was inspired by [1, 9]. We remark that this theorem seems not to hold, if one considers only the limits of critical points for f_n.

This approach has revealed particularly effective in the problem of the elastic bounce with fixed end-points, taking f_n to be the Lagrange integrals of the approximating problems. First of all, with suitable choices of the approximating force fields and of the underlying metric (with respect to which we are going to consider the gradients) it turns out that the asymptotically critical points are actually bounce trajectories. Moreover the "compactness like" assumptions which are required in the multiplicity asymptotic theorem are fulfilled. We may therefore say that also *the problem of elastic bounce trajectories with fixed end-points in an ideal billiard verifies a Hamilton principle suitable for multiplicity problems*: a sensitive asymptotic Hamilton principle. We remark that the metric we are induced to choose to obtain such a principle (the L_2 metric) is such that the Lagrange functionals I_n are not smooth, so we need to use some tools of subdifferential analysis.

Using this setting we can prove some multiplicity results, which are described in section 2. We point out that for the main result (see Theorem 4) the expected multiplicity is proven by means of a suitable adaptation of the ∇ -theorems, introduced in [13, 14], to the asymptotic framework.

We conclude by recalling some different approaches which were used, by different authours, for treating the bounce problem.

Among the known results concerning the number of bounce trajectories with fixed end points the first to be mentioned is the "Penrose counterexample" (see [17, 19]): in a "mushroom-shaped" billiard there are pairs of points A and B which are not connectable by any bounce trajectories. This counterexample poses some basic questions on which kind of assumptions one should consider for having multiplicity results.

On the contrary if the billiard is convex, then it was shown in [7] that for any pair of points A and B there are infinitely many bounce trajectories joining A and B. Also for proving this result a sequence of variational problem is used; the functionals however are very different from ours and the technique is very specifically related to the convexity of the billiard.

Some other results, concerning existence of bounce trajectories with few bounce points, were obtained in [2, 10].

Finally we recall that the problem of bounce trajectories with assigned initial position and velocity was studied in [3, 4, 18].

2. SETTING OF THE ELASTIC BOUNCE PROBLEM AND SOME RESULTS

Let Ω be a bounded subset of $\mathbb{R}^N$ with C^2 boundary. Moreover let $U_0 : \mathbb{R}^N \to \mathbb{R}$ be a C^2 potential. For x in $\partial\Omega$ let $\nu(x)$ denote the unit inward normal to $\partial\Omega$ at x.

2.1 Definition. *Let a, b with $a < b$ be real numbers. Let $\gamma \in W^{1,2}(a,b;\mathbb{R}^N)$. We say that γ is an elastic bounce trajectory in Ω, with respect to the potential field U_0, if*

$\gamma(t) \in \overline{\Omega}$ $\forall t \in [a,b]$ and there exists a positive Radon measure μ such that

$$\mathrm{spt}(\mu) \subset C(\gamma) := \left\{ t \in [a,b] \,|\, \gamma(t) \in \partial\Omega \right\} \tag{Eqn}$$

$$\int_a^b \dot{\gamma}\dot{\delta}\, dt - \int_a^b \nabla U_0(\gamma)\delta\, dt + \int_{[a,b]} \nu(\gamma)\delta\, d\mu = 0 \quad \forall \delta \in W_0^{1,2}(a,b;\mathbb{R}^N)$$

$$E(t) := \frac{1}{2} |\, \dot{\gamma}(t) \,|^2 + U_0(\gamma(t)) \text{ is constant on } [a,b] \tag{EC}$$

(Eqn) means that the equation $\ddot{\gamma} + \nabla U_0(\gamma) = \mu\nu(\gamma)$ is verified in the weak sense. $C(\gamma)$ is the *contact* set.

We say that γ is a true bounce trajectory, if $\mu \neq 0$.

2.2 Remark. Notice that the Energy Conservation law (EC) does not follow from the equation (Eqn). Actually the *reaction* given by μ could be too weak (not perfectly elastic walls) or too strong (pinball like behavior).
If t is an isolated contact point, then the conditions above imply

$$\dot{\gamma}(t^-)_{tan} = \dot{\gamma}(t^+)_{tan} \qquad , \qquad \dot{\gamma}(t^-)_{norm} = -\dot{\gamma}(t^+)_{norm}$$

where we are denoting by *tan* and *norm* the tangential and normal components to the boundary $\partial\Omega$ at the point $\gamma(t)$.

The multiplicity results we are going to present concern, for keeping this exposition simple, the bounce problem in the following situation: $0 \in \Omega$, $U_0(x) = \lambda q(x)$, where $\lambda \in \mathbb{R}$ and $q : \mathbb{R}^N \to \mathbb{R}$ is a symmetric quadratic form, $a = 0$, $b = 1$ and we look for bounce trajectories which start and end at the origin: $\gamma(0) = \gamma(1) = 0$.

To treat this problem we consider the "eigenvalues" λ of the problem:

$$\int_0^1 \dot{e}\dot{\delta}\, dt = \lambda \int_0^1 q'(e)(\delta)\, dt \qquad \forall \delta \in W_0^{1,2}(0,1;\mathbb{R}^N), \tag{EP}$$

for a suitable nontrivial function e (the eigenfunction) in $W_0^{1,2}(0,1;\mathbb{R}^N)$. It is simple to see that such eigenvalues do exist and that, if q is not trivial, there are infinitely many of them. We can describe the λ's as $(\lambda_i)_{i\in I}$ and the corresponding eigenfunctions as $(e_i)_{i\in I}$, where $I \subset \mathbb{Z}$ with $\lambda_i > 0$ for $i \geq 0$, $\lambda_i < 0$ for $i < 0$, $\lambda_i \leq \lambda_{i+i}$ for all i in I, $\lambda_n \to +\infty$ as $n \to +\infty$ (whenever $\sup I = +\infty$) and $\lambda_n \to -\infty$ as $n \to -\infty$ (whenever $\inf I = -\infty$).

For the main result we need Ω to be stricly star-shaped, with respect to zero, that is

$$\forall x \text{ in } \partial\Omega \quad (x, \nu(x)) > 0. \tag{Ω^*}$$

Now we can state the existence results we are able to prove.

2.3 Theorem (preliminary). *Let $0 \in \Omega$. For any λ there exists a true bounce trajectory y in Ω with respect to the potential field λq such that $\gamma(0) = \gamma(1) = 0$.*

The previous result actually holds under much more general assumptions on the potential.

2.4 Theorem. *Let* $0 \in \Omega$. *Let* λ_i *be a positive (resp. negative) eigenvalue of EP. Then there exists* $\eta > 0$ *such that for all* λ *in* $]\lambda_i - \eta, \lambda_i[$ *(resp.* $]\lambda_i, \lambda_i + \eta[$*), there exist two true bounce trajectories* γ_j *in* Ω, *with respect to the potential field* λq *such that* $\gamma_j(0) = \gamma_j(1) = 0 \quad j = 1, 2.$

2.5 Theorem (main result). *Let* $0 \in \Omega$ *and assume that* Ω^* *holds. Let* λ_i *be a positive (resp. negative) eigenvalue of* (EP). *Then there exists* $\eta > 0$ *such that for all* λ *in* $]\lambda_i - \eta, \lambda_i[$ *(resp.* $]\lambda_i, \lambda_i + \eta[$*), there exist three true bounce trajectories* γ_j *in* Ω, *with respect to the potential field* λq *such that* $\gamma_j(0) = \gamma_j(1) = 0 \quad j = 1, 2, 3.$

3. ASYMPTOTICALLY CRITICAL POINTS AND THEIR MULTIPLICITY

As we said in the introduction it will be useful to us to deal with suitable sequences of functionals. This section is devoted to a general notion of asympotical critical points for sequences of functions and to a corresponding asymptotic multiplicity theorem.

The metric we are induced to use in the application, to have that the asymptotically critical points are bounce trajectories, forces us to consider function which are not smooth in the classical sense. Therefore we need to recall some notions of subdifferential and some classes of nonsmooth functions. For all details we refer the reader to [8, 16, 5, 6].

Let H be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm. Let $f : H \to \mathbb{R} \cup \{+\infty\}$ be a function.

We set $\mathcal{D}(f) := \{u \mid f(u) < +\infty\}$; $\mathcal{D}(f)$; will be called the domain of f.

3.1 Definition. Let $u \in \mathcal{D}(f)$. We say that α in H is *a subdifferential* for f at u if

$$\liminf_{v \to u} \frac{f(v) - f(u) - \langle \alpha, v - u \rangle}{\|v - u\|} \geq 0$$

We denote by $\partial^- f(u)$ the set of all α's with the property above and we shall also say that $\partial^- f(u)$ is *the (Frechét) subdifferential* of f at u. Notice that $\partial^- f(u)$ is a (possibly empty) convex set. So if $\partial^- f(u) \neq \emptyset$ we can

define the (lower) gradient of f at u, denoted by $\operatorname{grad}^- f(u)$, as the element of minimal norm in $\partial^- f(u)$.

3.2 Definition. We say that a point u in $\mathcal{D}(f)$ is (lower) critical for f, if $0 \in \partial^- f(u)$ (i.e. $\operatorname{grad}^- f(u) = 0$).
We say that a real number c is a (lower) critical value for f, if there exists a (lower) critical point u such that $f(u) = c$.

As usual, for c in $\mathbb{R}$, we set $f^c := \{u \mid f(u) \le c\}$.

3.3 Definition. Let $p, q : \mathcal{D}(f) \to \mathbb{R}$ be two continuous functions. We say that f is in the class $C(p,q)$, if

$$f(v) \ge f(u) + \langle \alpha, v-u \rangle - (q(u) + p(u) \,\|\, \alpha \,\|) \,\|\, v-u \,\|^2$$
$$\forall u, v \text{ in } \mathcal{D}(f), \forall \alpha \text{ in } \partial^- f(u).$$

In the sequel we often say that some function f is of class $C(p,q)$, if there exist p and q such that the above conditions hold.

From now we consider a sequence $(f_n)_n$ of functions, $f_n : H \to \mathbb{R} \cup \{+\infty\}$. For the sake of simplicity we shall assume that all the domains $\mathcal{D}(f_n)$ are equal to a fixed set D.

Moreover we consider a function $f : D \to \mathbb{R} \cup \{-\infty\}$.

3.4 Definition. Let $u \in \mathcal{D}$. We say that u is asymptotically critical for the sequence $((f_n)_n, f)$, if there exist a strictly increasing sequence $(k_n)_n$ in $\mathbb{N}$ and two sequences $(u_n)_n$, $(\alpha_n)_n$ in H such that

$$\forall n \; u_n \in D(f_{k_n}) \;,\; u_n \to u, f_{k_n}(u_n) \to f(u) \;,\; \forall n \; \alpha_n \in \partial^- f_{k_n}(u_n) \;,\; \alpha_n \to 0$$

We say that a real number c is an asymptotically critical level for $((f_n)_n, f)$, if there exists an asymptotically critical point u such that $f(u) = c$.

3.5 Definition. Let $c \in \mathbb{R}$. We say that $(u_n)_n$ is a *nabla sequence* for $((f_n)_n, f)$, at level c, briefly a $\nabla(f_n, f, c)$-sequence, if there exists a strictly increasing sequence $(k_n)_n$ in $\mathbb{N}$ such that:

$$\forall n \; u_n \in D(f_{k_n}), \quad f_{k_n}(u_n) \to c, \quad \forall n \; \partial^- f_{k_n}(u_n) \neq \emptyset, \quad \operatorname{grad}^- f_n(u_n) \to 0.$$

We say that $((f_n)_n, f)$ verifies the *nabla condition* at level c, briefly $\nabla(f_n, f, c)$ holds, if any $\nabla(f_n, f, c)$-sequence admits a subsequence which converges in H to a point u in $\mathcal{D}$ such that $f(u) = c$.

3.6 Theorem (Asymptotic Multiplicity Theorem). *Assume that for all* n *in* $\mathbb{N}$ f_n *is lower semi-continuous and of class* $C(p,q)$ *(with* p, q *depending on* n*).*
Let $a,b \in \mathbb{R}$ *with* $a < b$ *and assume that* $\nabla(f_n, f, c)$ *holds for all* c *in* $[a,b]$.
Then

$$\#\Big\{u \text{ asymptotically critical for } ((f_n)_n, f), f(u) \in [a,b]\Big\} \geq \limsup_{n\to\infty} \operatorname{cat}_{\mathcal{D}}(f_n^b, f_n^a).$$

For the notion of *relative category* $\operatorname{cat}_X(A,B)$ we refer the reader to [9] and the references therein.

3.7 Remark. In the proof of the previous theorem a key fact is that, if $\mathcal{D}$ is the domain of a $C(p,q)$ function, then each point of $\mathcal{D}$ admits a neighborhood U which is contractible in $\mathcal{D}$.

4. A VARIATIONAL SETTING FOR THE BOUNCE PROBLEM WITH FIXED END POINTS

As announced in the introduction we now present a variational *asymptotic* setting for the elastic bounce problem with fixed end points.

Let Ω be a bounded subset of $\mathbb{R}^N$ with C^2 boundary and let A, B be two given points in Ω. We set

$$\mathbb{X}(A,B) := \Big\{\gamma \in \mathrm{W}^{1,2}(0,1;\mathbb{R}^N) \,|\, \gamma(0) = A, \gamma(1) = B\Big\}$$

Let $U_0 : \mathbb{R}^N \to \mathbb{R}$ be of class C^2.

We can introduce a C^2 function $G : \mathbb{R}^N \to \mathbb{R}$ such that $\Omega = \Big\{x \,|\, G(x) < 0\Big\}$ and $|\nabla G(x)| \geq \epsilon_0 > 0$ for all x in a neighborhood of $\partial\Omega$. Then $\nu(x) := -\dfrac{\nabla G(x)}{|\nabla G(x)|}$ is well defined in a neighborhood of $\partial\Omega$ and $\nu(x)$ is the unit inward normal if x is a boundary point. We set $U(x) := (G(x)^+)^p$, for a given $p > 1$.

For $\omega > 0,$, we define $g, f_\omega : \mathrm{L}^2(0,1;\mathbb{R}^N) \to \mathbb{R} \cup \{+\infty\}$ and $f_\infty : X(A,B) \to \mathbb{R} \cup \{-\infty\}$ by

$$g(\gamma) := \begin{cases} \displaystyle\int_0^1 \left(\frac{1}{2}|\dot{\gamma}|^2 - U_0(\gamma)\right) dt \text{ if } \gamma \in \mathbb{X}(A,B) \\ +\infty \qquad\qquad\qquad\qquad \text{otherwise} \end{cases}$$

$$f_\omega(\gamma) := g(\gamma) - \omega \int_0^1 U(\gamma)\, dt$$

$$f_\infty(\gamma) := \begin{cases} g(\gamma) & \text{if } \gamma \in \mathbb{X}(A,B), \gamma(t) \in \overline{\Omega}\ \forall t \in [0,1] \\ -\infty & \text{otherwise} \end{cases}$$

For techical reasons we also need (in some cases) another constraint: let $R \in \mathbb{R}$; We set

$$\mathbb{X}_R(A,B) := \left\{\gamma \in \mathbb{X}(A,B) \,|\, g(\gamma) \leq R\right\}$$

and define $f_{R,\omega} : \mathrm{L}^2(0,1;\mathbb{R}^N) \to \mathbb{R} \cup \{+\infty\}$, $f_{R,\infty} : X_R(A,B) \to \mathbb{R} \cup \{-\infty\}$ by

$$f_{R,\omega}(\gamma) : \begin{cases} g(\gamma) - \omega \displaystyle\int_0^1 U(\gamma)\, dt & \text{if } \gamma \in \mathbb{X}_R(A,B) \\ +\infty \qquad\qquad\qquad\qquad & \text{otherwise} \end{cases}$$

$$f_{R,\infty}(\gamma) := \begin{cases} g(\gamma) & \text{if } \gamma \in \mathbb{X}_R(A,B), \gamma(t) \in \overline{\Omega}\ \forall t \in [0,1] \\ -\infty & \text{otherwise} \end{cases}$$

The main fact we are going to discuss now is that bounce trajectories are asymptotically critical points for $((f_{R,\omega})_\omega, f_{R,\infty})$.

4.1 Proposition. *Let $R \in \mathbb{R}$ and $\gamma \in \mathbb{X}_R(A,B)$ be an asymptotically critical point for $((f_{R,\omega})_\omega, f_{R,\infty})$. Then γ is a bounce trajectory with respect to the potential U_0.*

4.2 Proposition. *The following alternative holds:*

- *either there exists γ such that $\ddot{\gamma} + \nabla U_0(\gamma) = 0$, $g(\gamma) = R$ and $\gamma([0,1]) \subset \overline{\Omega}$;*
- *or the condition $\nabla(f_{R,\omega}, f_{R,\infty}, c)$ is verified for all c in $\mathbb{R}$.*

4.3 Remark. In the case when Ω *has no holes* we can avoid using $((f_{R,\omega})_\omega, f_{R,\infty})$ and directly consider $(f_\omega, f_\infty)_\omega$: assume that

$$|\nabla G(x)| \geq \varepsilon_0 > 0 \qquad \forall x \in \mathbb{R}^N \setminus \Omega,$$

then for a suitable $p > 2$ the two propositions above hold replacing $((f_{R,\omega})_\omega, f_{R,\infty})$ by $((f_\omega)_\omega, f_\infty)$.
Such a **condition** is true if Ω is convex or more generally star-shaped.

4.4 Remark. The validity of the propositions above relies on the fact that we choose the asymptotically critical points in the $L^2([0,1])$-sense. If one takes $W^{1,2}(0,1;\mathbb{R}^N)$ as the underlying space, then asymptotically critical points are not, in general, bounce trajectories. In particular the energy conservation law can be lost when passing to the limit.

We conclude this section giving the idea for the proof of the three solutions theorem.

Sketch of proof of Theorem 2.5 First notice that under the given assumptions we have

$$\mathbb{X}(A,B) = W_0^{1,2}(0,1;\mathbb{R}^N), g(\gamma) = \begin{cases} \displaystyle\int_0^1 \left(\frac{1}{2}|\dot{\gamma}|^2 - \lambda q(\gamma)\right) dt & \text{if } \gamma \in W_0^{1,2}(0,1;\mathbb{R}^N) \\ +\infty & \text{otherwise} \end{cases}$$

We shall denote B_r and S_r the closed ball and sphere, centered at zero, with radius r in the $L^2([0,1])$ metric.

We consider the case $\lambda_i > 0$. We may suppose $\lambda_i < \lambda_{i+1}$ and we can take $j < i$ such that $\lambda_j < \lambda_{j+1} = \lambda_i$. We set

$X_1 := \mathrm{span}(e_0, \ldots, e_j)$ ($X_1 := \{0\}$, if $j = -1$),

$X_2 := \mathrm{span}(e_{j+1}, \ldots, e_i)$,

$X_3 := (X_1 \oplus X_2)^\perp$.

In what follows we take λ in $[\lambda_j, \lambda_i[$.

- For all ω we have

$$\sup f_\omega(X_1) = 0.$$

- If we set $b_\lambda := \sup_{\omega>0} \sup f_\omega(X_1 \oplus X_2)$, then one can show that

$$\lim_{\lambda \to \lambda_i^-} b_\lambda = 0$$

- There exist r and ω_0 positive such that for $\omega \geq \omega_0$

$$\sup f_\omega \left(S_r \cap (X_1 \oplus X_2)\right) \leq 0$$

- There exists a positive R_0 such that for all $R \geq R_0$

$$B_r \cap (X_1 \oplus X_2) \subset \mathbb{X}_R(0,0).$$

If R_0 is large we also get that $\mathbb{X}_R(0,0)$ and $X_1 \oplus X_3$ are not tangent at any γ with $g(\gamma) = R$.
We take $R > R_0$. Since $R > 0$ there are no γ's in $\mathbb{X}_R(0,0)$ with $\ddot{\gamma} + \nabla U_0(\gamma) = 0$, $g(\gamma) = R$ and $\gamma([0,1]) \subset \overline{\Omega}$.
Then, by 2, $\nabla(f_{R,\omega}, f_{R,\infty}, c)$ is verified for all c in $\mathbb{R}$.

- At this point we can find $\rho > 0$ such that

$$\forall \gamma \quad \gamma \in S_\rho \cap \mathbb{X}_R(0,0) \Rightarrow \gamma([0,1]) \subset \Omega$$

which implies

$$a_\lambda := \inf_{\omega>0} \inf f_{R,\omega}\left(S_\rho \cap (X_2 \oplus X_3)\right) > 0$$

Now, if

$$T := (B_r \cap X_1) \cup (S_r \cap (X_1 \oplus X_2)), \quad S_{23} := S_\rho \cap (X_2 \oplus X_3),$$
$$B_{12} := B_r \cap (X_1 \oplus X_2),$$

then for λ in $[\lambda_j \lambda_i[$ we have precisely the inequalities required in Theorem 4, uniformly with respect to ω large. Moreover it is possible to show that

- $\nabla(f_{R,\omega}, f_{R,\infty}, X_1 \oplus X_3, c)$ holds for all c in $\mathbb{R}$;

- possibly taking λ closer to λ_i there are no γ's with γ asymptotically constrained on $X_1 \oplus X_3$ for $f_{R,\omega}$, with $f_{R,\omega}(\gamma) \in]0, b_\lambda]$.

Then by Theorem (5.4) there are two asymptotically critical points γ_1, γ_2 with $g(\gamma_h) \in [a_\lambda, b_\lambda]$ $h = 1,2$.

We point out that the assumption Ω strictly star-shaped is only needed to prove $\nabla(f_{R,\omega}, f_{R,\infty}, X_1 \oplus X_3, c)$. If this assumption fails then we cannot prove that there are two solutions in the given interval. Nevertheless, due to the linking inequality found, we can prove that there is at least one solution γ with $g(\gamma) \in [a_\lambda, b_\lambda]$ for any λ in $[\lambda_j, \lambda_i[$.

Finally the existence of the third solution can be proved by considering the splitting given by $X'_1 := \text{span}(e_0, \ldots, e_i)$, $X_2 := \text{span}(e_{i+1}, \ldots, e_k)$ and $X_3 := (X_1 \oplus X_2)^\perp$, where $\lambda_{i+1} = \lambda_k < \lambda_{k+1}$ and taking λ possibly closet to λ_i. □

5. ∇ ASYMPTOTIC THEOREMS

In this section we present an asymptotic version of the ∇-theorems introduced in [13,14]. As we already showed, these theorems are used in the proof of Theorem (2.5).

As in section 2 we consider here a sequence $(f_n)_n$ of functions from a Hilbert space H into $\mathbb{R} \cup \{+\infty\}$ with a fixed domain $\mathcal{D}$ and a function $f : \mathcal{D} \to \mathbb{R} \cup \{-\infty\}$. Moreover all f_n are taken to be of class $C(p,q)$ (with p, q depending on n).

5.1 Definition. Let X be a closed linear subspace of H and let $c \in \mathbb{R}$. We say that a sequence $(u_n)_n$ in H is a $\nabla(f_n, f, X, c)$ sequence if

there exist (k_n) in $\mathbb{N}$ strictly increasing and $(\alpha_n)_n$ in H such that

$$\forall n \; u_n \in \mathcal{D} \;,\; \text{dist}(u_n, X) \to 0 \;,\; f_{k_n}(u_n) \to c$$

$$\forall n \; \alpha_n \in \partial^- f_{k_n}(u_n) \;,\; P_{X \oplus [u_n]} \alpha_n \to 0$$

where $P_{X \oplus [u_n]}$ denotes the orthogonal projection onto the space $X \oplus [u_n] := X \oplus \text{span}(u_n)$.

We say that the $\nabla(f_n, f, X, c)$-condition holds if any $\nabla(f_n, f, X, c)$-sequence admits a subsequence which converges to some point u in $\mathcal{D}$ such that $f(u) = c$.

5.2 Definition. Let X be a closed linear subspace of H and $u \in X$. We say that u is an asymptotically constrained-critical point for $((f_n)_n, f)$ on X, if there exist a strictly increasing sequence $(k_n)_n$ in $\mathbb{N}$ and two sequences $(u_n)_n$, $(\alpha_n)_n$ in H such that

$$\forall n\ u_n \in \mathcal{D}(f_{k_n})\ , u_n \to u,\ f_{k_n}(u_n) \to f(u)\ , \forall n\ \alpha_n \in \partial^- f_{k_n}(u_n)\ ,$$
$$P_{X \oplus [u_n]} \alpha_n \to 0$$

5.3 Definition. Let V_1, V_2 be two subsets of H and $u \in V_1 \cap V_2$. We say that V_1 and V_2 are tangent at u, if there exists α such that $\alpha \neq 0$, $\alpha \in \partial^- I_{V_1}$, $-\alpha \in \partial^- I_{V_2}$. Here we are denoting by I_V the indicator function such that $I_V(u) = 0$ whenever $u \in V$ and $I_V(u) = +\infty$ outside V.

5.4 Theorem. *Assume that* $H = X_1 \oplus X_2 \oplus X_3$, *where* X_i *are closed linear subspaces and* $\dim(X_1 \oplus X_2) < +\infty$. *Let* $R > \rho > 0$. *Set*

$$S_{12} := \left\{u \in X_1 \oplus X_2 \mid \| u \| = R\right\}\ , S_{23} := \left\{u \in X_2 \oplus X_3 \mid \| u \| = \rho\right\}$$
$$B_1 := \left\{u \in X_1 \mid \| u \| \leq R\right\}\ ,\ B_{12} := \left\{u \in X_1 \oplus X_2 \mid \| u \| \leq R\right\}\ ,$$
$$T := S_{12} \cup B_1$$

Assume that $a' < a < b < b'$ *and for all* n *large*

- *the following inequalities hold:*

$$\sup f_n(T) \leq a' < a \leq \inf f_n(S_{23}) \qquad \sup f_n(B_{12}) \leq b;$$

- $\mathcal{D}$ *and* $X_1 \oplus X_3$ *are not tangent at any* u *with* $f_n(u) \in [a', b']$;
- $\nabla(f_n, f, c)$ *holds for all* c *in* $[a, b]$;
- $\nabla(f_n, f, X_1 \oplus X_3, c)$ *holds for all* c *in* $[a, b]$;
- *there are no points* u *such that* $f(u) \in [a, b]$ *and* u *is asymptotically constrained-critical for* $((f_n)_n, f)$ *on* $X_1 \oplus X_3$.

Then there exist two asymptotically critical points u_1, u_2 *for* $((f_n)_n, f)$ *such that* $f(u_i) \in [a, b] \quad i = 1, 2$.

REFERENCES

[1] T. Bartsch and M. Clapp. Critical point theory for indefinite functionals with symmetries. *J. Funct. Anal.*, 138(1):107–136, 1996.

[2] V. Benci and F. Giannoni. Periodic bounce trajectories with a low number of bounce points. *Ann. Inst. H. Poincaré*, 6(1):73–93, 1989.

[3] G. Buttazzo and D. Percivale. On the approximation of the elastic bounc problem on riemanian manifolds. *J. Diff. Eq.*, 47:227–245, 1983.

[4] M. Carriero, A. Leaci, and E. Pascali. Convergenza per l'equazione degli integrali primi associati al problema del rimbalzo unidimensionale. *Ann. Mat. Pura Appl.*, 133:227–256, 1983.

[5] G. Chobanov, A. Marino, and D. Scolozzi. Evolution equation for the eigenvalue problem for the laplace operator with respect to an obstacle. *Rend. Accad. Naz. Sci. XL Mem. Mat.*, 14(5):139–162, 1990.

[6] G. Chobanov, A. Marino, and D. Scolozzi. Multiplicity of eigenvalues for the laplace operator with respect to an obstacle, and nontangency conditions. *Nonlinear Anal.*, 15(3):199–215, 1990.

[7] M. Degiovanni. Multiplicity of solutions for the bounce problem. *J. Diff. Eq.*, 54(3):414–428, 1984.

[8] M. Degiovanni, A. Marino, and M. Tosques. Evolution equations with lack of convexity. *Nonlinear Anal. T.M.A*, 9:1401–1443, 1985.

[9] G. Fournier, D. Lupo, M. Ramos, and M. Willem. Limit relative category and critical point theory. In U. K. C. K. R. T. Jones and H. O. Walther, editors, *Dynamics Reported*, pages 1–24. Springer, Berlin, 1994.

[10] F. Giannoni. Bounce trajectories with one bounce point. *Ann. Mat. Pura Appl.*, CLIX:101–115, 1991.

[11] Marino and D. Mugnai. Asymptotical multiplicity and some reversed variational inequalities. *Topol. Meth. Nonlinear Anal.*, 17:43–62, 2002.

[12] Marino and D. Mugnai. Asymptotically critical points and their multiplicity. *Topol. Meth. Nonlinear Anal.*, 20:29–38, 2002.

[13] Marino and C. Saccon. Some variational theorems of mixed type and elliptic problems with jumping nonlinearities. *Ann. Scuola Norm. Sup. Pisa*, XXV:631–665, 1997.

[14] Marino and C. Saccon. Nabla theorems and multiple solutions for some noncooperative elliptic systems. *Topol.Meth Nonlinear Anal.*, 17:213–237, 2001.

[15] Marino and M. C. Saccon. Multiplicity results for the elastic bounce problem. *to appear*, N.S.

[16] Marino and M. Tosques. Some variational problems with lack of convexity and some partial differential inequalities. In *Method of Nonconvex Analysis*, volume 1446 of *Lecture Notes in Math.*, pages 58–83, Berlin, 1990. Springer. Varenna, 1989.

[17] L. Penrose and R. Penrose. *Puzzles for Christmas*. New Scientist, 25 December 1958.

[18] D. Percivale. Uniqueness IN elastic bounce problems. *J. Diff. Eq.*, 54:1984, 1985.

[19] R J. Rauch. Illumination of bounded domains. *Amer. Math. Monthly*, pages 359–361, 1978.

A BRANCH-AND-CUT TO THE POINT-TO-POINT CONNECTION PROBLEM ON MULTICAST NETWORKS*

C.N. Meneses,** C.A.S. Oliveira and P.M. Pardalos
Dept. of Industrial and Systems Engineering, University of Florida, Gainesville, FL, USA

Abstract: In multicast routing, one of the basic problems consists of sending data from a set of sources to a set of destinations with minimum cost. A formalization of this problem using graph theory is given by the nonfixed Point-to-Point Connection (PPC) Problem. The optimization version of this problem is known to be NP-hard, and it can be also applied in areas such as circuit switching and VLSI design. We present a branch-and-cut approach to solve the PPC. Initially we describe a 0-1 integer programming formulation. Then, we prove that some of the constraints in this formulation are facet defining inequalities. Other valid inequalities, based on partitions of the set of vertices in the graph are also investigated. The proposed branch-and-cut algorithm is based on the previously discussed inequalities. Computational results of the branch-and-cut algorithm are presented, with comparisons to existing heuristic and approximation algorithms. The results show the effectiveness of the branch-and-cut method for instances of moderate size.

Key words: combinatorial optimization, point-to-point connection, integer programming

* This research is partially supported by US Air Force and NSF grants.
** This author is supported by the Brazilian Federal Agency for Post-graduate Education (CAPES), Grant No. 1797-99-9.

1. INTRODUCTION

Multicast networks are used to send information from one or more sources to a set of destinations. This type of networks has become increasingly important for applications that need to share resources in private networks, as well as in the Internet [2, 4, 9]. One useful formalization of multicast networks is provided by the point-to-point connection problem (PPC) [3,8,11]. In this problem, a set of source nodes must be connected to a set of destinations with minimum cost.

We are concerned with providing an optimal solution to the nonfixed PPC problem, which can be defined as follows. Let $G = (V, E)$ be an undirected connected graph, $S = \{s_1, \ldots, s_p\} \subset V$ a set of source nodes, and $D = \{d_1, \ldots, d_p\} \subset V$ a set of destination nodes, such that $S \cap D = \emptyset$. Let us define a nonnegative integer weight c_{uv} for each edge $(u, v) \in E$. A *point-to-point connection* is a subset $E' \subseteq E$ such that each source s_i is connected to at least one destination d_j, and conversely each destination d_j is connected to at least one source s_i, by an $(s_i\text{-}d_j)$-path in E' (see Fig. 1). The cost of a point-to-point connection is defined as $\sum_{(u,v)\in E'} c_{uv}$. The objective of the PPC problem is to find a minimum point-to-point connection $E' \subseteq E$. The PPC problem has additional applications in circuit switching and VLSI design [11].

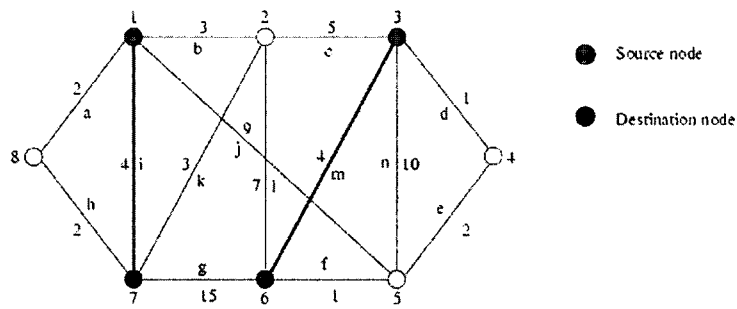


Figure 1. An instance of the PPC problem. A feasible solution with objective cost 8 is given by the edge set $\{i, m\}$. The edge set $\{a, d, e, f, h\}$ is an alternative solution to this instance.

We can derive four variants of the problem, depending on the type of graph (directed or undirected) and whether the source-destination pairs are fixed or nonfixed. That is, if s_i is required to be connected by a path to a

specific destination d_j, $i, j \in \{1, \ldots, p\}$, then we have the fixed destinations case; otherwise, we have the nonfixed destinations case. Throughout the text, given an input graph $G = (V, E)$, we adopt $n = |V|$, $m = |E|$ and $p = |S| = |D|$.

In [11], Li *et al.* proved that all four versions of the PPC problem are NP-hard, when p is given as input. In the same paper the authors gave a dynamic programming algorithm with time complexity $O(n^5)$ for the fixed destinations case on directed graphs when $p = 2$. Natu and Fang [13] proposed another dynamic programming algorithm for $p = 2$ with overall time complexity $O(mn + n^2 \log n)$, and they also showed an algorithm with time complexity $O(n^{11})$ for the fixed destinations case when the graph is directed and $p = 3$. Goemans and Williamson [7] presented an approximation algorithm for a class of forest constrained problems including the PPC problem that runs in $O(n^2 \log n)$, and gives its results within a factor of $2 - 1/p$ of an optimal solution.

Metaheuristic algorithms have also been proposed to PPC. An example is the Asynchronous Team (A-Team) metaheuristic, which is used to combine a number of fast and simple heuristics with the purpose of improving a pool of solutions. In [8] an A-Team was proposed for the PPC, and demonstrated to give near optimal results. The same methodology was extended in [3] to a parallel environment, where the constituent heuristics could run in separate processors.

In this paper we propose and implement a branch-and-cut approach for solving exactly the PPC problem. The algorithm is based on a particular integer formulation for the problem, presented in the next section. We derive some cutting-plane inequalities for this formulation, and the inequalities are used to define the branch-and-cut algorithm.

The remaining sections are organized as follows. In Section 2, we formulate the PPC problem as an Integer Programming (IP) problem and prove results concerning the facet structure of the polyhedron defined by this formulation. In Section 3 we present a branch-and-cut algorithm based on this formulation. In Section 4, computational results of the proposed algorithm are presented. Finally, concluding remarks are given in Section 5.

2. MATHEMATICAL MODEL

In this section we give an Integer Programming formulation for the PPC problem and show that the associated polyhedron is full dimensional. We also present facet defining inequalities for this polyhedron.

2.1 IP Formulation

A solution to the PPC problem can be described as a forest in G where, for each connected component the number of sources and destinations is equal. Also, if a set $U \subset V$ has a different number of sources and destinations, there must be an edge e going from U to $V \setminus U$. We note that in a feasible solution this condition must be satisfied by each subset $U \subset V$.

These conditions are expressed by the following model. Let

$$x_e = \begin{cases} 1 & \text{if edge } e \in E' \\ 0 & \text{otherwise} \end{cases}$$

The IP formulation is given by

$$\min \sum_{e \in E} c_e x_e$$

subject to

$$\sum_{e \in \delta(U)} x_e \geq 1 \quad \text{for all } U \subseteq V, |U \cap S| \neq |U \cap D| \tag{1}$$

$$0 \leq x \leq 1$$

x integer.

By $\delta(U)$ we denote the set of edges $(u, v) \in E$ such that $u \in U$ and $v \in V \setminus U$, i.e. $\delta(U)$ is a cut set. The inequalities (1) are subsequently called **point-to-point cut** inequalities. The theorem below establishes the correctness of the previous formulation.

Theorem 1 *The system above defines an IP formulation for the PPC problem.*

Proof: This is equivalent to showing that x satisfies inequalities (1) if and only if E' is a solution to an instance of the PPC problem. If x satisfies inequalities (1), then in each connected component of the forest defined by x, the number of source nodes is equal to the number of destination nodes. Therefore, according to the observations in the beginning of the section, x defines a solution to an instance of the PPC problem. Conversely, if E' is a feasible solution, then each connected component in E' has the same number of sources and destinations. Hence, x satisfies inequalities (1). □

2.2 Polyhedral Results

Consider an instance of the PPC problem and let $\mathcal{P}_C(G)$ be the convex hull of the incidence vectors of point-to-point connections, i.e., $\mathcal{P}_C(G) := Conv\{x^E \in \mathbb{R}^m \mid E' \text{ is a point-to-point connection }\}$. $\mathcal{P}_C(G)$ is called the **point-to-point connection polyhedron**.

As we wish to analyze the quality of the inequalities in the formulation, we need to know the dimension of $\mathcal{P}_C(G)$. We denote by $G \setminus \{e\}$ the graph obtained from graph $G = (V, E)$ when the edge $e \in E$ is removed. Let Q be the vertex set in $G \setminus \{e\}$, $e = (u, v) \in E$, reachable from u. Edge $e = (u, v)$ in G is called a **point-to-point bridge** if it is a bridge that separates nodes u and v such that $| Q \cap S | \neq | Q \cap D |$.

Let $x^E = (1, 1, \ldots, 1)$ and $x^{E\setminus\{e\}}$ denote the incidence vector obtained from x^E by setting the e-th component of x^E to zero for some edge $e \in E$.

Theorem 2 *Let $G = (V, E)$ be a connected graph, S a source-vertex set, D a destination-vertex set and $B(G)$ the point-to-point bridge set of G. Then*

$$dim(\mathcal{P}_C(G)) = | E | - | B(G) |.$$

Proof: If edge $e \in B(G)$, then every point-to-point connection of G has $x_e = 1$, i.e. $dim(\mathcal{P}_C(G)) \leq | E | - | B(G) |$. On the other hand, since by assumption G is connected, $x^E \in \mathcal{P}_C(G)$ and $x^{E\setminus\{e\}} \in \mathcal{P}_C(G)$ for some edge $e \notin B(G)$, so $dim(\mathcal{P}_C(G)) \geq | E | - | B(G) |$.

By Theorem 2, we infer that $\mathcal{P}_C(G)$ is full-dimensional if and only if G has no point-to-point bridges. Next, we present results concerning the facial structure of $\mathcal{P}_C(G)$. Theorems 3 to 5 summarize these results.

Theorem 3 *The inequality $x_e \geq 0$ for all $e \in E$ defines a facet of $\mathcal{P}_C(G)$ if and only if $G \setminus \{e\}$ has no point-to-point bridges.*

Proof: The validity follows from the observation that all variables are nonnegative. To show that this is a facet defining inequality, we need to show that the face $F = \{x \in \mathcal{P}_C(G) \mid x_e = 0\}$ has m affinely independent points. Note that if $G \setminus \{e\}$ has no point-to-point bridges, then for every edge $f \in E$, $f \neq e$, the points $x^{E\setminus\{e\}}$ and $x^{E\setminus\{e,f\}}$ belong to $\mathcal{P}_C(G)$. These points also belong to F and satisfy $x_e = 0$, and are affinely independent.

Thus, $x_e \geq 0$ defines a facet of $\mathcal{P}_C(G)$. Conversely, suppose that $f \in E$ is a point-to-point bridge in $G \setminus \{e\}$. Then, the inequality $x_e + x_f \geq 1$ is valid for all $x \in \mathcal{P}_C(G)$. Thus, $x_e \geq 0$ can be obtained by adding $x_e + x_f \geq 1$ and $x_f \leq 1$, and hence does not define a facet of $\mathcal{P}_C(G)$. □

Theorem 4 *Let $G = (V, E)$ be a connected graph. If G has no point-to-point bridges, then the inequality $x_e \leq 1$ for all $e \in E$ defines a facet of $\mathcal{P}_C(G)$.*

Proof: Since all x_e are binary variables, we must have $x_e \leq 1$. Let $F := \{x \in \mathcal{P}_C G) \mid x_e = 1\}$ be the face defined by $x_e \leq 1$. Since the points x_e and $x^{E\setminus\{f\}}$, for every $f \in E$, $f \neq e$, are affinely independent and they belong to F, then $dim(F) = \mid E \mid -1$. Therefore $x_e \leq 1$ defines a facet of $\mathcal{P}_C(G)$. □

Theorem 5 *Let $G = (V, E)$ be a connected graph. If G has no point-to-point bridges, then for all $\delta(U)$, $\varnothing \neq U \subset V$, $\mid U \cap S \mid \neq \mid U \cap D \mid$, the inequality $\sum_{e\in\delta(U)} x_e \geq 1$ defines a facet of $\mathcal{P}_C(G)$.*

Proof: The validity follows from the sufficiency condition of Theorem 1. To show that the inequality defines a facet, let us denote by $\omega^T x \geq \omega_0$ the inequality $\sum_{e\in\delta(U)} x_e \geq 1$ and let $F_\omega := \{x \in \mathcal{P}_C(G) \mid \omega^T x = \omega_0\}$ be the face defined by it. We note that $F_\omega \neq \mathcal{P}_C(G)$, since the point x^E is in $\mathcal{P}_C(G)$ but not in F_ω because, assuming that there are no bridges in G, there must be more than one edge between any set U and $V \setminus U$, $U \subset V$ (see Figure 2). Assume that $\pi^T x \geq \pi_0$ is a valid inequality for $\mathcal{P}_C(G)$ such that $F_\omega \subseteq F_\pi := \{x \in \mathcal{P}_C(G) \mid \pi^T x = \pi_0\}$. We want to show that $\pi = \alpha\omega$ and $\pi_0 = \alpha\omega_0$ for some positive real number α. Let $x_e = 1$ for some edge $e \in \delta(U)$, $x_i = 0$ for all $i \in \delta(U) \setminus \{e\}$, and $x_j = 1$ for all $j \in E \setminus \delta(U)$. We note that $x \in F_\omega$ and by hypothesis $x \in F_\omega$. Then

$$\pi_1 x_1 + \ldots + \pi_i x_i + \ldots + \pi_m x_m = \pi_0 \tag{2}$$

Let $\overline{x}$ be obtained from x by setting $x_i = 0$ for some i in $\mathcal{A} := \{(u, v) \in E \mid u, v \in U\} \cup \{(r, t) \in E \mid r, t \in V \setminus U\}$.

Observe that $\overline{x} \in F_\omega$, and by hypothesis $\overline{x} \in F_\pi$. Thus

$$\pi_1 \overline{x}_1 + \ldots + \pi_i \overline{x}_i + \ldots + \pi_m \overline{x}_m = \pi_0 \ . \tag{3}$$

By subtracting equation (3) from equation (2) we obtain $\pi_i x_i = 0$. Since $x_i = 1$, it follows that $\pi_i = 0$. Since $i \in A$ is arbitrary, we infer that $\pi_i = 0$ for all $i \in A$. As $x_i = 0$ for all $\delta(U) \setminus \{e\}$, $x_e = 1$ and $\pi_j = 0$ for all $j \in A$, it follows from equation (2) that $\pi_e = \pi_0$ for all $e \in \delta(U)$. □

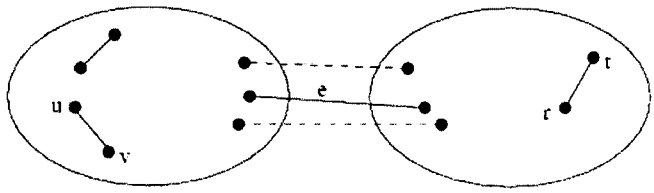
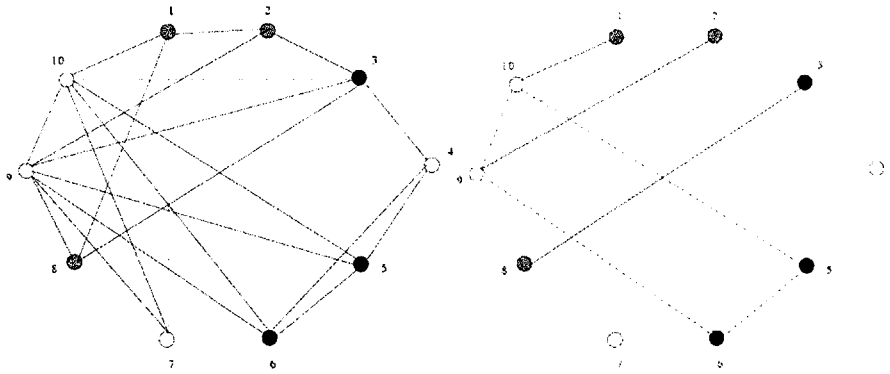


Figure 2. Subsets of graph G used in Theorem 5.



(a) (b)

Figure 3. (a) Problem instance. (b) Linear relaxation solution for instance in (a). The dotted lines in (b) have values 0.5, whereas solid lines have values 1.0. Nodes 1, 2 and 8 are sources, and nodes 3, 5 and 6 are destinations.

We shall now present another class of inequalities for the PPC problem. Let $G = (V, E)$ be a connected undirected graph and $P = (V_1, V_2, \ldots, V_k)$ be a k-partition of nodes of V satisfying $|V_i \cap S| \neq |V_i \cap D|$ with $|V_i \cap S| \neq 0$, $|V_i \cap D| \neq 0$ for $i = 1, \ldots, k$, and $\cup_{i=1}^{k} V_i = V$. In other words, the number of source nodes is not equal to the number of destination nodes in V_i and V_i has at least one source node and at least one destination node. A k-partition defines a multi-cut in G. We note that if $k = 2$, then P is a 2-partition cut. Figure 3 shows an example of a k-partition.

Theorem 6 *Given a k-partition P, the k-partition inequality*

$$\sum_{e\in\Delta(P)} x_e \geq \left\lceil \frac{k}{2} \right\rceil, \tag{4}$$

is valid for $\mathcal{P}_C(G)$, *where* $\Delta(P) = \{uv \in E \mid u \in V_i, v \in V_j$ *and* $i \neq j, 1 \leq i, j \leq k\}$.

Proof: Construct the graph G_P by shrinking each subset V_i into one node V'_i. Note that G_P has k nodes and its edge set is $\Delta(P)$. Now let T' be the restriction of any feasible solution of G onto $\Delta(P)$. Then, G_P is a forest and each component has the same number of source and destination nodes. This proves that inequality (4) is valid for $\mathcal{P}_C(G)$. □

As an example of a k-partition consider instance in Figure 3.a. We note that there is no 2-partition inequality violated by the linear relaxation solution in Figure 3.b. However, that linear relaxation solution violates the 3-partition inequality given by $P = (V_1, V_2, V_3)$, $V_1 = \{6\}$, $V_2 = \{3,4,5,8\}$ and $V_3 = \{1,2,9,10\}$, since $\sum_{e\in\Delta(P)} x_e = 1.5 \not\geq 2$, where $\Delta(P) = \{(5,6),(5,10),(6,9)\}$.

2.3 Separation Procedures

One of the difficulties in applying valid and facet defining inequalities to integer formulations is to define a proper routine capable of finding these inequalities, when given a feasible point for the linear relaxation of the problem. We present some of the separation procedures that can be used to find the described valid inequalities. We will use the following notation in the two separation heuristics described below. Let $U \subset V$ and $T \subset V$. We define $[U:T] = \{(u,t) \in E \mid u \in U,\ t \in T\}$. In other words, $[U:T]$ is the set of edges in the cut defined by node sets U and T in G. We also use $x([U:T])$ to denote $\sum_{e\in[U:T]} x_e$.

The first separation procedure is described in Algorithm 1. We observe that a k-partition obtained by the Separation Procedure 1 may not produce a k-partition inequality that is violated by the vector x'. Note, however, that if at the end of the first iteration of step 3 we have a partition P that violates a k-partition inequality, then we can keep P for the case when the final partition does not yield a violated k-partition.

A second separation procedure for the k-partition can be described as shown in Algorithm 2 (the idea is similar the one described in [5]).

Algorithm 1: Separation procedure 1.

Input: $G=(V,E)$, $x' \in [0,1]^m$

Step 1: Let $P=(V_0,V_1,\ldots,V_{2p})$ be a (2p+1) -partition of the nodes of the graph $G=(V,E)$ such that $V_0 = V \setminus (S \cup D)$, $V_i = S_i$, $i=1,\ldots,p$ and $V_{p+i} = D_i$, $i=1,\ldots,p$. Here, S_i and D_i correspond to the i th element in S and D.

Step 2: Compute $\sum_{e\in\Delta(P)} x'_e$.

Step 3: If $\sum_{e\in\Delta(P)} x'_e < \left\lceil \frac{2p+1}{2} \right\rceil$, then for each node $u \in V_0$ find a set V_i of P that maximize $x'([v:V_i])$, and insert this node in V_i.

Step 4: If $\sum_{e\in\Delta(P)} x'_e \geq \left\lceil \frac{2p+1}{2} \right\rceil$, then find an edge $e \in \Delta(P)$ that maximize x'_e. If e joins sets V_i and V_j, for $i,j \geq 1$, then construct partition P' obtained from P by removing sets V_i and V_j and inserting set $V_i \cup V_j$, replace P by P', and go to Step 2. If edge e joins a node u in V_0 and a set V_i, then assign P to P' and construct partition $P' = (V_0 \setminus \{u\}, V_1, \ldots, V_i \cup \{u\}, \ldots, V_k)$, and go to Step 2.

3. THE BRANCH-AND-CUT ALGORITHM

In order to derive an algorithm to solve the given IP formulation, we start by noting that the number of point-to-point cut inequalities (1) is exponential in n. Thus, to solve even instances of medium size using this formulation is impracticable. However, the facet defining inequalities in the proposed formulation can be used in a branch-and-cut framework to solve the PPC problem.

Branch-and-cut algorithms for combinatorial optimization problems have a vast literature (see [12] for a starting point). The basic idea is to combine branching rules, as in branch-and-bound algorithms, with cutting plane inequalities. By adding new valid inequalities to the problem the size of the original polytope is reduced in most cases, improving computational performance.

For the PPC problem, inequalities can be generated as needed in polynomial time by using an algorithm for the maximum flow problem. This is done as follows: let $x' \in [0,1]^m$ be the solution vector in the current node of the branch-and-cut tree. Construct a graph $G'=(V,E)$ with a capacity

function h associated with the edges of E such that $h(e) = x'_e$ for all $e \in E$. Now for each pair of nodes $(u,v) \in V$, with $u \in S$ and $v \in D$, compute the maximum flow, denoted by $f(u,v)$, between u and v.

Algorithm 2: Separation procedure 2.

Input: Input $G = (V,E)$, $x' \in [0,1]^m$

Step 1: Let $P = (V)$ and $k = 1$; that is P has just one set, the node set of G.

Step 2: Compute $\sum_{e \in \Delta(P)} x'_e$.

Step 3: If $\sum_{e \in \Delta(P)} x'_e < \lceil \frac{k}{2} \rceil$, then this partition induces a violated inequality. Return the violated inequality.

Step 4: If $\sum_{e \in \Delta(P)} x'_e \geq \lceil \frac{k}{2} \rceil$, then choose $i \in \{1,\ldots,k\}$ such that $|V_i \cap S| \geq 1$ and $|V_i \cap D| \geq 1$, compute, in the induced graph $G[V_i]$, a minimum capacity cut from a source node to a destination node, where the edge weights are those in the current x'. Suppose $[V_i^1 : V_i^2]$ is the cut that was found. Then replace P by $P' = (V_1,\ldots,V_i^1,V_i^2,V_k)$ and go to Step 2.

Let G_R be the residual network corresponding to the maximum flow $f(u,v)$, and Y the set of nodes in G_R reachable from u. By the Minimum Cut/Maximum Flow Theorem [6], there exists a cut $\delta(Y), Y \subset V$ with capacity $f(u,v)$ that separates u and v. Now if $f(u,v) < 1$ and $|Y \cap S| \neq |Y \cap D|$, given that all edges in E have integral weight, there can be no edge going from Y to $V \setminus Y$. This implies that

$$\sum_{e \in \delta(Y)} x_e \geq 1$$

is a point-to-point cut inequality violated by the current fractional solution x'. On the other hand, if $f(u,v) \geq 1$ for all pairs (u,v), $u \in S$ and $v \in D$, there are no point-to-point cut inequalities violated by x'. It is clear that the above checking can be done in $O(p^2n^3)$, using a standard preflow-push algorithm [1]. The branch-and-cut algorithm can be described as follows.

1. Start with an initial formulation consisting of $2p$ point-to-point cut inequalities, one for each $u \in S \cup D$. Note that these inequalities are always necessary in the formulation, since they must be valid for each feasible solution. Initialize the branch-and-cut tree to one node

with this formulation.

2. If the tree is not empty, take one of the available nodes and solve the LP relaxation using a LP solver. If all variables are integer and they define a feasible solution, we have an upper bound. Store the solution with smallest upper bound. Else, run the separation procedure.
3. If there is any violated inequality, include it in the formulation and continue on step 2. Else, branch on one of the fractional variables in the current solution of the relaxation.

Note that in the previous algorithm we can avoid to insert new nodes in the tree every time that the cost of the relaxation is greater than the upper bound. We summarize next some of the implementation decisions taken in the implementation of this branch-and-cut algorithm.

Node Selection Strategy We use the best first search strategy to choose the next unexplored node to process, i.e. we select the unexplored node with minimum objective function value.

Branching Strategy We adopt the traditional branching strategy: choose a fractional variable x_i with value closest to 0.5 and generate two new subproblems, one with $x_i = 0$ and another with $x_i = 1$.

Tailing-off Detection Tailing-off is a phenomenon that can appear in the cutting-plane phase in a branch-and-cut algorithm when, although violated inequalities are included to the problem, the profit in the objective function is very small. Some authors suggest that instead of trying to find new cuts, it is better to proceed with the branching phase [12, 14].

We detect the tailing-off phenomenon in each node of the branch-and-cut tree by checking the total improvement in the linear relaxation value during the last 10 iterations. If the total improvement is less than 0.01 percent, we force a branch, instead of generating new cuts.

Fixing variables by reduced cost If we know an upper bound for the value of an optimal integer solution, then it may be possible to set the values of some variables. This can be done by using the reduced costs of non-basic variables in the current linear relaxation, as shown in the following paragraph.

Let x' be the current optimal solution of the linear relaxation and z' its associated value. Suppose that x' is non-degenerate. Let $c_j = c_j - c_B B^{-1} N$ be the reduced cost of the non-basic variable x_j, where B is a basis matrix and N is the matrix consisting of the coefficient columns associated with the non-basic variables. Recall that the reduced cost of a variable gives the

improvement in the solution value when that variable enters the basis. Let z be an upper bound on the value of an optimal integer solution. Then, for each non-basic variable x'_j do:

- If $x'_j = 0$ and $z' + c_j > z$, then insert cut $0 \leq x_j \leq 0$. That is, if variable x_j enters the basis, then the value of the linear relaxation would exceed z.
- If $x'_j = 1$ and $z' - c_j > z$, then insert cut $1 \leq x_j \leq 1$.

Fixing Variables by Logic Implications Right after we apply the procedure for fixing variables by reduced cost, we may use the following ideas to fix some other variables. Let $G'(V,E)$ be the graph obtained from the graph $G'(V,E)$ by deleting the edges fixed at zero. Then we do:

- If there exists $v \in S \cup D$ with *degree* $(v) = 1$ in G', then fix $x_e = 1$ where $e = (v,u)$ or $e = (u,v)$ for some $v \in V$;
- If there exists $v \in V \setminus (S \cup D)$ with *degree* $(v) = 1$ in G', then fix $x_e = 0$ where $e = (v,u)$ or $e = (u,v)$ for some $v \in V$;
- If there exists a bridge joining two vertices $v \in S$, $u \in D$ with $e = (v,u) \in E$, then fix $x_e = 1$.

Selection of Violated Inequalities Let p be a parameter for the branch-and-cut algorithm. Then among all violated inequalities found by the separation procedures for the inequalities classes 2-partition and k-partition, we select the p most violated. By violation of an inequality we mean, given an inequality $a^T x \leq \alpha$ and a vector x', the violation of x with respect to $a^T x \leq \alpha$ is the value $\alpha - a^T x'$.

Pool of Inequalities The inequalities found by the separation procedures are kept in a data structure called the *inequality pool*. The policy used in this pool of inequalities is the following. Each inequality receives a time stamp, representing the time it entered the memory. Whenever k new inequalities are found, then the k oldest inequalities in the pool are deleted and the new inequalities take their places.

4. COMPUTATIONAL RESULTS

In this section we present results of computational experiments performed with the algorithm discussed. The algorithm was implemented in the C programming language and run in an IBM RISC 6000 system using

the AIX operating system, version 4.3. In the branch-and-cut implementation, the CPLEX 7.0 [10] linear programming solver was used.

The test instances were generated as follows: we construct a random Hamiltonian path over a vertex set of cardinality n. Further edges are then included in the graph at random. To each edge we assign a random weight chosen uniformly in the interval $[1,10]$.

Tables 1 to 3 summarize the tests done with the branch-and-cut. The columns *GW*, *LB*, *Optimal solution*, *Nodes*, *LPs*, *Cuts* and *CPU time* represent, respectively, the upper and lower bounds obtained by the Goemans-Williamson algorithm [7], the optimal solution value, the number of nodes in the branch-and-cut tree, the number of linear programs solved, the number of point-to-point cut inequalities and CPU time in hours, minutes and seconds needed in order to prove the optimal value of an instance.

The results presented in Tables 1 to 3 demonstrate the effectiveness of the proposed procedure for medium sized networks, with up to 400 edges, in a relatively small amount of time.

5. CONCLUDING REMARKS

In this paper we described and tested an exact approach to solve the point-to-point connection problem. This problem has interesting applications in routing, particularly for package delivering in multicast networks. We presented an IP formulation for the problem and proved the existence of cutting plane inequalities defined by this formulation. The cutting plane inequalities were used to describe a branch-and-cut algorithm to the PPC problem.

As questions for future research, other facet defining inequalities could be proposed for our formulation of PPC problem. Current work is being done on identifying and proving the validity of such additional cutting plane inequalities. Other formulations could also be employed to explore different aspects of the structure of the PPC polytope. It would also be interesting to develop exact or approximate algorithms for especial cases of the problem.

instance		GW		branch-and-cut				
m	p	Value	LB	Optimal solution	Nodes	LPs	Cuts	CPU time
100	8	40	18	29	208	184	349	1m24s
100	7	51	21	25	1	30	119	6s
100	6	48	22	30	90	112	288	35s
100	5	35	10	21	140	176	293	50s
100	4	24	9	18	6	21	57	2s
100	3	23	14	23	8	37	75	4s
200	8	32	17	21	96	123	340	2m05s
200	7	30	12	19	1426	1066	1098	38m12s
200	6	23	14	18	10	31	88	12s
200	5	31	16	22	12	60	194	29s
200	4	18	8	9	1	13	14	1s
200	3	6	4	6	1	14	13	1s
300	8	24	10	13	352	380	671	12m45s
300	7	22	11	13	36	90	281	2m05s
300	6	13	7	11	80	132	367	2m49s
300	5	17	9	11	46	79	216	1m06s
300	4	14	9	12	14	48	158	35s
300	3	12	4	7	10	32	36	9s
400	8	21	13	14	12	22	57	40s
400	7	16	9	14	638	719	1142	50m20s
400	6	17	9	12	162	204	523	8m11s
400	5	12	8	10	8	36	125	52s
400	4	11	7	8	1	24	37	12s
400	3	6	5	6	1	7	6	1s

Table 1. Results for instances with $n = 30$.

REFERENCES

[1] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin. *Network Flows: theory, algorithms and applications.* Prentice Hall, Eaglewood Cliffs, 1993.

[2] G.V. Chockler, Nabil Huleihel, Idit Keidar, and Danny Dolev. Multimedia multicast transport service for groupware. In *TINA Conference on the Convergence of Telecommunications and Distributed Computing Technologies,* pages 43-54, 1996.

[3] R. Correa, F. Gomes, C.A.S. Oliveira, and P.M. Pardalos. A parallel implementation of an asynchronous team to the point-to-point connection problem. *Parallelal Computing,* 29(4):447-466, 2003.

[4] H. Eriksson. MBONE: the multicast backbone. *Communications of* ACM, 37(8), 1994.

[5] C.E. Ferreira and Y. Wakabayashi. *Combinatória Poliédrica e Planos-de-Corte Faciais, volume 10, Escola de Computação, Unicamp.* Instituto de Computação, Unicamp, Campinas, São Paulo, Brazil, 1996.

[6] L.R. Ford and D.R. Fulkerson. Maximal Flow through a network. *Canadian Journal of Mathematics,* 8:399-404, 1956.

[7] M.X. Goemans and D.P. Williamson. A general Approximation Technique for Constrained Forest Problems. *SIAM J. Comp.,* 24, 1995.

[8] F.C. Gomes, C.N. Meneses, A.G. Lima, and C.A.S. Oliveira. Asynchronous organizations for solving the point-to-point connection problem. In *Proceedings of International Conference on Multi-Agent Systems (ICMAS-98), IEEE Computer Society,* pages 144-149,1998.

[9] L. Han and N. Shahmehri. Secure multicast software delivery. In *IEEE 9th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET ICE'00),* pages 207-212, 2000.

[10] ILOG Inc. ILOG CPLEX 7.0 user's manual. Manual, 2000.

[11] C.L. Li, S.T. McCormick, and D. Simchi-Levi. The Point-to-Point Delivery and Connection Problems: Complexity and Algorithms. *Discrete Applied Math.,* 36:267-292, 1992.

[12] J. Mitchell. Branch-and-cut methods for combinatorial optimisation problems. In P. M. Pardalos and M. G. C. Resende, editors, *Handbook of Applied Optimization.* Oxford University Press, 2002.

[13] M. Natu and S.C. Fang. Network Loading and Connection Problems. Technical Report 308, North Carolina State University, 1995.

[14] M. Padberg and G. Rinaldi. A branch-and-cut algorithm for the resolution of large-scale symmetric travelling salesman problems. *SIAM Review,* 33 (1):60-100, 1991.

instance		GW		branch-and-cut				
m	p	Value	LB	Optimal solution	Nodes	LPs	Cuts	CPU time
100	8	76	32	50	60	183	549	1m38s
100	7	54	27	48	118	159	451	1m49s
100	4	35	22	34	8	73	264	16s
100	3	26	15	26	1	19	22	1s
200	8	38	23	30	302	307	679	6m30s
200	7	25	13	19	20	44	121	24s
200	6	25	15	23	28	77	201	45s
200	5	19	13	16	2	23	50	5s
200	4	19	12	19	2	30	103	9s
200	3	12	5	12	2	29	67	6s
300	8	33	17	20	6	63	280	1m28s
300	7	18	11	14	32	59	194	1m02s
300	6	22	10	14	20	56	242	1m07s
300	5	26	11	20	168	216	424	5m06s
300	4	12	6	10	4	18	45	7s
300	3	12	7	12	8	51	140	26s
400	8	29	16	22	298	402	1171	31m47s
400	7	18	11	14	102	171	495	6m36s
400	6	25	12	17	38	72	203	1m41s
400	5	14	9	11	18	58	243	1m20s
400	4	13	5	10	22	86	230	1m36s
400	3	10	6	7	1	12	11	2s
500	8	28	15	17	24	55	266	3m12s
500	7	25	11.5	16	38	74	280	3m44s
500	6	18	11	14	30	75	233	2m51s
500	5	16	9	12	106	196	460	8m44s
500	4	13	6	11	74	169	397	6m37s
500	3	7	5	7	1	30	77	28s

Table 2. Results for instances with $n = 40$.

instance			GW		branch-and-cut				
n	m	p	Value	LB	Optimal solution	Nodes	LPs	Cuts	CPU time
50	100	12	134	51	84	716	647	1703	23m49s
50	100	11	126	45	78	390	540	1358	12m57s
50	100	10	138	53	85	4170	2643	1855	2h06m23s
50	100	9	80	33	55	58	96	434	44s
50	200	12	67	29	44	492	471	1337	20m06s
50	200	11	77	35	43	34	85	494	2m14s
50	200	10	63	32	48	320	312	905	10m39s
50	200	9	58	25	38	310	415	1271	17m36s
50	300	12	64	28	38	466	436	1233	31m58s
50	300	11	60	28	38	2228	1782	2025	3h56m51s
50	300	10	56	23	32	488	472	1061	26m38s
50	300	9	53	24	36	490	554	1550	46m42s
60	100	12	140	61	95	532	456	1330	13m34s
60	100	11	132	47	90	1516	1271	1877	1h10m57s
60	100	10	132	53	89	178	273	1185	6m33s
60	100	9	114	37	85	834	775	2031	51m57s
60	200	12	75	27	46	2176	2010	2780	4h51m05s
60	200	11	99	52	72	96	187	711	6m10s
60	200	10	73	37	44	40	85	324	1m17s
60	200	9	58	22	39	14	105	647	2m29s
60	300	12	51	29	39	252	314	1409	24m59s
60	300	11	50	20	28	32	48	217	1m16s
60	300	10	62	28	43	440	498	1428	45m59s
60	300	9	45	21	29	666	555	1476	47m06s

Table 3. Results for instances with $n = 50$.

VARIATIONAL INEQUALITY AND EVOLUTIONARY MARKET DISEQUILIBRIA: THE CASE OF QUANTITY FORMULATION

M. Milasi and C. Vitanza
Dept. of Mathematics, University of Messina, Messina, Italy

Abstract: We consider a time-dependent economic market in presence of excess on the supplies and on the demands and we assume that the demand and supply prices depend on the quantity of supplies and demands. This model generalizes the classic spatial price equilibrium problems and adopts, unchanged, the concept of the equilibrium, namely that at the same time the demand price is equal to the supply price plus the cost of transportation, if there is trade between the pair of supply and demand markets. The equilibrium conditions that describe this "disequilibrium" model are expressed in terms of a time-dependent Variational Inequality for which an existence theorem is shown. Moreover by means of the Lagrangean Theory we find the dual variables which have a remarkable economic meaning.

1. INTRODUCTION.

In this paper we are concerned with the spatial price equilibrium problem in the case of quantity formulation and in presence of excess on the supply and on the demand. We assume that the quantities of supplies and of demands evolve in the time and, as consequence, the supply prices and the demand prices, as well as the transportation costs and the commodity shipments, in turns depend on the time. The motivation for this time dependent approach is that, as in [5] P. Daniele writes, "we cannot avoid to consider that each phenomenon of our economic and physical world is not

stable with respect to the time and that our static models of equilibria are a first useful abstract approach" (for other contribution on this matter see [6], [7], [8], [4]). The case in which the excess on the supply and on the demand are not present has been studied by P. Daniele in [5]. The author gives the definition of time dependent market equilibrium and shows that it is equivalent to a Variational Inequality problem. Moreover the author provides existence theorems and performs a stability analysis of the equilibrium patterns. In this paper we follow a suggestion by P. Daniele (see Remark 1 in [5]) and consider a model with supply and demand excesses and with capacity constraints on prices and on transportation costs. The presence of the capacity constraints makes the model more realistic, because it is clear that the prices of supply and of demand are bounded by minimal and maximal prices. Also in this general case we give a time dependent market equilibrium definition and we show that it is equivalent to a Variational Inequality in a suitable Lebesgue space. Moreover we characterize the equilibrium solution by means of Lagrangean multipliers applying the Duality Theory in the case of infinite dimensional spaces.

2. SPATIAL PRICE EQUILIBRIUM PROBLEM. THE CASE OF TIME DEPENDING QUANTITY FORMULATION.

Let us consider n supply markets P_i, $i=1,2,..,n$, and m demand markets Q_j, $j=1,2,..,m$ involved in the production and in the consumption respectively of a commodity during a period of time $[0,T], T>0$. Let $g_i(t), t\in[0,T], i=1,2,..,n$ denote the supply of the commodity associated with supply market i at the time $t\in[0,T]$ and let $p_i(t), t\in[0,T], i=1,2,..,n$ denote the supply price of the commodity associated with supply market i at the time $t\in[0,T]$. A fixed minimal and maximum supply price $\underline{p}_i(t), \overline{p}_i(t)\geq 0$, respectively, for each supply market, are given. Let $f_j(t), t\in[0,T], j=1,2,..,m$ denote the demand associated with the demand market j at the time $t\in[0,T]$ and let $q_j(t), t\in[0,T], j=1,2,..,m$, denote the demand price associated with the demand market j at the time $t\in[0,T]$. Let $\underline{q}_j(t), \overline{q}_j(t)\geq 0$, for each demand market, be the fixed minimal and maximum demand price respectively. Since the markets are spatially separated, let $x_{ij}(t), t\in[0,T], i=1,2,..,n, j=1,2,...,m$ denote the nonnegative commodity shipment transportated from supply market P_i to demand market Q_j at the same time $t\in[0,T]$. Let $c_{ij}(t), t\in[0,T], i=1,2,..,n, j=1,2,..,m$ denote the nonnegative unit transportation cost associated with trading the commodity between (P_i,Q_j) at the same time $t\in[0,T]$. Let we suppose that we are in presence of excesses on the supply and on the demand. Let

$s_i(t), t \in [0,T], i = 1,2,...,n$ denote the supply excess for the supply market P_i at the time $t \in [0,T]$. Let $\tau_j(t), t \in [0,T], j = 1,2,..,m$ denote the demand excess for the demand market Q_j at the time $t \in [0,T]$. We assume that the following feasibility conditions must hold for every $i = 1,2,..,n$ and $j = 1,2,..,m$ a. e. in $[0,T]$:

$$g_i(t) = \sum_{j=1}^{m} x_{ij}(t) + s_i(t), \tag{1}$$

$$f_j(t) = \sum_{i=1}^{n} x_{ij}(t) + \tau_j(t). \tag{2}$$

Grouping the introduced quantities in vectors, we have the total supply vector $g(t) \in L^2([0,T],\mathbb{R}^n)$ and the total demand vector $f(t) \in L^2([0,T],\mathbb{R}^m)$. Furthermore in order to precise the quantity formulation, we assume that two mappings $p(g(t))$ and $q(f(t))$ are given:

$$p : L^2([0,T],\mathbb{R}^n) \to L^2([0,T],\mathbb{R}^n),$$
$$q : L^2([0,T],\mathbb{R}^m) \to L^2([0,T],\mathbb{R}^m).$$

The mapping p assigns for each supply $g(t)$ the supply price $p(g(t))$ and the mapping q assigns for each demand $f(t)$ the demand price $q(f(t))$. Analogously $x(t) \in L^2([0,T],\mathbb{R}^{mn})$ is the vector of commodity shipment and the mapping

$$c : L^2([0,T],\mathbb{R}^{mn}) \to L^2([0,T],\mathbb{R}^{mn})$$

assigns for each commodity shipment $x(t)$ the transportation cost $c(x(t))$. Moreover let $s(t) \in L^2([0,T],\mathbb{R}^n), \tau(t) \in L^2([0,T],\mathbb{R}^m)$ be the vectors of supply and demand excess. Denoting by $w(t) = (g(t), f(t), x(t), s(t), \tau(t))$, we set

$$\tilde{L} = \{w(t) = (g(t), f(t), x(t), s(t), \tau(t)) :$$
$$w(t) \in L^2([0,T],\mathbb{R}^n) \times L^2([0,T],\mathbb{R}^m) \times L^2([0,T],\mathbb{R}^{mn}) \times L^2([0,T],\mathbb{R}^n) \times L^2([0,T],\mathbb{R}^m)\},$$

and

$$\|w(t)\|_{\tilde{L}} = (\|g(t)\|^2_{L^2([0,T],R^n)} + \|f(t)\|^2_{L^2([0,T],R^m)} + \|x(t)\|^2_{L^2([0,T],R^{nm})} + \\ + \|s(t)\|^2_{L^2([0,T],R^n)} + \|\tau(t)\|^2_{L^2([0,T],R^m)})^{\frac{1}{2}}.$$

Furthermore we assume that the feasible vector $w(t) = (g(t), f(t), x(t), s(t), \tau(t))$ satisfies the condition

$$w(t) \geq 0 \quad \text{a.e.in}[0,T] \tag{3}$$

Taking into account conditions (1), (2) and (3), the set of feasible vectors $w(t)$ is:

$$\tilde{K} = \Big\{ w(t) = (g(t), f(t), x(t), s(t), \tau(t)) \in \tilde{L} : w(t) \geq 0,$$

$$g_i(t) = \sum_{j=1}^{m} x_{ij}(t) + s_i(t) \quad i = 1,2,...,n$$

$$f_j(t) = \sum_{i=1}^{n} x_{ij}(t) + \tau_j(t) \quad j = 1,2,...,m \quad a.e.\, in\,[0,T] \Big\}.$$

$\tilde{K}$ is a convex, closed, not bounded subset of the Hilbert space $\tilde{L}$.

Finally the presence of the capacity constraints on p, q, c can be expressed in the following way:

$$\underline{p}(t) \leq p(g(t)) \leq \bar{p}(t),$$
$$\underline{q}(t) \leq q(f(t)) \leq \bar{q}(t),$$
$$\underline{c}(t) \leq c(x(t)) \leq \bar{c}(t),$$

for each $w(t) = (g(t), f(t), x(t), s(t), \tau(t)) \in \tilde{K}$, where

$$\underline{p}(t) = (\underline{p}_1(t), \underline{p}_2(t), ..., \underline{p}_n(t)) \text{ and } \underline{q}(t) = (\underline{q}_1(t), \underline{q}_2(t), ..., \underline{q}_m(t)).$$

Then the time-dependent market equilibrium condition in the case of the quantity formulation takes the following form:

Definition 2.1. Let $w^*(t) = (g^*(t), f^*(t), x^*(t), s^*(t), \tau^*(t))$ in $\tilde{K}$. $w^*(t)$ is a market equilibrium if and only if for each $i = 1,2,..,n$ and $j = 1,2,..,m$ the following conditions hold a. e. in $[0,T]$:

$$\begin{cases} \text{if} \quad s_i^*(t) > 0 \Rightarrow p_i(g^*(t)) = \underline{p}_i(t) \\ \\ \text{if} \quad \underline{p}_i(t) < p_i(g^*(t)) \Rightarrow s_i^*(t) = 0 \end{cases} \tag{4}$$

$$\begin{cases} \text{if} \quad \tau_j^*(t) > 0 \Rightarrow q_j(f^*(t)) = \overline{q}_j(t) \\ \\ \text{if} \quad q_j(f^*(t)) < \overline{q}_j(t) \Rightarrow \tau_j^*(t) = 0 \end{cases} \tag{5}$$

$$\begin{cases} \text{if} \quad x_{ij}^*(t) > 0 \Rightarrow p_i(g^*(t)) + c_{ij}(x^*(t)) = q_j(f^*(t)) \\ \\ \text{if} \quad p_i(g^*(t)) + c_{ij}(x^*(t)) > q_j(f^*(t)) \Rightarrow \quad x_{ij}^*(t) = 0 \end{cases} \tag{6}$$

Condition (4) states that if in the supply market P_i there is supply excess in the time t, then the supply price in P_i must be equal to the supply minimal price in P_i at the time t. If the supply price in P_i is not equal to the minimal price at the time t, then in P_i at the time t there is not supply excess.

Conditions (5) have similar meaning. Condition (6) states that if the trade between a pair (P_i, Q_j) at the time t is greater to zero, then the supply price at supply market P_i plus the transportation cost between the pair of markets at the same time t must be equal to the demand price at demand market Q_j at the time t; whereas if the supply price plus the transportation cost at the same time t exceeds the demand price at the time t, then the trade between the supply and demand market pair at the time t will be equal to zero. Now let us consider the following Variational Inequality:

"Find $w^*(t) \in \widetilde{K}$ *such that:*

$$\int_0^T (p(g^*(t))(g(t) - g^*(t)) - q(f^*(t))(f(t) - f^*(t)) + c(x^*(t))(x(t) - x^*(t)) +$$
$$-\underline{p}(t)(s(t) - s^*(t)) + \overline{q}(t)(\tau(t) - \tau^*(t)))dt \geq 0 \tag{7}$$
$$\forall \; w(t) = (g(t), f(t), x(t), s(t), \tau(t)) \in \widetilde{K}."$$

If we denote

$$L = L^2([0,T],\mathbb{R}^{nm}) \times L^2([0,T],\mathbb{R}^n) \times L^2([0,T],\mathbb{R}^m),$$
$$K = \{u(t) = (x(t), s(t), \tau(t)) \in L : u(t) \geq 0\},$$

and

$$p(x(t),s(t)) = p(\sum_{j=1}^{m} x_{1j}(t) + s_1(t), \sum_{j=1}^{m} x_{2j}(t) + s_2(t), \ldots\ldots, \sum_{j=1}^{m} x_{nj}(t) + s_n(t)),$$
$$q(x(t),\tau(t)) = q(\sum_{i=1}^{n} x_{i1}(t) + \tau_1(t), \sum_{i=1}^{n} x_{i2}(t) + \tau_2(t), \ldots\ldots, \sum_{i=1}^{n} x_{im}(t) + \tau_m(t)),$$

taking into account the feasibility conditions, Variational Inequality (7) can also rewrite in the following form:

"Find $u^*(t) \in K$ *such that:*

$$\int_0^T \{(p(x^*(t),s^*(t)) - q(x^*(t),\tau^*(t)) + c(x^*(t)))(x(t) - x^*(t)) +$$
$$+(p(x^*(t),s^*(t)) - \underline{p}(t))(s(t) - s^*(t)) - (q(x^*(t),\tau^*(t)) - \overline{q}(t))(\tau(t) - \tau^*(t))\}dt \geq 0$$
$$\forall\ u(t) = (x(t),s(t),\tau(t)) \in K." \qquad (8)$$

In fact the following result holds:

Lemma 2.1. *Under assumptions (1) and (2), Variational Inequalities (7) and (8) are equivalent.*

Proof. Let $w^* \in \widetilde{K}$ a solution to the problem (7). Taking into account the conditions (1) and (2) we have:

$$g_i(t) = \sum_{j=1}^{m} x_{ij}(t) + s_i(t) \quad \forall i = 1,2,\ldots,n$$
$$f_j(t) = \sum_{i=1}^{n} x_{ij}(t) + \tau_j(t) \quad \forall j = 1,2,\ldots,m.$$

Then from (7) for all $w \in \widetilde{K}$ we have:

$$\int_0^T \{p(g^*(t))(g(t) - g^*(t)) - q(f^*(t))(f(t) - f^*(t)) + c(x^*(t))(x(t) - x^*(t)) +$$
$$-\underline{p}(t)(s(t) - s^*(t)) + \overline{q}(t)(\tau(t) - \tau^*(t))\}dt =$$

$$= \int_0^T \{\sum_{i=1}^{n} p_i(g^*(t))(\sum_{j=1}^{m} x_{ij}(t) + s_i(t) - \sum_{j=1}^{m} x_{ij}^*(t) - s_i^*(t)) +$$

$$-\sum_{j=1}^{m} q_j(f^*(t))(\sum_{i=1}^{n} x_{ij}(t) + \tau_j(t) - \sum_{i=1}^{n} x_{ij}^*(t) - \tau_j^*(t)) + \sum_{i,j} c_{ij}(x^*(t))(x_{ij}(t) - x_{ij}^*(t)) +$$

$$-\sum_{i=1}^{n} \underline{p}_i(t)(s_i(t) - s_i^*(t)) + \sum_{j=1}^{m} \overline{q}_j(t)(\tau_j(t) - \tau_j^*(t))\}dt =$$

$$= \int_0^T \{\sum_{i,j} (p_i(g^*(t)) - q_j(f^*(t)) + c_{ij}(x^*(t)))(x_{ij}(t) - x_{ij}^*(t)) +$$

$$+\sum_{i=1}^{n} (p_i(g^*(t)) - \underline{p}_i(t))(s_i(t) - s_i^*(t)) - \sum_{j=1}^{m} (q_j(f^*(t)) - \overline{q}_j(t))(\tau_j(t) - \tau_j^*(t))\}dt =$$

$$= \int_0^T \{(p(g^*(t)) - q(f^*(t)) + c(x^*(t)))(x(t) - x^*(t)) + (p(g^*(t))) - \underline{p}(t))(s(t) - s^*(t)) +$$

$$-(q(f^*(t)) - \overline{q}(t))(\tau(t) - \tau^*(t))\}dt =$$

$$= \int_0^T \{(p(x^*(t), s^*(t)) - q(x^*(t), \tau^*(t)) + c(x^*(t)))(x(t) - x^*(t)) +$$

$$+(p(x^*(t), s^*(t))) - \underline{p}(t)(s(t) - s^*(t)) - (q(x^*(t), \tau^*(t)) - \overline{q}(t))(\tau(t) - \tau^*(t))\}dt.$$

Then the above inequality holds for $u \in K$. If we find $w^* = (g^*, f^*, x^*, s^*, \tau^*) \in \widetilde{K}$ such that:

$$\int_T \{p(g^*(t))(g(t) - g^*(t)) - q(f^*(t))(f(t) - f^*(t)) + c(x^*(t))(x(t) - x^*(t)) +$$

$$-\underline{p}(t)(s(t) - s^*(t)) + \overline{q}(t)(\tau(t) - \tau^*(t))\}dt \geq 0$$

$$\forall \; w(t) = (g(t), f(t), x(t), s(t), \tau(t)) \in \widetilde{K}$$

then $u^* = (x^*, s^*, \tau^*) \in K$ verifies Variational Inequality (8) for all $u = (x, s, \tau) \in K$. □

If we consider the function:

$$v : L \to L$$

defined setting for each $u \in L$:

$$v(u) = (p(x^*, s^*) - q(x^*, \tau^*) + c(x^*), p(x^*, s^*) - \underline{p}, \overline{q} - q(x^*, \tau^*)),$$

Variational Inequality (8) can be rewrite as:

"Find $u^* \in K$ *such that:*

$$< v(u^*), u - u^* > \geq 0, \quad \forall u \in K."$$

Now let us characterize time dependent market equilibrium as a solution to Variational Inequality (8) (and hence (7)). In fact the following result holds:

Theorem 2.1. $u^* = (x^*, s^*, \tau^*) \in K$ *is a time dependent market equilibrium if and only if* u^* *is a solution to Variational Inequality (8).*

Proof. Let $u^* \in K$ be a market equilibrium.
For every $x(t) \geq 0$ we have:

$$\int_0^T \left(p(x^*(t), s^*(t)) - q(x^*(t), \tau^*(t)) + c(x^*(t))\right)(x(t) - x^*(t))dt \geq 0$$

In fact: if $x_{ij}^*(t) > 0$, by (6) we have $p_i(x^*(t), s^*(t)) - q_j(x^*(t), \tau^*(t)) + c_{ij}(x^*(t)) = 0$ and the product vanishes. Otherwise, if $x_{ij}^*(t) = 0$ then, from (6) we have $p_i(x^*(t), s^*(t)) - q_j(x^*(t), \tau^*(t)) + c_{ij}(x^*(t)) \geq 0$ and the product is nonnegative a. e. in $[0,T]$. Then the integral is nonnegative.

By similar case distinctions, one obtains that each product in the second and third sum is nonnegative. Then second and third integral are nonnegative. This proves that u^* is a solution to Variational Inequality (8).

Conversely, let Variational Inequality (8) holds. Let us prove that equilibrium conditions hold. First let us suppose that (6) is not verified, that is there exists a set $E \subset [0,T]$ such that $m(E) > 0$ and there exist i^* and j^* such that for each $t \in E$:

$$p_{i^*}(x^*(t), s^*(t)) - q_{j^*}(x^*(t), \tau^*(t)) + c_{i^*j^*}(x^*(t)) < 0.$$

Then if we choose $\overline{x} \in L^2([0,T], \mathbb{R}^{mn})$ such that

$$\overline{x}(t) = x^*(t) \quad \forall t \in [0,T] \setminus E,$$

and :

$$\overline{x}_{ij}(t) \begin{cases} = x_{ij}^*(t) & \forall (i,j) \neq (i^*, j^*),\ t \in E \\ > x_{i^*j^*}^*(t) & t \in E, \end{cases}$$

assuming in (8) $x(t)=\bar{x}(t)$, $s(t)=s^*(t)$, $\tau(t)=\tau^*(t)$, we get

$$< v(u^*), u-u^* >= \int_0^T (p(x^*(t),s^*(t)) - q(x^*(t),\tau^*(t)) + c(x^*(t)))(\bar{x}(t)-x^*(t))dt =$$
$$= \int_{[0,T]\setminus E} (p(x^*(t),s^*(t)) - q(x^*(t),\tau^*(t)) + c(x^*(t)))(\bar{x}(t)-x^*(t))dt +$$
$$+ \int_E (p(x^*(t),s^*(t)) - q(x^*(t),\tau^*(t)) + c(x^*(t)))(\bar{x}(t)-x^*(t))dt =$$
$$= \int_E (p_{i^*}(x^*(t),s^*(t)) - q_{j^*}(x^*(t),\tau^*(t)) + c_{i^*j^*}(x^*(t)))(\bar{x}_{i^*j^*}(t)-x^*_{i^*j^*}(t))dt < 0.$$

So we have proved that

$$p_i(x^*,s^*) - q_j(x^*,\tau^*) + c_{ij}(x^*) \geq 0 \quad \forall i,j \quad a.e.\ in\ [0,T].$$

Now let us prove that if $x^*_{ij} > 0$ then

$$p_i(x^*(t),s^*(t)) + c_{ij}(x^*(t)) = q_j(x^*(t),\tau^*(t)) \quad a.e\ in\ [0,T].$$

To this and, let us suppose ab absurdum that there exist a set $E \subset [0,T]$, i^*, j^* such that $m(E)>0$ and

$$p_{i^*}(x^*,s^*) - q_{j^*}(x^*,\tau^*) + c_{i^*j^*}(x^*) > 0.$$

Then if we choose

$$\bar{x}_{ij}(t)\begin{cases} = x^*_{ij}(t) & \text{in } [0,T]\setminus E \ \ \forall (i,j)\neq(i^*,j^*) \\ = x^*_{ij}(t) & \forall (i,j)\neq(i^*,j^*),\ t\in E \\ < x^*_{i^*j^*}(t) & t\in E, \end{cases}$$

assuming in (8) $x(t)=\bar{x}(t)$, $s(t)=s^*(t)$, $\tau(t)=\tau^*(t)$

$$< v(u^*, u-u^*) >=$$
$$\int_E (p_{i^*}(x^*(t),s^*(t)) - q_{j^*}(x^*(t),\tau^*(t)) + c_{i^*j^*}(x^*(t)))(\bar{x}_{i^*j^*}(t)-x^*_{i^*j^*}(t))dt < 0$$

It remain to prove that if, a. e. in $[0,T]$

$$p_{i^*}(x^*(t),s^*(t))+c_{i^*j^*}(x^*(t))>q_{j^*}(x^*(t),\tau^*(t))$$

then $x^*_{ij}(t)=0 \quad \forall i,j$. To this and, let us suppose ab absurdum that there exist a set $E\subset[0,T]$ and i^*, j^* such that $m(E)>0$

$$p_{i^*}(x^*(t),s^*(t))+c_{i^*j^*}(x^*(t))>q_{j^*}(x^*(t),\tau^*(t)) \quad \text{in } E$$

and $x^*_{i^*j^*}>0$.

Using the previous arguments it is easily to get a contradiction and hence x^*_{ij} must be zero for all i and j.

Now, let us prove that condition (4) is verified. Let us suppose ab absurdum that there exist a set $E\subset[0,T]$ and i^* such that $m(E)>0$:
$s^*_{i^*}>0$ and $p_{i^*}(x^*(t),s^*(t))>\underline{p}_{i^*}(t)$ in E.

If we choose the chooses:

$$\overline{s}_i(t)\begin{cases} s^*_i(t) & \forall t\in T\setminus E \quad \forall i=1,2,..,n \\ =s^*_i(t) & \forall i\neq i^*,\ \forall t\in E \\ <s^*_{i^*}(t) & i=i^*,\ \forall t\in E, \end{cases}$$

assuming $x(t)=x^*(t),\ s(t)=\overline{s}(t),\ \tau(t)=\tau^*(t)$, from (8) we get

$$<v(u^*,u-u^*)>=\int_E (p_{i_*}(x^*(t),s^*(t))-\underline{p}_{i^*})(\overline{s}_{i^*}-s_{i^*})dt<0.$$

Let us suppose ab absurdum that there exist a set $E\subset[0,T]$ and i^* such that $m(E)>0$ and:
if $s^*_{i^*}>0\Rightarrow p_{i^*}(x^*(t),s^*(t))>\underline{p}_{i^*}(t)$ in E. Analogously we proceed in order to obtain the another equilibrium conditions. □

3. EXISTENCE THEOREMS.

Let us recall some concepts that will be useful in the following. Let E be a real topological vector space, $K\subseteq E$ convex. Then $v:K\to E^*$ is said to be:

1. pseudomonotone if and only if

$$\forall u_1,u_2\in K \quad \langle v(u_1),u_2-u_1\rangle\geq 0 \Rightarrow \langle v(u_2),u_1-u_2\rangle\leq 0;$$

2. hemicontinuous if and only if

$\forall u \in K$ the function $z \to \langle v(z), u - z \rangle$

is upper semicontinuous on K;

3. *hemicontinuous along line segments* if and only if

$\forall u_1, u_2 \in K$ the function $z \to \langle v(z), u_2 - u_1 \rangle$

is upper semicontinuous on the line segment $[u_1, u_2]$.

Adapting a classical existence theorem for the solution of a variational inequality to our problem, we will have the following theorem, which provide existence with or without pseudomonotonicity assumptions. Moreover, since the convex K is unbounded, we need coercivity assumptions.

Theorem 1. *Each of the following conditions is sufficient to ensure the existence of the solution of (8):*

1. $v(u) = v(x(t), s(t), \tau(t))$ *is hemicontinuous with respect to the strong topology and there exist* $A \subseteq K$ *compact and* $B \subseteq K$ *compact, convex with respect to the strong topology such that*

$$\forall u_1 \in K \setminus A \quad \exists u_2 \in B : \langle v(u_1), u_2 - u_1 \rangle < 0;$$

2. v *is pseudomonotone,* v *is hemicontinuous along line segments and there exist* $A \subseteq K$ *compact and* $B \subseteq K$ *compact, convex with respect to the weak topology such that*

$$\forall p \in K \setminus A \quad \exists \tilde{p} \in B : \langle v(p), \tilde{p} - p \rangle < 0;$$

3. v *is hemicontinuous on* K *with respect to the weak topology,* $\exists A \subseteq K$ *compact,* $\exists B \subseteq K$ *compact, convex with respect to the weak topology such that*

$$\forall p \in K \setminus A \quad \exists \tilde{p} \in B : \langle v(p), \tilde{p} - p \rangle < 0.$$

4. LAGRANGEAN THEORY.

Now, our pourpose is to give a characterization to evolutionary market equilibrium conditions in terms of the Lagrangean multipliers, which, as it's very well know, play a very important role in economic theory. To this and we can prove the following result:

Theorem 4.1. *Let* $u^* \in K$ *be a solution to problem (8). Then there exist three functions* $\alpha^* \in L^2([0,T],\mathbb{R}^{nm})$, $\beta^* \in L^2([0,T],\mathbb{R}^n)$, $\gamma^* \in L^2([0,T],\mathbb{R}^m)$ *such that:*

$$\alpha^*(t),\beta^*(t),\gamma^*(t) \geq 0 \quad a.e.\ in\ [0,T];$$

$$\alpha^* \cdot x^* = 0,\ \beta^* \cdot s^* = 0,\ \gamma^* \cdot \tau^* = 0;$$

$$\begin{cases} p(x^*,s^*) - q(x^*,\tau^*) + c(x^*) = \alpha^*, \\ p(x^*,s^*) - \underline{p} = \beta^*, \\ \overline{q} - q(x^*,\tau^*) = \gamma^*. \end{cases}$$

In order to prove theorem (4.1) we need the following tools. We observe that if $u^* \in K$ is a solution to problem (8) then

$$\min_{u \in K} < v(u^*), u - u^* >= 0.$$

Let us introduce the functional

$$\psi_{u^*}(u) =< v(u^*), u - u^* >, \quad \forall u \in K$$

Let us observe that

$$\psi_{u^*}(u) \geq 0$$

and

$$\min_{u \in K} \psi_{u^*}(u) = 0.$$

We associate to Variational Inequality (8) the following Lagrangean function:

$$\forall u \in L,\ l = (\alpha,\beta,\gamma) \in C^*$$

$$L(u,l) = \psi_u(u) - (\int_0^T \alpha(t)x(t)dt + \int_0^T \beta(t)s(t)dt + \int_0^T \gamma(t)\tau(t)dt)$$

where

$$C^* = \{l(t) = (\alpha(t), \beta(t), \gamma(t)) \in L : \alpha(t) \geq 0, \beta(t) \geq 0, \gamma(t) \geq 0) \ \ a.e. \ in \ [0,T]\}$$

is the dual cone of L. We observe that the dual cone of L, from Riesz theorem is equal to the ordering cone of L: $C^* = C$.

For infinite dimensional convex optimization problems often the underlying constraint set has empty interior, so that the Slater constraint qualification condition cannot be applied. Following a suggestion by Borwein and Lewis [1], it is possible to overcome this difficulty replacing the Slater qualification condition by generalizing the notion of relative interiors as follows.

Definition 4.1. The quasi relative interior of a convex set K, which we denote by $qri\ K$, is the set of those x for which

$$Cl\ Cone(K - x)$$

is a subspace, where

$$Cone(K - x) = \{\lambda y : \lambda \geq 0, y \in K - x\}.$$

Then the generalized condition is the following

$$qri\ K = \varnothing.$$

In our case, the quasi relative interior of K is

$$qri\ K = \{(\alpha(t), \beta(t), \gamma(t)) \in K : \alpha(t) > 0, \beta(t) > 0, \gamma(t) > 0 \ \ a.e. \ in \ [0,T]\}.$$

Now, the infinite dimensional Lagrangean and duality theory follows from a separation theorem, proved in [2], in which the classical interior is replaced by the quasi-relative interior.

Then the Lagrangean and the duality theory can be adapted in the following way.

Proposition 4.1. *The problem*

$$\min_{u\in K} \psi_{u^*}(u). \tag{9}$$

is equivalent to the problem

$$\min_{u\in L} \sup_{l\in K} \left\{ \psi_{u^*}(u) - \left(\int_0^T \alpha(t)x(t)dt + \int_0^T \beta(t)s(t)dt + \int_0^T \gamma(t)\tau(t)dt \right) \right\}. \tag{10}$$

Proof. See [3].

Let us consider the dual problem

$$\max_{l\in K} \inf_{u\in L} \left\{ \psi_{u^*}(u) - \left(\int_0^T \alpha(t)x(t)dt + \int_0^T \beta(t)s(t)dt + \int_0^T \gamma(t)\tau(t)dt \right) \right\}. \tag{11}$$

and the associated problem

$$\max_{\Lambda\in\Delta} \Lambda. \tag{12}$$

where

$$\Delta = \left\{ \Lambda \in R : \psi_{u^*}(u) - \left(\int_0^T \alpha(t)x(t)dt + \int_0^T \beta(t)s(t)dt + \int_0^T \gamma(t)\tau(t)dt \right) \geq \Lambda \right\}.$$

It results:

Proposition 4.2. *$(\alpha^*, \beta^*, \gamma^*) \in K$ is a maximal solution to dual problem (11) if and only if Λ is a solution to (12).*

Proof. See [3].

Proposition 4.3. *If the problem (9) (or (10))is solvable, then dual problem (11) is also solvable and the extremal values of the two problems are equal.*

Finally, from the previous results easy follows:

Proposition 4.4 *u^* is a solution to (9) if and only if there exists $l^* \in K$ such that (u^*, l^*) is a saddle point of $L(u,l)$; namely:*

$$L(u^*,l^*) \le L(u,l^*) \quad \forall u \in L$$

$$L(u^*,l^*) \ge L(u^*,l) \quad \forall l \in K$$

Furthermore we are:

$$L(u^*,l^*) = \min_{u \in K} \psi_{u^*}(u) = 0$$

Proof of theorem (4.1).

Let u^* be a solution to minimal problem (9). Because $(\alpha^*,\beta^*,\gamma^*) \in K$ it follows that

$$\alpha^*(t) \ge 0,\ \beta^*(t) \ge 0,\ \gamma^*(t) \ge 0 \quad a.e.\ in\ [0,T].$$

Moreover taking into account that $L(u,l^*) = 0$, we get

$$0 = L(u^*,l^*) = \psi(u^*) - \int_0^T \alpha^*(t)x^*(t)dt - \int_0^T \beta^*(t)s^*(t)dt - \int_0^T \gamma^*(t)\tau^*(t)dt = 0$$
$$-\int_0^T \alpha^*(t)x^*(t)dt - \int_0^T \beta^*(t)s^*(t)dt - \int_0^T \gamma^*(t)\tau^*(t)dt = 0.$$

Being α^*, x^*, β^*, s^*, γ^*, τ^* nonnegative functions, we derive:

$$\alpha^* x^* = 0,\ \beta^* s^* = 0,\ \gamma^* \tau^* = 0 \quad a.e.\ in\ [0,T].$$

Finally, for all u in L we have:

$$L(u,l^*) = < v(u^*), u - u^* > - < l^*, u > + < l^*, u^* > =$$
$$= < v(u^*) - l^*, u - u^* >$$

Assuming:

$$u_1 = u^* + \varepsilon,\ u_2 = u^* - \varepsilon \quad \forall \varepsilon \in D([0,T])$$

we get, for all $\varepsilon \in D([0,T])$:

$$L(u_1,l^*) = < v(u^*) - l^*, u_1 - u^* > = < v(u^*) - l^*, \varepsilon >$$
$$L(u_2,l^*) = < v(u^*) - l^*, u_2 - u^* > = < v(u^*) - l^*, -\varepsilon >,$$

hence

$$v(u^*) - l^* = 0$$

namely:

$$\begin{cases} p(x^*, s^*) - q(x^*, \tau^*) + c(x^*) = \alpha^* \\ p(x^*, s^*) - \underline{p}(t) = \beta^* \\ \overline{q}(t) - q(x^*, \tau^*) = \gamma^* \end{cases} \tag{13}$$

□

Remark. The importance of functions $\alpha^*, \beta^*, \gamma^*$ derives from the fact that they are able to describe the behaviour of the evolutionary market. In fact the set $A_+^{ij} = \{t \in [0,T] : \alpha_{ij}^*(t) > 0\}$ indicates the time when there is not trade between the supply market i and the demand market j. Analogously $B_+^j = \{t \in [0,T] : \beta_i^*(t) > 0\}$ indicates the time when there is a zero supply excess of the market i. The same holds for γ_j^* which indicates when the demand market j has not demand excess. Moreover it is easy to show that if $u \in K$ and there exist $\alpha^*, \beta^*, \gamma^*$ as in theorem (4.1) such that conditions (13) are fulfilled, then u verifies Variational Inequality (8).

REFERENCES

[1] Borwein, J.M. and Lewis, A. S. (1989), Practical conditions for Fenchel duality in Infinite dimensions *Pitman Research Notes in Mathematics Series*, 252, 83–89.

[2] Cammaroto, F. and Di Bella, B., A separation theorem based on the quasi-relative interior and an application to the theory of duality , Preprint.

[3] Daniele, P. (1999), Lagrangean Function for Dynmic Variational Inequalities, *Rendiconti del Circolo Matematico di Palermo*, Vol. 58, pp. 101-119, 43–58.

[4] Daniele, P. (2001), Variational Inequalities for Static Equilibrium Market. Lagrangean Function and Duality, *Equilibrium Problems: Nonsmooth Optimization and Variational Inequality Models*, Kluwer Academic Publishers, F. Giannessi - A. Maugeri - P. Pardalos Eds., 43–58.

[5] Daniele, P., Time-Dependent spatial Price Equilibrium Problem: Existence and Stability results for the Quantity Formulation Model, *Journal of Global Optimization*, to Appear .

[6] Daniele, P. and Maugeri, A. (2001), On Dynamical Equilibrium Problems and Variational Inequalities, *Equilibrium Problems: Nonsmooth Optimization and Variational Inequality Models*, Kluwer Academic Publishers, F. Giannessi - A. Maugeri - P. Pardalos Eds., 59–69.

[7] Daniele, P., Maugeri, A. and Oettli, W. (1998), Variational Inequalities and time-dependent traffic equilibria, *C. R. Acad. Sci. Paris* t. 326, serie I, 1059–1062.

[8] Daniele, P., Maugeri, A. and Oettli, W. (1999), Time Dependent Traffic Equilibria, *Jou. Optim. Th. Appl.*, Vol. 103, No. 3, 543–555.

NUMERICAL APPROXIMATION OF FREE BOUNDARY PROBLEM BY VARIATIONAL INEQUALITIES. APPLICATION TO SEMICONDUCTOR DEVICES

M. Morandi Cecchi and R. Russo
University of Padova, Dept. of Pure and Applied Mathematics, Padova, Italy

Abstract: In this paper we treat problem arasing in semiconductor theory from a mathematical and numerical point of view, in particular we consider a boundary value problem with unknown interfaces arising by the determination of the depletion layer in the most basic semiconductor device namely the p-n junction diode. We present the numerical approximation of free boundary problem with double obstacle treated with quasi-variational inequalities. We deal with the L^∞ convergence of the standard finite element approximation of the system of quasi-variational inequalities.

1. INTRODUCTION

Problems in which the solution of a differential equation has to satisfy certain conditions on the boundary of a prescribed domain are referred to as boundary-value problems. In many important case, as free boundary problems, the boundary of the domain is not known in advance but has to be determinated as a part of the solution. Typically, a free boundary problem consist of a partial differential equations of elliptic type to be satisfied within a bounded domain together with necessary boundary conditions; one section of the boundary, the free boundary, is unknown and must be determined as part of the solution. These problems have been popular subject for research

in recent years, leading to a collection of new mathematical methods. Flow through porous media is an important source of free boundary problems [1], most frequently in relation to seepage phenomena that occur in nature. Examples are seepage throught earth dams; seepage out of open channels such as rivers, canals, ponds, and irrigation system. Practical interest in free boundary problems, however, is not confined to natural seepage but extends for example to topics in plasma physics, semiconductors, and electrochemical machinary. This work analyses a free boundary problem in semiconductors field, in particular the modelling of reverse-biased devices. In fact for the steady-state case of $p-n$ junction diode under reverse bias, after a singular perturbation analysis, the determination of the depletion layer leads to a free boundary problem.

For the case of $p-n$ junction diode under strong revers bias, an approximating problem which includes the same free-boundary for the potential and a mixed elliptic-hyperbolic problem for the analysis of current flow, has been derived and analyzed in a series of papers by Schmeiser [26],[27].

Without being derived as a limit of a singularly perturbed system, the double obstacle problem has already been formulated as a model for the potential distribution by Hunt and Nassif [16]. The free boundary model presented here differs from the previous, by the definition of the obstacles which are equal to the quasi-Fermi level, obtained as a solution to the continuity equations; we give a quasi-variational formulation of the model.

Then we deal with the L^{∞} convergence of the finite element approximation of the system of quasi-variational inequalities. The L^{∞}- error estimate is of particular interest not only for practical reasons but also due to its inherent difficulty of convergence in this norm. Moreover, the interest in using such a norm for the approximation of obstacle problems is that they are types of free boundary problems. This fact was validated by the paper of F. Brezzi; C. Caffarelli, [8] and later by that of Nochetto [20], on the convergence of the discrete free boundary to the continuous one.

A lot of results on error estimates for the classical obstacle problems and variational inequalities were achieved in this norm, (cf., e.g [2], [19], [12], [21]). However, very few works concerning quasi-variational inequalities are known on this subject. (cf., [14]), [6]), Under a $W^{2,p}(\Omega)$- regularity of the continuous solution, a quasi-optimal L^{∞}- convergence of finite element method is established, involving a monotone algorithm of Bensoussan-Lions type and standard L^{∞}- error estimates known for elliptic variational inequalities.

2. REVERSE BIASED p-n JUNCTION

One of the basic properties of semiconductors is the controlled implantation of impurity atoms into a semiconductor crystal; this process is usually called *doping*. It is possible to introduce into the crystal dopant atoms which can produce one or more excess conduction electrons (called donors), or dopant atoms which can accept electrons and thus produce holes (called acceptors). This process increases the conductivity significantly, and thus the electrical properties of the crystal can be controlled by doping. The performance of a semiconductor device is mainly determined by the distributions of donors and acceptors.

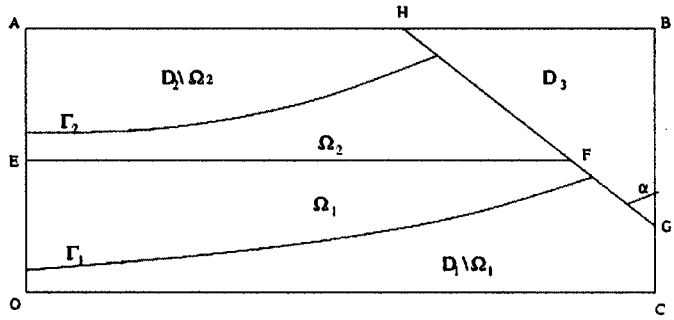


Figure 1. p-n junction

In the p-n junction the *p-side*, doped with acceptors, is positively charged and the *n-side*, doped with donors, is negatively charged. As a result of the tendency of holes to diffuse into the n region and of electrons into the p region, a nonconducting region is set up along the junction, called a *depletion layer.*

When a positive bias is applied to the junction, a large current flows through the diode, even if the voltage is small; a negative applied voltage widens the depletion layer. The unknown or free boundaries limiting the depletion layer are interfaces with another, dielectric region. The interfaces are determined by the concentrations of donors and acceptors and the potentials applied.

As in Fig. 1, let D_1 be the open boundary by the contour $OEFGC$, D_2 the one bounded $AHFE$, and D_3 the triangular domain bounded by HBG : D consist of the whole rectangular domain $OABC$, while Ω_1 and Ω_2 are the open sets which define the depletion layer. We also let $\Omega_1' = D_1/\overline{\Omega}_1$, $\Omega_2' = D_2/\overline{\Omega}_2$; $\Gamma_1 = D_1 \cap \overline{\Omega}_1$, $\Gamma_2 = D_2 \cap \overline{\Omega}_2$.

The model which describes potential distribution $u(x, y)$ in semiconductor device is the drift-diffusion one

$$\begin{cases} \nabla^2 u = \dfrac{q}{\varepsilon}(n - p - C) \\ \nabla \cdot (D_n \nabla n - n\mu_n \nabla \psi) = R_n \\ \nabla \cdot (D_p \nabla p + p\mu_p \nabla \psi) = R_p \end{cases} \tag{1}$$

where q and ε are the charge density and the dielectric permittivity, n and p are the concentrations of free carriers of negative and positive charge, electrons and holes, C is the predefined doping concentration, R_n and R_p the recombination-generation rate for hole and electron. We suppose $R_n = R_p = 0$. We rewrite the model for the two region remembering that Ω_1, Ω_2 are space charge regions, $C_2 = D_2/\Omega_2$ are charge neutral regions $\Gamma_1 = \bar{\Omega}_1 \cap C_1$ e $\Gamma_2 = \bar{\Omega}_2 \cap C_2$ are the free boundaries. In D_1 we have:

$$\begin{cases} \Delta u = \dfrac{q}{\varepsilon}(n - N_d) \\ J_n = D_n \nabla n - n\mu_n \nabla \psi \\ \nabla J_n = 0 \\ J_p = 0 \end{cases} \tag{2}$$

while in D_2

$$\begin{cases} \Delta u = \dfrac{q}{\varepsilon}(N_a - p) \\ J_p = D_p \nabla p + p\mu_p \nabla \psi \\ \nabla J_p = 0 \\ J_n = 0 \end{cases} \tag{3}$$

with the mixed boundary conditions

$$\begin{cases} u = u_1^D & \text{on } \Gamma_1^D \\ \dfrac{\partial u_1}{\partial n} = 0 & \text{on } \Gamma_1^N \end{cases} \qquad \begin{cases} u = u_2^D & \text{on } \Gamma_2^D \\ \dfrac{\partial u_2}{\partial n} = 0 & \text{on } \Gamma_2^N \end{cases} \tag{4}$$

where $\Gamma_i^D, i = 1,2$ are the Dirichlet part of boundary, $\Gamma_i^N, i = 1,2$ the Neumann ones.

At this point we use the quasi Fermi potential

$$n = N_d\, e^{\frac{q}{KT}(u-\phi_n)} \qquad p = N_a\, e^{\frac{q}{KT}(\phi_p - u)}$$

If $\frac{q}{KT} = k$, inserting the last in (2), (3) we obtain

$$\Delta u = \frac{q}{\varepsilon}(n - N_d) = \frac{q}{\varepsilon}(N_d\, e^{k(u-\phi_n)} - N_d) = \frac{q}{\varepsilon} N_d\, (e^{k(u-\phi_n)} - 1) \qquad \text{in } D_1$$

$$\Delta u = \frac{q}{\varepsilon}(N_a - p) = \frac{q}{\varepsilon}(N_a - N_a\, e^{k(\phi_p - u)}) = \frac{q}{\varepsilon} N_a\, (1 - e^{k(u-\phi_n)}) \qquad \text{in } D_2$$

To simplify the model we see that $D_1 = \Omega_1 \cup C_1$ and in the charge neutral region C_1, $n = N_d$ holds , so:

$$\Delta u = 0$$

and using the quasi Fermi potential we obtain

$$n = N_d\, e^{k(u-\phi_n)} = N_d$$

from the relation above follows that

$$e^{k(u-\phi_n)} = 1 \quad \Rightarrow \quad u = \phi_n$$

On the other hand in the space charge region Ω_1, $n = 0$ so

$$\Delta u = -\frac{q}{\varepsilon} N_d = -\xi_1$$

and using the quasi Fermi potential

$$n = N_d\, e^{k(u-\phi_n)} = 0$$

therefore

$$e^{k(u-\phi_n)} = 0 \quad \Rightarrow \quad u < \phi_n$$

In the same way for $D_2 = \Omega_2 \cup C_2$, since $p = N_a$ in C_2, we get

$$\Delta u = 0$$

and

$$p = N_a e^{k(\phi_p - u)} = N_a$$

therefore

$$e^{k(\phi_p - u)} = 1 \quad \Rightarrow \quad u = \phi_p$$

Since in Ω_2 we have $p = 0$, then

$$\Delta u = \frac{q}{\varepsilon} N_a = \xi_2$$

and

$$p = N_d e^{k(\phi_p - u)} = 0$$

As a result we obtain

$$e^{k(\phi_p - u)} = 0 \quad \Rightarrow \quad \phi_p < u.$$

Then we have a free boundary problem with double obstacle, with the free boundaries

$$\Gamma_1 = \overline{\Omega}_1 \cap C_1 \qquad \Gamma_2 = \overline{\Omega}_2 \cap C_2$$

and the obstacle are represented by the quasi Fermi level ϕ_p and ϕ_n.

3. QUASI-VARIATIONAL INEQUALITY FORMULATION

This section is devoted to define the functional spaces and variational problems. If $\mathcal{O}$ is an open bounded set of euclidean plane $\mathbb{R}^2$, we shall denote by $C^0(\overline{\mathcal{O}})$ the set of continuous functions on $\overline{\mathcal{O}}$, $C^k(\overline{\mathcal{O}})(k = 1, 2, \ldots)$ the set of all function defined on $\overline{\mathcal{O}}$ with continuous derivatives until the k order.

We denote by $D(\mathcal{O})$ the space of the functions of $C^\infty(\overline{\mathcal{O}})$, which are zero in a neighbourhood of $\partial\mathcal{O}$, the space $D'(\mathcal{O})$ of the distribtions on $\mathcal{O}$ is the dual of $D(\mathcal{O})$, and we denote by $L^p(O)(1 \leq p \leq +\infty)$ the usual space of the real functions, defined a.e. on $\mathcal{O}$, measurable and p- summable on $\mathcal{O}$

(or a.e. bounded on $\mathcal{O}$ if $p = \infty$); $W^{k,p}(O)(k = 1,2,\ldots;1 \leq p \leq \infty)$ denotes the Banach space:

$$\{f \in L^p(\mathcal{O}); D_x^h D_y^h f \in L^p(\mathcal{O}) \text{ per } h,l \geq 0, h + l \leq k\}$$

We have the following relations

$$(\Delta u + \xi_1)(u - \phi_n) = 0 \qquad \text{in } D_1 \tag{5}$$

because

$$\Delta u + \xi_1 = 0 \quad \text{e} \quad u < \phi_n \quad \text{in } \Omega_1$$
$$\Delta u \geq \xi_1 \quad \text{e} \quad u = \phi_n \quad \text{in } C_1$$

In equal manner in D_2 we will write

$$(\Delta u - \xi_2)(\phi_p - u) = 0 \qquad \text{in } D_2 \tag{6}$$

since

$$\Delta u - \xi_2 = 0 \quad \text{e} \quad u > \phi_p \quad \text{in } \Omega_2$$
$$\Delta u \leq \xi_2 \quad \text{e} \quad u = \phi_p \quad \text{in } C_2$$

Let now consider the following set:

$$U = \{v \in H^1(D), v = g \text{ on } \partial D\},$$

where $D = D_1 \cup D_2 \cup D_3$, e $g : \partial D \to \mathbb{R}$ a function with costant value on ∂D which satisfies the mathematical expression of the reverse biased conditions:

$$g = u_1^D \text{ on } \Gamma_1^D \qquad g = u_2^D \text{ on } \Gamma_2^D \tag{7}$$
$$u_2^D \leq g \leq u_1^D \quad \text{on } (HB) \cup (BG)$$
$$\sup_{\Gamma_1^D} g \leq 0 \leq \inf_{\Gamma_2^D} g$$

The potential u is related to ϕ_n, ϕ_p and the relation between u, ϕ_n and ϕ_p is given by a non linear operator which maps u in $M_1(u)$ and $M_2(u)$. This

operator is defined by logarithmic transformations of the solutions $w_1 = w_1(u)$ and $w_2 = w_2(u)$ of the following mixed boundary value problems in the Slotboom variables:

$$\begin{cases} \nabla \cdot (e^{ku} \nabla w_1) = 0, \\ w_1 = e^{kg_1} \text{ on } \Gamma_1^D, \quad \partial w_1 / \partial n = 0 \text{ on } \Gamma_1^N \end{cases} \tag{8}$$

$$\begin{cases} \nabla \cdot (e^{-ku} \nabla w_2) = 0, \\ w_2 = e^{-kg_2} \text{ on } \Gamma_2^D, \quad \partial w_2 / \partial n = 0 \text{ on } \Gamma_2^N \end{cases} \tag{9}$$

where the values $g_i = g\,|_{\Gamma_i^D}, i = 1,2$ are related with the potential at ohmic contacts; we set

$$V = \{v \in H^1(D), v = e^{-kg} \text{ on } \partial D\}.$$

We may write the quasi Fermi potentials as:

$$\phi_n = -\frac{1}{k} \ln w_1(u) = M_1(u) \qquad \phi_p = \frac{1}{k} \ln w_2(u) = M_2(u)$$

In order to give the classical formulation of the problem, we set $\mathcal{F} = \bigotimes_{i=1}^{3} H^2(D_i) \cup C^1(\overline{D}_i)$, and we have:

Problem 1. *Find* $(u, \varphi_1, \varphi_2)$ *such that* $u = (u_1, u_2, u_3) \in \mathcal{F}$, φ_1 *e* φ_2 *monotone nondecreasing functions (representing* Γ_1 *e* Γ_2*) satisfyng*

$$\Delta u_1 = \xi_1 \quad \text{in } \Omega_1 \qquad \text{where } u < M_1(u) \tag{10}$$

$$\Delta u_1 = 0 \quad \text{in } C_1 \qquad \text{where } u = M_1(u) \tag{11}$$

$$\Delta u_2 = \xi_2 \quad \text{in } \Omega_2 \qquad \text{where } u > M_2(u) \tag{12}$$

$$\Delta u_2 = 0 \quad \text{in } C_2 \qquad \text{where } u = M_2(u) \tag{13}$$

$$\Delta u_3 = 0 \quad \text{in } D_3 \tag{14}$$

with the free interface conditions

$$\frac{\partial u_i}{\partial n} = 0 \quad \text{on } \Gamma_i, \qquad i = 1,2 \tag{15}$$

as well as interface conditions

$$u_1 = u_2 \qquad \frac{\partial u_1}{\partial n} = \frac{\partial u_2}{\partial n} \quad \text{on } (EF), \tag{16}$$

$$u_1 = u_3 \qquad \frac{\partial u_1}{\partial n} = \frac{\partial u_3}{\partial n} \quad \text{on } (FG), \tag{17}$$

$$u_2 = u_3 \qquad \frac{\partial u_2}{\partial n} = \frac{\partial u_3}{\partial n} \quad \text{on } (HF), \tag{18}$$

and boundary conditions

$$u = u_1^D \quad \text{on } \Gamma_1^D \ (OC), \qquad u = u_2^D \quad \text{on } \Gamma_2^D \ (AH). \tag{19}$$

$$\frac{\partial u}{\partial n} = 0 \qquad \text{on } \Gamma_1^N \qquad \frac{\partial u}{\partial n} = 0 \qquad \text{on } \Gamma_2^N \tag{20}$$

Let the convex set

$$K(u) = \{\varphi \in U, \varphi \leq \phi_n = M_1(u) \text{ in } D_1, \phi_p = M_2(u) \leq \varphi \text{ in } D_2\} \tag{21}$$

We have the following:

Theorem 3.1 If u *is a solution of Problem 1 , then* $u \in K(u)$ *must satisfy the quasi-variational inequality*

$$\iint_D \nabla u \, \nabla(\varphi - u)\, dx\, dy + \iint_{D_1} \xi_1(\varphi - u) dx\, dy - \iint_{D_2} \xi_2(\varphi - u) dx\, dy \geq 0, \quad \forall \varphi \in K(u) \tag{22}$$

Proof. Let $\varphi \in K(u)$, being $D = \Omega \cup C$ with $\Omega = \Omega_1 \cup \Omega_2$ e $C = C_1 \cup C_2$, we have

$$\iint_D \Delta u\,(\varphi - u)\,dx\,dy = \iint_\Omega \Delta u\,(\varphi - u)\,dx\,dy + \iint_C \Delta u\,(\varphi - u)\,dx\,dy =$$
$$\iint_{\Omega_1} \Delta u\,(\varphi - u)\,dx\,dy + \iint_{\Omega_2} \Delta u\,(\varphi - u)\,dx\,dy +$$
$$\iint_{C_1} \Delta u\,(\varphi - u)\,dx\,dy + \iint_{C_2} \Delta u\,(\varphi - u)\,dx\,dy$$

but for (11) and (13) we have

$$\iint_D \Delta u\,(\varphi - u)\,dx\,dy = \iint_{\Omega_1} \Delta u\,(\varphi - u)\,dx\,dy + \iint_{\Omega_2} \Delta u\,(\varphi - u)\,dx\,dy \tag{23}$$

moreover from (10) and (12) it follows

$$\iint_D \Delta u\,(\varphi - u)\,dx\,dy = -\iint_{\Omega_1} \xi_1\,(\varphi - u)\,dx\,dy + \iint_{\Omega_2} \xi_2\,(\varphi - u)\,dx\,dy \tag{24}$$

being $\varphi \in K(u)$ will be $\varphi \le \phi_n = M_1(u)$ and $\varphi \ge \phi_p = M_2(u)$ therefore again thanks (11) and (13) $u = M_1(u)$ in C_1 and $u = M_2(u)$ in C_2 then $\varphi \le u$ in C_1 and $\varphi \ge u$ in C_2; this gives

$$\xi_1(\varphi - u) \le 0 \text{ in } C_1 \qquad \xi_2(\varphi - u) \ge 0 \text{ in } C_2$$

therefore in C_1 will be

$$\iint_{D_1} \xi_1\,(\varphi - u)\,dx\,dy = \iint_{C_1} \xi_1\,(\varphi - u)\,dx\,dy + \iint_{\Omega_1} \xi_1\,(\varphi - u)\,dx\,dy \le \iint_{\Omega_2} \xi_1\,(\varphi -$$

in equal manner for C_2

$$\iint_{D_2} \xi_2\,(\varphi - u)\,dx\,dy = \iint_{C_2} \xi_2\,(\varphi - u)\,dx\,dy + \iint_{\Omega_2} \xi_2\,(\varphi - u)\,dx\,dy \ge \iint_{\Omega_2} \xi_2\,(\varphi -$$

From (24) we obtain

$$-\iint_D \Delta u\,(\varphi - u)\,dx\,dy = -\iint_{D_1} \Delta u\,(\varphi - u)\,dx\,dy - \iint_{D_2} \Delta u\,(\varphi - u)\,dx\,dy$$

$$\iint_{\Omega_1} \xi_1\,(\varphi - u)\,dx\,dy - \iint_{\Omega_2} \xi_2\,(\varphi - u)\,dx\,dy \geq$$

$$\iint_{D_1} \xi_1\,(\varphi - u)\,dx\,dy - \iint_{D_2} \xi_2\,(\varphi - u)\,dx\,dy$$

For Green's theorem we have

$$\iint_D \Delta u\,(\varphi - u)\,dx\,dy =$$

$$-\iint_D \nabla u\,\nabla(\varphi - u)\,dx\,dy + \int_{\partial D} (\varphi - u)\frac{\partial u}{\partial n}\,ds = -\iint_D \nabla u\,\nabla(\varphi - u)\,dx\,dy$$

for the boundary conditions because $\varphi \in K(u)$.

Therefore

$$\iint_D \nabla u\,\nabla(\varphi - u)\,dx\,dy = -\iint_D \Delta u\,(\varphi - u)\,dx\,dy \geq$$

$$\iint_{D_1} \xi_1\,(\varphi - u)\,dx\,dy - \iint_{D_2} \xi_2\,(\varphi - u)\,dx\,dy$$

then is satisfied the quasi-variational inequality

$$\iint_D \nabla u\,\nabla(\varphi - u)\,dx\,dy + \iint_{D_1} \xi_1(\varphi - u)dx\,dy - \iint_{D_2} \xi_2(\varphi - u)dx\,dy \geq 0,$$

$\forall \varphi \in K(u)$ $\qquad \Diamond$

Let now

$$a(u, v) = \iint_D \nabla u\,\nabla v\,dx\,dy \qquad u, v \in U$$

we can rewrite the problem as

Problem 2.

$$\begin{cases} \text{Find } u \in K(u) \text{ such that} \\ a(u, \varphi - u) \geq (\zeta, \varphi - u), \quad \forall \varphi \in K(u) \\ \text{with } \zeta = -\xi_1 \text{ in } D_1 \quad \zeta = \xi_2 \text{ in } D_2 \quad \zeta = 0 \text{ in } D_3 \end{cases} \tag{25}$$

We can say that the (25) in general is not a variational inequality; it is a variational inequality only when $\forall \varphi \in U, K(\varphi) = K,$, with K being a non-empty closed convex set of $H^1(D)$. In fact it is a new type of entity, we will call it, according with Bensoussan-Goursat-Lions [3], a *quasi-variational inequality*. To the quasi-variational inequality (25) we can associate in a natural way a family of variational inequalities: for z fixed in U we will call *variational section* of the quasi-variational inequality (25) along z, the variational inequality

$$a(w, \varphi - w) \geq (\zeta, \varphi - w), \quad \forall \varphi \in K(z) \tag{26}$$

under the hypothesis (which is standard in the variational case, and which we will make here too) of the coerciveness of the form a

$$a(v, v) \geq \gamma_0 \parallel v \parallel^2_{1,D} \quad \gamma > 0 \tag{27}$$

$$\mid a(u, v) \mid \leq \gamma_1 \parallel u \parallel_{1,D} \parallel v \parallel_{1,D} \qquad u, v \in U \tag{28}$$

we can say that (26) has one and only one solution.

Therefore if $z \in U$, the application $S : U \to U$, such that $u_z = S(z)$ is a solution of (26),

$$u_z \in K(z) : a(u_z, \varphi - u_z) \geq (\zeta, \varphi - u_z), \quad \forall \varphi \in K(z)$$

We will call this application the *variational selection* associated with the quasi-variational inequality (25); under the hypothesis (27), (28), this selection is well defined.

It follows immediately that a solution of (25) is a fixed point for S. Therefore the basic idea to solve the **Problem 2** is to consider the variational selection of (25) and to find its fixed points; an inportant question is what type of fixed point theorem we can use. We do not expect a Lipschitz continuos or a monotonic situation, and thus the classical theorems are useless, more usefull is Schauder's theorem or the results of Joly and Mosco [22].

4. NUMERICAL APPROXIMATIONS

We have seen like some free boundary problem very complex in their structure can be solved through opportune modifications tied to the physical characteristics of the problem by means of variational and quasi-variational inequalities. From a numerical point of view the quasi-variational inequalities can be solved with the Bensoussan-Lions iterative scheme, which is a sequence of iterative variational inequalities, for a fixed obstacle. Quasi-variational inequalities and their applications in different areas have been investigated since the early eighties notably by Bensoussan, Lions, Mosco and Baiocchi. However, very little was known about the numerical methods for such problems till recently [10]. We show a technique for the approximation of quasi-variational inequalities.

To determinate the depletion region in a $p-n$ junction we have to solve the following model

$$\begin{cases} \text{Find } u \in K(u) \text{ such that} \\ a(u, \varphi - u) \geq (\zeta, \varphi - u), \quad \forall \varphi \in K(u) \\ \text{with } \zeta = -\xi_1 \text{ in } D_1, \quad \zeta = \xi_2 \text{ in } D_2, \zeta = 0 \text{ in } D_3 \end{cases} \tag{29}$$

with

$$K(u) = \{\varphi \in U, \varphi \leq \phi_n = M_1(u) \text{ in } D_1, \phi_p = M_2(u) \leq \varphi \text{ in } D_2\}$$

Where the obstacles $M_1(u)$ e $M_2(u)$ are defined resolving the two mixed boundary value problems

$$\begin{cases} \nabla \cdot (e^{ku} \nabla w_1) = 0, \\ w_1 = e^{-kg_1} \text{ on } \Gamma_1^D, \quad \partial w_1 / \partial n = 0 \text{ on } \Gamma_1^N \end{cases} \tag{30}$$

$$\begin{cases} \nabla \cdot (e^{-ku} \nabla w_2) = 0, \\ w_2 = e^{kg_2} \text{ on } \Gamma_2^D, \quad \partial w_2 / \partial n = 0 \text{ on } \Gamma_2^N \end{cases} \tag{31}$$

by a maximum principle we obtain $w_1, w_2 > 0$, thus we can compute the obstacles as follow

$$M_1(u) = -\frac{1}{k} \ln w_1(u) \qquad M_2(u) = \frac{1}{k} \ln w_2(u)$$

Consider a regular triangulation $\mathcal{T}_h$, established over the open polygonal $D \subset \mathbb{R}^2$ such that

$$D = \bigcup_{T \in \mathcal{T}_h} T$$

Let T a triangle in $\mathcal{T}_h$, and $\mathbb{P}_1(T)$ the space of all polynomials of degree ≥ 1 restricted to the set T. We associate with $\mathcal{T}_h$ the usual finite element spaces:

$$X_h = \{v_h \in C^0(\bar{D}), v_h|_T \in \mathbb{P}_1(T), \forall T \in \mathcal{T}_h\}, \qquad V_{0h} = \{v_h \in X_h, v_h = 0 \text{ on } \partial D\},$$
$$U_h = \{u_h \in X_h, u_h = g_h \text{ on } \partial D\}, \qquad V_h = \{v_h \in X_h, v_h = e^{kg_h} \text{ on } \partial D\}.$$

Then we define the obstacles as

$$M_{1h} : u_h \in U_h \longrightarrow M_{1h}(u_h) = r_h(-\frac{1}{k} \ln w_{1h})$$

$$M_{2h} : u_h \in U_h \longrightarrow M_{2h}(u_h) = r_h(\frac{1}{k} \ln w_{2h})$$

with $w_{1h}, w_{2h} \in V_h$ which satisfy

$$\nabla \cdot (e^{ku_h} \nabla w_{1h}) = 0, \quad \forall u_h \in V_h$$

$$\nabla \cdot (e^{-ku_h} \nabla w_{2h}) = 0, \quad \forall u_h \in V_h$$

We introduce the convex set

$$K_h(u_h) = \{\varphi_h \in U_h, \varphi_h \leq M_{1h}(u), M_{2h}(u) \leq \varphi_h\}.$$

We have the following finite element formulation of the problem

$$\begin{cases} \text{Find } u_h \in K_h(u_h) \text{ such that} \\ a(u_h, \varphi_h - u_h) \geq (\zeta, \varphi_h - u_h), \quad \forall \varphi_h \in K_h(u_h) \\ \zeta \in L^\infty(D) \end{cases} \tag{32}$$

To update the obstacles the continuity equations can be solved and then we have a system of quasi-variational inequality and we use a Bensoussan-Lions iterative scheme to solve the problem. We shall recall some result related to elliptic variational inequalities that are necessary to prove some useful qualititive properties.

5. ASSUMPTIONS AND NOTATIONS

In this section we are concerned with the standard finite element approximation of the system of quasi-variational inequalities (QVIs): Find a vector $U = (u^1, ..., u^M)$ satisfying

$$\begin{cases} a^i(u^i, v - u^i) \geq (f^i, v - u^i) \quad \forall v \in H^1(\Omega) \\ u^i \leq \psi u^i; \quad u^i \geq 0; \quad v \leq \psi u^i \end{cases} \tag{33}$$

where Ω is a bounded smooth domain of $\mathbb{R}^N$ with boundary $\partial\Omega$, $a^i(u, v)$ are bilinear forms defined on $H^1(\Omega) \times H^1(\Omega)$, $(\cdot, \cdot)$ is the inner product in $L^2(\Omega)$ and f^i are ψ regular functions. For sake of simplicity we will treat the case of one obstacle, considering the two obstacle problem a generalization in which we replace the constraint set of (21) with the following: $K = \{v \in H^1(\Omega) \text{ such that } v \leq \psi\}$

We are given functions

$$a^i_{jk}(x), a^i_k(x), a^i_0(x) \in C^2(\overline{\Omega}), \quad x \in \overline{\Omega}, \quad 1 \leq k, \quad j \leq N, \quad 1 \leq i \leq M,$$

sufficiently smooth such that:

$$\sum_{1 \leq j,k \leq N} a^i_{jk}(x)\xi_j\xi_k \geq \alpha \parallel \xi \parallel^2, \qquad \xi \in \mathbb{R}^N; \ \alpha > 0 \tag{34}$$

$$a_{jk} = a_{kj}, \quad a^i_0(x) \geq c_0 > 0; \tag{35}$$

We define the second-order , uniformly elliptic operator of the form

$$\mathcal{A}^i = \sum_{1 \leq j,k \leq N} a^i_{jk}(x) \frac{\partial^2}{\partial x_j \partial x_k} \sum_{k=1}^{N} b^i_k(x) \frac{\partial}{\partial x_k} + a^i_0(x) \tag{36}$$

and the bilinear forms associated with $\mathcal{A}^i$: for any $u, v \in H^1(\Omega)$

$$a^i(u,v) = \int_\Omega \left(\sum_{1\le j,k\le N} a^i_{jk}(x) \frac{\partial u}{\partial x_j} \frac{\partial v}{\partial x_k} + \sum_{k=1}^{N} a^i_k(x) \frac{\partial u}{\partial x_k} v + a^i_0(x) uv \right) dx \tag{37}$$

that we assume to be coercive, i.e., there exist $\gamma > 0$ such that

$$a^i(v,v) \ge \gamma \| v \|^2_{H^1(\Omega)}; \quad \forall v \in H^1(\Omega). \tag{38}$$

The right hand sides $f^1,\ldots,f^M$ are also given such that

$$f^i \in L^\infty(\Omega); \ f^i \ge 0 \tag{39}$$

We shall also need the following norm:

$$\forall W = (w^1,\ldots,w^M) \in \prod_{i=1}^{M} L^\infty(\Omega), \tag{40}$$

$$\| W \|_\infty = \max_{1\le i\le M} \| w^i \|_{L^\infty}, \tag{41}$$

where $\| \cdot \|_{L^\infty}$ denotes the classic L^∞ norm.

5.1 Elliptic Variational inequalities

Let f be a function in L^∞ and ψ an obstacle in $W^{2,\infty}$ such that $\psi \ge 0$ on $\partial\Omega$. Let also $\mathcal{A}$ be an elliptic operator and $a(\cdot,\cdot)$ its associated coercive bilinear form of the same forms as those defined in (36) and (37), respectively. We consider the following elliptic variational inequality (VI): Find $u \in K$ such that

$$a(u, v-u) \ge (f, v-u) \quad \forall v \in K \tag{42}$$

where $K = \{v \in H1(\Omega)$ such that $v \le \psi$ a.e.$\}$ Thanks to [23,5], the VI (42) has one and only one solution. Moreover, $u \in W^{2,p}, 1 \le p \le \infty$ and satisfies

$$\| u \|_{W^{2,p}} \le C(\| f \|_\infty + \| \mathcal{A}\psi \|_\infty) \tag{43}$$

Definition 1. *$z \in K$ is said to be a subsolution for VI (42) if*

$$a(z,v) \leq (f,v) \quad \forall v \in K, \qquad v \geq 0 \tag{44}$$

Let X denote the set of such subsolutions, then (see [5]) the solution of VI (42) is the maximum element of X.

Consider now the following mapping:

$$\sigma : L^{\infty}(\Omega) \longrightarrow L^{\infty}(\Omega)$$
$$\psi \longrightarrow \sigma(\psi) = u$$

where u is the solution to VI (42). The mapping σ is increasing, concave, and Lipschitz continuos with respect to ψ [7].

Existence of a unique solution to system (33) can be proved, adapting the approach developed in [4].

Indeed, let $H^{+} = (L^{\infty}_{+}(\Omega))^{M} = \{V = (v^1,\ldots,v^M) \text{ such that } v^i \in L^{\infty}_{+}(\Omega)\}$, equipped with the norm: $\| V \|_{\infty} = \max\limits_{1\leq i\leq M} \| v^i \|_{L^{\infty}(\Omega)}$ where $L^{\infty}_{+}(\Omega)$ is the positive cone of $L^{\infty}(\Omega)$. We consider the mapping

$$T : H^{+} \longrightarrow H^{+}$$
$$W \longrightarrow TW = \zeta = (\zeta^1,\ldots,\zeta^M)$$

where $\zeta^i = \sigma(\psi w^i) \in H^1(\Omega)$ is solution to the following VI:

$$\begin{cases} a^i(\zeta^i, v-\zeta^i) \geq (f^i, v-\zeta^i) \ \forall v \in H^1(\Omega) \\ \zeta^i \leq \psi w^i \quad ; \quad v \leq \psi w^i \end{cases} \tag{45}$$

Problem (45) being a coercive VI, thanks to [23], [5] has one and only one solution.

Consider now $\overline{U}^0 = (\overline{u}^{1,0},\ldots,\overline{u}^{M,0})$, where $\overline{u}^{i,0}$ is the solution to the following variational equation:

$$a^i(\overline{u}^{i,0}, v) = (f^i, v) \forall v \in H^1(\Omega) \tag{46}$$

Due to (39), problem (46) has a unique solution. Moreover, $\overline{u}^{i,0} \in W^{2,p}(\Omega)$; $2 \leq p < \infty$

Proposition 5.1 *Let* $\mathbb{C} = \{W \in H^+ \text{ such that } 0 \leq W \leq \overline{U}^0\}$, *then* T *maps* $\mathbb{C}$ *into itself. Moreover is* T *increasing, concave and Lipschitz continuous on* H^+.

We notice that the solutions $U = (u^1, \ldots, u^M)$ of system (33) correspond to fixed points of mapping T, that is $U = TU$. In this view it is natural to consider the following iterative scheme.

5.2 A Continuous Iterative Scheme of Bensoussan-Lions Type

An iterative scheme for the solution of system of QVIs is given as follows.

Starting from $\overline{U}^0$ defined in (46) (resp. $\underline{U}^0 = (0, \ldots, 0)$), we define the sequences

$$\overline{U}^{n+1} = T\overline{U}^n \; ; \; n = 0, 1, \ldots \tag{47}$$

respectively

$$\underline{U}^{n+1} = T\underline{U}^n; n = 0, 1, \ldots. \tag{48}$$

Making use of properties of mapping T we have the following convergence result.

Theorem 5.2 *The sequences* $(\overline{U}^n)$ *and* $(\underline{U}^n)$ *are monotone and well defined in* $\mathbb{C}$. *Moreover, they converge respectively from above and below to the unique solution of system (33), (cf. [4] p.453).*

The following estimations provide a rate of convergence for sequences.

Lemma 5.3 *There exist a constant* $\mathbb{C}$ *independent of* n *such that for any* $i = 1, 2, \ldots, M$, *[15]*

$$\max_{n \geq 0} \left(\| \overline{u}^{i,n} \|_{W^{2,p}(\Omega)}, \| \underline{u}^{i,n} \|_{W^{2,p}(\Omega)} \right) \leq C \; ; \; 2 \leq p < \infty$$

Theorem 5.4 *Assume* $a^i_{jk}(x)$ *in* $C^{1,\alpha}(\overline{\Omega})$, $a^i(x)$, $a^i_0(x)$ *and* f^i *in* $C^{0,\alpha}(\Omega)$. *Then* $(u^1, \ldots, u^M) \in (W^{2,p}(\Omega))^M$; $2 \leq p < \infty$.

Proposition 5.5 *There exist a positive constant* $0 \leq \mu \leq 1$ *such that*

$$\| \overline{U}^n - U \|_\infty \leq \mu^n \| \overline{U}^0 \|_\infty \tag{49}$$

$$\| \underline{U}^n - U \|_\infty \leq \mu^n \| \overline{U}^0 \|_\infty \tag{50}$$

5.3 The Discrete Problem

Let Ω be decomposed into triangles and let $\mathcal{T}_h$ denote the set of all those elements; $h > 0$ is the mesh size. We assume the family $\mathcal{T}_h$ is regular and quasi-uniform.

Let V_h denote the standard finite element space, $A^i, 1 \leq i \leq M$ be the matrices with generic coefficients $a^i(\varphi_l, \varphi_s)$, where $\varphi_s, s = 1,2,...m(h)$ are the nodal basis functions. Let also r_h be the usual interpolation operator.

In the sequel of the paper, we shall use the discrete maximum assumption (d.m.p.). Under the d.m.p., we shall achieve a similar study to that devoted to the continuos problem, therefore the qualitative properties and results stated in the continuous case are conserved in the discrete case.

The discrete system of QVIs is then defined as follows: Find $U_h = (u_h^1,..., u_h^M) \in (V_h)^M$ such that

$$\begin{cases} a^i(u_h^i, v - u_h^i) \geq (f^i, v - u_h^i) \quad \forall v \in V_h \\ \qquad u_h^i \leq r_h \psi u_h^i \ ; u_h^i \geq 0 \ ; \ v \leq r_h \psi u_h^i \end{cases} \tag{51}$$

Existence and uniqueness of a solution of system (51) can be shown similarly to that of the continuous case provided the discrete maximum principle is satisfied. Indeed, the idea for proving that consists of associating with the system (51) the following discrete fixed point mapping:

$$\begin{aligned} T_h : H^+ &\longrightarrow (V_h)^M \\ W &\longrightarrow T_h W = \zeta_h = (\zeta_h^1,..,\zeta_h^M) \end{aligned}$$

where $\zeta_h^i = \sigma_h(\psi w^i)$ is the solution of the following discrete VI:

$$\begin{cases} a^i(\zeta_h^i, v - \zeta_h^i) \geq (f^i, v - \zeta_h^i) \quad \forall v \in V_h \\ \qquad \zeta_h^i \leq r_h \psi w^i \ , \ v \leq r_h \psi w^i \end{cases} \tag{52}$$

Under the d.m.p the mapping T_h possesses analogous properties to that of mapping T.

Let $\overline{U}_h^0 = (\overline{u}_h^{1,0},..., \overline{u}_h^{M,0})$ be the discrete analogue to the solution of problem (46) :

$$a^i(\overline{u}_h^{i,0}, v) = (f^i, v) \forall v \in V_h \quad 1 \le i \le M \tag{53}$$

Proposition 5.6 *T_h maps $\mathbb{C}_h$ into itself, where $\mathbb{C}_h = \{W \in (L^\infty(\Omega))^M$ such that $0 \le W \le \overline{U}_h^0\}$, moreover T_h is increasing, concave and Lipschitz continuous on H^+.*

It is not hard to see that the solution of system of QVIs (51) is a fixed point of T_h, that is $U_h = T_h U_h$. Therefore, as in the continuous problem, one can define the following discrete iterative scheme.

Starting from $\overline{U}_h^0$ solution of (53) (resp. from $\underline{U}_h^0 = (0,\ldots,0)$), one can compute

$$\overline{U}_h^{n+1} = T_h\, \overline{U}_h^n \quad n = 0,1,\ldots \tag{54}$$

(resp.)

$$\underline{U}_h^{n+1} = T_h \underline{U}_h^n \quad n = 0,1,\ldots \tag{55}$$

Theorem 5.7 *Under the d.m.p. the sequences $(\overline{U}_h^n)$ and $(\underline{U}_h^n)$ are monotone and well defined in $\mathbb{C}_h$. Moreover, they converge respectively from above and below to the unique solution of system (51)*

Using the above result, we are able to establish the geometric convergence of sequence $(\overline{U}_h^n)$ and $(\underline{U}_h^n)$.

Proposition 5.8 *There exist a positive constant $0 \le \mu \le 1$ such that*

$$\| \overline{U}_h^n - U_h \|_\infty \le \mu^n \| \overline{U}_h^0 \|_\infty \tag{56}$$

$$\| \underline{U}_h^n - U_h \|_\infty \le \mu^n \| \overline{U}_h^0 \|_\infty \tag{57}$$

5.4 The Finite Element Error Analysis

We recall some known L^∞-error estimates result and introduce an auxiliary problem. From now on C will denote a constant independent of both h and n.

Theorem 5.9 *Let $\overline{u}^{i,0}$ (respectively, $\overline{u}_h^{i,0}$), be the solution of problem (46), (respectively (53)). Then (see [11,19]*

$$\| \overline{u}^{i,0} - \overline{u}_h^{i,0} \|_{L^\infty(\Omega)} \leq Ch^2 \mid log\ h \mid^{3/2} \quad \forall i = 1, 2, \ldots, M \tag{58}$$

Theorem 5.10 *Let the d.m.p. and regularity result (43) hold. Then (see [14])*

$$\| u - u_h \|_{L^\infty(\Omega)} \leq Ch^2 \mid log\ h \mid^2 \tag{59}$$

We introduce the following discrete sequence

$$\begin{cases} \widetilde{U}_h^{n+1} = T_h \overline{U}^n; \quad n = 0, 1, \ldots \\ \text{with } \widetilde{U}_h^0 = \overline{U}_h^0 \end{cases} \tag{60}$$

where $\overline{U}_h^0$ is defined in (53) and for any $n \geq 1$, $\tilde{u}_h^{i,n}$ is a solution to following discrete variational inequality:

$$\begin{cases} a^i(\tilde{u}_h^{i,n+1}{}_h, v - \tilde{u}_h^{i,n+1}) \geq (f^i, v - \tilde{u}_h^{i,n+1}) \quad \forall v \in V_h \\ \tilde{u}_h^{i,n+1} \leq r_h \psi \overline{u}^{i,n}\ ,\ v \leq r_h \psi \overline{u}^{i,n} \end{cases} \tag{61}$$

$\overline{U}^n = (\overline{u}^{1,n}, \ldots, \overline{u}^{M,n})$ being the sequence defined by (48). Again, thanks to [5], (61) has one and only one solution.

We notice that $\tilde{u}_h^{i,n}$ solution of (61) represents the standard finite element approximation of $\overline{u}^{i,n}$. Therefore, using the regularity result provided by Lemma 4.2 and next adapting [12], we have the following uniform error estimate.

Proposition 5.11

$$\| \overline{U}^n - \widetilde{U}_h^n \|_\infty \leq Ch^2 \mid log\ h \mid^2 \tag{62}$$

with the use of the result seen above we introduce the following :
Lemma 5.12

$$\| \overline{U}^n - \overline{U}_h^n \|_\infty \leq \sum_{p=0}^{n} \| \overline{U}^p - \widetilde{U}_h^p \|_\infty \tag{63}$$

Now guided by Propositions 5, 8, 11, Lemma 12 Theorem 9 we are in a position to demonstrate the main result.

Theorem 5.13

$$\| U - U_h \|_\infty \leq Ch^2 \mid log\ h \mid^3 \tag{64}$$

$$\| U - U_h \|_{1,\infty} \leq Ch \mid log\ h \mid^3 \tag{65}$$

where: $\| U \|_{1,\infty} = \max_{1 \leq i \leq M} \| u^i \|_{W^{1,\infty}(\Omega)}$

Proof. Using estimations (49), (56) we have:

$$\| U - U_h \|_\infty \leq \| U - \bar{U}^n \|_\infty + \| \bar{U}^n - \bar{U}_h^n \| + \| \bar{U}_h^n - U_h \|_\infty$$

$$\leq \| U - \bar{U}^n \|_\infty + \sum_{p=0}^{n} \| \bar{U}^p - \bar{U}_h^p \|_\infty + \| \bar{U}_h^n - U_h \|_\infty \leq$$

$$\| U - \bar{U}^n \|_\infty + \| \bar{U}^0 - \bar{U}_h^0 \|_\infty + \sum_{p=1}^{n} \| \bar{U}^p - \bar{U}_h^p \|_\infty + \| \bar{U}_h^n - U_h \|_\infty$$

$$\leq \mu^n \| \bar{U}^0 \|_\infty + \mu^n \| \bar{U}_h^0 \|_\infty + Ch^2 \mid log\ h \mid^{3/2} + nCh^2 \mid log\ h \mid^2$$

Finally, letting $\mu^n = h^2$ we get the desired result.

The $W^{1,\infty}$-error estimate (65) follows immediately from the standard inverse inequality (cf. [11]). It is important to notice that the error estimate obtained contains an extra power in $(log\ h)$ than expected, due to the approach followed.

6. RESULTS AND CONCLUSIONS

The variational method presented is an alternative approach to the classical drift-diffusion model which can be described by a nonlinear Poisson equation for the electrostatic potential coupled with a system of convection-diffusion equations for the transport of charge

$$\begin{cases} \nabla^2 \psi = \dfrac{q}{\varepsilon}(n - p - C) \\ \nabla \cdot (-n\mu_n \nabla \psi + D_n \nabla n) = R(\psi, n, p) \\ \nabla \cdot (p\mu_p \nabla \psi + D_p \nabla p) = R(\psi, n, p) \end{cases}$$

In the context of semiconductor device modelling, the presence of strong variation of the convection term $\nabla \psi$ is a source of numerical troubles since it give rise to sharp internal layers.

This equations can be solved with Gummel like process to decouple the system and Newton's method to obtain the resulting sequences of linear systems.

The Poisson problem leads to a symmetric, positive definite system which can be solved iteratively using BCG.

The transport equation leads to nonsymmetric indefinite systems; moreover their solutions exibit steep layers and are subject to numerical oscillation and instabilities if standard Galerkin-type discretization strategies are used.

We present numerical result for Variational Method and Drift Diffusion model for a two dimensional p-n junction with the following parameters: $\xi_1 = -\xi_2 = 4, u^D = |V_a|$ where V_a is the applied potential with value $-5V, -4V, -2V$.

Variational Method				Drift-Diffusion			
$h = 1/6 = 0.16$							
Appl.Pot.(V)	-5	-4	-2		-5	-4	-2
Iter. num.	253	175	120		296	225	167
Exec. time (sec)	25	18	13		42	31	23
Depletion layer size (μm)	1.19	1.01	0.78		1.18	1.02	0.79

Table 1. Numerical results with h = 1/6

Variational Method				Drift-Diffusion			
$h = 1/12 = 0.083$							
Appl.Pot.(V)	-5	-4	-2		-5	-4	-2
Iter. num.	615	524	392		712	638	453
Exec. time(sec)	224	192	133		370	321	235
Depletion layer size(μm)	1.22	1.00	0.76		1.18	1.01	0.80

Table 2. Numerical results with h = 1/2

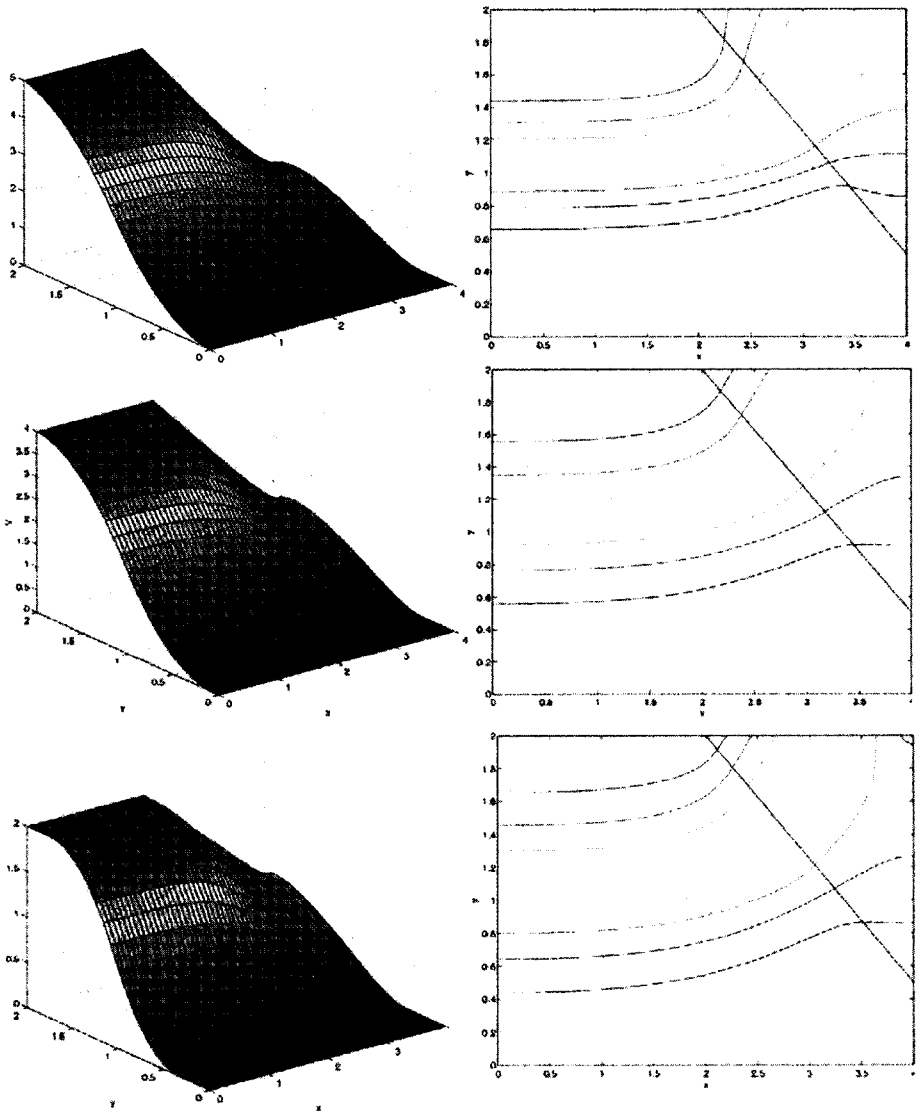


Figure 2. Varational Method. Numerical solution and depletion layer $V_a = -5V, -4V, -2V$

REFERENCES

[1] C. Baiocchi, V. Comincioli, E. Magenes, G. Pozzi *Free Boundary problems in the theory of fluid flow through porous media: existence and uniquess theorems,* Annali Mat. Pura e appl., (4) 97 (1973), pp 1-82.

[2] C. Baiocchi, *Estimation d'erreur dans L^∞ pour les inequations a obstacle,* In: I.Galligani, E. Magenes (eds.) Mathemtical Aspects of Finite Element M in Mathematics 606, 27-34 (1977).

[3] A. Bensoussan, Goursat M., J.-L. Lions, *Controle impulsionnel et inequations quasi-variationnelles stationnaires,* C.R. Acad. SC. Paris, Ser. A, Vol 276, (1973) pp 1279-1284.

[4] A. Bensoussan, J.-L. Lions, *Impulse control and quasi-variational inequalities*, Gauthier Villars, Paris (1984).

[5] A. Bensoussan, J.-L. Lions, *Applications des inequations variationnelles en controle stochastique.* Dunod , Paris, (1978).

[6] M. Boulbrachene, *The noncoercive quasi-variational inequalities related to impulse control problems,* computers Math, Applic, 101-108 (1998).

[7] M. Boulbrachene, M. Haiour *The Finite Element Approximation of Hamilton-Jacobi-Bellman Equations,* Comput. Math. Appl., 41 (2001) pp. 993-1007.

[8] F. Brezzi, L.A. Caffarelli, *Convergence of the discrete free boundaryfor finite element approximations*, R.A.I.R.O Anal. Numer. 17, 385-395 (1983).

[9] J.C. Jr. Bruch, C.A. Papadopoulos, J.M. Sloss, *Parallel computing used in solving wet chemical etching semiconductor fabrication problems,* Nonlinear mathematical problems in industry, I (Iwaki, 1992), pp. 281-292.

[10] J.C. Jr. Bruch, *A Survey of Free Boundary Value Problems in the Theory of Fluid Flow Through Porous Media: Variational Inequality Approach,* Advances in Water Resources, Part I, Vol.3- Part II Vol.3 (1980).

[11] P.G. Ciarlet, J.-L. Lions, *Handbook of Numerical Analysis* (Volume 2), Elsevier Science Publishers B.V. (North Holland), (1991).

[12] P. Cortey-Dumont, *On the finite element approximation in the L^∞ norm of variational inequalities with nonlinear operators*, Numer.Num., 47, 45-57 (1985).

[13] P.G. Ciarlet, P.A. Raviart, *Maximum principle and uniform convergence for the finite element method*, Comp. Meth. in Appl.Mech and Eng. 2, 1-20 (1973).

[14] P. Cortey-Dumont, *Approximation numerique d'une inequation qua quasi-variationnelles liees a des problemes de gestion de stock.* R.A.I.R.O Anal.Num. (1980).

[15] J. Hannouzet , P. Joly, *Convergence uniforme des iteres definissant la solution d'une inequation quasi-variationnelle,* C.R.Acad. Sci., Paris, Serie A, 286 (1978).

[16] C. Hunt and N.R. Nassif, *On a Variational inequality and its approximation, in the theory of semiconductors,* SIAM J. Numer. Anal. 12 (1975), pp 938-950.

[17] M. Morandi Cecchi, M.R. Russo, *The Error Analysis in a Free Boundary Problem in Semiconductors,* Proceedings of the Fifth World Congress on Computational Mechanics (WCCM V), July 7-12, 2002, Vienna, Austria.

[18] A.Nachaoui, N.R. Nassif, *Sufficient conditions for converging drift-diffusion discrete systems. Application to the finite element method* Math. Methods Appl. Sci. 19 (1996), n. 1, pp 33-51.

[19] J. Nitsche *L^∞ -convergence of finite element approximations, Mathematical aspects of finite element methods,* Lect. Notes Math, 606, 261-274 (1977).

[20] R.H. Nochetto *A note on the approximation of free boundaries by finite element methods* R.A.I.R.O Anal. Numer. 20, 355-368 (1986).

[21] R. H. Nochetto *Sharp L^∞ -Error estimates for semilinear elliptic problems with free boundaries*, Numer. Math. 54, 243-255 (1988).

[22] J.L. Joly,U. Mosco, *Sur les inequations quasi-variationnelles,* C.R. Acad. SC. Paris, 279 (1974) pp 499-502.

[23] D. Kinderlehrer, G. Stampacchia, *An introduction to Variational Inequalities and their applications*. Academic Press (1980).

[24] J.F. Rodrigues, *On a quasi-variational inequality arising in semiconductor theory,* Rev. Mat. Univ. Complut. Madrid, Vol.5 N.1 (1992), pp 137-51.

[25] M.R. Russo, *Un problema di frontiera libera nel campo dei semiconduttori,* Ph.D. Thesis, (2001).

[26] C. Schmeiser, *On strongly reverse biased semiconductor diodes,* SIAM J. Appl. Math., Vol 49 (1990), pp 1734-48.

[27] C. Schmeiser, *A singular perturbation analysis of reverse biased pn -junction,* SIAM J. Math. Anal., Vol 21 (1990), pp 313-26.

SENSITIVITY ANALYSIS FOR VARIATIONAL SYSTEMS

B.S. Mordukhovich
Dept. of Mathematics, Wayne State University, Detroit, Michigan USA

Abstract: The paper mostly concerns applications of the generalized differentiation theory in variational analysis to Lipschitzian stability and metric regularity of variational systems in infinite-dimensional spaces. The main tools of our analysis involve coderivatives of set-valued mappings that turn out to be proper extensions of the adjoint derivative operator to nonsmooth and set-valued mappings. The involved coderivatives allow us to give complete dual characterizations of certain fundamental properties in variational analysis and optimization related to Lipschitzian stability and metric regularity. Based on these characterizations and extended coderivative calculus, we obtain efficient conditions for Lipschitzian stability of variational systems governed by parametric generalized equations and their specifications.

Key words: Variational systems, Lipschitzian stability, variational analysis, generalized differentiation.

Mathematics Subject Classifications (2000): 49J52, 49K27, 90C48.

1. INTRODUCTION

This paper presents new results on sensitivity and stability analysis for parametric variational systems based on the application of generalized variational tools of variational analysis. Variational analysis has been recognized as a fruitful area of mathematics, which is mostly oriented on applications to optimization-related problems and also provides powerful

tools for the analysis of a broad spectrum of problems that may not be of a variational nature; see the book by Rockafellar and Wets [35] for a systematic exposition and thorough developments of the key feature of variational analysis in finite dimensions.

In this paper we concern the analysis of robust Lipschitzian stability for parametric variational systems described by *perturbed generalized equations*

$$0 \in f(x,y) + Q(x,y) \tag{1.1}$$

in the sense of Robinson [32], where $f : X \times Y \to Z$ is a single-valued mapping while $Q : X \times Y \rightrightarrows Z$ is a set-valued mapping between Banach spaces. For convenience we use the terms *base* and *field* referring to the single-valued and set-valued parts of (1.1), respectively, with the decision variable y and the parameter x. It has been well recognized that (1.1) provides an appropriate model for sensitivity analysis in a broad framework of constrained optimization and equilibria. In particular, generalized equations (1.1) cover classical *variational inequalities*

$$\text{find } y \in \Omega \text{ with } \langle f(x,y), v - y \rangle \geq 0 \text{ for all } v \in \Omega$$

and hence *complementarity problems* corresponding to the normal cone field $Q(y) = N(y;\Omega)$ in (1.1). Note that, in contrast to the standard framework, model (1.1) includes the case when the field Q may *depend on the perturbation parameter* x. The latter model is particularly convenient for describing stationary point maps and stationary point-multiplier maps in optimization problems with parameter-dependent constraints; see, e.g., [12].

By *robust Lipschitzian stability* we understand Lipschitzian behavior of the *solution map*

$$S(x) := \{y \in Y \mid 0 \in f(x,y) + Q(x,y)\} \tag{1.2}$$

to (1.1) around a reference point, which is *stable* with respect to perturbations of the initial data. The classical (Hausdorff) local Lipschitzian property of set-valued mappings is a good example of such behavior, but it is restricted to mappings with compact values. An adequate and non-restrictive property of this type was introduced by Aubin [1] under the name of "pseudo-Lipschitz" property, which concerns robust Lipschitzian behavior of a set-valued mapping around a given point $(\overline{x}, \overline{y})$ of its graph. In our opinion, it would be better to use the terms of *Aubin property* suggested in [7] and/or *Lipschitz-like* property emphasizing its Lipschitzian nature (while "pseudo" means "false"; see the discussion in [35]). Aubin's Lipschitz-like property is probably the most proper extension of the classical Lipschitz

continuity to set-valued mappings. On the other hand, for any $F: X \rightrightarrows Y$ it is equivalent to *metric regularity* and *linear openness* of the inverse $F^{-1}: Y \rightrightarrows X$.

The main tools of our analysis involve *coderivatives* of set-valued mappings that extend the classical concept of *adjoint derivative* operator to nonsmooth and multivalued frameworks, enjoy a comprehensive calculus, and play a crucial role in characterizations of Lipschitzian behavior, metric regularity, and covering/openness properties of general multifunctions; see [20] and the references therein. Applications of coderivative analysis to various problems related to Lipschitzian stability of variational systems in finite dimensions are given in [7,9,12,13,15,18,19,29], and other publications. The recent paper [22] contains coderivative-type results for robust Lipschitzian stability of solution maps (1.2) in infinite-dimensional spaces.

In this paper we develop another approach to Lipschitzian stability of parametric variational systems. In contrast to the one in [22], which is based on computing/estimating coderivatives of solution maps (1.2) and then on using coderivative criteria for the Lipschitz-like property, the approach of this paper involves a preliminary *first-order approximation* of the original variational system in the spirit of Robinson [33,34]; see also [6] and [5] for more recent developments. Then applying coderivative criteria for the approximation system, we derive *characterizations* as well as workable *sufficient conditions* for Lipschitzian stability of the original variational system. The latter approach is more efficient for the class of *canonically perturbed* variational systems given in the form

$$\Sigma(x,q) := \{y \in Y \mid q \in f(x,y) + Q(x,y)\} \tag{1.3}$$

with the pair of parameters $p := (x,q)$, where the canonical parameter q corresponds to the perturbation of the left-hand side of the generalized equation (1.1). One clearly has $S(x) = \Sigma(x,0)$ for the solution map (1.2). On the other hand, (1.3) can be viewed as a special case of (1.2) with respect to the parameter pair $p = (x,q)$. The stability results obtained below are generally *independent* of those in [22] even in finite dimensions.

The rest of the paper is organized as follows. Section 2 contains some preliminary material widely used in what follows. In Section 3 we derive characterizations and sufficient conditions of Lipschitzian stability of canonically perturbed variational systems (1.3). Section 4 is devoted to problems with *composite subdifferential structures* of the set-valued part in (1.3). It contains stability results expressed in terms of *second-order* subdifferentials of extended-real-valued functions that are derived on the base of second-order subdifferential calculus.

Throughout the paper we use standard notation, with special symbols introduced where they are defined. Unless otherwise stated, all spaces considered are Banach whose norms are always denoted by $\|\cdot\|$. For any space X we consider its dual space X^* equipped with the weak* topology w^*, where $\langle\cdot,\cdot\rangle$ means the canonical pairing. For multifunctions $F:X\rightrightarrows X^*$ the expression

$$\limsup_{x\to\bar{x}} F(x):=\{x^*\in X^*\mid \exists \text{ sequences } x_k\to\bar{x} \text{ and } x_k^*\xrightarrow{w^*}x^* \text{ with } x_k^*\in F(x_k) \text{ for all } k\in\mathbb{N}\}$$

signifies the *sequential* Painlevé-Kuratowski upper/outer limit with respect to the norm topology in X and the weak* topology in X^*, where $\mathbb{N}:=\{1,2,...\}$.

2. BASIC DEFINITIONS AND PRELIMINARIES

This sections contains basic definitions on Lipschitzian stability and generalized differentiation and review necessary preliminaries.

We say that a set-valued mapping $F:X\rightrightarrows Y$ has the *Aubin Lipschitz-like property* (or it is *Lipschitz-like*) around $(\bar{x},\bar{y})\in\operatorname{gph}F$ if there are neighborhoods U of $\bar{x}$ and V of $\bar{y}$, and a number $\ell>0$ satisfying

$$F(x)\cap V\subset F(u)+\ell\|x-u\|B_Y \text{ for all } x,u\in U, \tag{2.1}$$

where B_Y stands for the closed unit ball in Y. If $V=Y$ and the values of F are compact, the above property reduces to the local Lipschitz continuity of F around $\bar{x}$ with respect to the Pompieu-Hausdorff distance on 2^Y; for single-valued mappings $F=f:X\to Y$ it agrees with the classical local Lipschitz continuity. For general set-valued mappings F the (local) Lipschitz-like property can be viewed as a localization of Lipschitzian behavior not only relative to a point of the domain but also relative to a particular point of the image $\bar{y}\in F(\bar{x})$.

We are able to provide complete *dual characterizations* of the local Lipschitzian and Lipschitz-like properties of set-valued mappings using appropriate constructions of generalized differentiation. Let us recall the basic definitions referring the reader to [20,24,35] for more details, history, and discussions.

Given a nonempty subset Ω of a Banach space X and a number $\varepsilon\ge 0$, we first define the collection of ε *-normals* to Ω by

$$\hat{N}_\varepsilon(x;\Omega) := \left\{ x^* \in X^* \mid \limsup_{u \xrightarrow{\Omega} x} \frac{\langle x^*, u-x \rangle}{\|u-x\|} \le \varepsilon \right\}$$

and by $\hat{N}_\varepsilon(x;\Omega) := \varnothing$ for $x \notin \Omega$. Then the *basic normal cone* to Ω at $\bar{x} \in \Omega$ is defined by

$$N(\bar{x};\Omega) := \limsup_{\substack{x \to \bar{x} \\ \varepsilon \downarrow 0}} \hat{N}_\varepsilon(x;\Omega) \tag{2.2}$$

as the sequential Painlevé-Kuratowski upper limit of ε-normals at nearby points. When the space X is *Asplund* (i.e., its every separable subspace has a separable dual; see [31] for more information) and the set Ω is closed around $\bar{x}$, one can equivalently replace $\hat{N}_\varepsilon(\cdot;\Omega)$ in (2.1) with $\hat{N}(x;\Omega) := \hat{N}_0(x;\Omega)$; see [24, Theorem 2.9].

Given a set-valued mapping $F : X \rightrightarrows Y$, the (normal) *coderivative* of F at $(\bar{x},\bar{y}) \in \operatorname{gph} F$ is defined is a set-valued mapping $D^*F(\bar{x},\bar{y})Y^* \rightrightarrows X^*$ with the values

$$D^*F(\bar{x},\bar{y})(y^*) := \{x^* \in X^* \mid (x^*,-y^*) \in N((\bar{x},\bar{y});\operatorname{gph} F)\}. \tag{2.3}$$

When $F = f : X \to Y$ is single-valued and strictly differentiable at $\bar{x}$, the coderivative $D^*f(\bar{x})(y^*)$ reduces to the *adjoint* derivative operator

$$D^*f(\bar{x})(y^*) = \{\nabla f(\bar{x})^* y^*\} \text{ for all } y^* \in Y^*.$$

A mapping $F : X \rightrightarrows Y$ is *graphically regular* at $(\bar{x},\bar{y})$ if

$$D^*F(\bar{x},\bar{y})(y^*) = \hat{D}^*F(\bar{x},\bar{y})(y^*) := \{x^* \in X^* \mid (x^*,-y^*) \in \hat{N}((\bar{x},\bar{y});\operatorname{gph} F)\}, \quad y^* \in Y^*.$$

This class includes, in particular, strictly differentiable mappings and set-valued mappings with convex graphs, and it is stable with respect to various compositions.

Given an extended-real-valued function $\varphi : X \to \overline{\mathbb{R}} := [-\infty,\infty]$ finite at $\bar{x}$, we define its (first-order) *subdifferential* at $\bar{x}$ by

$$\begin{aligned}\partial\varphi(\overline{x}) &:= D^*E_\varphi(\overline{x},\varphi(\overline{x}))(1)\\ &= \{x^* \in X^* \mid (x^*,-1) \in N((\overline{x},\varphi(\overline{x}));\operatorname{epi}\varphi)\},\end{aligned} \tag{2.4}$$

and the *singular subdifferential*

$$\partial^\infty\varphi(\overline{x}) := D^*E_\varphi(\overline{x},\varphi(\overline{x}))(0) = \left\{x^* \in X^* \mid \left(x^*,0\right) \in N\left((\overline{x},\varphi(\overline{x})) : epi\varphi\right)\right\},$$

where $E_\varphi(x) := \{\nu \in \mathbb{R} \mid \nu \geq \varphi(x)\}$ is the corresponding epigraphical multifunction. Our basic subdifferential (2.4) is smaller than Clarke's generalized gradient $\overline{\partial}\varphi(\overline{x})$ [4] for every lower semicontinuous (l.s.c.) function on a Banach space. If X is Asplund and φ is Lipschitz continuous around $\overline{x}$, then we have the exact relationship [24, Theorem 8.11]:

$$\overline{\partial}\varphi(\overline{x}) = \mathrm{cl}^*\mathrm{co}\,\partial\varphi(\overline{x}),$$

where $\mathrm{cl}^*\mathrm{co}$ stands for the the convex closure in the weak* topology of X^*. Recall also the relationship between the basic subdifferential and coderivative via the *scalarization formula*

$$D^*f(\overline{x})(y^*) = \partial\left\langle y^*, f\right\rangle(\overline{x}) \neq \varnothing \text{ for all } y^* \in Y^* \tag{2.5}$$

established in [24, Theorem 5.2] for *strictly Lipschitzian* mappings $f : X \to Y$ on Asplund spaces X. The latter subclass is proved [36] to agree with compactly Lipschitzian mappings introduced earlier by Thibault; it reduces to the class of all locally Lipschitzian mappings when Y is finite-dimensional.

The *second-order subdifferential* of φ at $\overline{x}$ relative to $\overline{y} \in \partial\varphi(\overline{x})$ is defined as the coderivative of the first-order subdifferential mapping by

$$\partial^2\varphi(\overline{x},\overline{y})(u) := D^*(\partial\varphi)(\overline{x},\overline{y})(u), \quad u \in X^{**}. \tag{2.6}$$

If the function φ is twice continuously differentiable around $\overline{x}$, then

$$\partial^2\varphi(\overline{x})(u) = \{\nabla^2\varphi(\overline{x})^*u\} \text{ for all } u \in X^{**},$$

where $\nabla^2\varphi(\overline{x})$ stands for the classical second-order derivative operator.

The above first-order and second-order generalized differential constructions enjoy fairly rich calculi in both finite-dimensional and infinite-dimensional settings; see [3,10,11,16,20,21,23,24,27,28,35,37], and the

references therein. These calculi require certain qualification conditions and the so-called "normal compactness" conditions needed only in *infinite dimensions*; see [2,10,11,14,24,25,30] for the genesis of the later conditions and various applications. The following two properties formulated in [25] are of particular interest for applications in this paper.

A mapping $F: X \rightrightarrows Y$ is *sequentially normally compact* (SNC) at $(\bar{x},\bar{y}) \in \operatorname{gph} F$ if for any sequences $(\varepsilon_k, x_k, y_k, x_k^*, y_k^*) \in [0,\infty) \times (\operatorname{gph} F) \times X^* \times Y^*$ satisfying

$$\varepsilon_k \downarrow 0,\ (x_k,y_k) \to (\bar{x},\bar{y}),\ (x_k^*,y_k^*) \in \hat{N}_{\varepsilon_k}((x_k,y_k);\operatorname{gph} F) \tag{2.7}$$

one has $(x_k^*,y_k^*) \xrightarrow{w^*} (0,0) \Rightarrow \|(x_k^*,y_k^*)\| \to 0$ as $k \to \infty$. A mapping F is *partially sequentially normally compact* (PSNC) at $(\bar{x},\bar{y})$ if for any above sequences satisfying (2.7) one has

$$[x_k^* \xrightarrow{w^*} 0 \text{ and } \|y_k^*\| \to 0] \Rightarrow \|x_k^*\| \to 0 \text{ as } k \to \infty.$$

We may equivalently put $\varepsilon_k = 0$ in the above properties if both X and Y are Asplund while F is closed-graph around $(\bar{x},\bar{y})$. Finally, a set $\Omega \subset X$ is SNC at $\bar{x} \in \Omega$ if the constant mapping $F(x) \equiv \Omega$ satisfies this property.

Note that the SNC property of sets and mappings are closely related to the compactly epi-Lipschitzian property introduced by Borwein and Strojwas [2] but the latter may be more restrictive in both Banach and nonseparable Asplund spaces; see the recent papers [10] and [8] for comprehensive studies in this direction. Note also that *every Lipschitz-like mapping* $F: X \rightrightarrows Y$ between Banach spaces is PSNC at $(\bar{x},\bar{y})$, and hence it is SNC at this point when $\dim Y < \infty$; see [20, Theorem 3.3]. We refer the reader to [26,27] for extensive calculus results ensuring the preservation of the SNC and PSNC properties under various combinations and compositions of sets and set-valued mappings in general Banach and especially in Asplund spaces settings.

3. LIPSCHITZIAN STABILITY UNDER CANONICAL PERTURBATIONS

First recall the concept of *strong approximation* due to Robinson [34].

Definition 3.1 *Let $f: X \times Y \to Z$ be a mapping between Banach spaces. The mapping $h: Y \to Z$* STRONGLY APPROXIMATES *f in y at $(\bar{x},\bar{y})$ if*

$h(\overline{y}) = f(\overline{x},\overline{y})$ *and for each* $\varepsilon > 0$ *there are neighborhoods* U *of* $\overline{x}$ *and* V *of* $\overline{y}$ *such that*

$$\|[f(x,y_1) - h(y_1)] - [f(x,y_2) - h(y_2)]\| \le \varepsilon \|y_1 - y_2\|$$

whenever $x \in U$ *and* $y_1, y_2 \in V$.

This definition actually means that, although both f and h may not be differentiable in any sense, its difference $g(x,y) := f(x,y) - h(y)$ is *strictly differentiable in* y at $(\overline{x},\overline{y})$ in the sense of

$$\lim_{\substack{y,v\to\overline{y}\\ x\to\overline{x}}} \left[\frac{g(x,y) - g(x,v) - \nabla_y g(\overline{x},\overline{y})(y-v)}{\|y-v\|}\right] = 0 \tag{3.1}$$

with $\nabla_y g(\overline{x},\overline{y}) = 0$. Observe that (3.1) holds, in particular, when g is (Fréchet) differentiable in y around $(\overline{x},\overline{y})$ and $\nabla_y g$ is continuous in (x,y) at this point.

Note that any mapping f in the *separable form*

$$f(x,y) = f_1(x) + f_2(y)$$

admits an obvious strong approximation in y given by f_2. If f itself is strictly differentiable in y at $(\overline{x},\overline{y})$ in the sense of (3.1), its efficient strong approximation can be obtained by the *linearization*

$$h(y) := f(\overline{x},\overline{y}) + \nabla_y f(\overline{x},\overline{y})(y - \overline{y}). \tag{3.2}$$

Also one can check that the *composite mapping* $p(x,y) = f(x,s(y))$ admits a strong approximation in y at $(\overline{x},\overline{y})$ if $f(x,z)$ is strictly differentiable in z at $(\overline{x},\overline{z})$ with $\overline{z} := s(\overline{y})$ while s is Lipschitz continuous around $\overline{y}$.

Let $h: Y \to Z$ strongly approximate f in y at the point $(\overline{x},\overline{y})$ in the sense of Definition 3.1. Along with the original canonically perturbed generalized equation (1.3) we consider the following *approximating system*

$$\Xi(x,q) := \{y \in Y \mid q \in h(y) + Q(x,y)\}. \tag{3.3}$$

The next result, proved by Dontchev [5, Theorem 2.4] employing the Lyusternik-Graves iterative procedure, shows that the *Aubin Lipschitz-like property is preserved under first-order approximations*. Recall that $f: X \times Y \to Z$ is *locally Lipschitzian in* x *uniformly in* y around $(\overline{x},\overline{y})$ if there are neighborhoods U of $\overline{x}$ and V of $\overline{y}$ and a number $\ell \ge 0$ such that

$$\|f(x_1,y)-f(x_2,y)\|\le \ell\|x_1-x_2\|$$

whenever $x_1,x_2\in U$ and $y\in V$.

Lemma 3.2 *Let X,Y,Z be Banach, let Σ and Ξ be given in (1.3) and (3.3), and let $\overline{y}\in\Sigma(\overline{p})$ with $\overline{p}:=(\overline{x},\overline{q})$. Assume that both Σ and Ξ are closed-valued around $\overline{p}$, that f is locally Lipschitzian in x uniformly in y around $(\overline{x},\overline{y})$, and that h strongly approximates f in y at this point. Then the following are equivalent:*

(a) *Ξ is Lipschitz-like around $(\overline{p},\overline{y})$.*

(b) *Σ is Lipschitz-like around $(\overline{p},\overline{y})$.*

The above relationship between the Lipschitz-like property of Σ and Ξ allows us to obtain efficient coderivative conditions for Lipschitzian stability of the solution map (1.3) from those for the (apparently more simple) approximating system (3.3). Let us first derive in such a way *necessary and sufficient* conditions for Lipschitzian stability of the original system (1.3) in the case when f is *strictly differentiable* in y at the reference point. The following theorem unifies two results of this type. The first result concerns canonically perturbed generalized equations with parameter-independent fields $Q=Q(y)$, while the second one applies to the case of *regular* equations with $Q=Q(x,y)$. To formulate the first result, it is convenient to introduce the *partial adjoint generalized equation* to (1.1) involving the adjoint partial derivative of f and the coderivative of Q:

$$0\in\nabla_y f(\overline{x},\overline{y})^* z^* + D^*Q(\overline{y},\overline{s})(z^*), \tag{3.4}$$

where $\overline{s}\in Q(\overline{y})$ and $z^*\in Z^*$.

Theorem 3.3 *Let $\overline{y}\in\Sigma(\overline{x},\overline{q})$ for $\Sigma: X\times Z\rightrightarrows Y$ given in (1.3), where the spaces X,Y,Z are Asplund. Suppose that $f: X\times Y\to Z$ is strictly differentiable in y at $(\overline{x},\overline{y})$ and locally Lipschitzian in x uniformly in y around this point, and that $Q: X\times Y\rightrightarrows Z$ is closed-graph and SNC at $(\overline{x},\overline{y},\overline{s})$ with $\overline{s}:=\overline{q}-f(\overline{x},\overline{y})$. The following hold:*

(i) *Assume that $Q=Q(y)$. Then Σ is Lipschitz-like around $(\overline{x},\overline{q},\overline{y})$ if the partial adjoint generalized equation* (3.4) *has only the trivial solution $z^*=0$. This condition is also necessary for the Lipschitz-like property of Σ when either $\dim Y<\infty$ or Q is graphically regular at $(\overline{y},\overline{s})$.*

(ii) *Assume that $Q=Q(x,y)$ is graphically regular at $(\overline{x},\overline{y},\overline{s})$. Then the condition*

$$(x^*,-\nabla_y f(\overline{x},\overline{y})^* z^*) \in D^*Q(\overline{x},\overline{y},\overline{s})(z^*) \Rightarrow x^* = z^* = 0 \tag{3.5}$$

is necessary and sufficient for the Lipschitz-like property of Σ *around* $(\overline{x},\overline{q},\overline{y})$.

Proof. As mentioned above, if f is strictly differentiable in y at $(\overline{x},\overline{y})$, then its linearization $h(y)$ defined in (3.2) strongly approximates f in y at $(\overline{x},\overline{y})$. Note that $\nabla h(\overline{y}) = \nabla_y f(\overline{x},\overline{y})$. We know from Lemma 3.2 that the Lipschitz-like property of Σ around $(\overline{x},\overline{q})$ is equivalent to this property of Ξ in (3.3) with h defined by (3.2). Denoting $p := (x,q) \in P := X \times Z$, we observe that the approximating mapping $\Xi : P \rightrightarrows Y$ can be written in the form

$$\Xi(p) = \{y \in Y \mid 0 \in \tilde{h}(p,y) + \widetilde{Q}(p,y)\}, \tag{3.6}$$

where $\tilde{h} : P \times Y \to Z$ and $\widetilde{Q} : P \times Y \rightrightarrows Z$ are given by

$$\tilde{h}(p,y) := h(y) - q \quad \text{and} \quad \widetilde{Q}(p,y) := Q(x,y). \tag{3.7}$$

Clearly the strict derivative of $\tilde{h}$ at $(\overline{p},\overline{y})$ is surjective and the adjoint derivative operator is

$$\nabla\tilde{h}(\overline{p},\overline{y})^* z^* = (0,-z^*,\nabla_y f(\overline{x},\overline{y})^* z^*) \text{ for all } z^* \in Z^*.$$

Now we apply to (3.6) the results of [22, Theorem 4.2] based on the coderivative characterization of the Lipschitz-like property from [20, Theorem 3.3] and computing the coderivative of Ξ in terms of $\tilde{h}$ and $\widetilde{Q}$ via coderivative calculus. Taking into account the structures of $\tilde{h}$ and $\widetilde{Q}$ in (3.7), we arrive at the conclusions of the theorem. □

Note that Theorem 3.3 can be derived directly from [22, Theorem 4.2] in the case of canonical parameters provided that f is strictly differentiable at $(\overline{x},\overline{y})$ with respect to *both* variables x and y, while the preliminary strong approximation allows us to justify this result when f is strictly differentiable *only in* y. Taking into account the explicit coderivative representation

$$D^*F(\overline{x},\overline{y})(y^*) = \{x^* \in X^* \mid \langle x^*,\overline{x}\rangle - \langle y^*,\overline{y}\rangle = \max_{(x,y)\in \operatorname{gph} F}[\langle x^*,x\rangle - \langle y^*,y\rangle]\}$$

for *convex-graph* mappings $F: X \rightrightarrows Y$ that are graphically regular at every point of their graphs, one can deduce from Theorem 3.3 efficient characterizations of Lipschitzian stability for variational systems (1.3) with convex-graph fields Q.

Next we obtain sufficient conditions for Lipschitzian stability of canonically perturbed variational systems (1.3) with *nonsmooth and nonregular* data. In what follows the symbol $D_y^* F(\bar{x},\bar{y})$ stands for the *partial coderivative* of $F = F(x,y)$ with respect to y, i.e., for the coderivative (2.3) of the mapping $F(\bar{x},\cdot)$ at $\bar{y}$.

Theorem 3.4 *Let $\bar{y} \in \Sigma(\bar{x},\bar{q})$ for Σ given in (1.3) with $\bar{s} = \bar{q} - f(\bar{x},\bar{y})$. Assume that X, Y, Z are Asplund, that f admits a strong approximation in y at $(\bar{x},\bar{y})$, and that the following hold:*

(a) *f is continuous in (x,y) and locally Lipschitzian in x uniformly in y around $(\bar{x},\bar{y})$. Moreover, $f(\bar{x},\cdot)$ is PSNC at $\bar{y}$, which is automatic if $f(\bar{x},\cdot)$ is Lipschitz continuous around $\bar{y}$.*

(b) *Q is closed-graph around $(\bar{x},\bar{y},\bar{s})$ and SNC at this point.*

Then Σ is Lipschitz-like around $(\bar{x},\bar{q},\bar{y})$ provided the qualification condition

$$[y^* \in D_y^* f(\bar{x},\bar{y})(z^*), \quad (x^*,-y^*) \in D^* Q(\bar{x},\bar{y},\bar{s})(z^*)] \Rightarrow x^* = y^* = z^* = 0, \tag{3.8}$$

which is equivalent to

$$[y^* \in \partial_y \langle z^*, f \rangle (\bar{x},\bar{y}), \quad (x^*,-y^*) \in D^* Q(\bar{x},\bar{y},\bar{s})(z^*)] \Rightarrow x^* = z^* = 0 \tag{3.9}$$

if $f(\bar{x},\cdot)$ is strictly Lipschitzian around $\bar{y}$.

Proof. Let $hY \to Z$ strongly approximate f in y at $(\bar{x},\bar{y})$. By Lemma 3.2 it is equivalent to consider the Lipschitz-like property of the solution map Ξ defined in (3.6) in terms of the mappings $\tilde{h}$ and $\tilde{Q}$ from (3.7). Applying [22, Theorem 4.3] to (3.6), we get that Ξ is Lipschitz-like at $(\bar{p},\bar{y})$ provided that $\tilde{Q}$ is SNC at $(\bar{p},\bar{y},\bar{s})$, $\tilde{h}$ is PSNC at $(\bar{p},\bar{y})$, and one has the qualification conditions

$$[(p^*,0) \in D^* \tilde{h}(\bar{p},\bar{y})(z^*) + D^* \tilde{Q}(\bar{p},\bar{y},\bar{s})(z^*)] \Rightarrow p^* = 0, \tag{3.10}$$

$$(p^*,y^*) \in D^* \tilde{h}(\bar{p},\bar{y})(z^*) \cap (-D^* \tilde{Q}(\bar{p},\bar{y},\bar{s})(z^*))] \Rightarrow p^* = y^* = z^* = 0. \tag{3.10}$$

It follows from the scalarization formula (2.5) that the latter conditions are equivalent to

$$[(p^*,0)\in\partial\left\langle z^*,\tilde{h}\right\rangle(\bar{p},\bar{y})+D^*\widetilde{Q}(\bar{p},\bar{y},\bar{s})(z^*)]\Rightarrow p^*=z^*=0 \tag{3.11}$$

when $\tilde{h}$ is strictly Lipschitzian around $(\bar{p},\bar{y})$. It is obvious that the SNC property of $\widetilde{Q}$ at $(\bar{p},\bar{y},\bar{s})$ is equivalent to the one for Q at $(\bar{x},\bar{y},\bar{s})$. Since h strongly approximates f in y at $(\bar{x},\bar{y})$, the mapping $g(y):=f(x,y)-h(y)$ is strictly differentiable at $\bar{y}$ with $\nabla g(\bar{y})=0$. Elementary rules of coderivative and SNC calculi ensure that

$$D^*h(\bar{y})(z^*)=D^*_yf(\bar{x},\bar{y})(z^*) \text{ for all } z^*\in Z^*$$

and that h is PSNC at $\bar{y}$ in and only of $f(\bar{x},\cdot)$ is PSNC at this point. Furthermore, it follows from the structure of $\tilde{h}$ and $\widetilde{Q}$ in (3.7) that the qualification conditions in (3.10) and (3.11) are equivalent to (3.8) and (3.4), respectively, which completes the proof of theorem. □

To conclude this section, we present two consequences of Theorem 3.4 that give simplified sufficient conditions for Lipschitzian stability of canonically perturbed variational systems (1.3) in some settings important for applications. The first corollary concerns the case of perturbed generalized equations with parameter-independent fields. A mapping $F:X\rightrightarrows Y$ is said to be *strongly coderivatively normal* at $(\bar{x},\bar{y})\in\operatorname{gph}F$ if the coderivative (2.3) agrees with the so-called *mixed coderivative* $D^*_MF(\bar{x},\bar{y})$ of F at this point; see [22] for more details and sufficient conditions for the latter property, which always holds, in particular, when either $\dim Y<\infty$ or F is graphically regular at $(\bar{x},\bar{y})$, and also in various broad settings listed in [22, Proposition 3.2].

Corollary 3.5 *Let $Q=Q(y)$ under the assumptions of Theorem* 3.4 *Then Σ is Lipschitz-like around $(\bar{x},\bar{q},\bar{y})$ provided that*

$$[0\in D^*_yf(\bar{x},\bar{y})(z^*)+D^*Q(\bar{y},\bar{s})(z^*)]\Rightarrow z^*=0 \tag{3.12}$$

and that one has

$$D^*_yf(\bar{x},\bar{y})(0)\cap(-D^*Q(\bar{y},\bar{s})(0))=\{0\}. \tag{3.13}$$

The latter condition is automatic when either $f(\bar{x},\cdot)$ is strictly Lipschitzian around $\bar{y}$ or Q is Lipschitz-like around $(\bar{y},\bar{s})$ and strongly coderivatively normal at this point.

Proof. It is easy to check that that for $Q=Q(y)$ the qualification condition (3.8) of Theorem 3.4 is equivalent to the fulfillment of both qualification conditions (3.12) and (3.13) of the corollary. The last statement of the corollary follows from the coderivative scalarization (2.5) and from [20, Theorem 3.3], which ensures that the Lipschitz-like property of Q around $(\bar{y},\bar{s})$ yields the mixed coderivative conditions $D^*_M Q(\bar{y},\bar{s})(0)=\{0\}$. □

The next corollary gives sufficient conditions for Lipschitzian stability of solutions maps to canonically perturbed generalized equations with smooth bases. They are in the same form as in Theorem 3.5(ii) without imposing the regularity assumption on Q. □

Corollary 3.6 *In addition to the common assumptions of Theorem 3.3 suppose that the qualification condition (3.5) holds. Then Σ in (1.3) is Lipschitz-like around $(\bar{x},\bar{q},\bar{y})$.*

Proof. Follows from Theorem .4 taking into account that the base mapping f smooth in y always admits a strong approximation of form (3.2). □

Observe that for $Q=Q(y)$ the qualification condition (3.5) reduces to the triviality of solutions to the partial adjoint generalized equation (3.4), the sufficiency of which for the Lipschitz-like property of Σ has been established in Theorem 3.3(i). Note also that, since $S(x)=\Sigma(x,0)$ for the solution map (1.2), Corollary 3.5 *unreservedly* improves the sufficient conditions for the Lipschitz-like property of (1.2) in the case of smooth mappings f assuming the strict differentiability of f only in y but not in (x,y). In general the sufficient conditions for Lipschitzian stability of (1.2) obtained in Theorem 3.3 and [22, Theorem 4.3] are *independent*. Indeed, one can check that the qualification condition (3.4) always implies the one in [22, Theorem 4.3] for strictly Lipschitzian mappings. On the other hand, the results of [22] do not require the existence of strong approximations of f as in Theorem 3.3. Furthermore, Theorem 3.3 imposes the Lipschitz continuity of f in x, which is not generally assumed in [22, Theorem 4.3].

4. COMPOSITE VARIATIONAL SYSTEMS

In the concluding section of the paper we consider two classes of canonically perturbed variational systems (1.3) that are probably the most

interesting for applications. Such variational systems involve first-order *subdifferentials of extended-real-valued functions* to define fields of generalized equations in (1.3). In particular, variational and hemivariational inequalities, complementarity problems, and related models can be described in this way involving often the classical subdifferential and normal cone of convex analysis.

Let us first consider a broader class of variational systems defined via the *basic subdifferential* (2.4) of *composite functions* with no convexity assumptions:

$$\Sigma(x,q) := \{y \in Y \,|\, q \in f(x,y) + \partial(\varphi \circ g)(x,y)\}, \tag{4.1}$$

where $g : X \times Y \to W$, $\varphi : W \to \overline{\mathbb{R}}$, and $f : X \times Y \to X^* \times Y^*$ act generally in Banach spaces. Mappings (4.1) are a special case of (1.3) with the subdifferential fields $Q = \partial(\varphi \circ g)$. Employing the results of Section 3, we get conditions for Lipschitzian stability of (4.1) in terms of the *second-order subdifferential* (2.6) of compositions

$$\partial^2(\varphi \circ g)(\overline{x},\overline{y},\overline{s}) = D^*\partial(\varphi \circ g)(\overline{x},\overline{y},\overline{s}). \tag{4.2}$$

Thus one can derive efficient results for Lipschitzian stability of the composite systems (4.1) applying second-order chain rules available for (4.2); see [21,23,27]. Let us present some results in this direction. The next theorem gives *necessary and sufficient* conditions for Lipschitzian stability for the case of parameter-independent mappings $g = g(y)$ in (4.1) with surjective derivatives.

Theorem 4.1 *Let $\overline{y} \in \Sigma(\overline{x},\overline{q})$ for Σ given in* (4.1), *where $Y = \mathbb{R}^m$, X and W are Asplund, and where f is strictly differentiable in y at $(\overline{x},\overline{y})$ and locally Lipschitzian in x uniformly in y around this point. Assume that $g = g(y)$ is C^2 around $\overline{y}$ and the derivative operator $\nabla g(\overline{y})$ is surjective. Denote $\overline{s} := \overline{q} - f(\overline{x},\overline{y})$, $\overline{w} := g(\overline{y})$ and take a unique functional $\overline{v} \in W^*$ satisfying the relations*

$$\overline{s} = \nabla g(\overline{y})^* \overline{v}, \quad \overline{v} \in \partial\varphi(\overline{w}).$$

Then Σ is Lipschitz-like around $(\overline{x},\overline{q},\overline{y})$ if and only if the adjoint system

$$0 \in \nabla_y f(\overline{x},\overline{y})^* u + \nabla^2 \langle \overline{v}, g \rangle (\overline{y})u + \nabla g(\overline{y})^* \partial^2 \varphi(\overline{w},\overline{v})(\nabla g(\overline{y})u), \quad u \in \mathbb{R}^m, \tag{4.3}$$

has only the trivial solution $u = 0$.

Proof. Employing Theorem 3.3(i) with $Q(y) = \partial(\varphi \circ g)(y)$, one has that the mapping Σ in (4.1) is Lipschitz-like around $(\bar{x}, \bar{q}, \bar{y})$ if and only if the adjoint generalized equation

$$0 \in \nabla_y f(\bar{x}, \bar{y})^* u + \partial^2(\varphi \circ g)(\bar{y}, \bar{s})(u), \quad u \in \mathbb{R}^m, \tag{4.4}$$

has only the trivial solution $u = 0$. The second-order subdifferential chain rule of [21, Theorem 4.1] gives, under the assumptions made, that

$$\partial^2(\varphi \circ g)(\bar{y}, \bar{s})(u) = \nabla^2 \langle \bar{v}, g \rangle (\bar{y})^* u + \nabla g(\bar{y})^* \partial^2 \varphi(\bar{w}, \bar{v})(\nabla g(\bar{y}) u). \tag{4.5}$$

Substituting (4.5) into (4.4), we arrive at the triviality of solutions to (4.3) as a criterion of Lipschitzian stability for the canonically perturbed system (4.1). □

Our next theorem concerns *sufficient conditions* for Lipschitzian stability of composite systems (4.1) with a smooth inner mapping g that may depend on *both variables* (x, y) and whose derivative $\nabla g(\bar{y})$ may *not be surjective.* For simplicity we present an efficient result in the finite-dimensional setting in the case of *amenable potentials* $\psi := \varphi \circ g$ in (4.1).

Recall that a function $\psi Z \to \overline{\mathbb{R}}$ is *strongly amenable* at $\bar{z}$ if there is a neighborhood U of $\bar{z}$ on which ψ can be represented in the composition form $\psi = \varphi \circ g$ with a C^2 mapping $g : U \to \mathbb{R}^m$ and a proper l.s.c. convex function $\varphi : \mathbb{R}^m \to \overline{\mathbb{R}}$ satisfying the qualification condition

$$\partial^\infty \varphi(g(\bar{z})) \cap \ker \nabla g(\bar{z})^* = \{0\}.$$

Such functions, which are extensively studied in [35], play a major role in finite-dimensional variational analysis and optimization.

Theorem 4.2 *Let* $\bar{y} \in \Sigma(\bar{x}, \bar{q})$ *for* Σ *given in* (4.1)*, where* $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ *is Lipschitz continuous around* $(\bar{x}, \bar{y})$ *and admits a strong approximation in* y *at this point. Assume that the potential* $\psi = \phi \circ g$ *in* (4.1) *is strongly amenable at this point with* $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^l$, $\bar{w} := g(\bar{x}, \bar{y})$, *and* $\bar{s} := \bar{q} - f(\bar{x}, \bar{y})$. *Denoting*

$$M(\bar{x}, \bar{y}) := \{\bar{v} \in \mathbb{R}^l \mid \bar{v} \in \partial\varphi(\bar{w}), \quad \nabla g(\bar{x}, \bar{y})^* \bar{v} = \bar{s}\},$$

we assume the following second-order qualification conditions:

$$\partial^2\varphi(\overline{w},\overline{v})(0)\cap\ker\nabla g(\overline{x},\overline{y})^*=\{0\}\ \text{ for all }\ \overline{v}\in M(\overline{x},\overline{y}),\tag{4.6}$$

$$\begin{aligned}[y^*\in\partial_y\langle u,f\rangle(\overline{x},\overline{y}),\quad (x^*,-y^*)\in\bigcup_{\overline{v}\in M(\overline{x},\overline{y})}[\nabla^2\langle\overline{v},g\rangle(\overline{x},\overline{y})(u)\\ +\nabla g(\overline{x},\overline{y})^*\partial^2\varphi(\overline{w},\overline{v})(\nabla g(\overline{x},\overline{y})u)]]\Rightarrow x^*=u=0,\end{aligned}\tag{4.7}$$

where the latter reduces to

$$\begin{aligned}[0\in\partial_y\langle u,f\rangle(\overline{x},\overline{y})+\bigcup_{\overline{v}\in M(\overline{x},\overline{y})}[\nabla^2\langle\overline{v},g\rangle(\overline{y})(u)\\ +\nabla g(\overline{y})^*\partial^2\varphi(\overline{w},\overline{v})(\nabla g(\overline{y})u)]]\Rightarrow u=0\end{aligned}\tag{4.8}$$

if $g=g(y)$. *Then* Σ *is Lipschitz-like around* $(\overline{x},\overline{q},\overline{y})$.

Proof. Apply Theorem 3.4 with $Q(x,y)=\partial(\phi\circ g)(x,y)$ taking into account that the mapping Q is closed-graph around $(\overline{x},\overline{y},\overline{s})$, since $\varphi\circ g$ is amenable. The latter theorem, applied in finite-dimensions, ensures the Lipschitz-like property of (4.1) around $(\overline{x},\overline{q},\overline{y})$ if the qualification condition

$$[y^*\in D_y^*f(\overline{x},\overline{y})(z^*),\quad (x^*,-y^*)\in\partial^2(\varphi\circ g)(\overline{x},\overline{y},\overline{s})(z^*)]\Rightarrow x^*=y^*=z^*=0.\tag{4.9}$$

holds. Employing now [21, Corollary 4.3], one has the inclusion

$$\partial^2(\varphi\circ g)(\overline{x},\overline{y},\overline{s})(u)\subset\bigcup_{\overline{v}\in M(\overline{x},\overline{y})}[\nabla^2\langle\overline{v},g\rangle(\overline{x},\overline{y})^*u+\nabla g(\overline{x},\overline{y})^*\partial^2\psi(\overline{w},\overline{v})(\nabla g(\overline{x},\overline{y})u)]\tag{4.10}$$

for all $u\in\mathbb{R}^m$ provided the second-order condition (4.6). Substituting (4.10) into (4.9), we ensure the Lipschitz-like property of (4.1) with strongly amenable potentials under the conditions (4.6) and (4.7. The equivalence between (4.7) and (4.8) in the case of locally Lipschitzian functions f into finite-dimensional spaces follows from the scalarization formula (2.5). □

Finally in this paper we consider a class of canonically perturbed variational systems with another type of *subdifferential compositions*:

$$\Sigma(x,q):=\{y\in Y\,|\,q\in f(x,y)+(\partial\varphi\circ g)(x,y)\},\tag{4.11}$$

where $g: X \times Y \to W$, $\varphi W \to \overline{\mathbb{R}}$, and $f: X \times Y \to W^*$. The next theorem contains *sufficient* conditions, as well as *necessary and sufficient* conditions, for Lipschitzian stability of systems (4.11) in infinite-dimensions via their initial data.

Theorem 4.3 *Let* $\overline{y} \in \Sigma(\overline{x}, \overline{q})$ *with* $\overline{s} := \overline{q} - f(\overline{x}, \overline{y})$ *and* $\overline{w} := g(\overline{x}, \overline{y})$ *for* Σ *given in* (4.11), *where* X, Y, W *are Asplund and where* $\partial\varphi$ *is SNC at* $(\overline{w}, \overline{s})$. *The following assertions hold:*

(i) *Assume that* $g = g(y)$ *is strictly differentiable at* $\overline{y}$ *with the surjective derivative* $\nabla g(\overline{y})$, *and that* f *is strictly differentiable in* y *at* $(\overline{x}, \overline{y})$ *and locally Lipschitzian in* x *uniformly in* y *around this point. Then the condition*

$$[0 \in \nabla_y f(\overline{x}, \overline{y})^* u + \nabla g(\overline{y})^* \partial^2 \varphi(\overline{w}, \overline{s})(u)] \Rightarrow u = 0 \tag{4.12}$$

is necessary and sufficient for the Lipschitz-like property of Σ *around* $(\overline{x}, \overline{z})$ *provided that the space* Y *is finite-dimensional.*

(ii) *Assume that* W^* *is Asplund, that* g *is continuous around* $(\overline{x}, \overline{y})$ *and PSNC at this point, that the graph of* $\partial\varphi$ *is norm-closed around* $(\overline{w}, \overline{s})$, *and that* f *is strictly Lipschitzian around* $(\overline{x}, \overline{y})$ *and admits a strong approximation in* y *at this point. Assume also the qualification conditions*

$$\partial^2 \varphi(\overline{w}, \overline{s})(0) \cap \ker D^* g(\overline{x}, \overline{y}) = \{0\} \text{ and} \tag{4.13}$$
$$[y^* \in \partial_y \langle u, f \rangle (\overline{x}, \overline{y}), \quad (x^*, -y^*) \in D^* g(\overline{x}, \overline{y}) \circ \partial^2 \varphi(\overline{w}, \overline{s})(u)] \Rightarrow x^* = u = 0,$$

where the latter reduces to

$$[0 \in \partial_y \langle u, f \rangle (\overline{x}, \overline{y}) + D^* g(\overline{x}, \overline{y}) \circ \partial^2 \varphi(\overline{w}, \overline{s})(u)] \Rightarrow u = 0 \tag{4.14}$$

when $g = g(y)$. *Then* Σ *is Lipschitz-like around* $(\overline{x}, \overline{q}, \overline{y})$.

Proof. To prove (i), we first conclude from Theorem 3.3(i) with $Q = \partial\varphi \circ g$ and $\dim Y < \infty$ that the mapping Σ in (4.11) is Lipschitz-like at $(\overline{x}, \overline{q}, \overline{y})$ if and only of the adjoint generalized equation

$$0 \in \nabla_y f(\overline{x}, \overline{y})^* u + D^*(\partial\varphi \circ g)(\overline{y}, \overline{s})(u), \quad u \in W^{**}, \tag{4.15}$$

has only the trivial solution $u=0$, provided that the composition $\partial\varphi\circ g$ is SNC at $(\overline{y},\overline{s})$. It follows from the coderivative chain of [27, Theorem 3.10] that

$$D^*(\partial\varphi\circ g)(\overline{y},\overline{s})(u)=\nabla g(\overline{y})^*\partial^2\varphi(\overline{w},\overline{s})(u). \tag{4.16}$$

Furthermore, by [27, Corollary 5.4] the SNC property of $\partial\varphi\circ g$ at $(\overline{y},\overline{s})$ is equivalent to the one of $\partial\varphi$ at $(\overline{w},\overline{s})$. Substituting (4.16) into (4.15), we justify that (4.12) is necessary and sufficient for the Lipschitz-like property of (4.11) under the assumptions made in (i).

To prove assertion (ii) of the theorem, we use Theorem 3.4 and Corollary 3.5 with $Q=\partial\varphi\circ g$. Then applying the coderivative chain from [25] to $D^*(\partial\varphi\circ g)(\overline{x},\overline{y},\overline{s})$ and the result of [26, Theorem 5.4] on the preservation of the SNC property of the composition $\partial\varphi\circ g$, we conclude that the variational system (4.11) is Lipschitz-like around $(\overline{x},\overline{q},\overline{s})$ under the assumptions made in (ii). This completes the proof of the theorem. □

Let us present an efficient corollary of Theorem 4.3(ii) in the case when g is strictly differentiable at $(\overline{x},\overline{y})$ in both variables while f is strictly differentiable at this point in y.

Corollary 4.4 *In the notation of Theorem* 4.3, *assume that the spaces* X,Y,W,W^* *are Asplund, that the subdifferential mapping* $\partial\varphi$ *is SNC at* $(\overline{w},\overline{s})$ *and its graph is norm-closed around this point, that* g *is strictly differentiable at* $(\overline{x},\overline{y})$, *and that* f *is strictly differentiable in* y *at this point. Then the mapping* (4.11) *is Lipschitz-like around* $(\overline{x},\overline{q},\overline{y})$ *provided that* (4.13) *holds with* $D^*g(\overline{x},\overline{y})=\nabla g(\overline{x},\overline{y})^*$ *and that one has the qualification conditions*

$$\partial^2\varphi(\overline{w},\overline{s})(0)\subset\ker\nabla_x g(\overline{x},\overline{y})^* \text{ and} \tag{4.17}$$

$$[0\in\nabla_y f(\overline{x},\overline{y})^*u+\nabla_y g(\overline{x},\overline{y})^*\partial^2\varphi(\overline{w},\overline{s})(u)]\Rightarrow u=0, \tag{4.18}$$

Proof. Since f is strictly differentiable in y, the PSNC and strict approximation assumptions of Theorem 4.3 are automatic. Taking into account the coderivative representation for strict differentiable mappings, it is easy to observe that condition (4.14) is *equivalent* to the simultaneous fulfillment of conditions (4.17) and (4.18) in this case. □

Since the Lipschitz-like property of (4.11) obviously implies the one for

$$S(x) := \{y \in Y \mid 0 \in f(x,y) + (\partial\varphi \circ g)(x,y)\},$$

the result of Corollary 4.4 gives a corrected version of [23, Theorem 5.1] in finite dimensions.

Remark 4.5 The property of solution maps to parametric generalized equations to be *single-valued and Lipschitz continuous* around a reference point is known as *Robinson strong regularity* [33]. The results presented above allow us to obtain sufficient as well as necessary and sufficient conditions for this property in the case of *monotone* fields $Q = Q(y)$ in the original generalized equation (1.1), which particularly covers subdifferential operators $Q = \partial\varphi$ with a proper convex function φ (e.g., the classical variational inequalities and complementarity problems). This relates to the well-known fact that a monotone map has to be single-valued and continuous wherever it is lower/inner semicontinuous. Thus the above conditions for the Aubin Lipschitz-like property of solution maps to the variational systems under consideration ensure actually their strong regularity provided monotonicity. Such a monotonicity of solution maps follows from the monotonicity of Q and the corresponding monotonicity of a strong approximation to f in the sense of Definition 4.1; cf. [19, Section 7] for more discussions and coderivative conditions for strong regularity obtained in this way for generalized equations in finite dimensions. Note that in the case of mappings f strictly differentiable in y the monotonicity of strong approximations corresponds to the positive semidefiniteness of the partial derivative $\nabla_y f(\overline{x}, \overline{y})$.

If $Q = \delta(y;\Omega)$ is the indicator function of a *convex polyhedron* $\Omega \subset \mathbb{R}^n$ and f is smooth in y, efficient characterizations of strong regularity for *canonically perturbed variational inequalities* are obtained by Dontchev and Rockafellar [7] *with no positive semidefiniteness* assumption on $\nabla_y f(\overline{x}, \overline{y})$. Their main result establishes the equivalence between strong regularity of the original generalized equation and the Aubin property of the solution map to its linearization, for which a verifiable "critical face" condition is derived on the base of the coderivative criterion from [17].

REFERENCES

[1] Aubin, J.-P.: Lipschitz behavior of solutions to convex minimization problems, *Math. Oper. Res.* 9 (1984), 87–111.

[2] Borwein, J.M. and Strojwas H.M.: Tangential approximations, *Nonlinear Anal.* 9 (1985), 1347–1366.

[3] Borwein, J.M. and Zhu, Q.J.: A survey of subdifferential calculus with applications, *Nonlinear Anal.* 38 (1999), 687–773.

[4] Clarke, F.H.: *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
[5] Dontchev, A.L.: Chararacterization of Lipschitz stability in optimization, In: *Well-Posedness and Stability of Optimization Problems and Related Topics* (R. Lucchetti and J. Revalski, eds.), Kluwer, Dordrecht, 1995, pp. 95–116.
[6] Dontchev, A.L. and Hager, W.W.: Implicit functions, Lipschitz maps, and stability in optimization, *Math. Oper. Res.* 19 (1994), 753–768.
[7] Dontchev, A.L. and Rockafellar, R.T.: Characterizations of strong regularity for variational inequalities over polyhedral convex sets, *SIAM J. Optim.* 7 (1996), 1087–1105.
[8] Fabian, M. and Mordukhovich, B.S.: Sequential normal compactness versus topological normal compactness in variational analysis, *Nonlinear Anal.*, 54 (2003), pp. 1057-1067.
[9] Henrion, R. and Römisch, W.: Metric regularity and quantitative stability in stochastic programming with probabilistic constraints, *Math. Programming* 84 (1999), 55–88.
[10] Ioffe, A.D.: Coderivative compactness, metric regularity and subdifferential calculus, In: M. Théra (ed.), *Expremental, Constructive, and Nonlinear Analysis*, CMS Conference Proc. Vol. 27, American Mathematical Society, Providence, R.I., 2000, pp. 123–164.
[11] Jourani, A. and Thibault, L.: Coderivatives of multivalued mappings, locally compact cones and metric regularity, *Nonlinear Anal.* 35 (1999), 925–945.
[12] Levy, A.B. and Mordukhovich, B.S.: Coderivatives in parametric optimization, *Math. Programming*, in press.
[13] Levy, A.B., Poliquin, R.A., and Rockafellar, R.T.: Stability of locally optimal solutions, *SIAM J. Optim.* 10 (2000), 580–604.
[14] Loewen, P.D.: Limits of Fréchet normals in nonsmooth analysis, In: *Optimization and Nonlinear Analysis* (A.Ioffe et al., eds.), Pitman Research Notes Math. Ser. 244, 1992, pp. 178–188.
[15] Lucet, Y. and Ye, J.J.: Sensitivity analysis of the value function for optimization problems with variational inequality constraints, *SIAM J. Control Optim.* 40 (2001), 699–723.
[16] Mordukhovich, B.S.: *Approximation Methods in Problems of Optimization and Control*, Nauka, Moscow, 1988.
[17] Mordukhovich, B.S.: Complete characterization of openness, metric regularity, and Lipschitzian properties of multifunctions, *Trans. Amer. Math. Soc.* 340 (1993), 1–35.
[18] Mordukhovich, B.S.: Lipschitzian stability of constraint systems and generalized equations, *Nonlinear Anal.* 22 (1994), 173–206.
[19] Mordukhovich, B.S.: Stability theory for parametric generalized equations and variational inequalities via nonsmooth analysis, *Trans. Amer. Math. Soc.* 343 (1994), 609–658.
[20] Mordukhovich, B.S.: Coderivative of set-valued mappings: calculus and applications, *Nonlinear Anal.* 30 (1997), 3059–3070.
[21] Mordukhovich, B.S.: Calculus of second-order subdifferentials in infinite dimensions, *Control and Cybernetics* 31 (2002), 557–573.
[22] Mordukhovich, B.S.: Coderivative analysis of variational systems, *J. Global Optim.*, 28(2004).
[23] Mordukhovich, B.S. and Outrata, J.V.: On second-order subdifferentials and their applications, *SIAM J. Optim.* 12 (2001), 139–169.
[24] Mordukhovich, B.S. and Shao, Y.: Nonsmooth sequential analysis in Asplund spaces, *Trans. Amer. Math. Soc.* 348 (1996), 235–1280.
[25] Mordukhovich, B.S. and Shao, Y.: Nonconvex differential calculus for infinite-dimensional multifunctions, *Set-Valued Analysis* 4 (1996), 205–236.

[26] Mordukhovich, B.S. and Wang, B.: Calculus of sequential normal compactness in variational analysis, *J. Math. Anal. Appl.* 282 (2003), 63–84.
[27] Mordukhovich, B.S. and Wang, B.: Restrictive metric regularity and generalized differential calculus in Banach spaces, Preprint No. 15 (2002), Dept. of Math., Wayne State University, Detroit.
[28] Ngai, N.V. and Théra, M.: Metric regularity, subdifferential calculus and applications, *Set-Valued Anal* .9 (2001), 187–216.
[29] Outrata, J.V.: A general mathematical program with equilibrium constraints, *SIAM J. Control Optim.* 38 (2000), 1623–1638.
[30] Penot, J.-P.: Compactness properties, openness criteria and coderivatives, *Set-Valued Analysis* 6 (1998), 363–380.
[31] Phelps, R.R.: *Convex Functions, Monotone Operators and Differentiability*, 2nd edition, Springer, Berlin, 1993.
[32] Robinson, S.M.: Generalized equations and their solutions, part I: basic theory, *Math. Programming Study* 10 (1979), 128–141.
[33] Robinson, S.M.: Strongly regular generalized equations, *Math. Oper. Res.* 5 (1980), 43–62.
[34] Robinson, S.M.: An implicit-function theorem for a class of nonsmooth functions, *Math. Oper. Res.* 16 (1991), 292–309.
[35] Rockafellar, R.T. and Wets, R. J.-B.: *Variational Analysis*, Springer, Berlin, 1998.
[36] Thibault, L.: On compactly Lipschitzian mappings, In: P. Gritzmann et al.(eds.), *Recent Advances in Optimization*, Lecture Notes in Econ. Math. Syst. Ser. 456, Springer, Berlin, 1997, pp. 356–364.
[37] Vinter, R.B.: *Optimal Control*, Birkhäuser, Boston, 2000.

STABLE CRITICAL POINTS FOR THE GINZBURG LANDAU FUNCTIONAL ON SOME PLANE DOMAINS

M.K. Venkatesha Murthy
Dept. of Mathematics, University of Pisa, Pisa, Italy

We shall present a survey of some new developments in the study concerning the stable critical points for the Ginzburg - Landau energy functional in two dimensions. We shall consider only vortex free solutions of the problem. The Ginzburg - Landau functional is defined on the space of pairs (u, A), where u is a scalar complex valued function in $H^1(X, \mathrm{C})$ and A is a vector field in $H^1_{loc}(\mathrm{R}^2, \mathrm{R}^2)$ such that $\mathrm{rot}A$ is a square integrable field on R^2, by

$$\text{(GL)} \quad \mathcal{G}(u, A) = \frac{1}{2}\int_X \| (\nabla - iA)u \|^2 \, \mathrm{d}x + \frac{\kappa^2}{4}\int_X V(u)\mathrm{d}x + \frac{1}{2}\int_{\mathrm{R}^2} \| \mathrm{rot}A \|^2 \, \mathrm{d}x$$

where $V(u) = (|u|^2 - 1)^2$. The interesting feature of this functional is that the underlying topological and geometric structures of the domain play important and crucial roles in this study.

If X is a multiply connected domain in $\mathbb{R}^2$ or $\mathbb{R}^3$ it has been shown by Jimbo and his collaborators (see the references) that there is a stable critical point in each homotopy class of X. Infact, if X is a bounded domain in $\mathbb{R}^3$ with a Lipschitz boundary, which is topologically equivalent to a 3 - dimensional solid torus $\mathbb{T}^3$ then each 1-homotopy type of maps from X to

the unit circle S^1 contains a nontrivial critical point. Since these maps are a priori only H^1-maps, in order to make precise this notion, we recall the following fundamental approximation theorem of Sobolev maps between compact manifolds due to Bethuel:

Suppose M and N are two compact manifolds where N is a manifold without boundary. We assume that N is isometrically embedded in some Euclidean space $\mathbb{R}^d$. We introduce the space of Sobolev maps between M and N as follows:

$$W^{1,p}(M,N) = \{u \in W^{1,p}(M,\mathrm{R}^d); u(x) \in Na.e.\}$$

and this space is provided with the strong and weak topologies inherited from those of $W^{1,p}(M,\mathbb{R}^d)$

If $u : X \longrightarrow \mathbb{C}$ is a continuous map then its 1-homotopy type is defined as the 1-homotopy type of the restriction of u to the one dimensional skeleton (set of all simplices of the triangulation of dimensions ≤ 1) of any triangulation of the space X. This, in the case of the bounded Lipschitz domain topologically equivalent to the torus $\mathbb{T}^3$, is the same thing as the winding number of the restriction of u to any closed rectifiable curve which loops once around the hole in X (image of the hole in $\mathbb{T}^3$ by the homeomorphism of the equivalence).

Theorem 1. *(Bethuel) Density of smooth maps in Sobolev maps: Suppose M and N are two compact Riemannian manifolds where N is a manifold without boundary. If $1 \leq p < dimM = n$ then the set of smooth maps between M and N is dense in the Sobolev space of maps $W^{1,p}(M,N)$ if and only if the $[p]$ - th fundamental group $\pi_{[p]}(N) = 0$.*

In particular, taking $p = 2$ and $dimM = 3$ the set of smooth maps from M to S^1 is dense in $H^1(M,S^1)$. For example, we can take $M = \mathbb{T}^3$ and $N = S^1 \subset \mathbb{C} = \mathbb{R}^2$.

Now if u is a map belonging to $H^1(X,S^1)$ then in view of the result of Bethuel on the density of smooth maps in this space we define the 1-homotopy type of u as the 1-homotopy type of any approximating smooth map.

The result recalled above seems to suggest that, if X is a simply connected bounded domain then there might not be any non trivial critical points. However, it has been proved recently that there exist non trivial stable critical points in multiply connected domains and also in some simply connected domains which are in some sense perturbations of multiply connected domains. However, these critical points contain vortices, namely the zeros of the order parameter while in the multiply connected domains case there are no vortices.

Hence the question of existence or otherwise of non trivial critical points seems to be related not only to the underlying topological structure but also to the differential geometric properties of the domain. The existence of stable critical points therefore seems to be related to the geometry of the domain much more closely than with its fundamental group.

This survey is to illustrate one such connection.

Remarks on the physical interpretation - The Ginzburg - Landau functional is related to modelling superconducting materials in physics literature - super conductors are materials which have almost vanishing electrical resistivity (the resistivity is effectively zero). It is known that if a ring shaped superconducting material is subjected to an applied magnetic field which induces a current and the if the temperature is lowered below a certain critical value, the current persists even after the external field is removed for a very long time (even for some years). This phenomenon is known as persistence of permanent currents.

Mathematically, the stable critical points of the Ginzburg - Landau functional correspond to the existence of permanent currents.

In the funcional (GL) introduced above, $u(x)$ is the density of super conducting electron pairs , $A(x)$ denotes the magnetic vector potential and $V(u)$ is the energy density due to interaction. Since the electrons are confined to the super conducting material represented by the domain X the integral over the domain X represents the energy associated to the electron pair density $u(x)$ while the magnetic field is defined over the entire space the energy due to the current generated by the magnetic field is defined by the integral over the whole space $\mathbb{R}^2$.

The method of proof used is an extension of a known method in the study of critical points in the scalar case: If X is a bounded convex open set with smooth boundary, consider a functional of the form

$$u \longrightarrow \int_X [|| \nabla u ||^2 + F(u)] \mathrm{d}x$$

In order to study the critical points one explicitely computes the second variation about the critical points and then reduce the expression for the second variation to an integral over the boundary ∂X of the domain X. This allows one to use the convexity assumption on X.

In the case of the functional (GL) the explicite computation of the second variation and the reduction to boundary integrals leads to an expresion depending on the mean curvature of ∂X.

Infact, an identity involving the mean curvature of the boundary is proved from which, in particular, it follows that if the domain is smooth and

convex then the only stable critical point is the trivial pair $(1,0)$ upto gauge transformations.

1. NOTATION AND DEFINITIONS

Let X be a bounded open set in the plane $\mathbb{R}^2$ with $C^{5,\alpha}$ smooth boundary ∂X, and let $x=(x_1,x_2)\in X$. Let ν be the exterior normal vector field on ∂X and $K(x)$ be the mean curvature of ∂X. Consider the function space

$$E=H^1(X,\mathbb{C})\times\{A\in H^1_{loc}(\mathbb{R}^2,\mathbb{R}^2);\ \mathrm{rot}A\in L^2(\mathbb{R}^2,\mathbb{R}^2)\}$$

The Ginzburg - Landau energy functional $\mathcal{G}: E\longrightarrow\mathbb{R}$ is defined by

$$\mathcal{G}(u,A)=\frac{1}{2}\int_X\ ||\,(\nabla-iA)u\,||^2\ \mathrm{d}x+\frac{\kappa^2}{4}\int_X\ V(u)\mathrm{d}x+\frac{1}{2}\int_{\mathbb{R}^2}\ ||\,\mathrm{rot}A\,||^2\ \mathrm{d}x$$

where

$$V(u)=(|\,u\,|^2-1)^2$$

It is immediately seen that the functional $\mathcal{G}$ is invariant under the gauge transformations $(u,A)\longrightarrow(\mathrm{e}^{i\varphi}u,A+\nabla\varphi)$ in the sense that

$$\mathcal{G}(\mathrm{e}^{i\varphi}u,A+\nabla\varphi)=\mathcal{G}(u,A)$$

for all $\varphi\in H^2_{loc}(\mathbb{R}^2,\mathbb{R})$ such that $\nabla\varphi\in L^2(\mathbb{R}^2,\mathbb{R}^2)$

For the purposes of calculations it is convenient to make a suitable choice of the gauge:

First of all we may assume $\mathrm{div}A=0$ in $\mathbb{R}^2$. Infact, given $A\in H^1_{loc}(\mathbb{R}^2)$ consider the Poisson equation

$$-\Delta\varphi=\mathrm{div}A\in L^2_{loc}(\mathbb{R}^2)$$

There exists a solution $\varphi\in H^2_{loc}(\mathbb{R}^2)$. With this φ we have

$$\mathrm{div}(A+\nabla\varphi)=\mathrm{div}A+\Delta\varphi=0$$

Similarly, in the case of a bounded open set X we may assume

$$\mathrm{div}A = 0 \quad \text{in} \quad X \quad \text{and} \quad < A, \nu >= 0 \quad \text{on} \quad \partial X$$

Infact, since the boundary is asumed to be sufficiently smooth, the Neumann problem

$$-\Delta\varphi_0 = \mathrm{div}A \quad \text{in} \quad X, \frac{\partial\varphi_0}{\partial\nu} =< \nabla\varphi_0, \nu >= - < A, \nu > \quad \text{on} \quad \partial X$$

has a solution $\varphi_0 \in H^2_{loc}(X, \mathbb{R})$ and then we can take φ to be an extension of φ_0 in $H^2_{loc}(\mathbb{R}^2, \mathbb{R})$ in the gauge transformation.

Critical points of $\mathcal{G}$:

A pair $(u, A) \in E$ is said to be a critical point of the functional $\mathcal{G}$ if

$$(D\mathcal{G})(u, A) = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{G}(u + tv, A + tB)|_{t=0} = 0, \quad \text{where}(v, B) \in E$$

A critical point is said to be non trivial when

$(D\mathcal{G}(u, A) = 0$ and (u, A) is not equivalent to $(c, 0)$ for any $c \in \mathbb{C}$, a constant, under any gauge transformation; that is, (u, A) is not of the form $(ce^{i\varphi}, \nabla\varphi)$ for any $\varphi \in H^2_{loc}(\mathbb{R}^2, \mathbb{R})$ and any $c \in \mathbb{C}$.

Ginzburg Landau system of differential equations

Writing down explicitely the condition satisfied by a critical point of $\mathcal{G}$ we find the following coupled system of differential equations, satisfied in the weak sense:

$$(GL)\begin{cases} (\nabla - iA)^2 u + \kappa^2(|u|^2 - 1)u = 0 & \text{in } X \\ \mathrm{rot}(\mathrm{rot}A) + \left\{ i\dfrac{\kappa^2}{2}(u^*\nabla u - u\nabla u^*) + |u|^2 A \right\}\chi_X = 0 & \text{in } \mathbb{R}^2 \end{cases}$$

together with the following natural system of boundary conditions

$$(BC)\begin{cases} < (\nabla - iA)u, \nu >= 0, & \text{on } \partial X \\ \nu \wedge [\mathrm{rot}A] = 0 & \text{on } \partial X \end{cases}$$

where χ_X denotes the characteristic function of X and $[\mathrm{rot}A]$ denotes the jump of rotA across the boundary ∂X.

Since

$$\mathrm{rot}(\mathrm{rot}A) = -\Delta A + \nabla(\mathrm{div}A) = -\Delta A$$

because, by the chioce of the gauge namely, $\mathrm{div}A = 0$, the second system of equations can be written as

$$-\Delta A + \{i\frac{\kappa^2}{2}(u^*\nabla u - u\nabla u^*) + |u|^2 A\}\chi_X = 0 \quad \text{in}\mathbb{R}^2$$

Thus (GL) is an elliptic system for (u, A)
Choosing the gauge as indicated above, namely

$$\mathrm{div}A = 0 \quad \text{in} \quad X \quad \text{and} \quad < A, \nu >= 0 \quad \text{on} \quad \partial X$$

the first boundary condition becomes

$$\frac{\partial u}{\partial \nu} =< \nabla u, \nu >= 0 \quad \text{on } \partial X$$

Since $\mathrm{rot}A = (0, 0, \frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2})$ where $A = (a_1, a_2)$ we have

$$[\mathrm{rot}A] = (0, 0, \left[\frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2}\right]) \text{ where } A = (a_1, a_2)$$

Here $[\frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2}])$ denotes the jump of the real valued function $\frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2})$

Definition 1. Stable critical point. *A critical point $(u, A) \in E$ of $\mathcal{G}$ is said to be a stable critical point if the second variation of $\mathcal{G}$ is non-negative.* that is,

$$\begin{cases} (D\mathcal{G})(u, A) = 0 \\ (D^2\mathcal{G})(u, A; v, B) = \dfrac{d^2\mathcal{G}}{dt^2}(u + tv, A + tB)\,|_{t=0} \geq 0 \quad \text{for all } (v, B) \in E \end{cases}$$

An explicite computation shows that the second variation is given by the following expression

$$(D^2\mathcal{G})(u,A;v,B)=\frac{1}{2}\int_X \{|| \nabla v ||^2 + i < \nabla u, Bv^* > + i < \nabla v, Av^* + Bu^* >$$
$$- i < \nabla u^*, Bv > - i < \nabla v, Av + Bu > + || A ||^2 | v |^2 + || B ||^2 | u |^2$$
$$+2 < A, B > (uv^* + vu^*)\}dx + \frac{\kappa^2}{4}\int_X \{(uv^* + vu^*)^2 - 2(| u |^2 - 1) | v |^2\}dx$$
$$+\frac{1}{2}\int_{\mathbb{R}^2} || \mathrm{rot} B ||^2 \, dx$$

Two basic lemmas

We recall without proofs the following two properties of critical points of $\mathcal{G}$ in plane domains, which we shall use in the proof of the main result:

Lemma 1. *If $(u,A)\in E$ is a critical point of $\mathcal{G}$ then we have*

$$| u | \leq 1 \text{ in } \overline{X} \text{ and } \mathrm{rot} A = 0 \text{ in } \mathbb{R}^2 \setminus \overline{X}$$

This follows from the fact that the system is elliptic and the boundary condition is of Neumann type, using the first equation in (GL) to get a differential inequality for $| u |^2$ and then applying the maximum principle.

In $\mathbb{R}^2$, $\mathrm{rot}(\mathrm{rot} A) = 0$ in $\mathbb{R}^2 \setminus \overline{X}$ implies that

$$\mathrm{rot} A = (0, 0, \frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2})$$

is a constant. Since $\mathrm{rot} A \in L^2(\mathbb{R}^2, \mathbb{R}^2)$ this constant should be zero.

Lemma 2. *If a vector field $F \in C^1(X, \mathbb{R}^2)$ satisfies $< F, \nu >= 0$ and $\mathrm{rot} F = 0$ on ∂X then*

$$\frac{\partial}{\partial \nu}(|| F ||^2) = -2\sum_{j,k} \frac{\partial \nu^{(k)}}{\partial x_j} F^{(k)} F^{(j)} = -2K(x) || F ||^2$$

In particular, if $u \in C^2(\overline{X})$ and $\frac{\partial u}{\partial \nu} = < \nabla u, \nu >= 0$ on ∂X then

$$\frac{\partial}{\partial \nu}(|| \nabla u ||^2) = -2K(x) || \nabla u ||^2$$

where we recall that $K(x)$ denotes the mean curvature of ∂X

2. REGULARITY OF CRITICAL POINTS - PRELIMINARY RESULTS FROM ELLIPTIC REGULARITY THEORY

In the proof of the main result we shall use for (v, B) in the second variation the pair of derivatives $\left(\frac{\partial u}{\partial x_j}, \frac{\partial A}{\partial x_j}\right)(j = 1,2)$. For this end we require that this pair belongs to the space E, which is a regularity property of the critical point. In order to prove this crucial regularity of the critical point (u, A) we need the following regularity result which is a consequence of Schauder estmates for elliptic boundary value problems due to Agmon, Douglis and Nirenberg [1].

A critical point is a weak solution of the elliptic boundary value problem for the system (GL) with boundary conditions (BC). First of all, the system being elliptic the interior regularity for elliptic systems implies that such a weak solution (u, A) is infinitely differentiable in the two open sets X and $\mathbb{R}^2 \setminus \overline{X}$ and thus it is a classical solution of the system in the two open sets. However we shall need some kind of regularity upto the boundary. In view of the discontinuity across the boundary ∂X in the second set of equations in (GL) due to the presence of the characteristic function χ_X the boundary regularity is rather delicate. We do not have smoothness of high orders across the boundary. We have the following:

Theorem 2. *Suppose X is a bounded open set in R^2 with the boundary ∂X of class $C^{5,\alpha}$ with some $0 < \alpha < 1$. If we choose the appropriate gauge namely, div $A = 0$, then any critical point $(u, A) \in E$ belongs to the space $C^3(\overline{X}, \mathbb{C}) \times C^{1,\alpha}_{loc}(\mathbb{R}^2, \mathbb{R}^2)$. Moreover, A belongs to the space $C^{3,\alpha}(\overline{X}; \mathbb{R}^2)$, in the sense of (one sided) regularity from the interior of the domain X.*

Remark. The regularity properties remain invariant under gauge transformations

$$(u, A) \longrightarrow (ue^{i\varphi}, A + \nabla\varphi)$$

because the function φ, being a solution of a regular elliptic boundary value problem, is a smooth function and hence the gauge transformations are smooth. The first equation in (GL) is an elliptic equation for u with coefficients in $L^2(X, \mathbb{C})$.

Sketch of proof. We first note that as mentioned earlier, in view of the choice of the gauge div $A = 0$, we have the identity

$$\mathrm{rot}(\mathrm{rotA}) = -\Delta A + \nabla(\mathrm{div} A) = -\Delta A$$

so that A satisfies an elliptic system

$$-\Delta A = g(x) = \begin{cases} -\kappa^2 Im(u\nabla u^*) - |u|^2 A & \text{in } X \\ 0 & \text{in } \mathbb{R}^2 \setminus X \end{cases}$$

First of all we use the fact that the critical pair (u, A) belongs to the space E and has the minimum regularity. Writing the first equation of the system in the form

$$(\nabla - iA)^2 u = -\kappa^2 (|u|^2 - 1)u$$

we have a scalar elliptic (nonlinear) equation for u. The right hand side and the coefficients on the left hand side have a certain regularity which enables us to apply the elliptic regularity theorem to improve the regularity of u in $\overline{X}$. Then we use this in the system satisfied by the vector field A to obtain a further regularity of A. We continue this boot strap argument. In this procedure, we flatten the boundary locally and in neighbourhoods of boundary points we apply the regularity results of Agmon, Douglis and Nirenberg [1] on Schauder estimates and then patching up these we obtain first estmates for (u, A) on $\overline{X}$ and putting back into the system of equations (GL) we recover further regularity of the pair.

More precisely, we proceed as follows:

We first observe that, using the gauge such that $\text{div} A = 0$ in X and $< A, \nu >= 0$ on ∂X

$$\begin{cases} (\nabla - iA)^2 u = -\kappa^2 (1 - |u|^2)u \\ < (\nabla - iA)u, \nu >= 0, \quad \text{that is} \quad \frac{\partial u}{\partial \nu} = 0 \quad \text{on} \quad \partial X \end{cases}$$

is the Neumann problem for a second order semilinear elliptic equation. However, since $u \in H^1(X, \mathbb{C})$ and dim $X = 2$, we have in view of the Sobolev embedding theorem that $u \in L^\infty(X, \mathbb{R})$. We therefore consider this as a linear equation with L^2 coefficients in X with the nonlinear term on the right hand side as a known function in $L^2(X, \mathbb{C})$. It follows from the the regualrity theory for elliptic boundary value problems that $u \in H^2(X, \mathbb{C})$

Similarly, it follows from, applying the regularity theory for elliptic systems and using the regularity of u already obtained to consider the non linear term as a known vector field, to the following elliptic problem satisfied by the vector field A,

$$\begin{cases} \text{rot (rot)}A + \{\frac{i}{2}\kappa^2(u^*\nabla u - u\nabla u^*) + |u|^2 A\}\chi_A = 0 \\ \nu \wedge [\text{rot}A] = 0 \quad \text{on} \quad \partial X \end{cases}$$

we have $A \in H^2_{loc}(\mathbb{R}^2;\mathbb{R}^2)$

We have already remarked that due to the discontinuity of the vector field $g(x)$ across the boundary ∂X we are constrained to restrict the discussion of further regularity of the vector field A on $\overline{X}$ only from the interior of the domain X.

We flatten the boundary ∂X locally: if $x_0 \in \partial X$, then in some spherical neighbourhood $B(x_0,R)$ of $x_0 \in \partial X$ we may assume

$$X \cap B(x_0,R) \subset \{(x_1,x_2);|x_1|<l,\ x_2>0\}$$

and

$$\partial X \cap B(x_0,R) \subset \{(x_1,x_2);|x_1|<l,\ x_2=0\}.$$

We still denote by $g(x)$ the vector field

$$g(x) = \begin{cases} -\kappa^2 Im(u\nabla u^*) - |u|^2 A \quad \text{in} \quad B(x_0,R) \cap \{|x_1|<l,\ x_2>0\} \\ 0 \quad \text{in} \quad B(x_0,R) \cap \{|x_1|<l,\ x_2=0\} \end{cases}$$

Now take a test vector field $F \in H^1(\mathbb{R}^2,\mathbb{R}^2)$ with supp $F \subset B(x_0,R)$ and multiply the equation $-\Delta A = g$ by $\frac{\partial B}{\partial x_1}$ and integrating by parts we get

$$\int_{B(x_0,R)} <\nabla(\frac{\partial A}{\partial x_1}), \nabla F> \mathrm{d}x = \int_{B(x_0,R)} <\frac{\partial g}{\partial x_1}(x), F> \mathrm{d}x$$

that is,

$$-\Delta\left(\frac{\partial A}{\partial x_1}\right) = \frac{\partial g}{\partial x_1}, \text{weakly in } L^\infty(B(x_0,R);\mathbb{R}^2)$$

Here $\frac{\partial g}{\partial x_1} \in L^\infty(B(x_0,R);\mathbb{R}^2) \subset L^2(B(x_0,R);\mathbb{R}^2)$ and hence

$$\frac{\partial A}{\partial x_1} \in C^{1,\alpha}(\overline{X \cap B(x_0,R)};\mathbb{R}^2)$$

i.e. the tangential derivative of A along the boundary belongs to $C^{1,\alpha}$.

Then it follows that $g \in C^{1,\alpha}$ and A satisfies $-\Delta A = g$ in $X \cap B(x_0,R)$ and on the boundary $A \in C^{2,\alpha}(\partial X \cap B(x_0,R);\mathbb{R}^2)$. Again by the regularity of solutions of the Dirichlet problem for the elliptic system $-\Delta A = g$ we find that $A \in C^{2,\alpha}(\overline{X \cap B(x_0,R)};\mathbb{R}^2)$.

Next we differentiate the first equation for u in (GL) with respect to x_1 and also the corresponding boundary condition to obtain $u \in C^{3,\alpha}(\partial X \cap B(x_0,R);\mathbb{C})$ and once again the regularity theory gives $u \in C^{3,\alpha}(X \cap B(x_0,R);\mathbb{C})$

Similarly, differentiating the system $-\Delta A = g$ with respect x_1 and once again using the above argument we find that $A \in C^{3,\alpha}(\partial X \cap B(x_0,R);\mathbb{R}^2)$ and the regularity for elliptic systems shows

$A \in C^{3,\alpha}(X \cap B(x_0,R);\mathbb{R}^2)$, with $0 < \alpha < 1$

This completes the proof.

Remark. Taking into consideration of local flattenning of the boundary ∂X and the number of the local representations of ∂X we find that we require that ∂X should be of class $C^{5,\alpha}$, $(0 < \alpha < 1)$ in order to be able to apply the Schauder estimates for elliptic boundary value problems for elliptic systems due to Agmon, Douglis and Nirenberg.

Remark. In view of the above theorem we obtain the following important fact:

If $(u,A) \in E$ is a critical point of the functional $\mathcal{G}$ then the pair

$$\left(\frac{\partial u}{\partial x_j}, \frac{\partial A}{\partial x_j}\right) \in E, \text{for } j = 1,2$$

Infact, by the theorem we have $u \in C^{3,\alpha}(\overline{X},\mathbb{C})$ and $A \in C^{1,\alpha}_{loc}(\mathbb{R}^2,\mathbb{R}^2)$, which imply that

$$\left(\frac{\partial u}{\partial x_j}, \frac{\partial A}{\partial x_j}\right) \in C^{2,\alpha}(\overline{X},\mathbb{C}) \times C^{0,\alpha}_{loc}(\mathbb{R}^2,\mathbb{R}^2), \text{for } j = 1,2$$

Since $u \in H^2(X,\mathbb{C})$ we find $\frac{\partial u}{\partial x_j} \in H^1(X,\mathbb{C})$

We use a similar argument to show that A satisfies the required assumption. For this we first of all flatten the boundary locally: if $x_0 \in \partial X$ suppose that there is a ball $B(x_0,R)$ such that

$$\partial X \cap B(x_0,R) = \{(x_1,x_2) \in \mathbb{R}^2; |x_1| < l,\ x_2 = 0\}.$$

Define the vector field $g(x)$ in $B(x_0,R)$ by setting as before

$$g(x) = \begin{cases} -\kappa^2 Im(u\nabla u^*) - |u|^2 A & \text{in } B(x_0,R) \cap \{x_2 > 0\} \\ 0 & \text{in } B(x_0,R) \cap \{x_2 < 0\} \end{cases}$$

Hence, because $\frac{\partial g}{\partial x_1} \in L^\infty(B(x_0,\mathbb{R}^2)) \subset L^2(B(x_0,\mathbb{R}^2))$ it follows that $\frac{\partial A}{\partial x_1} \in C^{1,\alpha}(B(x_0,R);\mathbb{R}^2)$

Again using the elliptic system $-\Delta A = g$ we find $\frac{\partial A}{\partial x_j} \in C^{1,\alpha}(\overline{X};\mathbb{R}^2)$ so that $\frac{\partial A}{\partial x_j} \in H^1_{loc}(\mathbb{R}^2;\mathbb{R}^2)$ and moreover we have

$$\mathrm{rot}\left(\frac{\partial A}{\partial x_j}\right) = \begin{cases} \frac{\partial}{\partial x_j}\left(0,0,\frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2}\right) & \text{in } \overline{X} \\ 0 \quad \text{in } \mathbb{R}^2 \setminus \overline{X} \end{cases}$$

Hence $\frac{\partial A}{\partial x_j} \in H^1_{loc}(\mathbb{R}^2;\mathbb{R}^2)$ and $\mathrm{rot}\frac{\partial A}{\partial x_j} \in L^2(\mathbb{R}^2;\mathbb{R}^2)$

We also note that the trace on ∂X of $\mathrm{rot}(\frac{\partial A}{\partial x_j})$ from the interior of X is well defined by the elliptic regularity theory of Agmon, Douglis and Nirenberg.

3. CALCULATION OF $D^2\mathcal{G}\left(u,A;\frac{\partial u}{\partial x_j},\frac{\partial A}{\partial x_j}\right)$

We can now take $\left(\frac{\partial u}{\partial x_j},\frac{\partial A}{\partial x_j}\right) \in E$ to compute the second variation of the functional $\mathcal{G}$ at the critical point (u,A).

Differentiating the system (GL) with respect to x_1 and x_2 and then using the Green's formula we get

$$D^2\mathcal{G}\left(u,A;\frac{\partial u}{\partial x_j},\frac{\partial A}{\partial x_j}\right)$$

$$\int_{\partial X}\left\{\frac{1}{4}\frac{\partial}{\partial \nu}\left|\frac{\partial u}{\partial x_j}\right| + \frac{1}{2} < \nu \wedge \frac{\partial A}{\partial x_j}, \mathrm{rot}\frac{\partial A}{\partial x_j} > \right\} d\sigma$$

$$+\frac{i}{4}\int_X < \frac{\partial A}{\partial x_j}, \nu > \mathrm{div}\left(u^*\frac{\partial u}{\partial x_j} - u\frac{\partial u^*}{\partial x_j}\right) dx \text{ for } j = 1,2$$

Applying the divergence theorem to the last integral we obtain the following

Theorem 3. *Suppose $(u,A) \in E$ is a critical point of $\mathcal{G}$ and that the boundary ∂X is $C^{5,\alpha}$, $0 < \alpha < 1$. Then*

$$D^2\mathcal{G}(u,A;\frac{\partial u}{\partial x_j},\frac{\partial A}{\partial x_j})$$
$$=\int_{\partial X}\{\frac{1}{4}\frac{\partial}{\partial \nu}|\frac{\partial u}{\partial x_j}|^2+\frac{1}{2}<\nu\wedge\frac{\partial A}{\partial x_j},\mathrm{rot}\frac{\partial A}{\partial x_j}>$$
$$+\frac{i}{4}<\frac{\partial A}{\partial x_j},\nu>(u^*\frac{\partial u}{\partial x_j}-u\frac{\partial u^*}{\partial x_j})\}\mathrm{d}\sigma \text{ for } j=1,2$$

Thus the second variation of $\mathcal{G}$ is expressed as a sum of boundary integrals.

We deduce from this the following theorem expressing the second variation of $\mathcal{G}$ about a critical point (u,A) involving the mean curvature of the boundary ∂X.

Suppose $(u,A)\in E$ is a critical point of $\mathcal{G}$. We subdivide the boundary ∂X into two parts, namely $\partial X=\Sigma_1\cup\Sigma_2$ where

$$\Sigma_1=\{x\in\partial X;u(x)\neq 0\} \quad \text{and} \quad \Sigma_2=\{x\in\partial X;u(x)=0\}$$

Now if $x_0\in\Sigma_1$ then, in some contractible neighbourhood V of x_0, we can write

$$u(x)=w(x)\exp(i\varphi(x)) \quad \text{in } V\cap X$$

where φ and w are real valued functions ($w(x)\geq 0$) on $V\cap X$.

Theorem 4. *Assume ∂X is $C^{5,\alpha}$ $(0<\alpha<1)$ and let $K(x)$ denote the mean curvature of ∂X. If $(u,A)\in E$ is a critical point for $\mathcal{G}$ then we have*

$$\sum_{j=1}^{2} D^2\mathcal{G}(u,A;\frac{\partial u}{\partial x_j},\frac{\partial A}{\partial x_j})=$$
$$=-\frac{1}{2}\int_{\Sigma_1}\{||\nabla w||^2+w^2||\nabla\varphi-A||^2\}K(x)\mathrm{d}\sigma$$
$$-\frac{1}{2}\int_{\Sigma_2}||\nabla u||^2 K(x)\mathrm{d}\sigma$$

We evaluate the term (since $A(x)=(a_1(x),a_2(x))$)

$$Q=\sum_{j=1}^{2}<\nu\wedge\frac{\partial A}{\partial x_j},\mathrm{rot}\frac{\partial A}{\partial x_j}>=\sum_{l=1}^{2}\frac{\partial a_l}{\partial \nu}\Delta a_l$$

Infact, by explicite calculation we have $\mathrm{rot}(\mathrm{rot})A = -\Delta A$ since $\mathrm{div}A = 0$ in X. Then we get from the system (GL)

$$\Delta A = |u|^2 A + \frac{i}{2}(u^*\nabla u - u\nabla u^*) \quad \text{in} \quad X$$

and hence

$$Q = \sum_{l=1}^{2} \frac{\partial a_l}{\partial x_k}\Delta a_l = \sum_{l=1}^{2} \frac{\partial a_l}{\partial \nu}[|u|^2 a_l + \frac{i}{2}(u^*\nabla u - u\nabla u^*)]$$

Since $\mathrm{rot}A = 0$ on ∂X we find from the lemma 1 (section 2) that

$$\frac{\partial a_l}{\partial \nu} = \sum_k \frac{\partial a_l}{\partial x_k}\nu_k = \sum_k \frac{\partial a_k}{\partial x_l}\nu_k = <\frac{\partial A}{\partial x_l}, \nu>$$

We then obtain from theorem 3, (summing over $j = 1,2$)

$$\begin{aligned}
&\sum_{j=1}^{2} (D^2\mathcal{G})(u, A; \frac{\partial u}{\partial x_j}, \frac{\partial A}{\partial x_j}) \\
&= \frac{1}{4}\int_{\partial X} \frac{\partial}{\partial \nu} |\nabla u|^2 \, \mathrm{d}\sigma \\
&+ \frac{1}{4}\int_{\partial X} |u|^2 \frac{\partial}{\partial \nu} ||A||^2 \, \mathrm{d}\sigma \\
&+ \frac{i}{4}\int_{\partial X} \sum_{j=1}^{2} (u^* \frac{\partial u}{\partial x_j} - u\frac{\partial u^*}{\partial x_j})[\frac{\partial a_j}{\partial \nu} + < \frac{\partial A}{\partial x_j}, \nu >] \mathrm{d}\sigma
\end{aligned}$$

We calculate the integrand here seperately on the two subsets Σ_1 and Σ_2 seperately:

Let $x_0 \in \Sigma_1$ and V be a contractible neighbourhood of x_0 where we write as before $u(x) = w(x)\mathrm{e}^{i\varphi(x)}$. Then, we have, on $V \cap X$,

$$||\nabla u(x)||^2 = ||\nabla w(x)||^2 + w(x)^2 ||\nabla \varphi||^2$$

and

$$u^* \frac{\partial u}{\partial x_j} - u\frac{\partial u^*}{\partial x_j} = 2iw(x)^2 \frac{\partial \varphi}{\partial x_j}$$

Moreover u and φ satisfy the Neuman condition on $V \cap \Sigma_1$.

We can flatten $V \cap \Sigma_1$ locally and take $V \cap \Sigma_1 = \{(x_1, x_2); x_2 = 0\}$ and assume $\nu(x_0) = (0,1)$ and since $< A, \nu >= 0$ we find

$$\frac{\partial w}{\partial x_2}(x_0) = \frac{\partial \varphi}{\partial x_2}(x_0) = a_2(x_0) = 0$$

$$\frac{\partial \nu^{(1)}}{\partial x_1}(x_0) = K(x)$$

After some elementary calculations together with these conditions the integrand becomes

$$-2K(x)[\| \nabla w \|^2 + w^2 \| \nabla \varphi - A \|^2] \quad \text{in} \quad V \cap \Sigma_1$$

and on Σ_2 where $u(x) = 0$, the integrand becomes

$$\frac{\partial}{\partial \nu} \| \nabla u \|^2 = -2K(x) \| \nabla u \|^2$$

The assertion follows from these considerations.

4. THE MAIN RESULT

We obtain from the crucial identity proved in the previous paragraph our main result:

Theorem 5. *Suppose X is a bounded convex open set in $\mathbb{R}^2$ with the boundary ∂X of class $C^{5,\alpha}$ for some $0 < \alpha < 1$. If (u, A) is a (non vortex) stable critical point for the Ginzburg - Landau energy functional then (u, A) is gauge equivalent to the trivial one* $(1,))$.

i.e. there does not exist any non trivial non vortex stable critical points for the Ginzburg - Landau functional.

A sketch of the proof

By definition, if (u, A) is a stable critical point then the second variation gives

$$(D^2\mathcal{G})(u,A;\frac{\partial u}{\partial x_j},\frac{\partial A}{\partial x_j}) \geq 0$$

On the other hand, if $K(x) \geq 0$ then the expression in terms of the boundary integrals given by the theorem 4 for the second variation

$$(D^2\mathcal{G})(u,A;\frac{\partial u}{\partial x_j},\frac{\partial A}{\partial x_j}) \leq 0$$

and hence we find that

$$(D^2\mathcal{G})(u,A;\frac{\partial u}{\partial x_j},\frac{\partial A}{\partial x_j}) = 0$$

This means that the pair $\left(\frac{\partial u}{\partial x_j},\frac{\partial A}{\partial x_j}\right)$ is a minimizer for the second variation $(D\mathcal{G})$ for both $j = 1,2$. In particular, each is a critical point of $(D\mathcal{G})$ and hence satisfies the associated Euler - Lagrange equation and the natural boundary condition, namely,

$$< \nabla(\frac{\partial u}{\partial x_j}) - iu\frac{\partial A}{\partial x_j}, \nu >= 0 \quad \text{on} \quad \partial X \quad \text{for} \quad j = 1,2$$

Consider the subset Λ of ∂X :

$$\Lambda = \{x \in \partial X; K(x) > 0\}$$

Once again $(D^2\mathcal{G}) = 0$ and $K(x) > 0$ together imply, in view of the expression given by theorem 4, that, writing as before

$$u(x) = w(x)exp(i\varphi(x)) \text{ with } w(x) > 0 \text{ and } \varphi(x) \in \mathbb{R} \text{ on } \Lambda$$
$$\| \nabla u - iAu \|^2 = \| \nabla w \|^2 + w^2 \| \nabla \varphi - A \|^2 = 0 \quad \text{and} \quad \| \nabla u \|^2 = 0$$

In particular we have

$$\nabla u - iAu = 0 \text{ and } \| \nabla w \| = 0, \text{ that is } \nabla w = 0.$$

Decomposing further $\Lambda = \Lambda_1 \cup \Lambda_2$ where

$$\Lambda_1 = \{x \in \Lambda; u(x) \neq 0\} = \{x \in \partial X; K(x) > 0, u(x) \neq 0\}$$

and

$$\Lambda_2 = \{x \in \Lambda; u(x) = 0\} = \{x \in \partial X; K(x) > 0, u(x) = 0\}$$

we consider the two subsets Λ_1 and Λ_2 seperately.

Suppose $\Lambda_1 \neq \varnothing$ and let $x_0 \in \Lambda_1$. Writing again $u(x) = w(x)exp(i\varphi(x)$ in $\overline{X} \cap B(x_0, \varepsilon)$ with w and φ real valued, we find in view of lemma 1 (section 2) that $|u(x)| \leq 1$ and hence $|w(x)| \leq 1$ in $\overline{X} \cap B(x_0, \varepsilon)$. Since $\nabla w = 0$ on Λ_1 we have $w(x) = c$, a constant with $0 < c \leq 1$ in $\partial X \cap B(x_0, \varepsilon)$.

Next we claim that

$$\frac{\partial^2 w}{\partial x_j \partial x_k} = 0, \quad \text{on} \quad \partial X \cap B(x_0, \varepsilon), \forall j, k = 1, 2 .$$

For this, we calculate the normal and tangential derivatives:

$$< \nabla(\tfrac{\partial u}{\partial x_j}) - iu\tfrac{\partial A}{\partial x_j}, \nu >= 0 \text{ implies that } \begin{matrix} \sum \frac{\partial^2 w}{\partial x_j \partial x_k} \nu_k = 0, \\ \text{on} \quad \partial X \cap B(x_0, \varepsilon) \end{matrix} \quad \text{for } j = 1, 2 .$$

Differentiating $\nabla u - iAu = 0$ along the tangential direction $\tau = (\tau_1, \tau_2) = (-\nu_2, \nu_1)$ we find that

$$\sum \frac{\partial^2 w}{\partial x_j \partial x_k} \tau_k = 0, \quad \text{on} \quad \partial X \cap B(x_0, \varepsilon), \forall j = 1, 2 .$$

This proves our claim.

Now the Ginzburg - Landau system (GL)

$$(\nabla - iA)^2 u + \kappa^2 (|u|^2 - 1)u = 0 \quad \text{in} \quad X \cap B(x_0, \varepsilon)$$

and the boundary condition

$$< (\nabla - iA)u, \nu >= 0 \quad \text{on} \quad \partial X \cap B(x_0, \varepsilon)$$

together lead to the second order nonlinear elliptic equation for w with the Neumann boundary condition:

$$\begin{cases} \Delta w - \| \nabla\varphi - A \|^2 \, w + \kappa^2 w(1-w^2) = 0 \quad \text{in} \quad X \cap B(x_0, \varepsilon) \\ \frac{\partial w}{\partial \nu} = 0 \quad \text{on} \quad \partial X \cap B(x_0, \varepsilon) \end{cases}$$

By the elliptic regularity theorem on Schauder estimates of Agmon, Douglis and Nirenberg, it follows that $u \in C^2$ upto the boundary.

We next show that $w(x) = 1$ on $\Lambda_1 \cap B(x_0, \varepsilon)$. For this, we observe that letting $b(x) = -\kappa^2 w(1+w)$ we have the inequality

$$\Delta(1-w) + b(x)(1-w) = -\left\| \nabla\varphi - A \right\|^2 \le 0 \quad \text{in} \quad X \cap B(x_0, \varepsilon)$$

where since $0 < w(x) \le 1$ we have $1 - w \ge 0$. Now by the strong maximum principle applied to the non negative function $1-w$ on the smooth bounded open set $X \cap B(x_0, \varepsilon)$ we conclude that either $1 - w(x) = 0$ or $1 - w(x) > 0$ everywhere in $X \cap B(x_0, \varepsilon)$.

But, in view of the Neumann boundary condition $\frac{\partial w}{\partial \nu} = 0$ on $\partial X \cap B(x_0, \varepsilon)$, we can apply the maximum principle of Hopf and exclude the possibility that $1 - w(x) > 0$.

Hence $w(x) = 1$ in $X \cap B(x_0, \varepsilon)$ (and so $|u(x)|^2 = 1$).

Then the system of equations (GL) implies that we have

$$\| \nabla\varphi - A \| = 0 \quad \text{i.e.} \quad \nabla\varphi = A \quad \text{on} \quad \overline{X} \cap B(x_0, \varepsilon)$$

Since X is simply connected these relations extend to the whole of $\overline{X}$ by continuation. We have thus proved that

$$(u, A) = (e^{i\varphi}, \nabla\varphi) \quad \text{in} \quad X$$

that is, (u, A) is gauge equivalent to $(1, 0)$.

There remains to consider the case wherein $\Lambda_1 = \varnothing$ and hence $\Lambda = \Lambda_2 = \{x \in \partial X; \kappa(x) > 0, u(x) = 0\}$. In this case we write $u(x) = f(x) + ig(x)$ and we find that the real valued functions f and g satisfy near a point $x_0 \in \Lambda_2$ the elliptic system of equations

$$\begin{cases} \Delta f + 2 < A, \nabla g > - \| A \|^2 \, f + \kappa^2 (f^2 + g^2 - 1) f = 0 \\ \Delta g - 2 < A, \nabla f > - \| A \|^2 \, g + \kappa^2 (f^2 + g^2 - 1) g = 0 \\ \text{in} \quad X \cap B(x_0, \varepsilon) \end{cases}$$

(with some $\varepsilon > 0$) and the initial conditions

$$f = g = 0, < \nabla f, v >= \frac{\partial f}{\partial v} = 0, < \nabla g, v >= \frac{\partial g}{\partial v} = 0 \text{ on } \partial X \cap B(x_0, \varepsilon)$$

Using the theorem on uniqueness in the Cauchy problem due to Calderon [7] we find that $f(x) = g(x) = 0$, that is, $u(x) = 0$ in $\overline{X} \cap B(x_0, \varepsilon)$. Once again using the assumption that X is simply connected we find that $u = 0$ in X.

Finally using the system (GL), we have

$$\text{rot}(\text{rot}A) = 0 \quad \text{in} \quad \mathbb{R}^2$$

which implies, by the lemma, that $\text{rot}A = 0$ in $\mathbb{R}^2$. Hence A is of the form $A = \nabla h$ for a smooth real valued function $h : \mathbb{R}^2 \to \mathbb{R}$; that is $(u, A) = (0, \nabla h)$, which means that (u, A) is gauge equivalent to $(0,0)$. On the other hand $(0,0)$ is not a stable critical point for $\mathcal{G}$, which contradicts the assumption that (u, A) is a stable critic point. This completes the proof of the main theorem.

REFERENCES

[1] S. Agmon, A. Douglis, L. Nirenberg Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions I and II, Comm. Pure Appl. Math. 12 (1959), 623-727 and Comm. Pure Appl. Math. 17 (1964), 35 - 92

[2] F. Bethuel, Aproximation problem for Sobolev maps between two manifolds, Acta Math., 167 (1991), 153 - 206

[3] F. Bethuel, Some recent results for the Ginzburg Landau equations, Progr. Math., 168 (1998), Birkhauser, Basel

[4] F. Bethuel, H. Brezis, G. Orlandi, Asymototics for Ginzburg Landau equations on arbitrary domains, J. Funct. Anal., 186 (2001), 432 - 520

[5] F. Bethuel, H. Brezis, G. Orlandi, Small energy solutions of Ginzburg Landau equations, C.R. Acad. Sci., Paris, (2000)

[6] F. Bethuel, X. Zhang, Density of smooth maps between two manifolds in Sobolev spaces, Jr. Func. Anal. 80 (1988), 60-75

[7] A.P. Calderon, Uniqueness in the Cauchy problem for partial differential equations, Amer. J. of Math. 80 (1958), 16 - 36.

[8] V.L. Ginzburg, L.D. Landau, On the theory of super-condutivity J.E.T.P., 20 (1950)

[9] S. Jimbo, Y. Morita, Ginzburg - Landau equation and stable solutions in irrotational domain, SIAM Jr. Math.Anal. 27 (1996), 1360 - 1385

[10] S. Jimbo, J.Zhai, Ginzburg - Landau equation with magnetic effect: Non-simply-connected domains Jr. Math. Soc. Japan, 50 (1998), 663 - 684

[11] S. Jimbo, P. Sternberg, Non existence of permanent currents in convex planar samples, SIAM J. Math. Anal., 33 (2002), 1379 - 1392

[12] H. Matano, Asymptotic behavior and stability of solutions to semilinear diffusion equations, Pub. Res. Inst. Math. Sci. RIMS, Kyoto Univ. 15,(1979), 401 - 454

[13] J. Rubinstein, P.Sternberg, Homotopy classification of minimizers for Ginzburg Landau energy and existence of permanent currents, Comm. Math. Phys., 179 (1996), 257 - 263

[14] G. Stampacchia, Le problème de Dirichlet pour les equations elliptiques du second ordre à coefficients discontnus, Ann. Inst. Fourier, Grenoble, 15 (1965), 189-258

[15] G. Stampacchia, Équations elliptiques du second ordre a coefficients discontinus, Séminaires Mathématiques Supérieurs, Le Presses de l'Université de Montreal, Que. 16 (1966)

THE DISTANCE FUNCTION TO THE BOUNDARY AND SINGULAR SET OF VISCOSITY SOLUTIONS OF HAMILTON-JACOBI EQUATION

L. Nirenberg
Courant Institute, New York, New York, USA

This is a report on joint work with YanYan Li [4], concerning viscosity solutions of Hamilton-Jacobi (HJ) equations of the form

$$H(x,u,\nabla u)=1 \quad \text{in}\Omega, \tag{1}$$

a $C^{2,1}$ bounded domain in $\mathbb{R}^n$. One usually treats an initial value problem for u or a boundary value problem. We consider the latter, and seek positive solution u satisfying

$$u=0 \quad \text{on } \partial\Omega. \tag{2}$$

For definitions and properties of viscosity solutions see [5] and [1].

Near the boundary $\partial\Omega$, one may determine u using the method of characteristics, but these may then collide, and solutions develop singularities. Under rather standard conditions on $H(x,t,p)$ for $x\in\Omega$, $t\in\mathbb{R}$, $p\in\mathbb{R}^n$, such as convexity in p etc., one expects that the $(n-1)$-dimensional Hausdorff measure

$$H^{n-1}(\Sigma)$$

of the singular set Σ of a viscosity solution u is finite. We prove this for $H(x,p)$ independent of t, — under suitable conditions, and also treat a number of cases where H also depends on t.

We came to this problem by first studying the singular set Σ of the distance function $u(x)$ from x in Ω to $\partial\Omega$. It satisfies

$$|\nabla u| = 1 \quad \text{in } \Omega$$

$$u = 0 \quad \text{on } \partial\Omega.$$

These equations have, of course, many solutions. For instance, if Ω is an interval $(-a,a)$ in $\mathbb{R}$, then any jagged line with slopes ± 1, which vanishes at end points, is a solution. But the distance function, $u = a - |x|$, is the largest; it is the unique viscosity solution.

The singular set Σ of u is sometimes called the ridge, medial axis, or skeleton of Ω. We define Σ in the following way. Let G be the largest open subset of Ω such that every x in G has a unique closest point on $\partial\Omega$. We set

$$\Sigma := \Omega \setminus G;$$

so Σ is closed. It is easily seen that in G, the distance function u to the boundary is smooth—as smooth as the boundary permits ($C^{1,1}$ in our case, or C^∞ if $\partial\Omega$ is C^∞). It is well known that Σ is connected. We proved

Theorem 1. *$H^{n-1}(\Sigma)$ is finite.*

This follows directly from the following result

Theorem A *From any point y on $\partial\Omega$, go along the inner normal to $\partial\Omega$ until first hitting a point $m(y)$ on Σ. The length $\overline{s}(y)$ of the resulting segment is Lipschitz continuous in y.*

Remark *1 For Theorem A to hold the condition that $\partial\Omega \in C^{2,1}$ is sharp. This surprised us.*

For an unbounded domain Ω, with $\partial\Omega$ in $C^{2,1}_{loc}$, and G and Σ defined as above, the following form of Theorem A holds.

Theorem A' *For $y \in \partial\Omega$, let $\overline{s}(y)$ be defined as in Theorem A: it may be infinite. For any $N > 0$, $\min(N, \overline{s}(y))$ is locally Lipschitz in y.*

We then extended these results to any complete Riemannian manifolds (M^n, g):

Theorem A″ *For any domain* Ω *in* M, *with* $\partial\Omega$ *in* $C^{2,1}$, *the conclusion of Theorem A′ holds. Here* $\bar{s}(y)$ *represents the length of the geodesic going from* y *to* $\partial\Omega$, *normal to* $\partial\Omega$ *there, until it first hits* Σ.

Corollary 1 *For* Ω *as above in* (M^n, g), $H^{n-1}(\Sigma \cap B) < \infty$ *for any bounded set* B.

Li and I then discovered that Theorem A″ had been proved in 2001 by J.I. Itoh and M. Tanaka [3].

Cut point. In Theorem A″ we considered a geodesic from a point y on $\partial\Omega$ going into Ω (in a normal direction) until it first hits Σ at some point $x = m(y)$. The point x is called the cut point of y because if we go beyond it on the geodesic to any point x' then x' has a closer point on $\partial\Omega$ than y (this is not difficult to see). Thus Σ is the cut locus of $\partial\Omega$.

Walter Craig suggested to us that we try to extend Theorem A″ to HJ equations. From now on we consider viscosity solutions u of (1), (2). First, we wish to stress that the function H does not matter much. What really matters are the sets where $H(x,t,p) = 1$. For every x in $\overline{\Omega}$ we define

$$V_x = \{(t,p) \mid H(x,t,p) < 1\} \tag{3}$$

and

$$S_x = \{(t,p) \mid H(x,t,p) = 1\}. \tag{4}$$

In treating the problem we are free to change the function H provided the sets V_x are preserved. It is easy to verify that a viscosity solution for such a changed H is also a viscosity solution of the original H.

At this point we may formulate a general conjecture. We assume the following

(a) For every x in $\overline{\Omega}$, V_x is a convex set lying in a fixed downward cone

$$K = \{(t,p) \mid \ |p| \le k(C_1 - t), t < C_1\}, \quad k, C_1 > 0, \tag{5}$$

and for some fixed $r > 0$, the ball

$$B_r(0) = \{(t,p) \mid t^2 + |p|^2 < r^2\} \tag{6}$$

lies in V_x.

(b) For $t \ge -1$ assume that the S_x are smooth and have positive

principal curvatures bounded away from zero.

Conjecture 1 *Under conditions (a), (b), any viscosity solution of (1), (2) is in* $C^{1,1}$ *(or smoother if* $\partial\Omega$ *is smoother) on an open set* $G \subset \Omega$, *and, for* $\Sigma = \Omega \setminus G$,

$$H^{n-1}(\Sigma) < \infty.$$

We have proved this under various additional conditions. These results are derived using our principal result, Theorem B; it concerns viscosity solution u for $H = H(x,p)$, independent of t,

$$H(x, \nabla u) = 1. \tag{7}$$

Here we assume that $\forall x$ in $\overline{\Omega}$, the set

$$V_x = \{p \in \mathbb{R}^n \mid H(x,p) < 1\}$$

is a bounded closed convex set with smooth strictly convex boundary S_x, i.e. its principal curvatures are all positive — uniformly in x. In addition we assume that $\forall x$, the ball

$$B_r(0) = \{|\, p \,| < r\}, \quad 0 < r \text{ fixed},$$

lies in V_x.

Under the conditions above there exists a viscosity solution (see Theorem 5.3 in [5]), and the Conjecture holds for it:

Theorem B *The singular set* Σ *of the viscosity solution satisfies*

$$H^{n-1}(\Sigma) < \infty.$$

Theorem B is proved using the explicit formula for the viscosity solution given in Theorem 5.3 in [5]. It involves the support functions of the convex sets V_x. For fixed $x \in \Omega$, the support function of the set $\{p \mid H(x,p) = 1\}$ is defined, for $v \in \mathbb{R}^n$ by

$$\varphi(x; v) = \sup_{H(x,p)=1} v \cdot p. \tag{8}$$

Properties. φ is convex in v, positive homogeneous of degree 1 in v, and smooth in (x,v) for $v \neq 0$. Furthermore, φ satisfies the triangle inequality in v, and $\forall x \in \Omega$,

$$\{v \mid \varphi(x;v) = 1\}$$

is strictly convex with positive principal curvatures — uniformly in x.

For curve $\xi(t)$, $0 \leq t \leq T$ lying in Ω

$$\varphi(\xi(t);\dot{\xi}(t))dt$$

is a Finsler metric and the *viscosity solution* $u(x)$ *is given by the shortest distance from* $\partial\Omega$ *to* x *in this metric*, i.e.

$$u(x) = \inf_{y \in \partial\Omega} \inf \left\{ \int_0^T \varphi(\xi,\dot{\xi})dt \mid \xi(0) = y, \xi(T) = x, \xi(t) \in \bar{\Omega} \right\}. \tag{9}$$

Note. Since $\varphi(\xi;v)$ may not be symmetric in v, the length of a curve $\xi(t)$ depends on the direction it is transversed.

Thus the solution is given by a distance function — but in a Finsler metric — and we extend Theorem A′ to this situation. As before we set G = largest open subset of Ω such that for any x in G, there is a unique point y on $\partial\Omega$ which is closest *to* x in the metric. In G, u is smooth. We consider the singular set

$$\Sigma = \Omega \setminus G,$$

and for any $y \in \partial\Omega$ we consider the geodesic of the metric, going into Ω, "normally at y" until it hits a first point $m(y)$ of Σ.

Theorem A‴. *The length* $\bar{s}(y)$ *of the geodesic to* $m(y)$ *is locally Lipschitz continuous in* y.

The condition that the geodesic be "normal" at y is simply that for x lying on the geodesic, close to y, y is the closest point from $\partial\Omega$ to x.

Theorem A‴ implies Theorem B.

We do not give here our proof of Theorem A‴. It is not very simple — even in the case of Theorem A. It is of greater interest, I think, to describe our attempts to attack the general H depending also on t. But first we should mention that there are a number of papers treating the singular set Σ and the map $y \to m(y)$. References may be found in [4]. Here we call attention only the paper [6] of A.C. Mennucci in which it is shown that for viscosity solutions of (1), (2), with H independent of t, the singular set Σ

is a countable union of smooth $(n-1)$-dimensional hypersurfaces and a set having zero $(n-1)$-dimensional Hausdorff measure.See also C. Mantegazza and A.C. Mennucci [7]. In A. Cellina and S. Perrotta [2], the map $y \to m(y)$ enters.

Let us now consider the general problem (1), (2). What we try to do is to reduce it to a problem with a new H which is independent of t. We consider the situation described earlier, with conditions (a), (b).

We now make use of the fact that it is only the V_x that count, not the function H. But first, since we seek positive solutions, we alter the V_x below $t = -\frac{1}{2}$ by cutting it off and smoothing it out so that S_x is still uniformly convex, and lies in $t \geq -\frac{3}{4}$. Next, keeping the new V_x fixed, we change H by requiring that it is positive homogeneous of degree one in (t, p). For the new H, the equation still takes the form

$$H(x,u,\nabla u) = 1. \tag{10}$$

Now comes the main trick: We introduce a new independent variable $\tau \in \mathbb{R}$, and set

$$z(\tau, x) = e^{\tau} u(x).$$

Multiplying (10) by $e^{-\tau} \cdot e^{\tau}$ and using the homogeniety we obtain the following equation for z

$$e^{-\tau} H(x, z_{\tau}, \nabla_x z) = 1.$$

This is an equation of the form (7), but in a cylinder $Z := \mathbb{R} \times \Omega$. The boundary condition is

$$z = 0 \quad \text{on } \partial Z.$$

The formula (9), with the suitable support function gives a viscosity solution z. Furthermore, by our construction,

$$z(\tau, x) = e^{\tau} z(0, x),$$

and our desired solution is $u = z(0,x)$. The singular set $\tilde{\Sigma}$ of the solution z, say for $|\tau| \leq 1$ consists of vertical segments over the singular set Σ of u. Thus if $H^n\left(\tilde{\Sigma} \cap \{|\tau| \leq 1\}\right) < \infty$ it would follow that $H^{n-1}(\Sigma) < \infty$.

Thus we would like to apply Theorem A''' to the problem in the cylinder Z. This is, indeed, posible if the (reverse) geodesics from points $(0,x)$, $x \in \Omega$ all remain bounded in the τ-direction. Then we can cut down the cylinder to make it finite, and round it off. But this need not be the case — and when it is not, our method does not work.

Here are a few cases where the geodesics are bounded and thus for which the conjecture is true: In these, $h(x,p)$ is assumed to satisfy the conditions of Theorem A'''.

(i) There exists $\lambda_0 > 0$ depending on h and on Ω such that for any $0 < \lambda < \lambda_0$ for

$$H(x,t,p) = \lambda t + h(x,p),$$

the conjecture holds.

(ii) The same is true for

$$H(x,t,p) = \lambda t^2 + h(x,p), \quad 0 < \lambda < \lambda_0.$$

(iii) For H satisfying the earlier conditions (a), (b), the conjecture holds for narrow domains, i.e. there exists a number $d_0 > 0$ depending on H, such that if Ω' is a bounded subdomain of Ω, with $\partial\Omega' \in C^{2,1}$, and such that the Euclidean distance of any point x in Ω' to $\partial\Omega'$ is less than d_0 then the conjecture holds for Ω'.

(iv) Suppose H is independent of x,

$$H = H(t,p)$$

satisfying the condition above: $B_r(0) \subset V \subset K$. Let $\bar{t}$ be the positive number satisfying

$$H(\bar{t},0) = 1.$$

The conjecture holds in case

$$\bar{t} < \max_{H(t,p)=1} t =: \hat{t}.$$

In general, $\bar{t} \leq \hat{t}$. In case of equality our method of proof must fail. In fact, if

$$H(t,p)=(t^2+|p|^2)^{\frac{1}{2}}$$

the corresponding Finsler metric is Riemannian. But in case $n=1$ and $\Omega=(-R,R)$, for $R>\pi$, there is no (reverse) geodesic starting at $(0,0)$ going to the boundary of the strip $\mathbb{R}\times\Omega$. Neverthless for this H and Ω bounded, the function

$$u(x)=\begin{cases}1 & \text{if } d(x)\geq\dfrac{\pi}{2},\\ \sin(d(x)) & \text{if } d(x)\leq\dfrac{\pi}{2},\end{cases}$$

here $d(x)=$Euclidean $dist(x,\partial\Omega)$, is a viscosity solution and for its singular set Σ, $H^{n-1}(\Sigma)<\infty$.

REFERENCES

[1] M. Bardi and I. Capuzzo-Dolcetta, Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations, Birkhduser Boston, Inc., Boston, MA, 1997.

[2] A. Cellina and S. Perrotta, On the validity of the maximum principle and of the Euler-Lagrange equation for a minimum problem depending on the gradient, SIAM J. Control Optim. 36 (1998), 1987–1998.

[3] J.I. Itoh and M. Tanaka, The Lipschitz continuity of the distance function to the cut locus, Trans. Amer. Math. Soc. 353 (2001), 21–40.

[4] Y.Y. Li and L. Nirenberg, The distance function to the boundary, Finsler geometry and the singular set of viscosity solutions of some Hamilton-Jacobi equations, Comm. Pure Appl. Math., to appear.

[5] P.L. Lions, Generalized solutions of Hamilton-Jacobi equations, Research Notes in Mathematics; 69, Pitman (Advanced Publishing Program), Boston, Mass.-London, 1982.

[6] C. Mantegazza and A. C. Mennucci, Hamilton-Jacobi equations and distance functions on Riemannian manifolds, Applied Math. and Optim., 2002. DOI 10.1007/s00245-002-0736-4.

[7] A. C. Mennucci, Regularity and variationality of solutions to Hamilton-Jacobi equations. part I: regularity, preprint.

L^P-REGULARITY FOR POINCARÉ PROBLEM AND APPLICATIONS*

Dian K. Palagachev
Dept. of Mathematics, Technical University of Bari, Bari,Italy

Abstract: We improve the results from [10] on strong solvability and uniqueness for the oblique derivative problem

$$\begin{cases} a^{ij}(x)D_{ij}u + b^i(x)D_iu + c(x)u = f(x) & \text{a.a. in } \Omega, \\ \partial u/\partial \ell + \sigma(x)u = \varphi(x) & \text{on } \partial\Omega, \end{cases}$$

extending them to Sobolev's space $W^{2,p}(\Omega)$, for *any* $p > 1$. The vector field $\ell(x)$ tangent to $\partial\Omega$ at the points of $\mathcal{E} \subset \partial\Omega$ and directed outwards Ω on $\partial\Omega \setminus \mathcal{E}$.

Key words: Uniformly elliptic operator, Poincaré problem, Strong solutions

INTRODUCTION

The article deals with strong solvability in Sobolev spaces $W^{2,p}(\Omega)$, $\forall p \in (1,\infty)$, of the oblique derivative problem

* This is the definitive version of a lecture delivered at the *38th Workshop: "Variational Analysis and Applications", Erice, Sicily, 20 June-1 July, 2003.*

$$\begin{cases} \mathcal{L}u := a^{ij}(x)D_{ij}u + b^{i}(x)D_{i}u + c(x)u = f(x) \ \text{ a.e.}\Omega, \\ \mathcal{B}u := \partial u/\partial \ell + \sigma(x)u = \varphi(x) \qquad \text{on } \partial\Omega, \end{cases} \tag{$\mathcal{P}$}$$

where $\mathcal{L}$ is a uniformly elliptic operator with low regular coefficients and $\mathcal{B}$ is prescribed in terms of directional derivative with respect to a unit vector field $\ell(x) = (\ell_1(x), \ldots, \ell_n(x))$ defined on $\partial\Omega$. Precisely, we are interested in the Poincaré problem $(\mathcal{P})$, that is, a situation when $\ell(x)$ becomes tangential to $\partial\Omega$ at the points of a non-empty subset $\mathcal{E}$ of $\partial\Omega$. This way, $(\mathcal{P})$ is a *degenerate* oblique derivative problem because the Shapiro-Lopatinskii complementary condition is violated on $\mathcal{E}$ (cf.[8, 3, 11, 12, 17, 2, 5, 6, 7, 15, 13].

It is worth noting that $(\mathcal{P})$ arises naturally in problems of determining gravitational fields of celestial bodies. In fact, it was Poincaré the first to arrive at a problem of that type in his studies on tides([14]). The theory of stochastic processes is another area where $(\mathcal{P})$ models real phenomena. Now $\mathcal{L}$ describes analytically a strong Markov process with continuous paths in Ω (such as Bronian motion), while $\partial u/\partial \ell$ corresponds to reflection along ℓ on $\partial\Omega \setminus \mathcal{E}$ and to diffusion at the points of $\mathcal{E}$, and σu describes absorption phenomena.

The general qualitative properties of $(\mathcal{P})$ depend strongly on the behaviour of ℓ neat the tangency set $\mathcal{E}$. Let $\gamma(x)$ be the scalar product of $\ell(x)$ and the outward normal $\nu(x)$ to $\partial\Omega$. Depending on the way $\gamma(x)$ changes or no its sign on the trajectories of ℓ when these cross $\mathcal{E}$, $(\mathcal{P})$ may have either a kernel or a co-kernel of infinite dimension (see[8,3,11]). We are dealing here with the simplest case when γ preserves the sign on $\partial\Omega$ which means ℓ is either tangent to $\partial\Omega$ or directed outwards Ω. It means ℓ is of *neutral type* and, at least in the case of C^{∞} data, $(\mathcal{P})$ is of *quasi-Fredholm* type. In other words, $(\mathcal{P})$ has zero index, but the solution "loses" regularity from the data near the set $\mathcal{E}$, whence $(\mathcal{P})$ is a problem of *sub-elliptic* type. That loss of smoothness has been measured in terms of *order of contact* between ℓ and $\partial\Omega$ in case $\mathcal{E}$ is a sub-manifold of $\partial\Omega$, of co-dimension one with ℓ transversal to $\mathcal{E}$ (cf.[2,5,6,7,]). We deal here with the general situation when $\mathcal{E}$ can be a subset of $\partial\Omega$, of positive surface measure subject to a kind of *non-trapping* condition that all trajectories of ℓ through points of $\mathcal{E}$ leave $\mathcal{E}$ in a finite time.

The problem $(\mathcal{P})$ (with zero lower order terms in $\mathcal{L}$) has been studied in $W^{2,p}$-framework in case of low regular coefficients(see[10]). Indeed, the loss of smoothness already mentioned, imposes some more regularity of the data near the set $\mathcal{E}$. The approach used in [10] is based on elliptic regularization of $(\mathcal{P})$. That is, perturbing the boundary condition to

$\partial u/\partial(\ell+\varepsilon\nu)+\sigma u=\varphi$ one gets a regular oblique derivative problem for any $\varepsilon>0$, which admits a unique strong solution $u_\varepsilon \in W^{2,p}(\Omega)$. The non-trapping condition ensures the possibility to estimate $\| u_\varepsilon \|_{W^{2,p}(\Omega)}$ in terms of $\| u_\varepsilon \|_{L^\infty(\Omega)}$ *independently* of ε and therefore letting $\varepsilon \to 0$ would give a solution of $(\mathcal{P})$ once one disposes of a uniform estimate $\| u_\varepsilon \|_{L^\infty(\Omega)} \le C$. This last bound can be easily obtained from a variant of Aleksandrov-Bakelman-Pucci (ABP) maximum principle when $p>n$, and this naturally restricts solvability and uniqueness of $(\mathcal{P})$ ([10, Theorems 1.1,1.2]) to $W^{2,p}(\Omega)$ with $p>n$.

Our main purpose here is to improve the existence and solvability results from [10] extending them to $W^{2,p}(\Omega)$ for *any* $p\in(1,\infty)$. For this goal, we employ an approach completely different from that already used in [10], which fits better in the $W^{2,p}$-framework than elliptic regularization. In contrast to the *a' posteriori* estimate (see (1.5)) in [10]), we derive here an *a' priori* estimate (Theorem 4) for any $W^{2,p}(\Omega)$ solution to $(\mathcal{P})\forall p\in(1,\infty)$. To get uniqueness of solutions in $W^{2,p}(\Omega)$ for any $p>1$, consider a strong solution u to the homogeneous problem $(\mathcal{P})$ (i.e., $f\equiv 0, \varphi\equiv 0$). Now, if $(\mathcal{P})$ was a regular problem with smooth coefficients, the regularity of the right-hand sides would increase regularity of u at a level to be able to apply ABP's maximum principle. Unfortunately, this is not our case due to the above mentioned loss of smoothness near $\mathcal{E}$ and low regularity of the coefficients. However, even if $(\mathcal{P})$ is *degenerate* problem it behaves like an *elliptic* one for what concerns the degree of integrability p. It means the second derivatives of u have the same rate of integrability like f and φ (Proposition 8). For the solution u of the homogeneous problem this automatically implies $u\in W^{2,q}(\Omega)$ for any $q>1$ and this suffices to get $u=0$ through ABP. With the a'priori estimate and unicity at hand, it remains to apply Riesz-Schauder's theory in order to get strong solvability in $W^{2,p}(\Omega)$ for any $p>1$.

To complete this introduction it should be noted that, for the sake of conciseness, an additional assumption (11) is imposed to the non-trapping condition (4) that all arcs of ℓ-trajectories contained in $\mathcal{E}$ are of small enough length (which is, for instance, the case of codim $_{\partial\Omega}\mathcal{E}=1$ and ℓ transversal to $\mathcal{E}$). It is only a technical assumption which brings to light the *non-local* character of $(\mathcal{P})$ near $\mathcal{E}$ and simplifies the proofs. Anyway, the results hold true under the sole non-trapping condition (4) and the corresponding proofs will be published elsewhere.

1. ASSUMPTIONS AND AUXILIARY RESULTS

Let $\Omega \subset \mathbb{R}^n$, $n \geq 3$, be a bounded domain with reasonably smooth boundary. Denote by $\nu(x) = (\nu_1(x),\ldots,\nu_n(x))$ the unit *outward normal* to $\partial\Omega$, at $x \in \partial\Omega$, and let $\ell(x) = (\ell_1(x),\ldots,\ell_n(x))$ be a unit vector field defined on $\partial\Omega$. Decompose it into $\ell(x) = \tau(x) + \gamma(x)\nu(x) \forall x \in \partial\Omega$, where $\tau : \partial\Omega \to \mathbb{R}^n$ is the projection of $\ell(x)$ on the tangential hyperplane to $\partial\Omega$ at $x \in \partial\Omega$ and $\gamma : \partial\Omega \to \mathbb{R}$. Set

$$\mathcal{E} = \{x \in \partial\Omega : \gamma(x) = 0\}$$

for the subset of $\partial\Omega$ where the field $\ell(x)$ is tangential to the boundary.

Hereafter we set $\mathcal{N} \subset \overline{\Omega}$ to be a closed neighbourhood of $\mathcal{E}$ in $\overline{\Omega}$. Suppose $\mathcal{L}$ is a uniformly elliptic operator with measurable coefficients, satisfying

$$a^{ij}(x) = a^{ji}(x),\ \exists \lambda = \text{const} > 0 \text{ such that}$$
$$\lambda^{-1}|\xi|^2 \leq a^{ij}(x)\xi_i\xi_j \leq \lambda|\xi|^2 \quad \text{a.a. } x \in \Omega,\ \forall \xi \in \mathbb{R}^n; \tag{1}$$
$$a^{ij} \in VMO(\Omega) \cap C^{0,1}(\mathcal{N});\ b^i, c \in L^\infty(\Omega) \cap C^{0,1}(\mathcal{N}). \tag{2}$$

Here $VMO(\Omega)$ stands for functions of vanishing mean oscillation and $C^{0,1}$ is the class of Lipschitz continuous functions. It is to be noted that (2) and Rademacher's theorem ensure the coefficients of $\mathcal{L}$ belong to $W^{1,\infty}(\mathcal{N})$. Moreover, $a^{ij} \in L^\infty(\Omega)$ as consequence of (1).

Concerning the boundary operator $\mathcal{B}$, we suppose

$$\partial\Omega \in C^{1,1},\quad \partial\Omega \cap \mathcal{N} \in C^{2,1},$$
$$\ell_i,\ \sigma \in C^{0,1}(\partial\Omega) \cap C^{1,1}(\partial\Omega \cap \mathcal{N}), \tag{3}$$
$$\gamma(x) = \ell(x) \cdot \nu(x) \geq 0 \quad \forall x \in \partial\Omega.$$

The geometrical meaning of $\gamma(x) \geq 0$ is that $\ell(x)$ is either tangential to $\partial\Omega$ or is directed outwards Ω at each point $x \in \partial\Omega$. According to the physical interpretation of the problem $(\mathcal{P})$ in the theory of Brownian motion, that means ℓ is of *neutral type* on $\partial\Omega$ (cf. [2], [15], [13]). Finally, we impose a kind of *non-trapping* condition on the set of tangency

$$\begin{cases} \text{the arcs of } \tau \text{ - trajectories lying in } \mathcal{E} \text{ are} \\ \text{all non-closed and of finite length,} \end{cases} \tag{4}$$

which simply means the τ-trajectories (coinciding with these of ℓ on $\mathcal{E}$!) leave the set of tangency $\mathcal{E}$ in a *finite time* in both directions.

In what follows, we will use a suitable extension of the field ℓ in a neigh-bourhood of $\partial\Omega$. For this goal, for any point x sufficiently close to $\partial\Omega$ set $d(x)=\operatorname{dist}(x,\partial\Omega)$, $\Omega_0=\{x\in\Omega: d(x)\geq d_0>0\}$ with d_0 small enough. It is well known (see [4]) that to each $x\in\Omega\setminus\Omega_0$ there corresponds a unique $y(x)\in\partial\Omega$ closest to x, $d(x)$ has the regularity of $\partial\Omega$ at $y(x)\in\partial\Omega$ and $\nabla d(x)=\nu(y(x))$. This way, defining

$$L(x)=\ell(y(x))+d(x)\nabla d(x) \qquad \forall\in\Omega\setminus\Omega_0,$$

it is clear that $L\in C^{0,1}(\Omega\setminus\Omega_0)\cap C^{1,1}((\Omega\setminus\Omega_0)\cap\mathcal{N})$. Moreover, the following result holds true (see [10, Proposition 2.1], [15, Proposition 3.2.5]):

Proposition 1 *Assume* (3) *and* (4). *Then the field* $L(x)$ *is strictly transversal to* $\partial\Omega_0$ *and any point of* $\overline{\Omega\setminus\Omega_0}$ *can he reached from* $\partial\Omega_0$ *through an L-trajectory of length at most* $\kappa=\text{const}>0$.

Set further $\psi(\cdot,x):\mathbb{R}\to\mathbb{R}^n$ *for the parameterization of the L-integral curve passing through x and define* $\Omega_t=\Omega_0\cup\{\psi(s,x): s\in[0,t], x\in\partial\Omega_0\}$. *Then* $\{\Omega_t\}_{t\geq 0}$ *is a non-decreasing family and to each* $\delta>0$ *there corresponds a* $\theta(\delta)>0$, *independent of t, and such that dist* $(\Omega_t,\Omega\setminus\Omega_{t+\delta})\geq\theta$ *whenever* $\Omega\setminus\Omega_{t+\delta}\neq\emptyset$.

In what follows we will employ local a'priori estimates of special kind for the strong solutions to Dirichlet problem which take precise account of the distance to the boundary.

Proposition 2 *Assume* (1), (2) *and* $\partial\Omega\in C^{1,1}$. *Let* $\Omega_1\subset\Omega_2$ *be open subsets of* Ω *with* dist $(\Omega_1,\partial\Omega_2\setminus\partial\Omega)\geq\theta>0$ *if* $\Omega\not\equiv\Omega_2$ *and* dist $(\Omega_1,\partial\Omega)\geq\theta>0$ *when* $\Omega\equiv\Omega_2$. *Then* $\forall u\in W^{2,p}(\Omega)$, $p\in(1,\infty)$, *one has*

$$\|D^2u\|_{L^p(\Omega_1)}\leq C'\left(\|\mathcal{L}u\|_{L^p(\Omega_2)}+\|u\|_{W^{2-1/p,p}(\partial\Omega_2\cap\partial\Omega)}\right)$$
$$+C''(\theta)\left(\|u\|_{W^{1,p}(\Omega_2)}+\|u\|_{W^{1-1/p,p}(\partial\Omega_2\cap\partial\Omega)}\right)$$

where the constant C' depends on n,p,λ,Ω *and the coefficients of the operator* $\mathcal{L}$ *and C'' depends on* θ *in addition.*

Proof. Taking a cutoff function $\eta(x) \in C^\infty(\bar{\Omega})$ such that $\eta \equiv 1$ in Ω_1, supp $\eta \subset \Omega_2$, $\max_{\bar{\Omega}} |D^\alpha \eta| \le C\theta^{-|\alpha|}$, one gets

$$\mathcal{L}(\eta u) = \eta \mathcal{L} u + \left(a^{ij} D_{ij}\eta + b^i D_i \eta\right) u + 2a^{ij} D_j \eta D_i u =: F(x) \quad \text{a.e. } \Omega.$$

The choice of η and [1, Theorem 4.2] imply

$$\| D^2 u \|_{L^p(\Omega_1)} = \| D^2(\eta u) \|_{L^p(\Omega_1)} \le \| D^2(\eta u) \|_{L^p(\Omega)}$$
$$\le C\left(\| F \|_{L^p(\Omega)} + \| \eta u \|_{W^{2-1/p,p}(\partial\Omega)} \right).$$

Further,

$$||F||_{L^p(\Omega)} \le ||\eta \mathcal{L} u||_{L^p(\Omega_2)}$$
$$+ ||\left(a^{ij} D_{ij}\eta + b^i D_i \eta\right) u + 2a^{ij} D_j \eta D_i u||_{L^p(\Omega_2)}$$
$$\le ||\mathcal{L} u||_{L^p(\Omega_2)} + C(\theta) ||u||_{W^{1-p}(\Omega_2)},$$
$$\| \eta u \|_{W^{2-1/p,p}(\partial\Omega)} \le C \| u \|_{W^{2-1/p,p}(\partial\Omega_2 \cap \partial\Omega)} + C(\theta) \| u \|_{W^{1-1/p,p}(\partial\Omega_2 \cap \partial\Omega)}$$

and these give the desired estimate. □

The following Gronwall-type inequality (proved in [17, Proposition 4.1]) will be useful in the forthcoming consideration. We propose the proof here for reader's convenience.

Proposition 3 *Let $\zeta : [0,\infty) \to [0,\infty)$ be a bounded and continuous function and let there exist positive constants δ, A and C such that*

$$\zeta(t) \le A + C \int_0^t \zeta(s+\delta)\, ds \quad \forall t \ge 0.$$

If δ is so small that $C\delta e < 1$ then

$$\zeta(t) \le A\left(1 + \frac{e^{t/\delta}}{\sqrt{2\pi}(1 - C\delta e)}\right) \quad \forall t \ge 0.$$

Proof. Induction in $N \in \mathbb{N}$ gives

$$\zeta(t) \le \frac{C^N (t+N\delta)^N}{N!} \sup_{\mathbb{R}_+} \zeta(t) + A \sum_{k=0}^{N-1} \frac{C^k (t+k\delta)^k}{k!}.$$

Apply now the Stirling formula ($k! = \sqrt{2\pi k} e^{-k} k^k (1+\alpha_k)$ with $\alpha_k \to 0^+$ as $k \to \infty$) in order to get

$$\frac{C^k (t+k\delta)^k}{k!} \le \frac{C^k \delta^k e^k \left(1+\frac{t/\delta}{k}\right)^k}{\sqrt{2\pi k}} \le \frac{(C\delta e)^k e^{t/\delta}}{\sqrt{2\pi}}.$$

Thus, $C\delta e < 1$ implies

$$\frac{C^N (t+N\delta)^N}{N!} \le \frac{(C\delta e)^N e^{t/\delta}}{\sqrt{2\pi}} \to 0 \text{ as } N \to \infty$$

and therefore

$$\sum_{k=0}^{N-1} \frac{C^k (t+k\delta)^k}{k!} \le 1 + \frac{e^{t/\delta}}{\sqrt{2\pi}} \sum_{k=1}^{N-1} (C\delta e)^k = 1 + \frac{e^{t/\delta} C\delta e}{\sqrt{2\pi}(1 - C\delta e)}.$$

□

2. L^p-A PRIORI ESTIMATES

Consider the Banach spaces

$$\mathcal{F}^p(\Omega, \mathcal{N}) = \{f \in L^p(\Omega) : \partial f / \partial L \in L^p(\mathcal{N})\}$$

equipped with the norm $\|f\|_{\mathcal{F}^p(\Omega,\mathcal{N})} = \|f\|_{L^p(\Omega)} + \|\partial f/\partial L\|_{L^p(\mathcal{N})}$, and the fractional Sobolev space

$$\Phi^p(\partial\Omega, \mathcal{N}) = \{\varphi \in W^{1-1/p,p}(\partial\Omega) : \varphi \in W^{2-1/p,p}(\partial\Omega \cap \mathcal{N})\}$$

normed by $\|\varphi\|_{\Phi^p(\partial\Omega,\mathcal{N})} = \|\varphi\|_{W^{1-1/p,p}(\partial\Omega)} + \|\varphi\|_{W^{2-1/p,p}(\partial\Omega\cap\mathcal{N})}$.

Our first result concerns $W^{2,p}$-estimates for solutions to $(\mathcal{P})$.

Theorem 4 *Suppose* (1)-(4) *and let* $u \in W^{2,p}(\Omega)$ *be a strong solution of* $(\mathcal{P}), p \in (1,\infty)$, *with* $f \in \mathcal{F}^p(\Omega,\mathcal{N})$ *and* $\varphi \in \Phi^p(\partial\Omega,\mathcal{N})$. *Then* $\partial u/\partial L \in W^{2,p}(\mathcal{N})$ *and there is an absolute constant C such that*

$$\|u\|_{W^{2,p}(\Omega)}, \|\partial u/\partial L\|_{W^{2,p}(\mathcal{N})} \leq C\left(\|u\|_{L^p(\Omega)} + \|f\|_{\mathcal{F}^p(\Omega,\mathcal{N})} + \|\varphi\|_{\Phi^p(\partial\Omega,\mathcal{N})}\right). \tag{5}$$

Proof. Let $\mathcal{N}' \subset \mathcal{N}'' \subset \mathcal{N}$ be closed neighbourhoods of $\mathcal{E}$ in $\overline{\Omega}$. Bearing in mind (1)-(3) and $\gamma(x) > 0 \ \forall x \in \partial\Omega \setminus \mathcal{N}'$, we obtain that $(\mathcal{P})$ is a *regular* oblique derivative problem in $\Omega \setminus \mathcal{N}'$ and therefore the L^p-theory (see [9]) implies

$$\|u\|_{W^{2,p}(\Omega\setminus\mathcal{N}')} \leq C\left(\|u\|_{L^p(\Omega)} + \|f\|_{L^p(\Omega)} + \|\varphi\|_{W^{1-1/p,p}(\partial\Omega)}\right). \tag{6}$$

Further, (2) ensures that $\partial u/\partial L \in W^{2,p}(\mathcal{N})$ and it verifies

$$\begin{cases} \mathcal{L}(\partial u/\partial L) = \partial f/\partial L \\ -\left(\partial a^{ij}/\partial L\right) D_{ij}u - \left(\partial b^i/\partial L\right) D_i u - \left(\partial c/\partial L\right) u \\ +2a^{ij} D_j L_k D_{ki} u + a^{ij} D_{ij} L_k D_k u + b^i D_i L_k D_k u \quad a.e.\ \mathcal{N}, \\ \partial u/\partial L = \varphi - \sigma u \quad \text{on } \partial\Omega. \end{cases} \tag{7}$$

To estimate the $W^{2,p}$-norm of u in $\mathcal{N}''$, take arbitrary $x \in \mathcal{N}''$ and let $\psi(t,x)$ be the parameterization of the L-trajectory through x. Surely $\psi(t,\cdot) \in C^{1,1}(\mathcal{N})$. Without loss of generality, we may choose $\mathcal{N}''$ with $C^{1,1}$ smooth boundary and in such way that for any x in the interior of $\mathcal{N}''$ there corresponds a unique $\xi(x) > 0, \xi \in C^{1,1}(\mathcal{N}'')$ such that $\psi(-\xi(x),x) \in \partial\mathcal{N}'' \setminus \mathcal{N}'$. Thus, $\forall t \in [0,\kappa]$ and a.a. $x \in \Omega_t \cap \mathcal{N}''$ we have

$$u(x) = u \circ \psi(-\xi(x),x) + \int_{-\xi(x)}^{0} \frac{\partial u}{\partial L} \circ \psi(\mu,x)\, d\mu$$

$$= u \circ \psi(-\xi(x),x) + \int_{t-\xi(x)}^{t} \frac{\partial u}{\partial L} \circ \psi(s-t,x)\, ds.$$

The first term at the right-hand side regards values of u out of $\mathcal{N}'$ and (6) is applicable to it, while the integrand concerns values of $\partial u/\partial L$ in $\Omega_s \cap \mathcal{N}''$. Therefore, taking second derivatives and then L^p-norm lead to

$$\left\|D^2u\right\|^p_{L^p(\Omega_t\cap\mathcal{N}')} \le C\left(\|u\|^p_{W^{2,p}(\Omega\setminus\mathcal{N}')} + \left\|\partial u/\partial L\right\|^p_{W^{1,p}(\Omega\setminus\mathcal{N})}\right.$$

$$\left. + \int_0^t \left\|D^2\left(\partial u/\partial L\right)\right\|^p_{L^p(\Omega_s\cap\mathcal{N}')} ds\right). \tag{8}$$

Let $s\in[0,t]$ be arbitrary. By Proposition 1, for each $\delta>0$ there exists $\theta(\delta)>0$ such that dist $(\Omega_s,\Omega\setminus\Omega_{s+\delta})\ge\theta$ whenever $\Omega\setminus\Omega_{s+\delta}\neq\emptyset$. Employing the fact that $\partial u/\partial L$ solves locally the Dirichlet problem (7) and using Proposition 2 with $\Omega_1=\Omega_s\cap\mathcal{N}''$ and $\Omega_2=\Omega_{s+\delta}\cap\mathcal{N}''$, we obtain

$$\left\|D^2\left(\partial u/\partial L\right)\right\|^p_{L^p(\Omega_s\cap\mathcal{N}'')} \le C\left\|D^2u\right\|^p_{L^p(\Omega_{s+\delta}\cap\mathcal{N}'')} \tag{9}$$
$$+C'(\theta)\left(K^p+\|u\|^p_{W^{1,p}(\Omega)}+\left\|\partial u/\partial L\right\|^p_{W^{1,p}(\mathcal{N})}\right)$$

with C independent of δ and $K=\|f\|_{\mathcal{F}^p(\Omega,\mathcal{N})}+\|\varphi\|_{\Phi^p(\partial\Omega,\mathcal{N})}$. Further,

$$\left\|\partial u/\partial L\right\|^p_{W^{1,p}(\mathcal{N})} \le \left\|\partial u/\partial L\right\|^p_{W^{1,p}(\mathcal{N}\setminus\mathcal{N}')} + \left\|\partial u/\partial L\right\|^p_{W^{1,p}(\mathcal{N}')}$$
$$\le \|u\|^p_{W^{2,p}(\Omega\setminus\mathcal{N}')} + \varepsilon\left\|\partial u/\partial L\right\|^p_{W^{2,p}(\Omega\setminus\mathcal{N}')} + C(\varepsilon)\left\|\partial u/\partial L\right\|^p_{W^{1,p}(\Omega\setminus\mathcal{N}')}$$

after interpolating the L^p-norms ([4, Theorem 7.28]) with arbitrary $\varepsilon>0$ to be specified later. Remembering (7), we have

$$\left\|\partial u/\partial L\right\|_{W^{2,p}(\mathcal{N}')} \le C\left(\|u\|_{W^{2,p}(\mathcal{N}'')}+K\right) \le C\left(\|u\|_{W^{2,p}(\Omega)}+K\right)$$

and therefore (9) becomes

$$\left\|D^2\left(\partial u/\partial L\right)\right\|^p_{L^p(\Omega_s\cap\mathcal{N}'')} \le C\left\|D^2u\right\|^p_{L^p(\Omega_{s+\delta}\cap\mathcal{N}'')}$$
$$+C'(\theta)\left(K^p+\varepsilon\|u\|^p_{W^{2,p}(\Omega)}\right)+C''(\theta,\varepsilon)\|u\|^p_{W^{1,p}(\Omega)},$$

after applying (6). This way, (8) reads

$$\zeta(t) \le C\int_0^t \zeta(s+\delta)ds + C'(\theta)\left(K^p+\varepsilon\|u\|^p_{W^{2,p}(\Omega)}\right)+C''(\theta,\varepsilon)\|u\|^p_{W^{1,p}(\Omega)}$$

with $\zeta(t) = \|D^2u\|^p_{L^p(\Omega_t \cap \mathcal{N}'')}$ for $t \in [0,\kappa]$; $\zeta(t) = 0$ if $\Omega_t \cap \mathcal{N}'' = \emptyset$ and $\zeta(t) = \|D^2u\|^p_{L^p(\mathcal{N}'')}$ when $t > \kappa$. The multiplier C of the integral above is independent of δ and ε and therefore taking $\delta > 0$ (cf. Proposition 3) sufficiently small, we get

$$\|D^2u\|^p_{L^p(\mathcal{N}'')} \le C'\left(K^p + \varepsilon\|u\|^p_{W^{2,p}(\Omega)}\right) + C''(\varepsilon)\|u\|^p_{W^{1,p}(\Omega)}$$

which, coupled with (6), gives

$$\|D^2u\|^p_{L^p(\Omega)} \le C'\left(K^p + \varepsilon\|u\|^p_{W^{2,p}(\Omega)}\right) + C''(\varepsilon)\|u\|^p_{W^{1,p}(\Omega)}.$$

To get (5), it remains to choose ε so small that $C'\varepsilon < 1$ and then interpolate once again in order to move $\|u\|^p_{W^{1,p}(\Omega)}$ on the right. □

Remark 5 It is clear that the Lipschitz continuity in $\mathcal{N}$ of the coefficients of $\mathcal{L}$ can be relaxed to essential boundedness in $\mathcal{N}$ of their directional derivatives with respect to the extended field L. Moreover, instead of $b^i, c \in L^\infty(\Omega)$ one can ask $b^i \in L^q(\Omega)$ with $q > n$ if $p \le n, q = p$ otherwise and $c \in L^r(\Omega)$ with $r > n/2$ if $p \le n/2, r = p$ otherwise (see [9, Section 2.3]).

3. UNIQUENESS RESULTS

We start with the simpler case $p > n$.

Lemma 6 *Suppose* (1), (2), (3) *and assume* $c(x) \le 0$ *a.e. in* Ω *and* $\sigma(x) > 0$ *on* $\partial\Omega$. *Let* $p > n$ *and let* $u, v \in W^{2,p}(\Omega)$ *be two solutions of* $(\mathcal{P})$ *with* $f \in \mathcal{F}^p(\Omega,\mathcal{N}), \varphi \in \Phi^p(\partial\Omega,\mathcal{N})$. *Then* $u \equiv v$ *in* $\overline{\Omega}$.

Proof. The difference $w := u - v \in W^{2,p}(\Omega) \subset C^1(\overline{\Omega})$ (note $p > n$!) solves the homogeneous problem

$$\{\mathcal{L}w = 0 \text{ a.e. } \Omega, \quad \partial w/\partial\ell + \sigma(x)w = 0 \text{ on } \partial\Omega. \tag{10}$$

Suppose w assumes positive values in $\overline{\Omega}$ and set $w(x_0) = \max_{\overline{\Omega}} w(x) > 0$. Then $x_0 \notin \Omega$ as consequence of the strong Aleksandrov maximum principle. In fact, $x_0 \in \Omega$ and [4, Theorem 9.6] imply

$w(x) = \text{const} = w(x_0) > 0 \; \forall x \in \overline{\Omega}$ which is impossible in view of the boundary condition in (10) which holds in a classical sense.

Further, assuming $x_0 \in \partial\Omega$, one has $(\partial w/\partial \ell)(x_0) \geq 0$ (precisely, "> 0" if $x_0 \in \partial\Omega \setminus \mathcal{E}$ as it follows from the boundary point lemma (see [4, Lemma 3.4]), and "=0" when $x_0 \in \mathcal{E}$). Anyway,

$$(\partial w/\partial \ell)(x_0) + \sigma(x_0) w(x_0) > 0$$

which contradicts the boundary condition in (10). Therefore $w(x) \leq 0$ in $\overline{\Omega}$. Similarly, one obtains $w(x) \geq 0$ in $\overline{\Omega}$ whence $w \equiv 0$. □

Remark 7 Let us point out that Lemma 6 holds in a general situation when (4) is not necessarily verified. If, in addition, the non-trapping condition (4) holds then Lemma 6 remains still valid (see [13]) for $\sigma \geq 0$ on $\partial\Omega$ and such that either $c \not\equiv 0$ or $\sigma \not\equiv 0$.

To cover the case $p \leq n$ we will employ the regularizing properties of the couple $(\mathcal{B}, \mathcal{L})$. Roughly speaking, it means that even if $(\mathcal{P})$ is a degenerate problem and therefore the solution "loses" derivatives from the data, $(\mathcal{P})$ behaves like an elliptic BVP for what concerns the rate of integrability. That is, higher integrability of (f, φ) implies higher integrability of the second derivatives of solutions to $(\mathcal{P})$. We will restrict, however, (4) to the following *small non-trapping* condition

$$\begin{cases} \text{the arc-lengths of the } \tau\text{-trajectories lying in } \mathcal{E} \\ \text{are bounded by a sufficiently small number } \kappa_0. \end{cases} \tag{11}$$

For instance, (11) is surely verified when $\mathcal{E}$ is a submanifold of $\partial\Omega$ of co-dimension one and ℓ is transversal to it.

Proposition 8 *Suppose* (1)-(3), (11) *and let* $u \in W^{2,p}(\Omega)$, $p \in (1, \infty)$, *be a solution of* $(\mathcal{P})$ *with* $f \in \mathcal{F}^q(\Omega, \mathcal{N})$, $\varphi \in \Phi^q(\partial\Omega, \mathcal{N})$ *and* $q \geq p$. *Then* $u \in W^{2,q}(\Omega)$.[1]

Proof. Take the neighbourhoods $\mathcal{N}' \subset \mathcal{N}'' \subset \mathcal{N}$ of $\mathcal{E}$ in $\overline{\Omega}$ as in the proof of Theorem 4 and let $\mathcal{N}''$ be so "narrow" that $\mathcal{N}'' \subset \Omega \setminus \Omega_0$ (see Proposition 1). Employing the L^p-theory ([9]) of *regular* oblique derivative problems $(\ell(x) \cdot \nu(x) > 0 \; \forall x \in \partial\Omega \setminus \mathcal{N}')$, we get immediately $u \in W^{2,p}(\Omega \setminus \mathcal{N}')$.

[1] This result and forthcoming Theorems 9, 10 and 11 remain valid in the general situation when (4) holds instead of (11). The corresponding proof of Proposition 8, however, is rather technically complicated and will be published elsewhere.

To obtain higher integrability of the second derivatives in $\mathcal{N}''$ we will use the fact that $\partial u/\partial L$ solves locally the Dirichlet problem (7) but first of all we need to modify the original problem $(\mathcal{P})$. For, according to Proposition 1 for any $x\in\overline{\Omega}\setminus\Omega_0$ there is a unique $\eta(x)\in C^{0,1}(\Omega\setminus\Omega_0)\cap C^{1,1}(\mathcal{N})$ such that $\psi(-\eta(x),x)\in\partial\Omega_0$. Suppose the function σ is extended in $\Omega\setminus\Omega_0$ such that $\sigma\in C^{0,1}(\Omega\setminus\Omega_0)\cap C^{1,1}(\mathcal{N})$ and define

$$\Sigma(x)=\int_0^{\eta(x)}\sigma(\psi(t-\eta(x),x))\,dt.$$

Indeed, $(\partial\Sigma/\partial L)(x)=\sigma(x)\forall x\in\overline{\Omega}\setminus\Omega_0$ and $\Sigma,\partial\Sigma/\partial L\in C^{1,1}(\mathcal{N})$. At this point, the function $U(x)=u(x)e^{\Sigma(x)}\in W^{2,p}(\mathcal{N})\cap W^{2,q}(\mathcal{N}\setminus\mathcal{N}')$ solves the problem

$$\begin{cases}a^{ij}D_{ij}U+B^iD_iU+CU=\tilde{f}(x):=f(x)\mathrm{e}^{\Sigma(x)} & a.e.\,\mathcal{N},\\ \partial U/\partial L=\tilde{\varphi}(x):=\varphi(x)\mathrm{e}^{\Sigma(x)} & \text{on } \partial\Omega\cap\partial\mathcal{N},\end{cases}$$

with

$$B^i(x)=b^i(x)-2a^{ij}(x)D_j\Sigma(x),$$
$$C(x)=c(x)-a^{ij}(x)D_{ij}\Sigma(x)+a^{ij}(x)D_i\Sigma(x)D_j\Sigma(x)-b^i(x)D_i\Sigma(x).$$

Moreover $B^i,C,\partial B^i/\partial L,\partial C/\partial L\in L^\infty(\mathcal{N}),\tilde{f},\partial\tilde{f}/\partial L\in L^q(\mathcal{N})$ and $\varphi\in W^{2-1/q,q}(\partial\Omega\cap\partial\mathcal{N})$. It follows from Theorem 4 that and it satisfies

$$\begin{cases}a^{ij}D_{ij}V+B^iD_iV+CV=F(x):=\partial\tilde{f}/\partial L\\ \quad-(\partial a^{ij}/\partial L)D_{ij}U-(\partial B^i/\partial L)D_iU-(\partial C/\partial L)U\\ \quad+2a^{ij}D_jL_kD_{ki}U+a^{ij}D_{ij}L_kD_kU+B^iD_iL_kD_kU \quad \text{a.e.}\mathcal{N},\\ V=\tilde{\varphi} \quad \text{on} \quad \partial\Omega\cap\partial\mathcal{N}\end{cases}\tag{12}$$

We have, first of all, $F\in L^q(\mathcal{N}\setminus\mathcal{N}')$ and $V=\tilde{\varphi}\in W^{2-1/q,q}(\partial\Omega\cap\partial\mathcal{N})$ whence $V\in W^{2-1/q,q}(\partial\mathcal{N}'')$. Later on, for any $x\in\mathcal{N}''$ there is a unique $\xi(x)\in C^{1,1}$ such that $\psi(-\xi(x),x)\in\partial\mathcal{N}''$ and

$$U(x)=U\circ\psi(-\xi(x),x)+\int_0^{\xi(x)}V\circ\psi(t-\xi(x),x)\,dt\tag{13}$$

for a.a. $x \in \mathcal{N}''$. Since $U \circ \psi(-\xi(x), x)$ is a $W^{2,q}$-function $(\psi(-\xi(x), x) \in \mathcal{N}')$ we take the derivatives of $U(x)$ up to second order and substitute them into the right-hand side of the equation in (12), obtaining that V solves the *non-local* Dirichlet problem

$$\begin{cases} a^{ij} D_{ij} V + = \tilde{F}(x) + \displaystyle\int_0^{\xi(x)} \mathcal{L}_2(V) \circ \psi(t - \xi(x), x)\, dt \quad \text{a.e. } \mathcal{N}'' \\ V|_{\partial\Omega \cap \partial\mathcal{N}''} = \tilde{\varphi} \in W^{2-1/q,q}, \qquad V|_{\partial\mathcal{N}'' \setminus \partial\Omega} \in W^{2-1/q,q}, \end{cases} \tag{14}$$

where $\tilde{F}(x) := \partial \tilde{f}/\partial L + \mathcal{L}_2'(V)(x) + \mathcal{L}_2'(U) \circ \psi(-\xi(x), x)$ and $\mathcal{L}_i'$ and $\mathcal{L}_2$ are linear differential operators with L^∞ coefficients, ord $\mathcal{L}_i' = i$, ord $\mathcal{L}_2 = 2$. Therefore, $V \in W^{2,p}(\mathcal{N}'')$ and Sobolev's imbedding theorem imply $\tilde{F} \in L^{q'}(\mathcal{N}'')$ with $q' = \min\{q, np/(n-p)\}$ if $p < n, q' = q$ otherwise. Indeed $q' \geq p$.

We will prove now $V \in W^{2,q'}(\mathcal{N}'')$ by means of the contraction mapping principle. If $q' = p$ then we are done. Otherwise, take any $r \in p, q'$ and define the operator $\mathcal{T}: W^{2,r}(\mathcal{N}'') \to W^{2,r}(\mathcal{N}'')$ follows: for any $w \in W^{2,r}(\mathcal{N}'')$ the image $\mathcal{T}w \in W^{2,r}(\mathcal{N}'')$ is the unique solution of the Dirichlet problem

$$\begin{cases} a^{ij} D_{ij}(\mathcal{T}w) = \tilde{F} + \displaystyle\int_0^{\xi(x)} \mathcal{L}_2(w) \circ \psi(t - \xi(x), x)\, dt \in L^r \quad \text{a.e. } \mathcal{N}'' \\ (\mathcal{T}w) = V \in W^{2-1/r,r} \text{ on } \partial\mathcal{N}''. \end{cases}$$

Indeed, for any $w_1, w_2 \in W^{2,r}(\mathcal{N}'')$ the difference $\mathcal{T}w_1 - \mathcal{T}w_2$ solves

$$\begin{cases} a^{ij} D_{ij}(\mathcal{T}w_1 - \mathcal{T}w_2) = \displaystyle\int_0^{\xi(x)} \mathcal{L}_2(w_1 - w_2) \circ \psi(t - \xi(x), x)\, dt \text{ a.e. } \mathcal{N}'' \\ (\mathcal{T}w_1 - \mathcal{T}w_2) = 0 \text{ on } \partial\mathcal{N}'' \end{cases}$$

and [1] implies

$$\begin{aligned} \| \mathcal{T}w_1 - \mathcal{T}w_2 \|_{W^{2,r}(\mathcal{N}'')} &\leq C \left\| \int_0^{\xi(x)} \mathcal{L}_2(w_1 - w_2) \circ \psi(t - \xi(x), x)\, dt \right\|_{L^r(\mathcal{N}'')} \\ &\leq C \max_{\overline{\mathcal{N}''}} \xi(x) \|w_1 - w_2\|_{W^{2,r}(\mathcal{N}'')} \end{aligned}$$

with a constant C depending on the coefficients of $\mathcal{L}$, and their derivatives in direction of L. It is clear now that if both κ_0 from (11) and $d_0 = \operatorname{dist}(\Omega_0, \partial\Omega)$ were small enough, then $C \max_{\overline{\mathcal{N}''}} \xi(x)$ can be made less than 1 and therefore $\mathcal{T}$ will be a contraction from $W^{2,r}(\mathcal{N}'')$ into itself for *any* $r \in [p, q']$. In particular, there is a unique fixed point of $\mathcal{T}$ lying in $W^{2,r}(\mathcal{N}'')$ for any $r \in [p, q']$. Since $V \in W^{2,p}(\mathcal{N}'')$ solves (14) and thus is already such a point, we conclude $V \in W^{2,q'}(\mathcal{N}'')$. Hence, $U \in W^{2,q'}(\mathcal{N}'')$ in view of (13) and therefore $u \in W^{2,q'}(\mathcal{N}'')$. To complete the proof of Proposition 8 it remains to repeat the above procedure finitely many times until q' becomes equal to q.

□

The general uniqueness result is contained in the following

Theorem 9 *Under the assumptions of Lemma 6, let* $p \in (1, \infty)$ *and suppose* (11) *in addition if* $p \leq n$. *Let* $u, v \in W^{2,p}(\Omega)$ *be two solutions of* $(\mathcal{P})$ *with* $f \in \mathcal{F}^p(\Omega, \mathcal{N}), \varphi \in \Phi^p(\partial\Omega, \mathcal{N})$. *Then* $u \equiv v$ *in* $\overline{\Omega}$.

Proof. We have to treat only the case $p \leq n$. The difference $w = u - v$ solves the homogeneous problem (10) with zero, and therefore C^∞, right-hand sides. According to Proposition 8 one has $w \in W^{2,q}(\Omega)$ for any $q > 1$ and therefore application of Lemma 6 is possible. □

4. REFINED L^p - *A* PRIORI ESTIMATE AND EXISTENCE

Under the uniqueness hypotheses of previous section we are able to drop out the norm $\|u\|_{L^p(\Omega)}$ from (5).

Theorem 10 *Suppose* $p \in (1, \infty)$, (1)-(4) *and* (11) *in addition if* $p \leq n$. *Assume moreover* $c \leq 0$ *a.e.* Ω *and* $\sigma > 0$ *on* $\partial\Omega$. *Let* $u \in W^{2,p}(\Omega)$ *be a strong solution of* $(\mathcal{P})$ *with* $f \in \mathcal{F}^p(\Omega, \mathcal{N})$ *and* $\varphi \in \Phi^p(\partial\Omega, \mathcal{N})$. *Then there is a constant C independent of u and such that*

$$\|u\|_{W^{2,p}(\Omega)} \leq C\left(\|f\|_{\mathcal{F}^p(\Omega,\mathcal{N})} + \|\varphi\|_{\Phi^p(\partial\Omega,\mathcal{N})}\right). \tag{15}$$

Proof. Note first of all that if $p>n$ then (15) follows immediately from (5) and [16, Theorem 2.6.2]. In fact, $u\in L^\infty(\Omega)$ and $\|u\|_{L^\infty(\Omega)}$, and therefore $\|u\|_{L^p(\Omega)}$ also, is estimated in terms of the respective norms of φ and f.

Thus, let $p\le n$ and suppose (15) is false. Then there exists a sequence $\{u_k\}\in W^{2,p}(\Omega)$ such that

$$\|u_k\|_{L^p(\Omega)}=1,\ \|\mathcal{L}u_k\|_{\mathcal{F}^p(\Omega,\mathcal{N})}\to 0,\quad \|\mathcal{B}u_k\|_{\Phi^p(\partial\Omega,\mathcal{N})}\to 0\ \text{ as }\ k\to\infty.$$

By means of the a'priori estimate (5), $\|u_k\|_{W^{2,p}(\Omega)}$ is bounded and therefore there is a subsequence, still denoted $\{u_k\}$, such that $u_k\rightharpoonup u\in W^{2,p}(\Omega)$ *weakly* as $k\to\infty$. Therefore,

$$\int_\Omega(\mathcal{L}u_k)v\,dx\to\int_\Omega(\mathcal{L}u)v\,dx\quad\forall v\in L^{p/(p-1)}(\Omega)$$

whence $\mathcal{L}u=0$ a.e. in Ω. Moreover, the compactness of the imbeddings $W^{2,p}(\Omega)\to W^{1,p}(\Omega)\to L^p(\Omega)$ ensures $\mathcal{B}u=0$ on $\partial\Omega$ and $\|u\|_{L^p(\Omega)}=1$. This last is, however, impossible in view of the uniqueness assertion (Theorem 9) which gives $u=0$. □

We are in a position now to prove solvability of $(\mathcal{P})$ for any $p>1$ generalizing thus [10, Theorem 1.1].

Theorem 11 *Let $p\in(1,\infty)$ and assume (1)-(4) and (11) in addition when $p\le n$. Suppose further $c(x)\le 0$ a.e. in Ω, $\sigma(x)>0$ on $\partial\Omega$. Then the Poincaré problem $(\mathcal{P})$ is uniquely solvable in $W^{2,p}(\Omega)$ for any $f\in\mathcal{F}^p(\Omega,\mathcal{N})$, $\varphi\in\Phi^p(\partial\Omega,\mathcal{N})$.*

Proof. Fix $q>n$ if $p\le n$ and $q=p$ otherwise, and consider sequences $\{f_k\}\in W^{1,q}(\Omega)$, $\{\varphi_k\}\in W^{2-1/q,q}(\partial\Omega)$ such that $f_k\to f$ in $\mathcal{F}^p(\Omega,\mathcal{N})$, $\varphi_k\to\varphi$ in $\Phi^p(\partial\Omega,\mathcal{N})$ as $k\to\infty$. Noting that [10, Theorems 1.1, 1.2] still hold true for the operator $\mathcal{L}$ with lower-order coefficients satisfying (2), we get from these results and [10, Remark 1.1] that there exists a unique solution $\{u_k\}\in W^{2,q}(\Omega)$ of the problem

$$\{\mathcal{L}u_k=f_k(x)\ \text{ a.e.}\Omega,\quad \partial u_k/\partial\ell+\sigma(x)u_k=\varphi_k(x)\quad\text{on }\partial\Omega.$$

Further, $p \le q$ anyway, and therefore Theorem 10 implies

$$\| u_k - u_m \|_{W^{2,p}(\Omega)} \le C\left(\| f_k - f_m \|_{\mathcal{F}^p(\Omega,\mathcal{N})} + \| \varphi_k - \varphi_m \|_{\Phi^p(\partial\Omega,\mathcal{N})} \right).$$

It follows $\{u_k\}$ is a Cauchy sequence in $W^{2,p}(\Omega)$ and therefore converges to a strong $W^{2,p}(\Omega)$ solution of $(\mathcal{P})$. The unicity of that solution is a consequence of Theorem 9. □

REFERENCES

[1] F. Chiarenza, M. Frasca and P. Longo, $W^{2,p}$-solvability of the Dirichlet problem for nondivergence elliptic equations with VMO coefficients, *Trans. Amer. Math. Soc.* 336 (1993), 841-853.

[2] Y.V. Egorov, *Linear Differential Equations of Principal Type*, Contemporary Soviet Mathematics, New York, 1986.

[3] Y.V. Egorov and V. Kondrat'ev, The oblique derivative problem, *Math. USSR Sbornik* **7** (1969). 139-169.

[4] D. Gilbarg and N.S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, 2nd ed., Springer-Verlag, Berlin, 1983.

[5] P. Guan, Hölder regularity of subelliptic pseudodifferential operators, *Duke Math.* J. **60** (1990), 563-598.

[6] P. Guan and E. Sawyer, Regularity estimates for the oblique derivative problem, *Ann. Math.* **137** (1993), 1-70.

[7] P. Guan and E. Sawyer, Regularity estimates for the oblique derivative problem on non-smooth domains I, *Chinese Ann. Math., Ser. B* **16** (1995), No. 3, 1-26; II *ibid.* **17** (1996), No. 1, 1-34.

[8] L. Hörmander, Pseudodifferential operators and non-elliptic boundary value problems, *Ann. Math.* **83** (1966), 129-209.

[9] A. Maugeri, D.K. Palagachev and L.G. Softova, *Elliptic and Parabolic Equations with Discontinuous Coefficients*, Wiley-VCH, Berlin, 2000.

[10] A. Maugeri, D.K. Palagachev and C. Vitanza, A singular boundary value problem for umformly elliptic operators, *J. Math. Anal. Appl.* **263**(2001), 33-48.

[11] V. Maz'ya and B.P. Paneah, Degenerate elliptic pseudodifferential operators and oblique derivative problem, *Trans. Moscow Math. Soc.* **31** (1974), 247-305.

[12] A. Melin and J. Sjöstrand, *Fourier Integral Operators with Complex- Valued Phase Functions*, in: Lect. Notes Math., Vol. 459, pp. 120-223,Springer-Verlag, Berlin, 1975.

[13] B.P. Paneah, *The Oblique Derivative Problem. The Poincaré Problem*, Wiley-VCH, Berlin, 2000.

[14] H. Poincaré, *Lecons de Méchanique Céleste, Tome III, Théorie de Marées*, Gauthiers-Villars, Paris, 1910.

[15] P.R. Popivanov and D.K. Palagachev, *The Degenerate Oblique Derivative Problem for Elliptic and Parabolic Equations*, Wiley-VCH (Akademie-Verlag), Berlin, 1997.

[16] N.S. Trudinger, *Nonlinear Second Order Elliptic Equations,* in: Lecture Notes of Math. Inst. of Nankai Univ., Tianjin, China, 1986.

[17] B. Winzell, A boundary value problem with an oblique derivative, *Commun. Partial Differ. Equations* **6** (1981), 305-328.

MINIMAL FRACTIONS OF COMPACT CONVEX SETS

D. Pallaschke[1] and R. Urbański[2]
Institut für Statistik und Mathematische Wirtschaftstheorie, Universität Karlsruhe, Karlsruhe, Germany;[1] Wydzal Matematyki i Informatyki, Uniwersytet im Adama Mickiewicza, Poznań, Poland[2]

Abstract: Pairs of compact convex sets naturally arise in quasidifferential calculus as sub- and super-differentials of a quasidifferentiable function (see [1]). Since the sub- and superdifferential are not uniquely determined, minimal representations are of special importance. In this paper we show that the problem of finding minimal representatives for the elements of pairs of compact convex sets is a special case of the more general problem of determining minimal fractions in ordered commutative semigroups which satisfy the order cancellation law. All the material of this paper is taken from the recently published textbook on pairs of compact convex sets ([11]).

Key words: quasidifferentiable function, pairs of compact convex sets.

AMS(MOS) Subject Classification: 26A27, 90C30.

1. NOTATIONS AND PRELIMINARIES

For a topological vector space $X = (X, \tau)$ let us denote by $\mathcal{A}(X)$ the set of all nonempty subsets of X, by $\mathcal{B}^*(X)$ the set of all nonempty bounded subsets of X, by $\mathcal{C}(X)$ the set of all nonempty closed convex subsets of X, by $\mathcal{B}(X) = \mathcal{B}^*(X) \cap \mathcal{C}(X)$ the set of all bounded closed convex sets of X and by $\mathcal{K}(X)$ the set of all nonempty compact convex subsets of X. For $A, B \in \mathcal{A}(X)$ the *algebraic sum* is defined by

$A + B = \{x = a + b \mid a \in A \text{ and } b \in B\}$ and for $\lambda \in \mathbb{R}$ and $A \in \mathcal{A}(X)$ the *multiplication* is defined by $\lambda A = \{x = \lambda a \mid a \in A\}$.

The *Minkowski sum* for $A, B \in \mathcal{A}(X)$ is defined by

$$A \dot{+} B = \mathrm{cl}(\{x = a + b \mid a \in A \text{ and } b \in B\}),$$

where $\mathrm{cl}(A) = \overline{A}$ denotes the closure of $A \subset X$ with respect to τ. With $\mathrm{relint}(A)$ we denote the relative interior of $A \subset X$ with respect to τ.

For $A, B \in \mathcal{A}(X)$ we define:

$A \mathring{\vee} B = \mathrm{conv}(A \cup B), A \vee B = \overline{A \mathring{\vee} B} = \mathrm{cl}\ \mathrm{conv}(A \cup B)$ and by

$A \veebar B = \bigcup_{\alpha,\beta \geq 0, \alpha+\beta=1} (\alpha A + \beta B)$ the *skeleton* of A and B. It is easy to observe that $A \veebar B \subset A \mathring{\vee} B \subset A \vee B$. In the case when A and B are convex sets then $A \veebar B = A \mathring{\vee} B$. For two elements $a, b \in X$ the interval with end points a and b will be denoted by $[a, b] = \{a\} \vee \{b\}$.

For compact convex sets, the Minkowski sum coincides with the algebraic sum, i,e., for $A, B \in \mathcal{K}(X)$ we have $A \dot{+} B = A + B$ and also $A \mathring{\vee} B = A \vee B$. We will use the abbreviation $A \dot{+} B \vee C$ for $A \dot{+} (B \vee C)$ and $C + d$ instead of $C + \{d\}$ for all bounded closed convex sets $A, B, C \in \mathcal{A}(X)$ and a point $d \in X$.

A convex subset B of convex set $A \subseteq X$ is called an *extreme subset* if for every $x, y \in A$ and some $t \in (0,1)$ the condition $tx + (1-t)y \in B$ implies that $x, y \in B$. An extreme subset which consists of a single point only is called an *extreme point* and $\mathcal{E}(A)$ denotes the set of extreme points of A.

A convex set which is the convex hull of finitely many points is called a *polytope*. The set of all polytopes of a vector space X is denoted by $\mathcal{P}(X)$. An extreme subset of a polytope is called a *face* and a one-dimensional extreme set of a polytope is called an *edge*.

If (X, τ) is a topological vector space and X^* its dual space, then we denote for $A \in \mathcal{K}(X)$ and $f \in X^*$ by

$$H_f(A) = \left\{z \in A \mid f(z) = \max_{y \in A} f(y)\right\}$$

the *(maximal) face* of A with respect to f.

Finally, we will call a set $A \in \mathcal{B}(X)$ a *summand* of $B \in \mathcal{B}(X)$ if there exists a set $C \in \mathcal{B}(X)$ such that $A \dot{+} C = B$.

The following statements hold for convex sets:

Addition of maximal faces:

Proposition 1.1. *Let X be a topological vector space, $f \in X^*$ and $A, B \in \mathcal{K}(X)$. Then*

$$H_f(A+B) = H_f(A) + H_f(B).$$

Proof: Assume that $x = a + b \in H_f(A+B)$ with $a \in A$ and $b \in B$. Then $a \in H_f(A)$ and $b \in H_f(B)$. Indeed, assume for instance that $a \notin H_f(A)$. Since $A \in \mathcal{K}(X)$ is compact, there exists an element $a' \in A$ with $f(a) < f(a')$. From this it follows that

$$f(x) = f(a) + f(b) < f(a') + f(b) = f(a' + b) \leq \sup_{\substack{u \in A \\ v \in B}} f(u+v) = f(x)$$

because $x \in H_f(A+B)$. This implies the inclusion

$$H_f(A+B) \subseteq H_f(A) + H_f(B).$$

The reverse inclusion can be proved in the same way. Assume that $a \in H_f(A)$ and $b \in H_f(B)$. Then $x = a + b \in H_f(A+B)$. Let us assume that this is not true. Then there exists an element $x' = a' + b' \in A + B$ with $f(x) < f(x')$. But this implies:

$$f(a) + f(b) = f(x) < f(x') = f(a') + f(b')$$

and hence $f(a) < f(a')$ or $f(b) < f(b')$ which completes the proof.

The additivity of the convex hull:

Proposition 1.2. *Let X be a vector space and $A, B \subset X$. Then*

$$\text{conv } A + \text{conv } B = \text{conv } (A+B).$$

Proof: First observe that

$$\operatorname{conv} A + B = \bigcup_{b\in B} (\operatorname{conv} A + b) = \bigcup_{b\in B} [\operatorname{conv} (A + b - b) + b]$$
$$\subseteq \bigcup_{b\in B} (\operatorname{conv} [\operatorname{conv} (A + B) - b] + b) = \operatorname{conv} (A + B).$$

Since $A + B \subset \operatorname{conv} A + \operatorname{conv} B$ we have $\operatorname{conv} (A + B) \subset \operatorname{conv} A + \operatorname{conv} B$. Now it follows from the above observation that

$$\operatorname{conv} (A + B) \subseteq \operatorname{conv} A + \operatorname{conv} B$$
$$\subseteq \operatorname{conv} (A + \operatorname{conv} B)$$
$$\subseteq \operatorname{conv} [\operatorname{conv} (A + B)] = \operatorname{conv} (A + B).$$

□

2. THE ORDERED SEMIGROUP OF CONVEX SETS

We state two fundamental properties about closed bounded convex sets in topological vector spaces, namely the *order cancellation law* [13], [16] and *Pinker's formula* [12].

The order cancellation law:

Theorem 2.1. *Let X be a topological vector space. Then for any $A \in \mathcal{A}(X)$, $B \in \mathcal{B}^*(X)$ and $C \in \mathcal{C}(X)$ the inclusion*

$$A + B \subseteq C \dot{+} B \text{ implies } A \subseteq C. \qquad \textbf{(olc)}$$

Proof: Let $\mathcal{U}$ be a base of neighborhoods of zero in the topological vector space X. Given any neighborhood $U \in \mathcal{U}$ we define a sequence $(V_n)_{n\in\mathbb{N}}$ such that:

$$V_0 + V_0 \subseteq U \text{ and } V_{n+1} + V_{n+1} \subseteq V_n.$$

From $A + B \subseteq C \dot{+} B$ it follows that for every $V \in \mathcal{U}$ we have

$$A + B \subseteq C + B + V,$$

and hence for every $n \in \mathbb{N}$ we have:

$$A + B \subseteq C + B + V_n.$$

Now let $a \in A$ and $b_1 \in B$. Then

$$a + b_1 = c_1 + b_2 + v_1 \qquad \text{for some} \qquad c_1 \in C, b_2 \in B, v_1 \in V_1,$$
$$a + b_2 = c_2 + b_3 + v_2 \qquad \text{for some} \qquad c_2 \in C, b_3 \in B, v_2 \in V_2,$$

and in general, for every $n \in \mathbb{N}$:

$$a + b_n = c_n + b_{n+1} + v_n \qquad \text{for some} \qquad c_n \in C, b_{n+1} \in B, v_n \in V_n.$$

Hence

$$a = \frac{1}{n}(c_1 + \ldots + c_n) + \frac{1}{n}(b_{n+1} - b_1) + \frac{1}{n}(v_1 + \ldots + v_n), \quad n \in \mathbb{N}$$

and thus by the convexity of C and the boundedness of B we get for sufficiently large $n \in \mathbb{N}$ that

$$a \in C + V_0 + V_1 + \ldots + V_n \subseteq C + U.$$

Thus $A \subseteq C + U$ for every $U \in \mathcal{U}$, and therefore, $A \subset C$.

□

The implication $A + B \subseteq C + B \Rightarrow A \subseteq C$ is called the *order cancellation law* and the weaker implication $A + B = C + B \Rightarrow A = C$ is called the *cancellation law.*

The Pinker formula:

Next we prove an identity for bounded closed convex sets, which was first observed by A. G. Pinsker [12] for locally convex vector spaces and will be called the *Pinsker formula.* For its proof we need the following three lemmas:

Lemma 2.2. *Let X be a vector space and $A, B, C \subset X$ subsets. Then*

$$A \cup B + C = (A + C) \cup (B + C).$$

Proof: For $x \in A \cup B + C$, there exist $c \in C$ and $d \in A \cup B$ such that $x = c + d$. Hence $x \in (A + C) \cup (B + C)$, i.e. $A \cup B + C \subseteq (A + C) \cup (B + C)$.

Conversely, for $x \in (A + C) \cup (B + C)$ there exist elements $c \in C$ and $d \in A$ or $d \in B$ such that $x = c + d$. Hence $x \in A \cup B + C$, i.e. $(A + C) \cup (B + C) \subseteq A \cup B + C$.

□

Lemma 2.3. *Let X be a vector space and $A, B, C \in \mathcal{A}(X)$ and C a convex set. Then*

$$\operatorname{conv}(A \cup B) + C = \operatorname{conv}[(A + C) \cup (B + C)].$$

Proof: From Lemma 2.2 and Proposition 1.2 it follows that

$$\operatorname{conv}[(A + C) \cup (B + C)] = \operatorname{conv}[(A \cup B) + C] = \operatorname{conv}(A \cup B) + C.$$

Lemma 2.4. *Let X be a topological vector space, and $A, B, C \in \mathcal{A}(X)$ and C be a convex set. Then*

$$((A \dot{+} C) \vee (B \dot{+} C)) = C \dot{+} (A \mathbin{\mathring{\vee}} B).$$

Proof: By Lemma 2.3 we have:

$$\begin{aligned} C \dot{+} \operatorname{conv}(A \cup B) &= \operatorname{cl}\,(\operatorname{cl}\,(\operatorname{conv}(A \cup B) + \operatorname{cl}(C)) \\ &= \operatorname{cl}\,(\operatorname{conv}(A \cup B) + C) \\ &= \operatorname{cl}\,\operatorname{conv}((A + C) \cup (B + C)) \\ &= \operatorname{cl}\,\operatorname{conv}(\operatorname{cl}\,((A + C) \cup (B + C))) \\ &= \operatorname{cl}\,\operatorname{conv}(\operatorname{cl}\,(A + C) \cup \operatorname{cl}\,(B + C)), \end{aligned}$$

since for every $D \subseteq X$ we have $\operatorname{cl}\,\operatorname{conv}(D) = \operatorname{cl}\,\operatorname{conv}(\operatorname{cl}\,D)$.

□

This implies the Pinsker formula:

Proposition 2.5. *Let (X, τ) be a topological vector space, $A, B, C \in \mathcal{A}(X)$ and C be a convex set. Then*

$$(A \dot{+} C) \vee (B \dot{+} C) = C \dot{+} (A \vee B) \qquad \text{(Pinsker formula).}$$

From the algebraic point of view the set $\mathcal{B}(X)$ of all nonempty closed bounded convex subsets of a real topological vector space (X, τ), endowed with the Minkowski addition is a commutative semigroup with unit $\mathbf{1} = \{0\}$ (i.e. a set endowed with a group operation, without having inverse elements) with cancellation property which contains $\mathcal{K}(X)$, i.e. the set all nonempty compact convex subsets, as a sub-semigroup. With respect to the order which is given by the inclusion,i.e. for $A, B \in \mathcal{B}(X)$ holds $A \leq B$ if and only if $A \subseteq B$, both semigroups $\mathcal{B}(X)$ and $\mathcal{K}(X)$ are ordered. Obviously, the maximum of two elements $A, B \in \mathcal{B}(X)$ exists and is given by $A \vee B = \operatorname{cl} \operatorname{conv}(A \cup B)$. All together we have

Theorem 2.6. *Let (X, τ) be a topological vector space. Then $(\mathcal{B}(X), \dot{+}, \leq)$ is a commutative ordered semigroup with unit $\mathbf{1} = \{0\}$ which satisfies the order cancellation law and contains $\mathcal{K}(X)$ as a sub-semigroup. Moreover the distributivity law holds for maximum operation and the Minkowski addition.*

3. SEMIGROUPS WITH CANCELLATION PROPERTY

Let $(S, \cdot, \leq)$ be a ordered commutative semigroup. We say that S satisfies the *order cancellation law* if

$$as \leq bs \text{ for some } s \in S, \text{ then } a \leq b \qquad \text{(S1)}$$

holds.

The weaker condition that for $a, b, s \in S$ the equation $as = bs$ implies $a = b$ is called the *the cancellation law*. For $a, b \in S$ we call a a *divisor* of b if there exists an element $c \in S$ with $ac = b$.

Since we will only consider commutative semigroups in this book, the word *"commutative"* will be omitted.

A pair of elements of S, i.e. an element $(a, b) \in S^2 = S \times S$ is called a *fraction* and we write a/b or $\frac{a}{b}$ for the order pair (a, b) i.e. $a/b = \frac{a}{b} = (a, b)$. We call two fractions a/b and c/d *equivalent*

$$\frac{a}{b} \sim \frac{c}{d},$$

($a/b \sim c/d$ for short), if $ad = bc$ holds. Note that this is an equivalence relation on the set S^2 and we denote by

$$[a/b] = \{c/d \in S^2 \mid a/b \sim c/d\} \subseteq S^2$$

the equivalence class which contains $\frac{a}{b}$.

It is well known that

$$\tilde{S} = S^2/_{\sim} = \{[a/b] \mid a/b \in S^2\}$$

is a commutative group with the multiplication defined by

$$[a/b][c/d] = [(a/b)(c/d)] = [ac/bd].$$

The inverse element of $[a/b] \in \tilde{S}$ is $[b/a]$.

Moreover, the ordering "$\leq$" on S can be extended to an ordering on $\tilde{S}$ by:

$$[a/b] \leq [c/d] \Longleftrightarrow ad \leq bc.$$

These definitions are independent of the choice of representatives. We have:

Proposition 3.1. *Let $(S, \cdot, \leq)$ be an ordered semigroup which satisfies the order cancellation law. Then for every $c \in S$, the mapping*

$$h : S \longrightarrow \tilde{S} = S^2/_{\sim} \quad \text{with} \quad s \mapsto [sc/c]$$

is an isomorphic order preserving embedding of S into $\tilde{S}$.

For the investigation of minimal representatives we have to introduce a further ordering "$\preceq$" on S^2. For $a'/b', a/b \in S^2$ we define:

$$a'/b' \preceq a/b \Longleftrightarrow a' \leq a \quad \text{and } b' \leq b.$$

We denote by $a \vee b = \sup\{a, b\}$ and $a \wedge b = \inf\{a, b\}$.

Definition 3.2. An ordered semigroup $(S, \cdot, \leq)$ which satisfies the order cancellation law is called *regular* if the following conditions are satisfied:

If $a \leq b$, then $ac \leq bc$ for every $c \in S$, (S2)

if $a \leq s$, and $b \leq s$ for some $s \in S$, then $a \vee b$ exists, (S3)

if $s \leq a$, and $s \leq b$ for some $s \in S$, then $a \wedge b$ exists, (S4)

if $a \vee b$, exist, then $(a \vee b)c \leq ac \vee bc$ for every $c \in S$. (S5)

For an ordered semigroup $(S, \cdot, \leq)$ which satisfies the order cancellation law we define:

Definition 3.3. A fraction $a/b \in S^2$ is called *minimal*, if for any fraction c/d with $c/d \sim a/b$ and $c/d \preceq a/b$ it follows that $a = c$ and $b = d$.

4. AMOUNT OF MINIMAL FRACTIONS

Let $(S, \cdot, \leq)$ be an ordered semigroup which satisfies the order cancellation law. By $m(S)$ we denote the set of minimal elements in S and by $n(S)$ the set of non-minimal elements. Moreover, by $m(S^2)$ and $n(S^2)$ the set of minimal fractions in S^2 and the set of non-minimal fractions is denoted respectively:

$$m(S) = \{a \in S \mid a \text{ is minimal in } S\},$$
$$n(S) = S \setminus m(S),$$
$$m(S^2) = \{a/b \in S^2 \mid a/b \text{ is minimal fraction}\},$$
$$n(S^2) = S^2 \setminus m(S^2).$$

For a nonempty $T \subseteq S^2$ and an element $x \in S$ let us define the mapping

$$f_x : T \to (x/x)T$$

by

$$f_x(a/b) = ax/bx, \quad \text{where} \quad (x/x)T = \{ax/bx \mid a/b \in T\}.$$

Note that the mapping f_x is injective, since $f_x(a/b) = f_x(c/d)$ implies $ax/bx = cx/dx$ and, by the cancellation law we have $a = c$ and $b = d$.

Proposition 4.1. *Let $(S, \cdot, \leq)$ be an ordered semigroup, which satisfies the order cancellation law. If the set $n(S)$ of non-minimal elements of S is nonempty, then for every $a/b \in S^2$ holds*

$$\text{card } m(S) \leq \text{card } n(S) \leq \text{card } n([a/b]) = \text{card } [a/b],$$

$$\text{card } m([a/b]) \leq \text{card } n([a/b]).$$

Proof: Since $n(S) \neq \emptyset$ there exists $x \in S$ which is not minimal. Since $(x/x)[a/b] \subset [a/b]$, it follows from the injectivity of the function $f_x : [a/b] \to [a/b]$ that card $[a/b] \leq$ card $n([a/b])$. But $n([a/b]) \subset [a/b]$, hence card $n\ [a/b] \leq$ card $([a/b])$ and therefore, card$n\ [a/b] =$ card $([a/b])$.

Define the function $g_{a/b} : S \to [a/b]$ by $g_{a/b}(x) = f_x(a/b)$. Since $g_{a/b}$ is injective, we have card $(S) \leq$ card $[a/b]$. If $m(S) = \emptyset$, then the inequality card $m(S) \leq$ card $n(S)$ is obvious. Now assume that $m(S) = \emptyset$. Then for any $c \in n(S)$ the function $h_c(s) = cs$ maps $m(S)$ injectively into $n(S)$. Hence card $m(S) \leq n(S)$.

The second inequality can be proved similarly. If $m([a/b]) = \emptyset$, then the inequality is obvious. Now suppose that $m([a/b]) = \emptyset$. Take any $x \in n(S)$. Since for $T = m([a/b])$, we have $(x, x)T \subset n([a/b])$ the mapping f_x maps $m([a/b])$ into $n([a/b])$. Since f_x is injective we obtain card $m([a/b]) \leq n([a/b])$.

Proposition 4.2. *Let $(S, \cdot, \leq)$ be an ordered semigroup, which satisfies the order cancellation law. If the sets $m(S)$ and $m([a/b])$ are nonempty and if $m(S)$ is a group, then*

$$\text{card } m(S) \leq \text{card } m([a/b]).$$

Proof: We can assume that a/b is a minimal fraction. Take any $s \in m(S)$ and consider the fraction $as/bs \in [a/b]$. Suppose that there exists a fraction $a'/b' \sim as/bs$ such that $a'/b' \preceq as/bs$ holds. This implies that $a' \leq as$ and $b' \leq bs$. By assumption $m(S)$ is a group and therefore, $s^{-1} \in m(S)$. Hence we obtain $a's^{-1} \leq ass^{-1} = a$ and analogously $bs^{-1} \leq b$. It follows from the minimality of the fraction a/b that $a's^{-1} = a$ and $b's^{-1} = b$. Hence $a' = as$ and $b' = bs$ and the fraction as/bs is minimal.

The above calculation show that by $g_{a/b}(s) = as/bs$ an injective mapping $g_{a/b} : m(S) \to m([a/b])$ is defined. Hence card $m(S) \leq$ card $m([a/b])$.

Theorem 4.3. Let $(S, \cdot, \leq)$ be an ordered semigroup, which satisfies the order cancellation law. If the sets of minimal and of non-minimal elements of a semigroup S are nonempty and if *card* $S \notin \mathbb{N}$, then the sets of minimal and non-minimal fractions are of equipotential.

Proof: Take any $x \in n(S)$ and put $T = m(S^2)$. Since $(x/x)m(S^2) \subset n(S^2)$, the assignment $f_x(a/b) = ax/bx$ defines an injective mapping

$f_x : m(S^2) \to n(S^2)$ and therefore, card $m(S^2) \leq$ card $n(S^2)$. Since $n(S^2) \subset S^2$ one has card $n(S^2) \leq$ card S^2.

Now given any $c \in m(S)$. The assignment $g_c(s) = s/c$ defines an injective mapping $g_c : S \to m(S^2)$ and therefore, card $S \leq$ card $m(S^2)$. Since card $S \notin \mathbb{N}$, we have card $S =$ card S^2 (see for instance [5]; Theorem 1 p. 267). Therefore, card $S =$ card $n(S^2) =$ card $m(S^2) =$ card S^2.

5. PAIRS OF CLOSED BOUNDED CONVEX SETS

We will now consider the ordered commutative semigroup $(\mathcal{B}(X), \dot{+}, \leq)$ with unit $1 = \{0\}$ of pairs of nonempty closed bounded convex sets in locally convex topological vector spaces (X, τ). Let us recall that an equivalence relation between pairs $(\mathrm{A}, \mathrm{B}), (C, D) \in \mathcal{B}^2(X)$ of closed bounded convex sets is given by the relation $(\mathrm{A}, \mathrm{B}) \sim (C, D)$ if and only if $A \dot{+} D = B \dot{+} C$ and the ordering in $\mathcal{B}(X)$ is extended to pairs by $(\mathrm{A}, \mathrm{B}) \leq (C, D)$ with $A \subseteq C, B \subseteq D$. From the order cancellation law it follows that "$\sim$" is a relation of equivalence in $\mathcal{B}^2(X)$. The equivalence class $(A, B) \in \mathcal{B}^2(X)$ is denoted by $[A, B]$.

For compact convex sets we have the following result:

Theorem 5.1 *Let (X, τ) be a topological vector space. Then for any pair $(A, B) \in \mathcal{K}^2(X)$ there exists a pair $(C, D) \in [A, B]$ which is minimal.*

Proof: Using the Kuratowski-Zorn Lemma it is sufficient to show that for any totally ordered subset $\Sigma = \{(C, D) \in [A, B] \mid (C, D) \leq (A, B)\}$ of $[A, B]$ there exists an element $(A^*, B^*) \in [A, B]$ such that for any $(C, D) \in \Sigma$ the relation $(A^*, B^*) \leq (C, D)$ holds.

For any $\sigma = (C, D) \in \Sigma$ we will denote by A_σ the set C and by B_σ the set D. The ordering on Σ yields that $\sigma_1 \leq \sigma_2$ if and only if $A_{\sigma_1} \subset A_{\sigma_2}$ and $B_{\sigma_1} \subset B_{\sigma_2}$.

Now we fix $\sigma_0 \in \Sigma$ and define the sets $A^* = \bigcap_{\sigma \in \Sigma_0} A_\sigma$ and $B^* = \bigcap_{\sigma \in \Sigma_0} B_\sigma$, where $\Sigma_0 = \{\sigma \in \Sigma \mid \sigma \leq \sigma_0\}$. By Cantor Intersection Theorem the set A^* is nonempty. Moreover A^* is a closed subset of A_{σ_0} and hence it is compact. The convexity of A^* follows immediately from the convexity of A_σ for $\sigma \in \Sigma_0$. Since the same arguments hold for B^* it follows that $(A^*, B^*) \in \mathcal{K}^2(X)$.

It remains to show that $(A^*, B^*) \in [A, B]$. By definition of the equivalence relation, for any pair $(C, D) \in [A, B]$ and for any $\sigma \in \Sigma_0$ the

equation $A_\sigma + D = B_\sigma + C$ holds. This implies that $A^* + D \subseteq B_\sigma + C$ for every $\sigma \in \Sigma_0$ Hence for any $z \in A^* + D$ and any $\sigma \in \Sigma_0$ we can find a representation of the form $z = b_\sigma + c_\sigma$, where $b_\sigma \in B_\sigma$ and $c_\sigma \in C$. Since the net $\{b_\sigma \mid \sigma \in \Sigma_0\}$ is contained in the compact set B_{σ_0} there exists a subnet $\{b_{\sigma_\delta} \mid \delta \in \Delta\}$ converging to some $b_0 \in B_{\sigma_0}$. Hence for any neighborhood $U(b_0)$ of $b_0 \in B_{\sigma_0}$ there exists an index $\delta_0 \in \Delta$ such that for any $\sigma_\delta \leq \sigma_{\delta_0}$ we have $b_{\sigma_\delta} \in U(b_0)$ and therefore $B_{\sigma_\delta} \cap U(b_0) \neq \emptyset$. Now let $\sigma \in \Sigma_0$ be an arbitrary element. Since the set Σ_0 is totally ordered we have $\sigma_{\delta_0} \leq \sigma$ or $\sigma \leq \sigma_{\delta_0}$. In the first case $\sigma_{\delta_0} \leq \sigma$ we have $B_{\sigma_{\delta_0}} \subseteq B_\sigma$ and hence $B_\sigma \cap U(b_0) \neq \emptyset$. In the other case where $\sigma \leq \sigma_{\delta_0}$. we can find an index $\delta_1 \in \Delta$ such that $\sigma_{\delta_1} \leq \sigma$ and for any $\sigma_\delta \leq \sigma_{\delta_1}$ we have $B_{\sigma_\delta} \cap U(b_0) \neq \emptyset$ and hence $B_\sigma \cap U(b_0) \neq \emptyset$. Thus we have shown that for any neighborhood $U(b_0)$ and any $\sigma \in \Sigma_0$ the set $B_\sigma \cap U(b_0)$ is not empty. Since the sets B_σ are compact, it follows that $b_0 \in B_\sigma$ for any $\sigma \in \Sigma_0$ and consequently $b_0 \in B^*$. The subnet $\{c_{\sigma_\delta} \mid \delta \in \Delta\}$ converges to the point $z - b_0$ which by the compactness of C belongs to C Thus $A^* + D \subseteq B^* + C$ and by a similar argument we get $B^* + C \subseteq A^* + D$. Hence it follows that $(A^*, B^*) \in [A, B]$. The Kuratowski-Zorn Lemma yields now that $[A, B]$ has a minimal element. □

This is not longer true for closed bounded convex sets. Here we have:

Theorem 5.2 *Let (X, τ) be a reflexive locally convex vector space. Then every class $[A, B] \in \mathcal{B}^2(X)_{/\sim}$ contains a minimal element $(C, D) \in [A, B]$.*

Proof: In the case of finite-dimensional vector spaces, bounded closed sets are compact, and the theorem follows from Theorem 5.1. Let us denote by $\tau^* = \sigma(X, X^*)$ the weak topology for X To avoid confusion, we will indicate during this proof the topology under consideration by an index at $\mathcal{B}$ and $\mathcal{K}$. In a relexive locally convex vector space every bounded closed convex set $A \in \mathcal{B}_\tau(X)$ is compact in the topology τ^* and consequently belongs to $\mathcal{K}_{\tau^*}(X)$. Observe that every $A \in \mathcal{K}_{\tau^*}(X)$ is also closed in τ since $\tau^* \subset \tau$. Take any $(A, B) \in \mathcal{B}^2_\tau(X) \subset \mathcal{K}^2_{\tau^*}(X)$. Then

$$A + B \in \mathcal{K}_{\tau^*}(X) \text{ and } A \dot{+} B \in \mathcal{K}_\tau(X).$$

Therefore, the convex set $A+B$ is closed in τ and contained in $A \dot{+} B$, which is a bounded set in X with respect to τ. This implies that $A+B \in \mathcal{B}_\tau(X)$ and consequently $A \dot{+} B = A+B$ holds in all reflexive topological vector spaces (X,τ). Hence $[A,B] \subset [A,B]_{\tau^*} \in \mathcal{K}^2_{\tau^*}(X)_{/\sim}$, where $[A,B]_{\tau^*}$ is the class of equivalent pairs of compact convex sets in the space (X,τ^*) which contains (A,B). According to Theorem 5.1, the equivalence class $[A,B]_{\tau^*}$ contains a minimal element $(C,D) \in \mathcal{K}^2_{\tau^*}(X)$, such that $C \subset A$ and $D \subset B$. Since C,D are closed in τ convex and contained in bounded sets it follows that $(C,D) \in \mathcal{B}^2_\tau(X)$. Moreover, $(C,D) \in [A,B] \subset [A,B]_{\tau^*}$. Therefore, (C,D) is a minimal element in $[A,B]$ and, of course, $(C,D) \leq (A,B)$.

Example 5.3 Let l^∞ be the Banach space of all bounded real sequences endowed with the supremum-norm $\|(x_n)\| = \sup_n |x_n|$ and let c and c_0 be the subspaces of all convergent sequences resp. all sequences convergent to zero of l^∞. Obviously $c_0 \subset c \subset l^\infty$. Note that all three spaces are Banach spaces and that none of them is reflexive.

Let $\mathbb{B}(0,1)$ be the unit ball in c_0 and $A = \{a \in \mathbb{B}(0,1) \mid a_n \geq 0,$ for all $n \in \mathbb{N}\}$. Put $B = -A$ and $A_m = \{a \in A \mid a_1 = \ldots = a_m = \frac{1}{2}\}$ and $B_m = -A_m$ for $m \in \mathbb{N}$. Then $(A_m, B_m) \in \mathcal{B}^2(c_0)$ and $A + B_m = A_m + B$ for all $m \in \mathbb{N}$ and $A + B = \mathbb{B}(0,1)$. Thus (A_m, B_m) is a chain of decreasing pairs in $[A,B]$ i.e.

$$(A,B) \geq (A_1,B_1) \geq \ldots \geq (A_m,B_m) \geq \ldots$$

with an empty intersection, i.e. $\bigcap_m A_m = \bigcap_m B_m = \emptyset$. Now observe that the proof of Theorem 5.1 on the existence of minimal pairs of compact convex sets is based on the Cantor intersection property for compact sets. Therefore, we have:

Theorem 5.4 *For each of the spaces $X = c_0, c,$ and l^∞ there exists a class $[A,B] \in \mathcal{B}^2(X)_{/\sim}$ which contains no minimal element.*

Open question:

The following question remains open: Given any non-reflexive topological vector space X Does there exist an equivalence class $[A,B] \in \mathcal{B}^2(X)_{/\sim}$ which contains no minimal elements?

Next we present sufficient conditions for minimality:

Let (X,τ) be locally convex topological vector space. For $A \in \mathcal{K}(X)$ we consider a set $\mathcal{S} \subseteq X^* \setminus \{0\}$ such that

$$\overline{\text{conv}(\bigcup_{f\in\mathcal{S}} H_f(A))} = A.$$

The sets $\mathcal{S} \subseteq X^* \setminus \{0\}$ of this type can be ordered by inclusion. A minimal element will be called a *shape of* A and will be denoted by $\mathcal{S}(A)$. For a shape $\mathcal{S}(A)$. we consider subsets

$$\mathcal{S}_p(A) := \{f \in \mathcal{S}(A) \mid \text{card}(H_f(A)) = 1\}$$

which may be empty and

$$\mathcal{S}_l(A) := \mathcal{S}(A) \setminus \mathcal{S}_p(A).$$

The criteria presented here are of two different types: The first type of criteria uses conditions which ensure that a two compact convex sets are in a certain *"general position"*, while the second type of criteria uses information about exposed points of the Minkowski sum of compact convex sets.

We begin with a criterium for minimality which is of the first type:

Theorem 5.5 *Let X be a locally convex vector space, and let $A, B \subset X$ be nonempty compact convex sets. Let us assume that there is a shape $\mathcal{S}(A)$ of A which satisfies the following conditions:*

i) for every $f \in \mathcal{S}(A)$, $\text{card}\,(H_f(B)) = 1$,

ii) for every $f \in \mathcal{S}_l(A)$ and every $b \in B$, the condition $\mathcal{S}_l(A) + (b - H_f(B)) \subseteq A$ implies $b = H_f(B)$,

iii) for every $f \in \mathcal{S}_p(A)$, $H_f(A) - H_f(B) \in \mathcal{E}(A - B)$

or conversely, by interchanging A and B.

Then the pair $(A,B) \in \mathcal{K}^2(X)$ is minimal.

Proof: Let us assume that $A' \subseteq A$ and $B' \subseteq B$ are nonempty compact convex sets such that

$$A + B' = B + A'.$$

Choose an element $f \in \mathcal{S}(A)$. Since

$$H_f(A) + H_f(B') = H_f(B) + H_f(A')$$

and since $H_f(B) = \{b\}$, this can be written as

$$H_f(A) + H_f(B') = b + H_f(A').$$

Now choose an element $b' \in H_f(B')$ and determine, for every extreme point $e \in \mathcal{E}(H_f(A))$, an element $a_e \in H_f(A')$ such that

$$e + b' = b + a_e.$$

Now the following two cases are possible:

p) Let us assume that $f \in \mathcal{S}_p(A)$. Then $e - b = a_e - b'$. Since, by condition iii), $e - b \in \mathcal{E}(A - B)$, we have $a_e = e$ and $b' = b$. Hence $H_f(B') = H_f(B) = b$ and therefore,

$$H_f(A') = H_f(A).$$

l) Now we assume that $f \in \mathcal{S}_l(A)$. In this case we have for an arbitrary $b' \in H_f(B')$ that

$$H_f(A) + b' \subseteq b + H_f(A').$$

Therefore,

$$H_f(A) + (b' - b) \subseteq A' \subset A$$

and condition ii) gives $b = b'$. Hence

$$H_f(A') = H_f(A).$$

Thus for all $f \in \mathcal{S}(A)$ we have

$$H_f(A') = H_f(A)$$

and therefore,

$$A' \supseteq \text{cl conv}\left(\bigcup_{f\in S} H_f(A')\right) = \text{cl conv}\left(\bigcup_{f\in S} H_f(A)\right) = A,$$

i.e. $A' = A$. Now from the equality $A + B' = B + A'$ we get by the cancellation law that $B' = B$, which completes the proof. □

The next criterium for minimality is based on a sufficient condition on the indecomposability of a nonempty compact convex set and is formulated in terms of its exposed points. It uses a modified version of the Krein-Milman Theorem [4].

Theorem 5.6 *Let X be a Banach space and let $(A, B) \in \mathcal{K}^2(X)$. If for every exposed point $a + b \in \mathcal{E}_0(A + B)$ with $a \in \mathcal{E}_0(A)$, $b \in \mathcal{E}_0(B)$ there exists $b_1 \in \mathcal{E}_0(B)$ or $a_1 \in \mathcal{E}_0(A)$ such that $a + b_1 \in \mathcal{E}_0(A + B)$ and $a - b_1 \in \mathcal{E}(A - B)$ or $a_1 + b \in \mathcal{E}_0(A + B)$ and $a_1 - b \in \mathcal{E}(A - B)$, then (A, B) is minimal.*

Proof: Let $(A, B) \in \mathcal{K}^2(X)$. By Proposition 1.1 for every $f \in X^*$ holds

$$H_f(A + B) = H_f(A) + H_f(B).$$

This implies the unique representation of every exposed point of $A + B$ as the sum of exposed points of A and B.

Let us show that the pair $(A, B) \in \mathcal{K}^2(X)$ is minimal. Therefore, we choose a pair $(A', B') \in \mathcal{K}^2(X)$ with $A' \subseteq A$, $B' \subseteq B$ and $A + B' = B + A'$. For $a + b \in \mathcal{E}_0(A + B)$ we can assume without loss of generality that for $a \in \mathcal{E}_0(A)$ there exists $b_0 \in \mathcal{E}(B)$ such that $a + b_0 \in \mathcal{E}_0(A + B)$ and $a - b_0 \in \mathcal{E}(A - B)$. Hence there exists a continuous linear functional $f_0 \in X^*$ such that

$$H_{f_0}(A + B) = \{a + b_0\}.$$

By Proposition 1.1 we have $H_{f_0}(A) = \{a\}$ and $H_{f_0}(B) = \{b_0\}$. From

$$A + B' = B + A' = Y$$

it follows that

$$H_{f_0}(A) + H_{f_0}(B') = H_{f_0}(B) + H_{f_0}(A').$$

Hence there exist elements $a' \in H_{f_0}(A') \subseteq A$ and $b' \in H_{f_0}(B') \subseteq B$ such that

$$a + b' = b_0 + a'.$$

Since $a - b_0 \in \mathcal{E}(A - B)$ it follows that $a = a', b_0 = b'$. From the equality $a = a'$ it follows that

$$B + a \subseteq B + A' = Y.$$

Hence

$$a + b \in Y,$$

and since $a + b \in E_0(A + B)$ it follows from V. Klee's modification of the Krein-Milman Theorem (see [4]) that $A + B = Y$.

Hence by the cancellation law we have

$$A + B = A' + B, \quad \text{i.e.} \quad A = A'$$

and

$$A + B = A + B', \quad \text{i.e.} \quad B = B'.$$

Therefore, $(A, B) \in \mathcal{K}^2(X)$. is minimal.

Example 5.7 To illustrate these criteria, we will give two typical examples for $X = \mathbb{R}_2$.

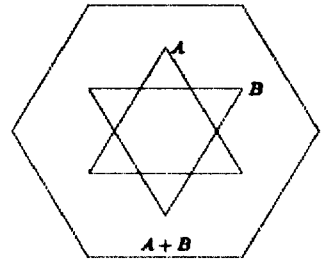


Figure 5.1

i) Let R be a positive real number and put $x = \frac{1}{2}\sqrt{3}R$, $y\frac{1}{2}R$, $a_1 = (0, R)$, $a_2 = (x, -y)$, $a_3 = (-x, -y)$ and let $A = a_1 \vee a_2 \vee a_3$ and $B = -A$. It follows from Theorem 5.5 that the pair (A, B) i.e. the *Star of David* (see Fig. 5.1) is minimal.

ii) Let $R > 0$ be given and define the linear map

$$T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2 \qquad \text{by} \qquad T(x_1, x_2) = (-x_2, x_1).$$

For $x_0 = \left(\frac{1}{2}\sqrt{2}R, 0\right)$ take the balls $K_1 = \mathbb{B}(x_0, R)$, $K_2 = \mathbb{B}(-x_0, R)$. Put $A = K_1 \cap K_2$, $B = T(A)$. Then $A + B = A - B = B((0,0), R)$. It is easy to see that the conditions stated in Theorem 5.6 give the minimality of the pair (A, B) of *orthogonal lenses* (see Fig. 5.2).

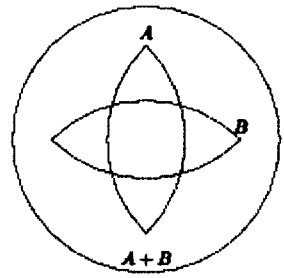


Figure 5.2:

It was proved by S. Scholtes and J. Grzybowski that minimal pairs in the two-dimensional space are unique up to translation. This is not true for higher dimensions as shown by a counter-example of J. Grzybowski (cf. [2],[15]).

In December 2000 S. Rolewicz posed the question, whether the set of equivalent minimal pairs, which are not related by translation may be finite and greater than one.

Recently J. Grzybowski and R. Urbański gave a negative answer to this question.

Theorem 5.8 *Let* (X,τ) *be a topological vector space and* $(A_1,B_1),(A_2,B_2)\in\mathcal{K}^2(X)$ *be two equivalent minimal pairs which are not related by translation. Then there exists a non-countable family* (A_λ,B_λ), $\lambda\in\Lambda$ *of minimal pairs that are all equivalent to* (A_1,B_1) *and no* (A_λ,B_λ) *is a translate of* (A_μ,B_μ) *for* $\lambda\neq\mu$.

6. EXAMPLES

i) Let $X=(X,\tau)$ be a topological vector space and $\mathcal{B}(X)$ be the set of all nonempty closed bounded convex subset of X endowed with the Minkowski sum $\dot{+}$ given by $A\dot{+}B=\mathrm{cl}(A+B)$ and ordered by inclusion, i.e. $A\preceq B\Longleftrightarrow A\subseteq B$. Then $S=(B(X),\dot{+},\preceq)$ is a regular semigroup with unit $1=\{0\}$. For $A,B\in\mathcal{B}(X)$ we have $A\vee B=\mathrm{cl\,conv}(A\cup B)$ and $A\wedge B=A\cap B$ if $A\cap B\neq\emptyset$. Moreover, $m(S)=\{x\mid x\in X\}$ and $n(S)=\{A\mid A\in\mathcal{B}(X)$ with card $A\geq 2\}$.

ii) Let $S=\mathbb{N}$ be the semigroup of integers with the usual addition $+$ and the usual ordering $\leq$. For every $n,m\in\mathbb{N}$ we have $n\vee m=\max\{n,m\}$, $n\wedge m=\min\{n,m\}$, and $n+m=\max\{n,m\}+\min\{n,m\}$. The fraction n/m is minimal if and only if $n=1$ or $m=1$.

iii) Let $S=\mathbb{N}$ the semigroup of integers endowed with the usual multiplication. For $n,m\in\mathbb{N}$ we define $n\leq m$ if n divides m, i.e. $n\mid m$ holds. In this case $n\vee m=w(n,m)$ and $n\wedge m=d(n,m)$, where $w(n,m)$ is the *least common multiple* and $d(n,m)$ is the *greatest common divisor.* Observe that for every $n,m\in\mathbb{N}$ the equation $nm=w(n,m)d(n,m)$ holds. A fraction n/m is minimal if and only if $d(n,m)=1$.

iv) Let $S=\mathbb{N}$ be endowed with the usual multiplication. For $n,m\in\mathbb{N}$ we define $n\leq m$ if $m=n+2k$ for some $k\in\mathbb{N}\cup\{0\}$. If $n/m\in(2\mathbb{N})\times(2\mathbb{N}-1)$, then the fraction n/m is minimal. If $n/m\in(2\mathbb{N}-1)^2$, then the fraction n/m is minimal if and only if $d(n,m)=1$.

v) Let us now consider the *Hilbert semigroup* given by $(S_H,\cdot)$, where $S_H=\{4k+1\mid k\in\mathbb{N}\}$ and $\cdot$ denotes the usual multiplication of numbers. It

is clear that $(S_H, \cdot)$ is a semigroup which satisfies the cancellation law. Now we introduce on S_H the following ordering:

$$a \preceq b \quad \text{if and only if there exists a } \ c \in \mathbb{N} \text{ such that } b = a \cdot c \quad \text{i.e. } a \mid b.$$

Let us denote by $m(S_H) = \{a \in S_H \mid a \text{ is minimal with respect to } \preceq\}$ the $\preceq$-minimal elements of S_H and put $P = \{p \in \mathbb{N} \mid p \neq 2 \text{ and prime}\}$. It follows from a straightforward calculation that P is the union of the following disjoint sets $P_0 = P \cap S_H$ and $P_1 = P \setminus P_0 = \{p \in P \mid p = 4l - 1, \ l \in \mathbb{N}\}$. Let us note that card $P_0 = \aleph_0$. Again by a straightforward calculation we get that for all $p, q \in P_1$ the product $p \cdot q \in S_H$, and that for all $p \in P_0$ and $q \in P_1$ we have $p \cdot q \notin S_H$. The reader can verify that $m(S_H) = P_0 \cup P_1 \cdot P_1$, where $P_1 \cdot P_1 = \{p \cdot q \mid p, q \in P_1\}$. The inclusion $P_0 \cup P_1 \cdot P_1 \subset m(S_H)$ is clear. Now suppose that $4k + 1 \in m(S_H)$. Then let $4k + 1 = p_1 \cdot p_2 \cdot \ldots \cdot p_r$ be a prime factor decomposition of $4k + 1$. Observe that from the decomposition of $P = P_0 \cup P_1$ it follows that for $r \geq 3$ the element $4k + 1$ is not minimal. For $r = 1$ it follows that $4k + 1 \in P_0$. Now assume that $r = 2$. In this case one factor is in P_0 and the other factor is in P_1 and therefore, the product can not be in S_H. Hence $4k + 1 \in P_1 \cdot P_1$.

Now let $p, q \in P_1$ and assume that $p \neq q$. Then obviously

$$(p \cdot p) \cdot (q \cdot q) = (p \cdot q) \cdot (p \cdot q).$$

Next observe that $p \cdot p, \ q \cdot q, \ p \cdot q \in P_1 \cdot P_1 \subset m(S_H)$. Hence

$$\frac{p \cdot p}{p \cdot q} \sim \frac{p \cdot q}{q \cdot q}$$

and we see that both fractions are minimal and that there does not exist an $s \in S_H$ such that

$$\frac{p \cdot p}{p \cdot q} = \frac{s \cdot p \cdot q}{s \cdot q \cdot q}.$$

This means, for example that for $p = 3$ and $q = 7$ the equivalent fractions $\frac{9}{21} \sim \frac{21}{49}$ are minimal. Analogously, this holds for the fractions $\frac{9}{33} \sim \frac{33}{121}$ or $\frac{49}{77} \sim \frac{77}{121}$ etc.

Open question:

It is not known whether for every pair of polytopes there exists also an equivalent minimal pair of polytopes. For instance, it is known that if

$(A,B) \in K^2(\mathbb{R}^n)$, $n = 1,2$, is a minimal pair, which is equivalent to a pair of polytopes, then the sets A and B are also polytopes. Is this still true for dimension $n > 2$?

REFERENCES

[1] V.F. Demyanov and A.M. Rubinov, *Quasidifferential calculus*, Optimization Software Inc., Publications Division, New York, 1986.

[2] J. Grzybowski, *Minimal pairs of compact convex sets, Archiv der Mathematik* **63** (1994), 173-181.

[3] L. Hörmander, *Sur la fonction d' appui des ensembles convexes dans un espace localement convex,* Arkiv för Matematik 3 (1954), 181-186.

[4] V. Klee, *Extremal structure of convex sets II,* Math. Zeitschrift 69 (1958), 90-104.

[5] K. Kuratowski and A. Mostowski, (1966): *Teoria Mnogości,* PWN-Polish Scientific Publishers, Warszawa.

[6] G. Köthe, *Topologische Lineare Räume,* Grundlehren der mathematischen Wissenschaften, Band **107**, Springer Verlag, Berlin, Heidelberg, New York, 1966.

[7] D. Pallaschke, P. Recht, R. Urbański, *On locally Lipschitz quasidifferentiable functions in Banach spaces,* Optimization 17 (1986), 287-295.

[8] D. Pallaschke, S. Scholtes, R. Urbański, *On minimal pairs of compact convex sets,* Bull. Acad. Polon. Sci. Ser. Math. 39 (1991), pp 1-5.

[9] D. Pallaschke and R. Urbański, *Some criteria for the minimality of pairs of compact convex sets,* Zeitschrift für Operations Research 37 (1993), 129-150.

[10] D. Pallaschke and R. Urbański, *Reduction of quasidifferentials and minimal representations, Mathem. Programming,* (Series A) (1994) **66**, 161-180.

[11] Pallaschke, D. and Urbański, R., *Pairs of Compact Convex Sets —Fractional Arithmetic with Convex Sets,* Mathematics and its Applications, Vol. 548, Kluwer Acad. Publ. Dordrecht, 2002.

[12] A. G. Pinsker, *The space of convex sets of a locally convex space,* Trudy Leningrad Engineering-Economic Institute 63 (1966), 13-17.

[13] H. Rådström, *An embedding theorem for spaces of convex sets,* Proc. Amer. Math. Soc. 3 (1952), 165-169.

[14] S. Rolewicz, *Metric linear spaces*, PWN and D.Reidel Publ. Company, Warszawa-Dordrecht, 1984.

[15] S. Scholtes, *Minimal pairs of convex bodies in two dimensions,* Mathematika 39 (1992), 267-273.

[16] R. Urbański, *A generalization of the Minkowski-Rådström-Hörmander Theorem,* Bull. Acad. Polon. Sci. Math. Astr. Phys. 24 (1976), 709-715.

ON GENERALIZED VARIATIONAL INEQUALITIES

B. Panicucci[1] and M. Pappalardo[2]
Dept. of Mathematics, University of Pisa, Pisa, Italy;[1] Dept. of Applied Mathematics, University of Pisa, Pisa, Italy[2]

Abstract: In this paper we illustrate the connections between generalized variational inequalities (GVI) and other mathematical models: optimization, complementarity, inclusions, dynamical systems. In particular, we analyse relationships between existence theorems of solutions of GVI and existence theorems of equilibrium points of inclusions and projected differential inclusions.

Mathematics Subject Classification (2000): 49J40, 47H04, 49J53.

Key words: generalized variational inequalities, differential inclusions, projected differential inclusions.

1. INTRODUCTION

Competitive phenomena in diverse disciplines are often characterized by the specific equilibrium state. Some well-known equilibrium problems are oligopolistic market equilibrium problems, traffic network equilibrium problems, general economic equilibrium problems, spatial price equilibrium problems and so on. In recent years, variational inequality theory has emerged as a very useful tool for the qualitative analysis and computation of various equilibrium problems. The other traditional approach for solving equilibrium models is differential inclusion (DI), in particular when we look

for the constant trajectory. This paper proposes to study the relationships between different mathematical models used for the study of equilibrium theory and in particular the relationships between existence theorems for GVI and for stationary points of DI.

The paper is organized as follows. In Section 2 we present definitions and notations needed in addressing our study. In Section 3 we introduce GVI and complementarity. In Section 4 we state a general existence theorem for GVI and a general existence theorem for inclusions, In Section 5 we explore the connections between these two results. In Section 6 we introduce the projected differential inclusion and relationships with GVI and in Section 7 we deal with algorithms for GVI.

2. PRELIMINARIES

We give in this section some important facts and results which will be needed. The inverse of any multivalued operator always exists and is denoted by $A^{-1}(y) := \{x \in \mathbb{R}^n \mid y \in A(x)\}$. The *domain* and *range* of A are taken to be the sets

$$domA := \{x \mid A(x) \neq \varnothing\}, \qquad rgeA := \{y \mid \exists x \text{ with } y \in A(x)\}.$$

The *graph* of A is

$$graph(A) = \{(x, x^*) \mid x^* \in A(x)\}.$$

Definition 2.1. A set valued map $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is called *upper semicontinuous at* $x \in dom(A)$ if for each open set $V \supseteq A(x)$, there exists a neighborhood U of x such that $A(x) \subseteq V$ for all $x \in U$.

Definition 2.2. A set valued map $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be *monotone on* $K \subseteq \mathbb{R}^n$ if

$$\langle x_1^* - x_2^*, x_1 - x_2 \rangle \geq 0 \quad \forall x_1^* \in A(x_1), x_2^* \in A(x_2), \quad \forall x_1, x_2 \in K;$$

where $\langle \cdot, \cdot \rangle$ denotes the usual inner product in $\mathbb{R}^n$.

Definition 2.3. A set valued map $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be *pseudomonotone on* $K \subseteq R^n$ if for all $x_1, x_2 \in K, x_1^* \in A(x_1), x_2^* \in A(x_2)$,

$$\langle x_1^*, x_2 - x_1 \rangle \geq 0 \Rightarrow \langle x_2^*, x_2 - x_1 \rangle \geq 0.$$

A monotone operator is said to be *maximal* if its graph is not properly contained in the graph of any other monotone operator, in other words, if the following statements are equivalent:

1. For every $(x, x^*) \in graph(A), \langle y^* - x^*, y - x \rangle \geq 0$
2. $y^* \in A(y)$.

The following properties [2] will be useful:

Theorem 2.1.

1. *A^{-1} is maximal monotone if and only if A is maximal monotone.*
2. *Let A_1, A_2 be maximal monotone, then $A_1 + A_2$ is also maximal monotone if $ri(domA_1) \cap ri(domA_2) \neq \varnothing$, where ri stands for relative interior.*
3. *If A is maximal monotone, the images of A, $A(x) \forall x$, are convex and closed, and the graph of A is closed.*

If K is a closed convex set, consider the normal cone operator $N_K = \partial\delta(\cdot \mid K)$ where $\delta(\cdot \mid K)$ is a closed proper convex function definited by $\delta(x \mid K) = 0$ if $x \in K$ and $+\infty$ otherwise, and ∂h denotes the subdifferential of a proper closed convex funcion h. It is well known that N_K is a maximal monotone operator on $\mathbb{R}^n$. In particular, if K is a closed convex subset of a finite dimensional space, then $x \to N_K(x)$ has a closed graph.

We recall now some definition of convex analysis.

Definition 2.4. Let h be a convex function from $\mathbb{R}^n$ to $\mathbb{R} \cup \{\infty\}$. The *epigraph* of h is the convex set

$$epi(h) := \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}, \textit{ such that } h(x) \leq \lambda\}.$$

Definition 2.5. The *tangent cone* $T_K(x)$ to a convex subset K at $x \in K$ is the closed cone spanned by $K - x$, which is convex:

$$T_K(x) = \overline{\bigcup_{h>0} \frac{K - x}{h}}$$

The *polar cone* of the tangent cone $T_K(x)$ to a convex subset K, is called the *normal cone* to K at x and is denoted by

$$N_K(x) := (T_K(x))^+ = \{d \in \mathbb{R}^n : \langle d, z - x\rangle \leq 0, \forall z \in K\}.$$

Definition 2.6. For a convex set K, *the recession cone* K_∞ is a convex closed cone

$$K_\infty = \{d \in \mathbb{R}^n : \overline{x} + td \in cl(K) \quad \forall t \geq 0\},$$

with $\overline{x}$ any element of K and where $cl(K)$ is the closure of K.

For a closed and proper convex function $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, $epi(h)$ is a convex closed set and $(epi(h))_\infty$ is a closed convex cone of $\mathbb{R}^n \times \mathbb{R}$ and it itself an epigraph:

$$(epi(h))_\infty = epi(h_\infty)$$

with $h_\infty(x) = inf\{\alpha : (x, \alpha) \in (epi(h))_\infty\}$. h_∞ is called the recession function of h.

Theorem 2.2. [7] *Let K be a closed convex cone. For the three element x, x_1, x_2 in $\mathbb{R}^n$, the properties below are equivalent:*

1. $x = x_1 + x_2$ *with* $x_1 \in K, x_2 \in K^+$ *and* $\langle x_1, x_2\rangle = 0$,
2. $x_1 = Pr_K(x)$ *and* $x_2 = Pr_{K^+}(x)$,

where Pr_K is the operator projection onto the set K.

3. GVI AND COMPLEMENTARITY

The *variational inequality problem* $VI(K, F)$ is a problem of finding $x^* \in K$ such that

$$\langle F(x^*), y - x^*\rangle \geq 0 \text{ for all } y \in K,$$

where F is a map from $\mathbb{R}^n$ into $\mathbb{R}^n$, and K is a nonempty, closed and convex subset of $\mathbb{R}^n$. In what follows, we consider the case where the set K is defined by

$$K = \{x \in \mathbb{R}^n : g(x) \leq 0, h(x) = 0\}, \tag{1}$$

where the given functions $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^s$ are continuously differentiable. To solve $VI(K, F)$ there are several approaches. One of them

transforms $VI(K,F)$ in a system of nonlinear equations, via a generalization of the KKT conditions, or in a complementarity problem. Let us show this briefly. A KKT-system similar to the KKT optimality conditions for the standard non linear program has been formulated for $VI(K,F)$ [7], under mild constraint qualification.

So, from now on, we suppose that Mangasarian-Fromovitz constraint qualification (in short, MFCQ) holds.

Theorem3.1. *If $x \in K$ solves $VI(K,F)$ then there exist multipliers (μ,λ) such that*

$$F(x) + \lambda^T \nabla g(x) + \mu^T \nabla h(x) = 0;$$
$$\lambda \geq 0, \qquad \lambda^T g(x) = 0.$$

Definition 3.1. Let C be a convex cone in $\mathbb{R}^n$ and let T be a mapping from $\mathbb{R}^n$ into itself. The *complementarity problem,* denoted by CP(C, T), is to find a vector $x^* \in K$ such that

$$T(x^*) \in C^* \text{ and } \langle T(x^*), x^* \rangle = 0,$$

where C^* denotes the *polar cone* of C, i.e.

$$C^* = \{y \in R^n : \langle y, x \rangle \geq 0, \forall x \in C\}.$$

$VI(K,F)$ can be converted into a complementarity problem. The following result which summarizes this conversion has been used in different contexts by several authors.

Theorem 3.2. *If x solves $VI(K,F)$ then for some $\lambda \in R^m$ and $\mu \in R^s$, x solves the $CP(\mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^s, H)$ where $H : \mathbb{R}^{n+m+s} \to \mathbb{R}^{n+m+s}$ is defined* by

$$H(x,\lambda,\mu) = \begin{pmatrix} F(x) + \lambda^T \nabla g(x) + \mu^T \nabla h(x) \\ -g(x) \\ h(x) \end{pmatrix}.$$

Now we pass to GVI.

Definition 3.2. Given $F : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$, the *GVI problem* denoted by $GVI(K, F)$ consists of finding $x \in K$ such that there exists $x^* \in F(x)$, satisfying

$$\langle x^*, y - x \rangle \geq 0, \quad \forall y \in K.$$

From now on F denotes a multivalued function. The KKT conditions for VI have been generalized to GVI.

Theorem 3.3. *If $x \in K$ solves $GVI(K, F)$ then there exist multipliers $\lambda = (\lambda_1, \cdots, \lambda_m)$ and $\mu = (\mu_1, \cdots, \mu_s)$ such that*

$$0 \in F(x) + \sum_i \lambda_i \nabla g_i(x) + \sum_j \mu_j \nabla h_j(x),$$
$$\lambda_i g_i(x) = 0 \qquad i = 1, \cdots, m, \qquad \lambda_i \geq 0.$$

Proof. We note that $GVI(K, F)$ is equivalent to the problem

$$\min_{y \in K} J(y), \qquad J(y) = \langle x^*, y - x \rangle$$

We observe that J is linear in y so, applying Kuhn-Tucker theorem, necessary condition of minimality of J on K, is that there exist vectors $\lambda = (\lambda_1, \cdots, \lambda_m)$ and $\mu = (\mu_1, \cdots, \mu_s)$ such that

$$0 = x^* + \lambda^T \nabla g(x) + \mu^T \nabla h(x),$$

where $x^* \in F(x)$, and then

$$0 \in F(x) + \lambda^T \nabla g(x) + \mu^T \nabla h(x).$$

□

We can also generalize Theorem 3.2.

Definition 3.3. Given $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$, the generalized complementarity problem ($GCP(C, T)$) over a convex cone C consists of finding $x \in R^n$ such that there exists $x^* \in T(x)$ satisfying

$$x^* \in C^*, \quad x \in C, \quad \langle x^*, x \rangle = 0$$

$GVI(K,F)$ can be converted into a GCP.

Theorem 3.4. *If* $x \in K$ *solves* $GVI(K,F)$ *then, for some* λ *and* μ, x *solves* $GCP(\mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^s, H)$ *where*

$$H(x,\lambda,\mu) = \begin{pmatrix} F(x) + \lambda^T \nabla g(x) + \mu^T \nabla h(x) \\ -g(x) \\ h(x) \end{pmatrix}.$$

Proof. $GCP(\mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^s, H)$ consists of finding (x,λ,μ) with $\lambda \geq 0$ such that there exists $(u^*, z^*, w^*) \in H(x,\lambda,\mu)$, i.e. $u^* = x^* + \lambda^T \nabla g(x) + \mu^T \nabla h(x)$, with $x^* \in F(x), z^* = -g(x), w^* = h(x)$, satisfying

$$(u^*, z^*, w^*) \in (\mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^s)^* = \{0\} \times \mathbb{R}^m_+ \times \{0\},$$

i.e.

$$0 \in F(x) + \lambda^T \nabla g(x) + \mu^T \nabla h(x), \quad g(x) \leq 0, \quad h(x) = 0,$$

and

$$\langle (u^*, z^*, w^*), (x,\lambda,\mu) \rangle = 0,$$

i.e.

$$\lambda^T g(x) = 0.$$

Thesis follows from Theorem 3.3. □

One drawback with the conversion of a GVI into a GCP is the increase in the number of variables from n to $n + m + s$.

4. EXISTENCE THEOREMS FOR GVI AND DI

We denote by S the set of vectors x that are solutions of GVI. The following classical result holds:

Theorem 4.1. [6] *Assume that*:

1. K *is a nonempty, compact and convex set in* $\mathbb{R}^n$,
2. F *is an upper semicontinuous set valued map on* K,
3. $F(x)$ *is a nonempty, compact and convex set in* $\mathbb{R}^n$, $\forall x \in K$, *Then* S *is nonempty.*

In the case where the set K is not bounded, in order to establish the existence of a solution, we have to consider some additional properties for F, Now we recall an existence result under monotonicity condition.

Theorem 4.2 [5] *Assume that:*

1. K *is a nonempty, closed and convex set in* $\mathbb{R}^n$,
2. F *is an upper semicontinuous set valued map on* K,
3. $F(x)$ *is a nonempty, compact and convex set in* $\mathbb{R}^n$, $\forall x \in K$,
4. F *is* pseudomonotone *on* K.

Then S *is nonempty and compact if and only if*

$$K_\infty \cap (F(K))^+ = \{0\}.$$

Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$; fixed $x_0 \in K$ consider the following $DI(F, K, x_0)$

$$\begin{cases} x'(t) \in F(x(t)) \\ x(0) = x_0 \end{cases}$$

In what follows we shall deal with the existence of an *equilibrium* (or a *stationary solution*) of the dynamical system, i.e., a solution $\overline{x} \in K$ to the inclusion

$$0 \in F(\overline{x}). \tag{2}$$

Theorem 4.3 [1] *Let*

1. K *be a nonempty, convex and compact set in* $\mathbb{R}^n$,
2. F *be an upper semicontinuous set valued map on* K,
3. $F(x)$ *be a nonempty, closed and convex set in* $\mathbb{R}^n$, $\forall x \in K$,
4. *the viability condition*

$$\forall x \in K \qquad F(x) \cap T_K(x) \neq \emptyset$$

hold true.

Then, there exists a solution $\overline{x} \in K$ *to (2).*

So, under the above assumptions, viability implies the existence of an equilibrium. Since a GVI problem can be reduced to an inclusion (see section 5) the following theorem will be useful:

Theorem 4.4. [2] *Let*

1. K *be a nonempty, convex, compact set in* $\mathbb{R}^n$,
2. F *be an upper semicontinuous set valued map on* K,
3. $F(x)$ *be a nonempty, compact and convex set in* $\mathbb{R}^n$, $\forall x \in K$.

Then there exists a solution $\overline{x} \in K$ *to the inclusion*

$$0 \in F(\overline{x}) - N_K(\overline{x}).$$

When K is no longer compact, we consider some monotonicity assumption and we have the following theorem:

Theorem 4.5. *Given a maximal monotone operator* $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, *the solution set* $A^{-1}(0)$ *of the inclusion* $0 \in A(x)$ *is closed and convex. Moreover* $A^{-1}(0)$ *is nonempty and bounded if and only if* $0 \in int(domA^{-1}) = int(rgeA)$.

We can observe that when A is a maximal monotone map, then $F = -A$ satisfies the viability condition when K is the domain of A.

5. RELATIONSHIPS BETWEEN GVI AND DI

In this section, we investigate a way of establishing the existence of a solution to GVI via the DI and viceversa.

Lemma 5.1. *The problem* $GVI(K, -F)$ *i.e.:*

1. $\overline{x} \in K$,
2. $\exists \overline{x}^* \in F(\overline{x})$ *such that* $\langle \overline{x}^*, y - \overline{x} \rangle \leq 0, \quad \forall y \in K.$

is equivalent to the inclusion

$\overline{x} \in K$ *such that* $0 \in F(\overline{x}) - N_K(\overline{x})$.

Proof. It follows from the definition of the normal cone to K at $\overline{x}$, $N_K(\overline{x})$. □

Since $T_K(x)$ is the polar cone of $N_K(x)$, Theorem 2.2 implies that any element $x^* \in F(x)$ decomposes into the form $x^* = t + n$ where $t \in T_K(x)$,

$n \in N_K(x)$ and $\langle t, n\rangle = 0$. Thus, for any $x^* \in F(x)$, the element $x^* - n = t$ belong to $(F(x) - N_K(x)) \cap T_K(x)$, which shows that the set valued map $F - N_K$ satisfies the viability condition.

The only trouble is that when F is upper semicontinuous, $F - N_K$ does not inherit this property; hence we cannot apply Theorem 4.3 to deduce the existence of solutions of $GVI(K, F)$. We can overcome this drawback, considering the following theorem, considered in [10] :

Theorem 5.1. *Let K, F as in Theorem 4.1. Let $m(F(x))$ denotes the element of minimal norm of $F(x)$, i.e.*

$$m(F(x)) = Pr_{F(x)}(0),$$

and

$$c = \sup_{x \in K} \|m(F(x))\|$$

Let B the unit ball and

$$H(x) = F(x) - (cB \cap N_K(x)).$$

Then, there exists an equilibrium $x^* \in K$ such that $0 \in H(x^*)$.

But $H(x) \subseteq F(x) - N_K(x), \forall x$, so under condition of Theorem 4.1 there exists $x^* \in K$, such that $0 \in F(x^*) - N_K(x^*)$, i.e. x^* solves $GVI(K, F)$. So this theorem gives us a direct relationship between Theorem 3.1 and Theorem 3.3.

Now we investigate the relation between Theorem 4.2 for GVI and Theorem 4.5 for DI.

In what follows we consider the case where $-F(x)$ is a maximal monotone set valued map. This includes the case where $F(x) = -\partial f(x)$ is the subdifferential of a convex lower semi-continuous function. We are, obviously, interested in the inclusion $0 \in F(x) - N_K(x)$. The following theorem [10] gives us the desired relationship:

Theorem 5.2. $(-F + N_K)^{-1}(0)$ *is nonempty and bounded if and only if*

$$K_\infty \cap (F(K))^+ = \{0\}.$$

6. PROJECTED DIFFERENTIAL INCLUSION

Relationships between VI and dynamical systems have been recently developed in literature (see [8,9,11]). If we want to study dynamical behaviour in the framework of GVI we propose, in this section, to consider projected differential inclusions. Consider the $DI(F,K,x_0)$; it is known [1] that a necessary and sufficient condition for a trajectory of this differential inclusion to remain in K is that F satisfies the viability condition

$$\forall x \in K, \qquad F(x) \cap T_K(x) \neq \emptyset.$$

When this assumption is no longer satisfied we can replace $F(x)$ by its projection onto the tangent cone $T_K(x)$, and we consider the so called *projected differential inclusion* (see also [12]):

$$\begin{cases} x'(t) \in Pr_{T_K(x)}F(x(t)) \\ x(0) = x_0 \end{cases}$$

where $Pr_{T_K(x)}(F(x)) := \bigcup_{x^* \in F(x)} Pr_{T_K(x)}(x^*)$.

Theorem 6.1. *The solutions to the inclusion* $0 \in Pr_{T_K(x)}F(x)$ *are the solutions to the inclusion* $0 \in F(x) - N_K(x)$ *and conversely.*

Proof. It is known that $Pr_{T_K(x)}(F(x)) \subseteq F(x) - N_K(x)$. So solutions to the projected inclusion are solutions to the differential VI.

It remains to prove that any solution to the differential VI is a solution of the projected inclusion .

Suppose $\overline{x} \in K$ is such that $0 \in F(\overline{x}) - N_K(\overline{x})$ then, there exists $\overline{x^*} \in F(\overline{x})$ such that $\overline{x^*} \in N_K(\overline{x})$. Since the normal cone is the polar of the tangent cone,

$$\overline{x^*} \in N_K(\overline{x}) \text{ if and only if } \left\langle \overline{x^*}, z \right\rangle \leq 0 \qquad \forall z \in T_K(\overline{x})$$

i.e. $\left\langle \overline{x^*} - 0, z - 0 \right\rangle \leq 0$, hence 0 is the projection of $\overline{x^*}$ onto $T_K(\overline{x})$, i.e. $0 \in Pr_{T_K(\overline{x})}(F(\overline{x}))$. □

7. COMMENTS ON METHODS AND ALGORITHMS

An important problem in GVI is the development of an efficient iterative algorithm to compute solutions.

Looking at Theorem 3.3 we will see that we are interested in solving the inclusion

$$0 \in A + B \tag{3}$$

where $A = F$ and $B = \sum_i \lambda_i \nabla g_i(x) + \sum_j \mu_j \nabla h_j(x)$. We shall assume that A and B are maximal monotone on K. A is maximal monotone if and only if its *resolvents* $J_\lambda^A = (I + \lambda A)^{-1}$ with $\lambda > 0$ is a single valued nonexpansive map from H into H.

In the case of linear operators A and B, there is a standard algorithm [4] for solving (3):

$$z_{k+1} = (I + \lambda B)^{-1}(I - \lambda A)z_k, \tag{4}$$

which converges to the solution for λ sufficiently small if A is Lipschitz continuous. When A and B are set valued map we can try to generalize this procedure letting:

$$z_{k+1} = (I + \lambda_k H_k^{-1} B)^{-1}(I - \lambda_k H_k^{-1} A)z_k \tag{5}$$

i.e.

$$z_{k+1} = (H_k^{-1}(H_k + \lambda_k B))^{-1} H_k^{-1}(H_k - \lambda_k A)z_k = (H_k + \lambda_k B)^{-1}(H_k - \lambda_k A)z_k.$$

First we observe that if $H_k = I$, we have $(I + \lambda_k B)^{-1}$, which is a single valued map. But in general, $(I + \lambda_k H^{-1} B)^{-1}$ has not this property.

Because A and B are set valued map, we need to make precise the definition of the algorithm (5).

For $z_0 \in dom(A)$ given, we choose $a_0 \in A(z_0)$ and set $w_0 = H_0 z_0 - \lambda_0 a_0$, then $z_1 = (H_0 + \lambda_0 B)^{-1} w_0$, i.e. $w_0 \in H_0 z_1 + \lambda B(z_1)$, and so on.

When, for simplicity we choose $H_k = 0, \lambda_k = \lambda$, we have:

$$z_{k+1} = (I + \lambda N_K)^{-1}(I + \lambda F)z_k.$$

We can observe that $(I+\lambda N_K)^{-1} = Pr_K$. In effect $(I+\lambda N_K)^{-1} y = x$ which is equivalent to $y \in x + \lambda N_K(x)$, i.e. $\langle y-x, v-x\rangle \leq 0$, for all $v \in K$ i.e. $x = Pr_K y$. So $z_{k+1} = Pr_K(I+\lambda F)z_k$.

Remark 7.1. *When* $F = \nabla f$, *we have* $z_{k+1} = Pr_K(z_k - \nabla f(z_k))$ *and we obtain the projected gradient method.*

Remark 7.2. *For the general case when* $A = 0$,

$$z_{k+1} = (I+\lambda B)^{-1} z_k$$

which is the classical proximal point algorithm. And if $B = 0$ $z_{k+1} = (I+\lambda A)z_k$. *So, if* A *is single valued and* $A = \nabla f$, $z_{k+1} = (I-\lambda\nabla f)z_k$, *i.e. the steepest decent method.*

Remark 7.3. *If we consider the typical Fenchel problem:*

$$\min f(x) + g(Dx)$$

we can write optimality condition

$$0 \in \partial f(\overline{x}) + D^T \partial g(D\overline{x}).$$

Putting $\overline{y} \in \partial g(D\overline{x})$ *we have*

$$0 \in \partial f(\overline{x}) + D^T \overline{y}$$

and therefore

$$D\overline{x} \in \partial g^{-1}(\overline{y})$$

or

$$D\overline{x} \in \partial g^*(\overline{y})$$

(g^* is the Fenchel-conjugate of g).

The inclusion is:

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & D^T \\ -D^T & 0 \end{bmatrix} \begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix} + \begin{bmatrix} \partial f(\overline{x}) \\ \partial g(\overline{y}) \end{bmatrix} \tag{6}$$

Putting $\overline{z} = (\overline{x}, \overline{y})$, *from (6) we get*

$$0 \in T_1(\overline{z}) + T_2(\overline{z}),$$

where $T_1 = \begin{bmatrix} 0 & D^T \\ -D^T & 0 \end{bmatrix}$ and $T_2 = \begin{vmatrix} \partial f \\ \partial g^* \end{vmatrix}$.

REFERENCES

[1] Aubin J. P., Cellina A., *Differential inclusion*. Springer-Verlag, 1984.

[2] Aubin J. P., Frankowska H., *Set-Valued Analysis*. Birkhauser Boston, 1990.

[3] Aubin J. P., *Optima and Equilibria*. Springer-Verlag, 1993.

[4] Auslender A., Teboulle M., *Lagrangian duality and related multiplier methods for variational inequality problems*. Siam Journal of Optimization Vol. 10, No 4 (2000), 1097-1115.

[5] Crouzeix J. P., *Pseudomonotone variational inequality problems: existence of solutions*. Mathematical Programming 78 (1997), 305-314.

[6] Fang S. C., Peterson E. L., *Generalized Variational Inequalities*. Journal of Optimization Theory and Applications Vol. 38, No. 3 (1982), 363-383.

[7] Hiriart-Urruty J.B., Lemarechal C., *Convex analysis and minimization algorithms*. Springer-Verlag, 1993.

[8] Kinderlehrer D., and Stampacchia G., *An Introduction to Variational Inequality and Their Application*, Academic Press, New York, New York, 1980.

[9] Nagurney A., and Zhang D., *Projected Dynamical Systems and Variational Inequalities with Applications*, Kluwer Academic Publishers, Dordrecht, Holland, 1996.

[10] Panicucci B., and Pappalardo M., *Generalized variational inequalities and differential inclusions*, Technical Report, Department of Applied Mathematics, University of Pisa, 2003.

[11] Pappalardo M. and Passacantando M., *Stability for Equilibrium Problems: from Variational Inequalities to Dynamical Systems*, Journal of Optimization Theory and Applications, Vol.113, n.3 pp.567-582.

[12] Pappalardo M., *Equilibrium in Variational Inequalities, Games and Dynamical Systems*. In Optimization in Economics, Finance and Industry, Datanova, 47-60.

[13] Rockafellar R. T., Wets R. J-B., *Variational Analysis*. Springer, 1998.

[14] Xia Y.S., and Wang J., *On the Stability of Globally Projected Dynamical Systems*, Journal of Optimization Theory and Applications, Vol. 106, pp. 129-150, 2000.

[15] Yao J.-C., *Variational Inequalities with Generalized Monotone Operators*, Mathematics of Operations Research, Vol. 19, pp. 691-705, 1994.

BOUNDED (HAUSDORFF) CONVERGENCE: BASIC FACTS AND APPLICATIONS

Jean-Paul Penot[1] and Constantin Zălinescu[2]
Laboratoire de Mathématiques appliquéès, Faculté des Sciences, PAU, France ;[1] Faculty of Mathematics, University "Al. I.Cuza" Iasi, Iasi, Rumania[2]

Abstract: We present a survey of some uses of a remarkable convergence on families of sets or functions. We evoke some of its applications and stress some calculus rules. The main novelty lies in the use of a notion of "firm" (or uniform) asymptotic cone to an unbounded subset of a normed space. This notion yields criteria for the study of boundedness properties.

AMS 2000 Subject Classification : 54A20, 54B99, 26B12

Key words: Apart sets, asymptotic cone, asymptotic function, boundedness, bounded convergence, closedness, convergence, convolution, expansive map, Fenchel transform, firm asymptotic cone.

1. INTRODUCTION

It is the purpose of this paper to survey some properties of a convergence on sets and functions which has received a great deal of interest during the last two decades. We review some of its applications and show why this convergence is convenient. However, we leave apart the application to Hamilton–Jacobi equations which are dealt with in [63]. We also observe that when restricted to the space of continuous linear functions on a normed vector space X the convergence we consider reduces to convergence for the

dual norm; this fact (and the abundance of terminologies) suggests to call this convergence "bounded convergence" or, in short, "b-convergence".

One of the reasons of the success of this convergence lies in its compatibility with the usual operations, provided some technical assumptions reminiscent to constraint qualification conditions in mathematical programming are satisfied. Such conditions already appeared in [44] in the finite dimensional case and in our very first investigations about this question which motivated our interest ([19], [53]); see also [4], [22], [25], [26], [38], [55], [56], [69], [72]. These assumptions involve openness or boundedness conditions. This fact justifies the focus we give to such questions.

The main novelty of the present paper is in the use of a concept of asymptotic cone introduced in [60] which bears some uniformity with respect to directions in a way reminiscent of the uniformity with respect to directions which is involved in the notion of Fréchet derivative (or semi-derivative [45], [50], also called B-derivative) or in the notion of Fréchet cone in the sense of [31], [33]. This concept replaces asymptotic compactness conditions which were used in [62].

As in [62], our methods are essentially geometric. Given an operation $*$ and some sort of variational convergence, in order to prove that $(f_n * g_n) \to f * g$ whenever the sequences of functions (f_n) and (g_n) are such that $(f_n) \to f$, $(g_n) \to g$, we reduce this question to several problems of set convergence: images, intersections, products. Each of these set-theoretical results yields a rule for convergence of functions. In particular, convergence of performance functions and of infimal convolutions are deduced from convergence of images (or sums) of sets. Such a study may have been conducted for other convergences, for instance the ones considered in [4], [9], [20], [24], [38], [42], [70], [72]. However, we believe bounded convergence is appropriate in such a respect and we do not look for completeness.

Other applications could benefit from our analysis. Regularization properties and well-posedness results are already considered in [26], [57]–[59], [61]; more attention could be given to nonconvex cases and to asymptotic methods.

The paper is organized as follows. The next section is devoted to preliminary material about convergences. The main novelties are contained in Section 4: conical enlargements, an expansion property and a notion of disjointness at infinity for non convex sets. Section 4 is also focused on the new notion of firm asymptotic cone to a subset of a normed vector space (n.v.s.). There this tool is applied to boundedness properties. These properties may play a role in obtaining a priori estimates for solving equations. They are crucial for ensuring that convergence properties of

families of sets or functions are preserved under usual operations; a short account of this topic is given in section 5. Such properties are used in [63] to obtain stability and persistence properties of explicit solutions to first order Hamilton–Jacobi equations. Other applications to the convergence of functions are presented in [62] and in [73] where integral functionals and well-posedness questions are considered. In section 3 we evoke some other applications.

2. BOUNDED CONVERGENCE

Throughout this paper, unless otherwise stated, X and Y are real normed vector spaces (n.v.s.), U_X (resp. B_X) is the open (resp. closed) unit ball of X and S_X is the unit sphere in X. The closed (resp. open) ball with center x and radius r is denoted by $B(x,r)$ (resp. $U(x,r)$). For a subset A of X, $intA$, $c\ell A$ stand for the interior and the closure of A respectively. The product space $X\times Y$ is equipped with the max norm. In particular, one has $U_{X\times Y}=U_X\times U_Y$, $B_{X\times Y}=B_X\times B_Y$. The distance of $x\in X$ to a subset E of X is $d(x,E):=\inf\{d(x,w):w\in E\}$, with $d(x,\varnothing):=\infty$. The *remoteness* of E is $d(0,E)$. We denote by $\mathbb{P}$ (resp. $\mathbb{R}_+$) the set of positive (resp. nonnegative) numbers.

Recall (see [3], [13], [24], [69]...) that a sequence (A_n) of subsets of X is said to *converge* to a subset A of X in the sense of *Painlevé–Kuratowski* if $\limsup_n A_n = A = \liminf_n A_n$, where $\limsup_n A_n$ is the set of limits of sequences (x_n) such that $x_k\in A_k$ for k in an infinite subset K of $\mathbb{N}$ and $\liminf_n A_n$ is the set of limits of sequences (x_n) such that $x_n\in A_n$ for each $n\in\mathbb{N}$. We write $(A_n)\to A$. Here we focus our attention to a somewhat stronger notion. It requires the definition of the *excess* of a subset A of X over another subset B of X which is given by

$$e(A,B):=\sup_{a\in A} d(a,B) \quad \text{if } A,B\neq\varnothing,$$

with $e(A,\varnothing)=\infty$ if $A\neq\varnothing$ and $e(\varnothing,B)=0$ for any B. Then, for $p\in\mathbb{P}$, we set

$$e_p(A,B):=e(A\cap pU_X,B),\quad d_p(A,B):=\max(e_p(A,B),e_p(B,A)).$$

It is convenient to write symbolically $A\subset b-\liminf_n A_n$ if, for each $p\in\mathbb{P}$, $\big(e_p(A,A_n)\big)_n\to 0$ as $n\to\infty$ and $A\supset b-\limsup_n A_n$ if $\big(e_p(A_n,A)\big)_n\to 0$ for each $p\in\mathbb{P}$. We write $(A_n)\xrightarrow{\ b\ }A$ and we say that (A_n) *boundedly converges* (or b-converges) to A or that (A_n) converges to A for the

bounded (Hausdorff) topology if $A \subset b-\liminf_n A_n$ and b–$\limsup_n A_n \subset A$. Let us note that $c\ell A \subset \liminf_n A_n$ whenever $A \subset b$–$\liminf_n A$ since then $A \subset \liminf_n A_n$ and since $\liminf_n A_n$ is closed. On the other hand, when $A \supset b$–$\limsup_n A_n$ then $c\ell A \supset \limsup_n A_n$. Thus, we get that $(A_n) \to c\ell A$ when $(A_n) \xrightarrow{b} A$. If X is finite dimensional, the reverse implication holds. The choice of the open unit ball of X in what precedes, rather than the closed unit ball, enables one to use the equalities

$$e_p(c\ell A, B) = e_p(A, B) = e_p(A, c\ell B) = e_p(c\ell A, c\ell B).$$

These equalities show that we could restrict our attention to the case the limit set is closed; then we get uniqueness of the set A such that $(A_n) \xrightarrow{b} A$ and we can write $A = b$–$\lim_n A_n$.

As for other variational convergences, one can pass from these convergences of sets to convergences of functions. Denoting by $\operatorname{epi} f$ the epigraph of f, we set $e_p(f,g) := e_p(\operatorname{epi} f, \operatorname{epi} g)$. Accordingly, for a sequence (f_n) of functions from X to $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ and a function f on X, we write $f \geq b-\limsup_n f_n$ if $\operatorname{epi} f \subset b-\liminf_n (\operatorname{epi} f_n)$ and $f \leq b-\liminf_n f_n$ if $\operatorname{epi} f \supset b$–$\limsup (\operatorname{epi} f_n)$. Of course, writing $(f_n) \xrightarrow{b} f$ when $(\operatorname{epi} f_n) \xrightarrow{b} \operatorname{epi} f$ means that $f \leq b$–$\liminf_n f_n$ and $f \geq b$–$\limsup_n f_n$; we say that (f_n) *b-converges* to f. This type of convergence which has been thoroughly studied in [4]-[6], [8]-[12], [18]-[26], [32], [38], [43], [54]–[58], [68]–[72]... is also called the *Attouch–Wets convergence, the bounded Hausdorff convergence* and the *epidistance convergence*; this last term is justified by the fact that b-convergence on the space $\mathcal{P}_c(X)$ of closed nonempty subsets of X arises from the distance d given by

$$d(A,B) := \sum_{p=1}^{\infty} 2^{-p} \min\{d_p(A,B), 1\}, \qquad A, B \in \mathcal{P}_c(X),$$

where $d_p(A,B) := \max(e_p(A,B), e_p(B,A))$ (see [5], [24]). This convergence has been studied (in Hilbert spaces) in analytical terms through the Moreau regularization in [8]. Pioneering contributions in this vein are due to Choquet, Moreau [46], Mosco [47]; the case of cones is considered in [28], [31], [33], [35].

A convenient way of expressing that a sequence (A_n) of subsets of X b-converges to A is: for any bounded sequence (a_n) of A one has $(d(a_n, A_n)) \to 0$ and for any bounded sequence (a_n) of X such that $a_n \in A_n$ for n large enough one has $(d(a_n, A)) \to 0$ (see [71]).

The following result shows how natural bounded convergence is; it also justifies the simplification of terminology we suggest.

Proposition 1. *Let* $f, f_n \in X^*$ $(n \in \mathbb{N})$. *Then*

$$\operatorname{epi} f \subset b\text{–}\liminf(\operatorname{epi} f) \Leftrightarrow \|f - f_n\| \to 0 \Leftrightarrow \operatorname{epi} f = b\text{–}\lim(\operatorname{epi} f_n).$$

Proof. Assume that $\operatorname{epi} f \subset b\text{–}\liminf(\operatorname{epi} f_n)$. Let $0 < \varepsilon < \rho < 1$. For every $n \in \mathbb{N}$ there exists $x_n \in U_X$ such that $\rho\|f_n\| \le \langle x_n, f_n\rangle$. Because the sequence $((x_n, \langle x_n, f\rangle))$ is bounded, it follows that $d((x_n, \langle x_n, f\rangle), \operatorname{epi} f_n) \to 0$. Hence there exists $n_\varepsilon \in \mathbb{N}$ such that for every $n \ge n_\varepsilon$ there exists $(u_n, t_n) \in \operatorname{epi} f_n$ with $\|x_n - u_n\| \le \varepsilon$ and $|\langle x_n, f\rangle - t_n| \le \varepsilon$. It follows that

$$\begin{aligned}\rho\|f_n\| &\le \langle x_n, f_n\rangle \le \langle x_n, f_n\rangle - \langle u_n, f_n\rangle + t_n - \langle x_n, f\rangle + \langle x_n, f\rangle \\ &\le \|f_n\| \cdot \|x_n - u_n\| + \varepsilon + \|f\| \le \varepsilon\|f_n\| + \varepsilon + \|f\|,\end{aligned}$$

and so $(\rho - \varepsilon)\|f_n\| \le \varepsilon + \|f\|$ for $n \ge n_\varepsilon$. Hence $(\rho - \varepsilon)\limsup\|f_n\| \le \varepsilon + \|f\|$. As ε and ρ are arbitrary such that $0 < \varepsilon < \rho < 1$, we obtain that $\limsup\|f_n\| \le \|f\|$. Now, let $(\rho_n) \uparrow 1$ and $(x_n) \subset U_X$ be such that $\rho_n\|f_n - f\| \le (f_n - f)(x_n)$ for every n. Once again, because the sequence $((x_n, \langle x_n, f\rangle))$ is bounded, we have that $d((x_n, \langle x_n, f\rangle), \operatorname{epi} f_n) \to 0$; there exists $((u_n, t_n)) \subset X$ such that $\langle u_n, f_n\rangle \le t_n$ for every n, $(\|x_n - u_n\|) \to 0$ and $(\langle x_n, f\rangle - t_n) \to 0$. But

$$\begin{aligned}\rho_n\|f_n - f\| &\le \langle x_n, f_n - f\rangle \le \langle x_n, f_n\rangle - \langle u_n, f_n\rangle + t_n - \langle x_n, f\rangle \\ &\le \|f_n\| \cdot \|x_n - u_n\| + t_n - \langle x_n, f\rangle.\end{aligned}$$

Since (f_n) is bounded, it follows that $(\|f_n - f\|) \to 0$. Assume now that $(\|f_n - f\|) \to 0$. Let $((x_n, t_n)) \subset \operatorname{epi} f$ be bounded; in particular, (x_n) is bounded. Let $s_n := \max\{t_n, \langle x_n, f_n\rangle\}$; of course, $(x_n, s_n) \in \operatorname{epi} f_n$. Then

$$\begin{aligned}d((x_n, t_n), \operatorname{epi} f_n) &\le \|(x_n, t_n) - (x_n, \Delta_n)\| = s_n - t_n = (\langle x_n, f_n\rangle - t_n)_+ \\ &\le (\langle x_n, f_n\rangle - \langle x_n, f\rangle)_+ + (\langle x_n, f\rangle - t_n)_+ \le \|f - f_n\| \cdot \|x_n\| \to 0.\end{aligned}$$

Hence $\operatorname{epi} f \subset b\text{–}\liminf(\operatorname{epi} f_n)$. Let now $((x_n, s_n))$ be bounded such that $(x_n, s_n) \in \operatorname{epi} f_n$ for every n; in particular (x_n) is bounded. Let $t_n := \max\{s_n, \langle x_n, f\rangle\}$; of course, $(x_n, t_n) \in \operatorname{epi} f$. Then

$$d((x_n,s_n),\operatorname{epi} f) \le \|(x_n,s_n)-(x_n,t_n)\| = t_n - s_n = \big(\langle x_n, f\rangle - s_n\big)_+$$
$$\le \big(\langle x_n, f\rangle - \langle x_n, f_n\rangle\big)_+ + \big(\langle x_n, f_n\rangle - s_n\big)_+ \le \|f - f_n\|\cdot\|x_n\| \to 0.$$

Hence $\operatorname{epi} f \supset b - \limsup(\operatorname{epi} f_n)$. □

The preceding result can be transposed to a somewhat more general (and in fact different) case. Here b-convergence of a sequence of operators means b-convergence of their graphs and $e_p(S,T) := e_p(\operatorname{gph} S, \operatorname{gph} T)$.

Proposition 2. *Let X,Y be normed vector spaces and $T,T_n : X \to Y$ $(n \in \mathbb{N})$ be continuous linear operators. Then*

$$\operatorname{gph} T \subset b - \liminf(\operatorname{gph} T_n) \Leftrightarrow \|T_n - T\| \to 0 \Leftrightarrow \operatorname{gph} T = b\text{–}\lim(\operatorname{gph} T_n).$$

Proof. As elsewhere in the paper, the product space $X \times Y$ is endowed with the box norm. Assume that $\operatorname{gph} T \subset b - \liminf(\operatorname{gph} T_n)$. Let $0 < \varepsilon < \rho < 1$. For every $n \in \mathbb{N}$ there exists $x_n \in U_X$ such that $\rho\|T_n\| \le \|T_n x_n\|$. Because the sequence $((x_n, Tx_n))$ is bounded, it follows that $d((x_n,Tx_n),\operatorname{gph} T_n) \to 0$. Hence there exists $n_\varepsilon \in \mathbb{N}$ such that for every $n \ge n_\varepsilon$ there exists $u_n \in X$ with $\|x_n - u_n\| \le \varepsilon$ and $\|Tx_n - T_n u_n\| \le \varepsilon$. It follows that

$$\rho\|T_n\| \le \|T_n x_n\| \le \|T_n x_n - T_n u_n\| + \|T_n u_n - Tx_n\| + \|Tx_n\|$$
$$\le \|T_n\|\cdot\|x_n - u_n\| + \|T_n u_n - Tx_n\| + \|T\| \le \varepsilon\|T_n\| + \varepsilon + \|T\|,$$

and so $(\rho - \varepsilon)\|T_n\| \le \varepsilon + \|T\|$ for $n \ge n_\varepsilon$. Hence $(\rho - \varepsilon)\limsup\|T_n\| \le \varepsilon + \|T\|$. Since ε and $1-\rho$ are arbitrarily close to 0, we obtain that $\limsup\|T_n\| \le \|T\|$. Now, let $(\rho_n) \uparrow 1$ and (x_n) in U_X be such that $\rho_n\|T_n - T\| \le \|(T_n - T)x_n\|$ for every n. Once again, because the sequence $((x_n,Tx_n))$ is bounded, we have that $d((x_n,Tx_n),\operatorname{gph} T_n) \to 0$; there exists $(u_n) \subset X$ such that $(\|x_n - u_n\|) \to 0$ and $(\|Tx_n - T_n u_n\|) \to 0$. But

$$\rho_n\|T_n - T\| \le \|(T_n - T)x_n\| \le \|T_n x_n - T_n u_n\| + \|T_n u_n - Tx_n\|$$
$$\le \|T_n\|\cdot\|x_n - u_n\| + \|T_n u_n - Tx_n\|.$$

Since (T_n) is bounded, it follows that $(\|T_n - T\|) \to 0$.

Now assume that $(\|T_n - T\|) \to 0$. Let $((x_n,Tx_n))$ be bounded (or equivalently, (x_n) be bounded). Then

$$d((x_n,Tx_n),\text{gph } T_n) \le \|(x_n,Tx_n)-(x_n,T_nx_n)\| = \|Tx_n - T_nx_n\| \le \|T-T_n\|\cdot\|x_n\| \to 0.$$

Hence $\text{gph } T \subset b\text{–lim inf}(\text{gph } T_n)$. Let now $((x_n,T_nx_n))$ be bounded (or equivalently, (x_n) be bounded). Then

$$d((x_n,T_nx_n),\text{gph } T) \le \|(x_n,T_nx_n)-(x_n,Tx_n)\| = \|T_nx_n - Tx_n\| \le \|T_n-T\|\cdot\|x_n\| \to 0.$$

Hence $\text{gph } T \supset b\text{–lim sup}(\text{gph } T_n)$. □

As noted in [62], b-convergence is a stringent condition. Therefore, it may be advisable to use compromises with weaker convergence notions, as done in [53], [4], [38]. For simplicity, we do not do that here.

3. APPLICATIONS

We devote the present section to some illustrations of the uses of bounded convergence; we just give a sample. We refer to [6], [7], [22], [24], [30], [49], [63], [73] for other applications.

3.1 Reinforced tangency

In [2] and its references, approximations of a subset E of a n.v.s. X around one of its points are considered. Outer firm approximations C of E at $e \in E$ are obtained in requiring that

$$C \supset b\text{–}\limsup_{t\to 0_+} \frac{1}{t}(E-e).$$

Clearly, such a set C, when closed, contains the tangent cone $T(E,e) = \limsup_{t\to 0_+} t^{-1}(E-e)$; but it enjoys better properties. In [17] (see also [16]), a notion of equicirca-tangent cone is introduced in order to prove open mapping theorems for multimappings. It involves a notion akin to

$$b\text{–}\liminf_{(t,e')\to(0_+,e),\, e'\in E} \frac{1}{t}(E-e').$$

Reinforced asymptotic approximation properties which bear some analogy with the preceding reinforced tangency will be considered later on.

Similar notions of approximations for functions can be defined and used.

3.2 Nonlinear conditioning and perturbations

It is not difficult to see that the functional $f \mapsto m_f := \inf f(X)$ from $\overline{\mathbb{R}}^X$ to $\overline{\mathbb{R}}$ is upper semicontinuous when $\overline{\mathbb{R}}^X$ is endowed with the topology associated with b-convergence. A more precise and quantitative result can be given. Given $f : X \to \overline{\mathbb{R}}$ such that $m_f := \inf f(X) \in \mathbb{R}$ and $S_f := \arg\min f \neq \varnothing$, a nondecreasing function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+ \cup \{\infty\}$ is said to be a *conditioner* for f if $\varphi(0) = 0$ and

$$\forall x \in X : d(x, S_f) \leq \varphi(f(x) - m_f).$$

f is said to be well-set if it has a conditioner which is a modulus (i.e. $\varphi(t) \to 0$ as $t \to 0$).

The following statement shows that one only gets a one-sided perturbation result for the set of minimizers. Other results are given in [11].

Theorem 3. ([57]) *Suppose S_f is nonempty and bounded. Suppose f is well-set, with an usc conditioner φ. Then there exists $r > 0$ and $\delta > 0$ such that for any function $g : X \to \mathbb{R} \cup \{\infty\}$ whose sublevel sets are connected satisfying $d_r(f,g) < \delta$ one has*

$$\left|m_g - m_f\right| \leq d_r(f,g), \quad e\left(S_g, S_f\right) \leq d_r(f,g) + \varphi\left(2d_r(f,g)\right).$$

3.3 Convergence of fixed points

Following Aubin, given $\lambda \in \mathbb{P}$, a complete metric space (X,d), and a nonempty subset U of X, one says that $F : X \rightrightarrows X$ is *pseudo-λ-Lipschitzian with respect to U* if

$$e(F(x) \cap U, F(x')) \leq \lambda d(x,x') \quad \forall x, x' \in U.$$

The following existence result is close to the Nadler fixed point theorem [48]. However, here we use the preceding weakening of the notion of Lipschitzian multimapping.

Proposition 4. ([21]) *Let $F : X \rightrightarrows X$ be a multimapping with closed values which is assumed to be pseudo-λ-Lispchitzian with respect to some ball $U(x_0, r)$ with $\lambda \in (0,1)$, $r > (1-\lambda)^{-1} d(x_0, F(x_0))$. Then the set Φ_F of fixed points of F is nonempty and*

$$d(x_0,\Phi_F) \le (1-\lambda)^{-1} d(x_0,F(x_0)).$$

The following result gives a measure of the variation of the sets of fixed points of multimappings in terms of the variation of the graphs. Again it is a one-sided result. In [21] this result is applied to the variations of the sets of solutions to a differential inclusion.

Proposition 5. ([21]) *Let $F: X \rightrightarrows X$ be a multimapping with closed values which is pseudo-λ-Lispchitzian with respect to $U(x_0,r)$, and $\lambda \in (0,1)$. Then for any $s \in (0,r)$ and for any $G: X \rightrightarrows X$ with $e_s(G,F) < (1-\lambda)(1+\lambda)^{-1}(r-s)$, one has*

$$e_s(\Phi_G,\Phi_F) \le (1-\lambda)^{-1}(1+\lambda) e_s(G,F).$$

3.4 Continuity of the Fenchel transform

In the sequel we denote by $\mathcal{F}(X)$ the set of proper lsc functions on X with values in $\mathbb{R} \cup \{+\infty\}$. The Fenchel–Legendre conjugate of $f \in \mathcal{F}(X)$ is

$$f^*: X^* \to \overline{\mathbb{R}}, \quad f^*(x^*) = \sup_{x \in X}(\langle x^*, x \rangle - f(x)),$$

where X^* is the topological dual of X. The continuity of the transform $f \mapsto f^*$ is important for a number of applications ([22], [27], [32], [63]...). It has been mostly studied under convexity assumptions.

Theorem 6. ([23], [54], [64]) *Let $f, f_n, g, g_n \in \mathcal{F}(X)$ $(n \in \mathbb{R})$, with f_n, g convex.*

(a) $f \le b\text{-}\liminf_n f_n \Rightarrow f^* \ge b - \limsup_n f_n^*$ *if* $\sup_n d((0,0),\operatorname{epi} f_n) < \infty$.

(b) $g \ge b - \limsup_n g_n \Rightarrow g^* \le b - \liminf_n g_n^*$.

(c) $(f_n) \xrightarrow{b} f \Rightarrow (f_n^*) \xrightarrow{b} f^*$.

However some conclusions can be drawn without convexity assumptions; note that the following statement can be converted into a continuity result in terms of uniform convergence on bounded subsets of the transforms.

Theorem 7. ([64]) *Let $f \in \mathcal{F}(X)$ be hypercoercive (i.e. $\lim_{\|x\| \to \infty} f(x)/\|x\| = \infty$) and bounded below. Then, for all $q, \varepsilon \in \mathbb{P}$ there exist $r, \delta \in \mathbb{P}$ such that*

$$e_r(f,g) < \delta \Rightarrow [\forall x^* \in qU_{X^*},\ \ g^*(x^*) > f^*(x^*) - \varepsilon] \Rightarrow [e_q(g^*, f^*) \le \varepsilon].$$

In particular, if $f \le b\text{–}\liminf_n f_n$ *then* $f^* \ge b\text{–}\limsup_n f_n^*$.

4. BOUNDEDNESS PROPERTIES

We devote the present section to some concepts which will be used as key ingredients in some boundedness properties we need.

4.1 Apart subsets

Given a nonempty subset E of X and $\varepsilon \in \mathbb{P}$, the *conical* ε*-enlargement of* E is the set

$$C_\varepsilon(E) := \{x \in X : d(x,E) < \varepsilon \| x \|\} \cup \{0\}.$$

For $\alpha, \beta \in]0,1[$ and $\gamma := \alpha + \beta + \alpha\beta$ one has, whenever $0 \in E$,

$$C_\beta(C_\alpha(E)) \subset C_\gamma(E). \tag{1}$$

When $E \ne \{0\}$ is a cone, for $\alpha, \beta \in (0,1)$, one has the following inclusions:

$$\mathbb{R}_+(E \cap S_X + \alpha U_X) \subset C_{\alpha(1-\alpha)^{-1}}(E), \tag{2}$$

$$C_\beta(E) \subset \mathbb{R}_+(E \cap S_X + \beta(1-\beta)^{-1}U_X). \tag{3}$$

The notion of conical enlargement is thus especially useful when dealing with cones; for such subsets it is related to the notion of plastering due to Krasnoselski ([37]; see also [28], [31], [33], [35]). But it can be used for any subset.

The following definition recalls a notion introduced and used in [41], [60] which will be much used in the sequel.

Definition 8. *Two nonempty subsets* E, F *of* X *are said to be (asymptotically) apart if there exists* $\varepsilon \in \mathbb{P}$ *such that* $C_\varepsilon(E) \cap C_\varepsilon(F)$ *is bounded.*

Equivalently, the nonempty subsets E, F of X are apart if, and only if, there is no sequence (x_n) such that $(\|x_n\|) \to \infty$, $(\|x_n\|^{-1} d(x_n, E)) \to 0$,

$(\|x_n\|^{-1} d(x_n, F)) \to 0$. In the case E and F are cones, several other characterizations are given in [41] and [60]; we recall them for the reader's convenience. Their simple proofs are consequences of relations (1)–(3).

Lemma 9. *([41]) Given two cones P, Q in X, the following assertions are equivalent and hold if and only if P and Q are apart:*

a) there exist $\alpha, \beta > 0$ such that $C_\alpha(P) \cap C_\beta(Q) = \{0\}$;

b) there exists $\gamma > 0$ such that $P \cap C_\gamma(Q) = \{0\}$;

c) there exists $\delta > 0$ such that $P \cap (Q \cap S_X + \delta U_X) = \varnothing$;

d) there exists $\varepsilon > 0$ such that $(P \cap S_X + \varepsilon U_X) \cap (Q \cap S_X + \varepsilon U_X) = \varnothing$;

e) there exists $\kappa > 0$ such that $\max(d(x,P), d(x,Q)) \geq \kappa \|x\|$ for each $x \in X$.

These assertions are satisfied when P, Q are closed, $P \cap Q = \{0\}$ and one of the following conditions is satisfied:

i) P (or Q) is locally compact (in particular if span P *is finite dimensional);*

ii) P (or Q) is weakly locally compact and P and Q are convex.

When P and Q are convex, dual properties can be given in terms of polar cones.

4.2 Boundedness and expansion properties

The preceding notions can be used for studying boundedness questions. Let us recall that a multimapping $M : W \rightrightarrows X$ between two n.v.s. is said to be *bounding* if it transforms any bounded set into a bounded set (sometimes M is said to be bounded, but we prefer to avoid any confusion with the case the image of M is bounded). Let us say it is *quasi-bounding* if the remoteness of M is bounded over any bounded subset of its domain. It is easy to give examples showing that the latter condition is less exacting than the former one; in particular, the notion of bounding multimapping cannot be used when the values of M are unbounded, in particular when they are epigraphs. The following concepts have been used repeatedly but implicitly in [53], [62] and explicitly in [60]. In this last reference, by analogy with the case of proper maps, a quasi-expanding map was called *boundedly proper on E*. There is also a certain analogy between expansive maps and expanding maps as any expansive map is expanding (but the converse is not true).

Definition 10. *A map F from X to a normed vector space Y is said to be expanding (resp. quasi-expanding) on a subset E of X if the multimapping $M: y \rightrightarrows F^{-1}(y) \cap E$ is bounding (resp. quasi-bounding) from Y to X. It is said to be linearly expanding on E if there are $\alpha \in \mathbb{P}$, $\rho \in \mathbb{R}_+$ such that*

$$\|F(x)\| \geq \alpha \|x\| \text{ for all } x \in E \setminus \rho U_X.$$

It is said to be linearly quasi-expanding on E if there are $\alpha \in \mathbb{P}$, $\rho \in \mathbb{R}_+$ such that

$$F(E) \cap \alpha r U_Y \subset F(E \cap r U_X) \text{ for all } r > \rho.$$

Let us state easy characterizations of these properties.

Proposition 11. *The map $F: X \to Y$ is expanding on $E \subset X$ if, and only if,*

$$\forall r \in \mathbb{P}, \exists q \in \mathbb{P}: E \cap F^{-1}(r U_Y) \subset q U_X.$$

It is quasi-expanding on E if, and only if,

$$\forall r \in \mathbb{P}, \exists q \in \mathbb{P}: F(E) \cap r U_Y \subset F(E \cap q U_X). \tag{4}$$

Moreover, the mapping $F: X \to Y$ is expanding on a subset E of X if, and only if, any sequence (x_n) in E is bounded when $(F(x_n))$ is bounded. It is quasi-expanding on E if, and only if, for any bounded sequence (y_n) in $F(E)$ there exists a bounded sequence (x_n) in E such that $y_n = F(x_n)$ for each $n \in \mathbb{N}$.

We also have the following immediate implications.

Proposition 12.

(a) If F is expanding on E then it is quasi-expanding on E.

(b) If F is linearly expanding on E then it is expanding on E and linearly quasi-expanding on E.

(c) If F is linearly quasi-expanding on E then it is quasi-expanding on E.

For positive homogeneous maps, more can be said.

Proposition 13. *Suppose E is a cone and F is positively homogeneous. Then*

(a) *F is linearly expanding on E if, and only if, it is expanding if, and only if, there exists some $c \in \mathbb{P}$ such that $E \cap F^{-1}(U_Y) \subset cU_X$ if, and only if, there exists some $\alpha \in \mathbb{P}$ such that $\|F(x)\| \geq \alpha \|x\|$ for all $x \in E$. In such a case one has $F^{-1}(0) \cap E = \{0\}$.*

(b) *F is linearly quasi-expanding on E if, and only if, it is quasi-expanding if, and only if, there exists some $c \in \mathbb{P}$ such that $F(E) \cap U_Y \subset F(E \cap cU_X)$.*

Moreover, when $0 \in E$, F is linearly quasi-expanding if, and only if, F is open at 0 at a linear rate from E onto $F(E)$.

Example 1. *The preceding notions can be illustrated by the case $X = Y = \mathbb{R}$. In such a case, F is expanding if, and only if, F is coercive in the sense that $|F(x)| \to \infty$ when $|x| \to \infty$.*

Example 2. *Suppose $A : X \to Y$ is a linear isomorphism, $h : \mathbb{R}_+ \to \mathbb{R}$ is a function and $F(x) = h(\|x\|)A(x)$ for $x \in X$. If $\liminf_{r \to \infty} h(r) > 0$, then F is linearly expanding on X.*

The linear expansion property enjoys a useful stability property detected in [41] and [60].

Lemma 14. *If $F : X \to Y$ is Lipschitzian and linearly expanding on a subset E of X, then there is a positive number δ such that F is linearly expanding on $C_\delta(E)$. Moreover, for any $\varepsilon > 0$ there exist $\delta, \sigma > 0$ such that $F(C_\delta(E) \setminus \sigma U_X) \subseteq C_\varepsilon(F(E))$.*

Let us quote some criteria from [41] and [60, Lemma 8].

Lemma 15. *Let P be a cone in X and let F be a continuous linear map from X to Y, with $N := \ker F$. Each of the following conditions is sufficient for F to be linearly expanding on P:*

a) *F is open onto its image and N and P are apart;*

b) *F is quasi-expanding on P and N and P are apart;*

c) *F is quasi-expanding on P, P is closed and N is finite dimensional with $N \cap P = \{0\}$;*

d) *P is closed, locally compact and $N \cap P = \{0\}$;*

e) *P is closed, P has a weakly compact base and $N \cap P = \{0\}$.*

Another connection between the two concepts introduced above is the following one (see [41, Lemma 2.2 c] for a quantitative proof in the case E and F are cones and [60] in the general case).

Lemma 16. *The subsets E and F of X are apart if and only if the map $L:(x,y)\mapsto x-y$ is linearly expanding on $E\times F$.*

We also need a notion which is a global variant of a property which has been widely used in nonsmooth analysis since its introduction in [34] and its use in [50], [52] in which the terminology has been coined.

Definition 17. *A mapping $F:X\to Y$ between two normed vector spaces is said to be metrically regular (resp. asymptotically metrically regular) on a subset E of X if there exists $\gamma>0$ such that $d(x,N)\le\gamma\|F(x)\|$ for $x\in E$ (resp. for $x\in E$ with $\|x\|$ large enough), where $N:=F^{-1}(0)$.*

When the closure of N contains 0 (in particular when N is nonempty and F is positively homogeneous), F is asymptotically metrically regular on E whenever F is linearly expanding on E. When X and Y are Banach spaces and F is linear, continuous and surjective F is metrically regular on X. Let us note the following simple facts which clarify some relationships between the preceding concepts.

Lemma 18. *Suppose $F:X\to Y$ is positively homogeneous. Let $N:=F^{-1}(0)$ and let C be a cone of X.*

(a) F is metrically regular on C if, and only if, $(d(x_n,N))\to 0$ whenever $(F(x_n))\to 0$ with $x_n\in C$ for each n.

(b) Suppose $C-N\subset C$ and $F(x-w)=F(x)$ for any $w\in N$, $x\in C$. If F is metrically regular on C then F is linearly quasi-expanding on C.

(c) Suppose $C-N\subset C$ and $x'-x\in N$ whenever x, $x'\in C$ and $F(x)=F(x')$. If F is quasi-expanding on C then F is metrically regular on C.

Proof.

(a) If F is not metrically regular on C there exists a sequence (x_n) in C such that $d(x_n,N)>n\|F(x_n)\|$ for each n. Since F is positively homogeneous and $F(x_n)\neq 0$, we may suppose that $n\|F(x_n)\|=1$ for each n. Then $(F(x_n))\to 0$ and $(d(x_n,N))$ does not converges to 0. The converse is obvious.

(b) Suppose there exists $\gamma > 0$ such that $d(x,N) \le \gamma \|F(x)\|$ for $x \in C$. Then for any $q \in \mathbb{P}$, $y \in F(C) \cap qU_Y$ and any $x \in F^{-1}(y) \cap C$ one can find $w \in N$ such that $\|w - x\| < \gamma q$, so that $y = F(x-w) \in F(\gamma q U_X \cap C)$ by the assumption $C - N \subset C$.

(c) Suppose F is quasi-expanding on C. Let $p \in \mathbb{P}$ be such that $F(C) \cap U_Y \subset F(C \cap pU_X)$. For each $x \in C$ and each $q > \|F(x)\|$ one has $F(q^{-1}x) = F(pu)$ for some $u \in C \cap U_X$, hence $q^{-1}d(x,N) = d(q^{-1}x,N) \le \|q^{-1}x - (q^{-1}x - pu)\| < p$ as $q^{-1}x - pu \in N$, and one gets $d(x,N) \le p\|F(x)\|$.

□

Part (a) of the preceding lemma can be used to show that if $F : X \to \mathbb{R}$ is positively homogenous and if $N_- := \{x \in X : F(x) \le 0\}$, then F satisfies $d(x,N_-) \le \gamma F(x)_+$ for some $\gamma > 0$ and each $x \in C$ if, and only if, $(d(x_n,N_-)) \to 0$ whenever $(F(x_n)_+) \to 0$ with $x_n \in C$ for each n, where $r_+ := \max(r,0)$: it suffices to replace $F(\cdot)$ by $F(\cdot)_+$.

4.3 Firm asymptotic cones

Let E be a nonempty subset of the normed vector space X. We recall that the *asymptotic cone* (sometimes called the recession cone) of E is the cone $E_\infty := \limsup_{t \to +\infty} t^{-1}E$, consisting of all limits of sequences $(t_n^{-1}x_n)$, where $x_n \in E$ and $t_n \in \mathbb{P}$ with $(t_n) \to \infty$ (see [14], [15], [39]–[41], [69], [75] for the study of related properties).

The following definition, which is the central concept of [60], will be used here instead of the concept of asymptotic compactness used in [62] as a boundedness criteria. Recall that E is said to be *asymptotically compact* if for any sequence (x_n) of E such that $(\|x_n\|) \to \infty$ the sequence $(\|x_n\|^{-1}x_n)$ has a converging subsequence (see [29], [51], [76] for preliminary definitions).

Definition 19. *A cone C of X is a firm (outer) asymptotic cone of a subset E of X if for any $\varepsilon > 0$ there exists some $r > 0$ such that $E \setminus rU_X \subset C_\varepsilon(C)$.*

The following characterizations may be convenient.

Proposition 20. *For a subset E of X and a closed cone C in X, the following assertions are equivalent:*

a) C is a firm asymptotic cone of E;

b) $d(x,C)/\|x\| \to 0$ *as* $\|x\| \to \infty$ *with* $x \in E$;

c) there exists a map $h : E \to C$ *such that* $d(x,h(x))/\|x\| \to 0$ *as* $\|x\| \to \infty$ *with* $x \in E$.

Proof. The implications a)$\Rightarrow$b), c)$\Rightarrow$a) are direct consequences of the definitions. To prove that b)$\Rightarrow$c), given $c > 1$, for $x \in E$ we pick $h(x) \in C$ such that $\|h(x) - x\| \le c\,d(x,C)$ (considering separately the case $d(x,C) = 0$ and the case $d(x,C) > 0$). □

An interpretation of the preceding conditions in terms of bounded convergence can be given.

Proposition 21. *A cone* C *of* X *is a firm asymptotic cone of a subset* E *of* X *if, and only if,* $b\text{-}\limsup_{t\to\infty} t^{-1}E \subset C$.

Proof. Suppose C is a firm asymptotic cone of E. Given $p \in \mathbb{P}$ we have $e_p(t^{-1}E,C) \to 0$ as $t \to \infty$: otherwise, we could find $c > 0$, a sequence $(t_n) \to \infty$ and $x_n \in E$ such that $\|t_n^{-1}x_n\| < p$ and $d(t_n^{-1}x_n,C) > c$ and then we would have $\|x_n\| \ge ct_n \to \infty$, and $d(x_n,C) > ct_n > cp^{-1}\|x_n\|$, a contradiction.

Conversely suppose $e_p(t^{-1}E,C) \to 0$ as $t \to \infty$ for each $p \in \mathbb{P}$. Given $\varepsilon > 0$, let $t_\varepsilon > 0$ be such that $e_1(t^{-1}E,C) < \varepsilon$ for $t > t_\varepsilon$. Then, for $x \in E \setminus t_\varepsilon U_X$ and for $t > \|x\|$ we have $t^{-1}x \in U_X$ hence $d(t^{-1}x,C) < \varepsilon$ and $d(x,C) < \varepsilon t$. Since t is arbitrarily close to $\|x\|$, we get $d(x,C) \le \varepsilon\|x\|$ and $x \in C_\varepsilon(C)$. □

Of course, the preceding definition does not determine C uniquely: any cone D containing C is also a firm asymptotic cone. Thus, one is led to take as a firm asymptotic cone a cone which is as small as possible. The following result shows a limitation in this direction.

Proposition 22. *If* C *is a closed firm asymptotic cone of* E, *then* C *contains the asymptotic cone* E_∞ *of* E. *If* E *is asymptotically compact, then* E *is firmly semi-asymptotable in the sense that* E_∞ *is a firm asymptotic cone of* E.

Proof. Let $v \in E_\infty \setminus \{0\}$: there exists a sequence (e_n) in E and a sequence $(t_n) \to \infty$ in $\mathbb{P}$ such that $(t_n^{-1}e_n) \to v$. Then $(\|e_n\|) \to \infty$ and

$$d(v,C) = \lim_n d(t_n^{-1}e_n,C) = \lim_n t_n^{-1}d(e_n,C) = \|v\|\lim_n \|e_n\|^{-1}d(e_n,C) = 0,$$

so that $v \in C$. Suppose E is asymptotically compact and the asymptotic cone E_∞ of E is not a firm asymptotic cone of E. Then there exist $\varepsilon > 0$

and a sequence (e_n) of E such that $(\|e_n\|) \to \infty$ and $d(e_n, E_\infty) \geq \varepsilon \|e_n\|$. Since E is asymptotically compact, taking a subsequence if necessary, we may suppose that $(e_n/\|e_n\|)$ has a limit v. Then $v \in E_\infty$ and we get a contradiction: $d(\|e_n\|^{-1} e_n, v) \geq d(\|e_n\|^{-1} e_n, E_\infty) \geq \varepsilon$. □

Thus, in any finite dimensional space, the asymptotic cone is a firm asymptotic cone. On the other hand, in any infinite dimensional normed vector space X there exists a set E whose asymptotic cone is not a firm asymptotic cone.

Example 3. *Let E be the epigraph of a function $f : W \to \overline{\mathbb{R}}$ with nonempty domain in a n.v.s. W which is bounded below on bounded sets, and let $X = W \times \mathbb{R}$. If f is hypercoercive* (i.e., $f(x)/\|x\| \to \infty$ *as* $\|x\| \to 0$), *then E is firmly asymptotable and $E_\infty = \{0\} \times \mathbb{R}_+$.*

Example 4. *Let E be the epigraph of a function $f : W \to \overline{\mathbb{R}}$ which is bounded below on bounded sets, such that $\liminf_{\|w\| \to \infty} (f(w) - p(w))/\|w\| \geq 0$, where $p : W \to \mathbb{R}$ is a positively homogeneous function and let C be the epigraph of p in $X = W \times \mathbb{R}$. Then C is a firm asymptotic cone of E. In particular, if $c_\infty := \liminf_{\|w\| \to \infty} f(w)/\|w\| \in \mathbb{R}$, the set $C := \operatorname{epi} c_\infty \|\cdot\|$ is a firm asymptotic cone of E. When $c_\infty \in \mathbb{P} \cup \{+\infty\}$, f is said to be super-coercive.*

Example 5. *Suppose there exist a bounded subset B of X and a closed cone C such that $E \subset B + C$. Then C is a firm asymptotic cone to E.*

Some calculus rules can be given (see [60, Prop. 13]).

A connection between the concept of firm asymptotic cone and the notion of apart subsets is as follows.

Proposition 23. ([60, Prop. 17]) *Let P and Q be firm asymptotic cones of subsets E and F of X respectively. If P and Q are apart, then E and F are apart.*

4.4 Applications to boundedness properties

Let us now show that the preceding concepts can be used for the study of boundedness properties. We only deal with mappings; boundedness properties of correspondences could be dealt with similarly. The first result we give is a simple consequence of Lemmas 12-14.

Lemma 24. *Let $F : X \to Y$ be a Lipschitzian, positively homogeneous map between two normed vector spaces and let E be a subset of X. Suppose K*

is a firm asymptotic cone of E and that F is expanding on K. Then F is linearly expanding on E.

The following proposition is closely related.

Proposition 25. *Under the following assumptions, a positively homogeneous mapping $F: X \to Y$ is expanding on a subset E of X:*

(a) E has a firm asymptotic cone K;

(b) K and $N := F^{-1}(0)$ are apart;

(c) F is asymptotically metrically regular on E.

In fact assumption (a) can be replaced with the following weaker condition:

(a') there exists $\alpha \in \mathbb{P}$ such that $E \cap C_\alpha(N)$ has a firm asymptotic cone K.

Proof. If the conclusion does not hold, one can find $r \in \mathbb{P}$, a sequence (x_n) of E such that $\|F(x_n)\| < r$ and $(\|x_n\|) \to \infty$. In view of (c), we have $(\|x_n\|^{-1} d(x_n, N)) \to 0$. Then, dropping a finite number of terms if necessary, we have $x_n \in E \cap C_\alpha(N)$ for each $n \in \mathbb{N}$. Then, by assumption (a'), we have $(\|x_n\|^{-1} d(x_n, K)) \to 0$. In view of the characterization given after Definition 8 of the property that K and N are apart, we get a contradiction. □

The preceding result can be specialized to the case E is a sub-level set $[f < q]$ of some function f on X. It can also be adapted to the case the function f has a *firm asymptotic approximation* φ on some subset S of X; by this we mean that $\liminf_{x \in S, \|x\| \to \infty} (f(x) - \varphi(x))/\|x\| \geq 0$.

Corollary 26. *Under the following assumptions, the map F is expanding on the sub-level set $[f < q]$:*

(a) there exists $\beta \in \mathbb{P}$ such that f has a firm asymptotic approximation φ on $[f < q] \cap C_\beta(N)$ which is positively homogeneous;

(b) there exists $\gamma \in \mathbb{P}$ such that $\varphi(x) \geq \gamma \|x\|$ for each $x \in C_\gamma(N)$;

(c) F is asymptotically metrically regular on $[f < q]$.

Proof. Let $\alpha \in (0, \min\{\beta, \gamma\})$ and let

$$K := \{x \in C_\beta(N) : \varphi(x) \leq \alpha \|x\|\}.$$

Since $K \cap C_\gamma(N) = \{0\}$ the sets K and N are apart. It remains to show that K is a firm asymptotic cone to $[f < q] \cap C_\alpha(N)$. If it is not the case, one can find $\delta \in \mathbb{P}$ and a sequence (x_n) in $[f < q] \cap C_\alpha(N)$ such that

$(\|x_n\|) \to \infty$ and $x_n \notin C_\delta(K)$ for each $n \in \mathbb{N}$. Then, by (a) and (b), there exists a sequence $(\varepsilon_n) \to 0_+$ such that

$$q > f(x_n) \geq \varphi(x_n) - \varepsilon_n \|x_n\| \geq (\gamma - \varepsilon_n)\|x_n\|,$$

a contradiction. □

The preceding proposition can be applied to the case of the sum $S : X^2 \to X$ given by $S(x,y) = x + y$. We note that S is metrically regular: for any $(x,y) \in X^2$ we have

$$d((x,y),N) \leq d((x,y),\frac{1}{2}(x-y,y-x)) = \frac{1}{2}\|x+y\| = \frac{1}{2}\|S(x,y)\|.$$

Since when P (resp. Q) is a firm asymptotic cone of A (resp. B), the cone $P \times Q$ is a firm asymptotic cone of $A \times B$, and since $P \times Q$ is apart from $N := \ker S$ when P and $-Q$ are apart, as easily seen, we get the following result. Another (simple, direct) proof is provided in [60, Prop. 20]; still another proof can be derived from Lemmas 14 and 16.

Proposition 27. *Let A and B be two nonempty subsets of X and let P (resp. Q) be a firm asymptotic cone of A (resp. B). If P and $-Q$ are apart then the mapping $S : (x,y) \mapsto x + y$ is expanding on $A \times B$.*

5. CONTINUITY OF SOME OPERATIONS

We are in a position to give some persistence and stability results for usual operations on sets and functions.

5.1 Continuity of some operations with sets

The most obvious results concern products and unions for which a direct easy analysis leads to the following statement.

Proposition 28. ([62, Lemma 21 (e)]) *Suppose $(A_n) \xrightarrow{b} A$, $(B_n) \xrightarrow{b} B$. Then $(A_n \times B_n) \xrightarrow{b} A \times B$. If A, B, A_n, B_n are subsets of the same space then $(A_n \cup B_n) \xrightarrow{b} A \cup B$.*

For intersections, a convexity argument and a qualification condition ([50], [53], [66], [67], [75]) have to be used.

Proposition 29. *([62, Prop. 27 (e)]) Suppose* $(A_n) \xrightarrow{b} A$, $(B_n) \xrightarrow{b} B$ *where* A, A_n, B, B_n *are closed convex subsets of a Banach space* X *and* $X = \mathbb{R}_+(A - B)$. *Then* $(A_n \cap B_n) \xrightarrow{b} A \cap B$.

In fact a quantitative result can be given. Assuming that

$$sU_X \subset A \cap rU_X - B \cap rU_X \tag{5}$$

for some $r, s \in \mathbb{P}$ (what occurs when $X = \mathbb{R}_+(A - B)$), we will show that for each $p \in \mathbb{P}$ and any $A', B' \subset X$ we have

$$e_p(A' \cap B', A \cap B) \leq \frac{p + r + s}{s + \max(e_p(A', A), e_p(B', B))}(e_p(A', A) + e_p(B', B)), \tag{6}$$

and that if $p \geq r$ and if $e_p(A, A') + e_p(B, B') < s$, we have

$$d_p(A' \cap B', A \cap B) \leq s^{-1}(p + r + s)(d_p(A', A) + d_p(B', B)) \tag{7}$$

Proof. Let $x' \in A' \cap B' \cap pU_X$ and let $t > e_p(A', A) + e_p(B', B)$. We can find $y \in A$, $z \in B$ such that $\|y - x'\| + \|z - x'\| < t$. Relation (5) ensures that there exists $(a, b) \in (A \cap rU_X) \times (B \cap rU_X)$ such that

$$st^{-1}(z - y) = a - b.$$

Then $x := (s + t)^{-1}(sy + ta) = (s + t)^{-1}(sz + tb)$ belongs to $A \cap B$ and since $\|a - x'\| < p + r$

$$\|x - x'\| \leq (s + t)^{-1} s \|y - x'\| + (s + t)^{-1} t \|a - x'\| < (s + t)^{-1} t(s + p + r).$$

Since t is arbitrarily close to $e_p(A', A) + e_p(B', B)$, we obtain (6).

Now let us suppose A' and B' are such that $e_p(A, A') + e_p(B, B') < t < s$. Then, for $s' \in (t, s)$ we have

$$s'B_X \subset A \cap rU_X - B \cap rU_X \subset A' \cap (r + t)U_X - B' \cap (r + t)U_X + tU_X.$$

Using the Rådström's cancellation rule, we get

$$(s' - t)B_X \subset \mathrm{cl}\left(A' \cap (r + t)U_X - B' \cap (r + t)U_X\right).$$

Then the openness result of [67, Lemma 1.0] ensures that

$$(s'-t)U_X \subset A' \cap (r+t)U_X - B' \cap (r+t)U_X .$$

Thus the first part of the proof can be applied with A, B interchanged with A', B' and r, s replaced by $r+t$ and $s'-t$ respectively:

$$e_p(A \cap B, A' \cap B') \leq \frac{p+r+s'}{s'-t+e_p(A,A')+e_p(B,B')}(e_p(A,A')+e_p(B,B')).$$

Since s' and t can be chosen arbitrarily close to s and $e_p(A,A')+e_p(B,B')$ respectively, we get (7). □

Using the diagonal mapping, and the product rule, the preceding statement can be considered as a special case of a result about inverse images under a continuous linear map (see [62, Lemma 24]). On the other hand, the inverse image by a linear continuous map $L: X \to Y$ of a subset D of Y is obtained as the projection on X of the intersection $L \cap (X \times D)$, where L is identified with its graph and $p_X|_L$ is an isomorphism from L onto X. In fact, a direct analysis yields a quantitative result which reveals a kind of Lipschitzian behavior.

Proposition 30. ([19, Cor. 2.4], [62, Lemma 24]) *Let D be a closed convex subset of Y. Assume*

$$sU_Y \subset L(rU_X) - D$$

for some $r,s>0$; this condition is satisfied when X, Y are complete and $Y = \mathbb{R}_+(L(X)-D)$. Then, for $t \in (0,s)$, $p>0$, $q > \max(p\|L\|, r\|L\|+s)$ and D', D'' closed convex subsets of Y with $d_q(D,D') < t$, $d_q(D,D'') < t$ one has

$$d_p(L^{-1}(D'), L^{-1}(D'')) \leq \frac{p+r}{s-t} d_q(D', D'').$$

In particular, for a sequence (D_n) of closed convex subsets of Y one has

$$(D_n) \xrightarrow{b} D \Rightarrow (L^{-1}(D_n)) \xrightarrow{b} L^{-1}(D).$$

The study of b-convergence of images can be eased by the use of the criteria for the expansion property we displayed above. Here we introduce a slight refinement of condition (4), and of conditions (14) and (15) of [62]. We say that a map $F: X \to Y$ is *approximately quasi-expanding on a sequence* (E_n) of subsets of X if for each $\varepsilon > 0$ and each $q \in \mathbb{P}$ there exist

$p \in \mathbb{P}$ and $k \in \mathbb{N}$ such that $F(D_m) \cap qU_Y \subset F(D_m \cap pU_X) + \varepsilon U_Y$ for each $m \geq k$, where $D_m := \bigcup_{n \geq m} E_n$. This condition is satisfied if F is *expanding on* (E_n) in the following sense: for each $q \in \mathbb{P}$ there exist $p \in \mathbb{P}$ and $k \in \mathbb{N}$ such that $D_k \cap F^{-1}(qU_Y) \subset pU_X$. We also need an extension of the notion of firm asymptotic cone: we say that K is a *firm asymptotic cone to a sequence* (E_n) *of subsets of* X if for each $\varepsilon > 0$ there exist $r \in \mathbb{P}$ and $k \in \mathbb{N}$ such that $E_n \setminus rU_X \subset C_\varepsilon(K)$ for $n \geq k$. When the sequence (E_n) is constant, we recover the definition above.

Proposition 31. *Let* E, E_n $(n \in \mathbb{N})$ *be subsets of* X *and let* $F : X \to Y$ *be Lipschitzian on bounded sets.*

(a) Suppose that $b\text{–}\limsup_n E_n \subset E$. *Then* $b\text{–}\limsup_n F(E_n) \subset F(E)$ *provided that the map* F *is approximately quasi-expanding on* (E_n).

(b) Suppose that $E \subset b\text{–}\liminf_n E_n$. *Then* $F(E) \subset b\text{–}\liminf_n F(E_n)$ *provided* F *is quasi-expanding on* E.

(c) $F(E) \subset b\text{–}\liminf_n F(E_n)$ *provided that* $E \subset b\text{–}\liminf_n E_n$, F *is positively homogeneous, asymptotically metrically regular on* E *and* E *has a firm asymptotic cone* K *which is apart from* $N := F^{-1}(0)$.

(d) If $E = b\text{–}\lim_n E_n$ *and if* F *is positively homogeneous, metrically regular on* $\bigcup_n E_n$ *and* E, *if* E *and* (E_n) *have a firm asymptotic cone* K *which is apart from* $N := F^{-1}(0)$, *then* $F(E) = b\text{–}\lim_n F(E_n)$.

Proof. (a) (Compare with [62, Prop. 8 (d)] when F is linear.) Let $q \in \mathbb{P}$ and $\varepsilon > 0$ be given. Since F is approximately quasi-expanding on (E_n), setting $D_m := \bigcup_{h \geq m} E_m$, we can find $p \in \mathbb{P}$ and some $k \in \mathbb{N}$ such that $F(D_m) \cap qU_Y \subset F(D_m \cap pU_X) + \frac{1}{2}\varepsilon U_Y$ for each $m \geq k$. Let κ be the Lipschitz rate of F on $(p+1)U_X$ and let $\delta := \min(\varepsilon/2\kappa, 1)$. Let $m \geq k$ be such that $E_n \cap pU_X \subset E + \delta U_X$ for $n \geq m$. Then, for $n \geq m$ we have

$$F(E_n) \cap qU_Y \subset F(D_m) \cap qU_Y \subset F(D_m \cap pU_X) + \frac{1}{2}\varepsilon U_Y$$

$$\subset F(E + \delta U_X) + \frac{1}{2}\varepsilon U_Y \subset F(E) + \varepsilon U_Y.$$

The proof of part (b) is simpler and is omitted (see [62, Prop. 1 (b)]). Part (c) is a consequence of part (b) and of Proposition 25.

(d) It is a consequence of parts (a) and (c) and of an adaptation of Proposition 25 which consists in proving that under the assumptions of (d) the map F is expanding on (E_n), and of course, on E. If the conclusion does not hold, one can find $q \in \mathbb{P}$, a sequence (x_p) of X such that $x_p \in E_{n(p)} \setminus pU_X$, $\|F(x_p)\| < q$ for each $p \in \mathbb{N}$, with $n(p) \to \infty$ as $p \to \infty$. Let $\alpha \in (0,1)$ be such that $C_\alpha(K) \cap C_\alpha(N) = \{0\}$. Since F is metrically regular on $\bigcup_n E_n$, we have $(\|x_p\|^{-1} d(x_p, N)) \to 0$, hence $x_p \in C_\alpha(N)$ for $p \in \mathbb{N}$ large enough. On the other hand, since K is a firm asymptotic cone to (E_n), we have $x_p \in C_\alpha(K)$ for $p \in \mathbb{N}$ large enough. This is a contradiction. □

The convergence of sums of sets is a special case of the preceding statement.

Corollary 32. *Let A, A_n, B, B_n $(n \in \mathbb{N})$ be subsets of X such that $(A_n) \xrightarrow{b} A$, $(B_n) \xrightarrow{b} B$. Suppose P (resp. Q) is a firm asymptotic cone to A and (A_n) (resp. B and (B_n)) and P and $-Q$ are apart. Then $(A_n + B_n) \xrightarrow{b} A + B$.*

5.2 Continuity of some operations on functions

The preceding results can be adapted to epigraphs of functions in order to get results about usual operations. The most immediate application concerns composition.

Proposition 33. *Let W, Z be two Banach spaces, let $A : W \to Z$ be a continuous linear map and let g be a closed proper convex function on Z such that $Z = \mathbb{R}_+ \operatorname{dom} g + A(W)$. If (g_n) is a sequence of closed proper convex functions on Z which b-converges to g, then $(g_n \circ A)$ b-converges to $g \circ A$.*

Proof. Let $X := W \times \mathbb{R}$, $Y := Z \times \mathbb{R}$, let D (resp. D_n) be the epigraph of g (resp. g_n) and let $L : X \to Y$ be given by $L(x,r) := (A(x), r)$. Then the epigraph of $g \circ A$ (resp. $g_n \circ A$) is $L^{-1}(D)$ (resp. $L^{-1}(D_n)$). Since the qualification condition of the statement easily implies that $Y = \mathbb{R}_+ D + L(X)$ there exist $r, s > 0$ such that $sU_Y \subset L(rU_X) - D$. Thus the conclusion follows from Proposition 30. □

The case of marginal functions can be deduced from the case of images of sets; in particular the convergence of the infimal convolution of two functions can be derived from the convergence of the sequence of the sum of two sets.

As a sample of what can be obtained with functions, let us give a result for infimal convolutions and recall the following result for sums (see [18], [19], [26], [53], [62], [74], [77]...); in the finite dimensional case such a result has been obtained by McLinden–Bergstrom [44].

Proposition 34. *Suppose that f, f_n, g, g_n are closed proper convex functions on the Banach space X satisfying*

$$X = \mathbb{R}_+(\operatorname{dom} f - \operatorname{dom} g).$$

Then, if $(f_n) \xrightarrow{b} f$, $(g_n) \xrightarrow{b} g$ one has $(f_n + g_n) \xrightarrow{b} f + g$.

For the infimal convolution of two functions given by $(f \square g)(x) := \inf_{w \in X}(f(w) + g(x - w))$ we devise a direct proof inspired by Corollary 32.

Proposition 35. *Let f, f_n, g, g_n be functions on the normed vector space X. Suppose that f, g are bounded below on bounded subsets and have asymptotic firm approximations p, q respectively which are positively homogeneous and for which there exist $\alpha, \beta \in \mathbb{P}$ such that*

$$p(u) + q(v) \geq \alpha \min(\|u\|, \|v\|) - \beta \|u + v\| \qquad \forall u, v \in X. \tag{8}$$

If $f \geq b\text{–}\limsup_n f_n$ and $g \geq b\text{–}\limsup_n g_n$, then $f \square g \geq b\text{–}\limsup_n f_n \square g_n$.

Let us note that relation (8) is satisfied whenever there exist $\gamma, \lambda \in \mathbb{P}$ such that q is λ-Lipschitzian and

$$p(u) + q(-u) \geq \gamma \|u\| \qquad \forall u \in X.$$

In fact, in such a case, for any $u, v \in X$ we have

$$p(u) + q(v) \geq p(u) + q(-u) - \lambda \|u + v\| \geq \gamma \|u\| - \lambda \|u + v\|.$$

Proof. Let F, F_n, G, G_n be the strict epigraphs of f, f_n, g, g_n respectively, so that $F + G$ (resp. $F_n + G_n$) is the strict epigraph of $f \square g$ (resp. $f_n \square g_n$). The assertion amounts to show that $F + G \subset b\text{–}\liminf_n (F_n + G_n)$. In view of Proposition 31 it suffices to show that $S : (x, r, y, s) \mapsto (x + y, r + s)$ is expanding on $F \times G$. Suppose, on the contrary, that there exist $s \in \mathbb{P}$ and a sequence $((x_n, r_n, y_n, s_n))$ in $F \times G$ such that $(\|(x_n, r_n, y_n, s_n)\|) \to \infty$ and $(\|(x_n + y_n, r_n + s_n)\|)$ is bounded. Since f and g are bounded below on

bounded subsets, the sequence $(\|(x_n, y_n)\|)$ cannot have a bounded subsequence. Thus $(\|x_n\|) \to \infty$ and $(\|y_n\|) \to \infty$. Then there exist $(\varepsilon_n) \to 0_+$ such that

$$r_n + s_n \geq f(x_n) + g(y_n) \geq p(x_n) + q(y_n) - \varepsilon_n \|x_n\| - \varepsilon_n \|y_n\|$$
$$\geq \alpha \min(\|x_n\|, \|y_n\|) - \beta \|x_n + y_n\| - \varepsilon_n \|x_n\| - \varepsilon_n \|y_n\|.$$

Because $\|x_n\| / \|y_n\| \to 1$ we obtain the contradiction $(r_n + s_n) \to \infty$. □

REFERENCES

[1] S. Adly, E. Ernst and M. Théra, Stability of the solution set of non-coercive variational inequalities, Commun. Contemp. Math. 4 (2002), 145–160.

[2] A. Agadi and J.-P. Penot, New asymptotic cones and usual tangent cones, submitted.

[3] H. Attouch, Variational Convergence for Functions and Operators, Pitman, Boston, (1984).

[4] H. Attouch, D. Azé and G. Beer, On some inverse stability problems for the epigraphical sum, Nonlinear Anal. 16 (1991), 241–254.

[5] H. Attouch, R. Lucchetti and R. J.-B. Wets, The topology of the ρ-Hausdorff distance, Ann. Mat. Pura Appl. (4) 160 (1991), 303–320.

[6] H. Attouch, A. Moudafi and H. Riahi, Quantitative stability analysis for maximal monotone operators and semi-groups of contractions, Nonlinear Anal. 21 (1993), 697–723.

[7] H. Attouch, J. Ndoutoume and M. Théra, Epigraphical convergence of functions and convergence of their derivatives in Banach spaces, Sém. Anal. Convexe 20 (1990), Exp. No. 9, 45 pp.

[8] H. Attouch and R. J.-B. Wets, Isometries for the Legendre–Fenchel transform, Trans. Amer. Math. Soc. 296 (1986), 33–60.

[9] H. Attouch and R. J.-B. Wets, Epigraphical analysis, Ann. Inst. H. Poincaré Anal. Non Linéaire 6 (suppl.) (1989), 73–100.

[10] H. Attouch and R. J.-B. Wets, Quantitative stability of variational systems: I. The epigraphical distance, Trans. Amer. Math. Soc. 328 (1991), 695–729.

[11] H. Attouch and R. J.-B. Wets, Quantitative stability of variational systems. II. A framework for nonlinear conditioning, SIAM J. Optim. 3 (1993), 359–381.

[12] H. Attouch and R. J.-B. Wets, Quantitative stability of variational systems. III: ε-approximate solutions, Math. Programming 61A (1993), 197–214.

[13] J.-P. Aubin and H. Frankowska, Set-Valued Analysis, Birkhäuser, Basel, (1990).

[14] A. Auslender, How to deal with the unboundedness in optimization: theory and algorithms, Math. Programming ser. B 31 (1997), 3–19.

[15] A. Auslender and M. Teboulle, Asymptotic cones and functions in optimization and variational inequalities, Springer, New York, 2002.

[16] D. Azé, An inversion theorem for set-valued maps, Bull. Aust. Math. Soc. 37, No.3 (1988), 411-414.

[17] D. Azé and C.C. Chou, On a Newton type iterative method for solving inclusions, Math. Oper. Res. 20, No.4 (1995), 790-800.

[18] D. Azé and J.-P. Penot, Recent quantitative results about the convergence of convex sets and functions, Functional analysis and approximation (Bagni di Lucca, 1988), Pitagora, Bologna, (1989), 90–110.
[19] D. Azé and J.-P. Penot, Operations on convergent families of sets and functions, Optimization 21 (1990), 521–534.
[20] D. Azé and J.-P. Penot, Qualitative results about the convergence of convex sets and convex functions, Optimization and nonlinear analysis (Haifa, 1990), Longman Sci. Tech., Harlow, (1992), 1–24.
[21] D. Azé and J.-P. Penot, On the dependence of fixed point sets of pseudo-contractive multimappings. Applications to differential inclusions, submitted.
[22] D. Azé and A. Rahmouni, On primal dual stability in convex optimization, J. Convex Anal. 3 (1996), 309–329.
[23] G. Beer, Conjugate convex functions and the epi-distance topology, Proc. Amer. Math. Soc. 108 (1990), 117–126.
[24] G. Beer, Topologies on Closed and Convex Sets, Kluwer, Dordrecht, (1993).
[25] G. Beer and R. Lucchetti, Convex optimization and the epi-distance topology, Trans. Amer. Math. Soc. 327 (1991), 795–813.
[26] G. Beer and R. Lucchetti, The epi-distance topology: continuity and stability results with applications to convex optimization problems, Math. Oper. Res. 17 (1992), 715–726.
[27] L. Contesse and J.-P. Penot, Continuity of the Fenchel correspondence and continuity of polarities, J. Math. Anal. Appl. 156 (1991), 305–328.
[28] J. Daneš and J. Durdill, A note on the geometric characterization of differentiability, Comm. Math. Univ. Carolin. 17 (1976), 195–204.
[29] J.-P. Dedieu, Cône asymptote d'un ensemble non convexe. Application à l'optimisation, C. R. Acad. Sci. Paris 287 (1977), 501–503.
[30] A. Dontchev and T. Zolezzi, Well-posed Optimization Problems, Lecture Notes in Maths 1543, Springer-Verlag, Berlin, (1993).
[31] J. Durdill, On the geometric characterization of differentiability I, Comm. Math. Univ. Carolin. 15 (1974), 521–540; II, idem, 727–744.
[32] A. Eberhard and R. Wenczel, Epi-distance convergence of parametrised sums of convex functions in non-reflexive spaces, J. Convex Anal. 7 (2000), 47–71.
[33] M. Fabian, Theory of Fréchet cones, Casopis Pro Pěstivani Mat., 107 (1982), 37–58.
[34] A. D. Ioffe, Regular points of Lipschitz functions, Trans. Amer. Math. Soc. 251 (1979), 61–69.
[35] T. Kato, Perturbation theory for linear operators, Springer Verlag, Berlin (1966).
[36] D. Klatte, On quantitative stability for non-isolated minima, Control Cybern. 23 (1994), 183–200.
[37] M. A. Krasnoselskii, Positive solutions of operator equations, Noordhoff, Groningen (1964).
[38] J. Lahrache, Stabilité et convergence dans les espaces non réflexifs, Sém. Anal. Convexe 21 (1991), Exp. No. 10, 50 pp.
[39] D. T. Luc, Recession maps and applications, Optimization 27 (1993), 1–15.
[40] D. T. Luc, Recessively compact sets: properties and uses, Set-Valued Anal. 10 (2002), 15–35.
[41] D. T. Luc and J.-P. Penot, Convergence of asymptotic directions, Trans. Amer. Math. Soc. 353 (2001), 4095–4121.
[42] R. Lucchetti and A. Pasquale, A new approach to a hyperspace theory, J. Convex Anal. 1 (1994), 173–193.

[43] R. Lucchetti and A. Torre, Classical convergences and topologies, Set-Valued Anal. 2 (1994), 219–241.
[44] L. McLinden and R. C. Bergstrom, Preservation of convergence of convex sets and functions in finite dimensions, Trans. Amer. Math. Soc. 268 (1981), 127–142.
[45] F. Mignot, Contrôle dans les inéquations variationelles elliptiques, J. Funct. Anal. 22 (1976), 130–185.
[46] J.-J. Moreau, Intersection of moving convex sets in a normed space, Math. Scand. 36 (1975), 159–173.
[47] U. Mosco, Convergence of convex sets and of solutions of variational inequalities, Adv. Math. 3 (1969), 510–585.
[48] S. B. Nadler, Multivalued contraction mappings, Pacific J. Math. 30 (1969), 475–488.
[49] T. Pennanen, R. T. Rockafellar, M. Théra, Graphical convergence of sums of monotone mappings, Proc. Amer. Math. Soc. 130 (2002), 2261–2269.
[50] J.-P. Penot, On regularity conditions in mathematical programming, Math. Prog. Study 19 (1982), 167–199.
[51] J.-P. Penot, Compact nets, filters and relations, J. Math. Anal. Appl., 93 (1983), 400–417.
[52] J.-P. Penot, Differentiability of relations and differential stability of perturbed optimization problems, SIAM J. Control Optim. 22 (1984), 529–551.
[53] J.-P. Penot, Preservation of persistence and stability under intersections and operations, Part I: Persistence, J. Optim. Theory Appl. 79 (1993), 525–550; Part II: Stability, idem, 551–561.
[54] J.-P. Penot, The cosmic Hausdorff topology, the bounded Hausdorff topology and continuity of polarity, Proc. Amer. Math. Soc., 113 (1991), 275–285.
[55] J.-P. Penot, Topologies and convergences on the space of convex functions, Nonlinear Anal. 18 (1992), 905–916.
[56] J.-P. Penot, On the convergence of subdifferentials of convex functions, Nonlinear Anal. 21 (1993), 87–101.
[57] J.-P. Penot, Conditioning convex and nonconvex problems, J. Optim. Theory Appl. 90 (1996), 535–554.
[58] J.-P. Penot, Metric estimates for the calculus of multimappings, Set-Valued Anal. 5 (1997), 291–308.
[59] J.-P. Penot, What is quasiconvex analysis? Optimization 47 (2000), 35–110.
[60] J.-P. Penot, A metric approach to asymptotic analysis, Bull. Sci. Maths,.
[61] J.-P. Penot and C. Zălinescu, Approximation of functions and sets, in Approximation, Optimization and Mathematical Economics, M. Lassonde ed., Physica-Verlag, Heidelberg, (2001), 255–274.
[62] J.-P. Penot and C. Zălinescu, Continuity of usual operations and variational convergences, Set-Valued Anal. 11 (2003), 225–256.
[63] J.-P. Penot and C. Zălinescu, Persistence and stability of solutions to Hamilton–Jacobi equations, preprint, Univ. of Pau, June 2000.
[64] J.-P. Penot and C. Zălinescu, Fenchel–Legendre transform and variational convergences, preprint, 2003.
[65] H. Rådström, An embedding theorem for spaces of convex sets, Proc. Amer. Math. Soc. 3 (1952), 165–169.
[66] S. M. Robinson, Stability theory for systems of inequalities, Part I : linear systems, SIAM J. Numer. Anal., 12 (1975), 754–769.
[67] S. M. Robinson, Regularity and stability for convex multivalued functions, Math. Oper. Res. 1 (1976), 130–143.

[68] R. T. Rockafellar and R. J.-B. Wets, Cosmic convergence, in: Optimization and Nonlinear Analysis, A. Ioffe et al. eds., Pitman Notes 244, Longman, Harlow, 1992, 249–272.
[69] R. T. Rockafellar and R. J.-B. Wets, Variational Analysis, Springer, Berlin, 1997.
[70] Y. Sonntag and C. Zălinescu, Set convergences. An attempt of classification, Trans. Amer. Math. Soc. 340 (1993), 199–226.
[71] Y. Sonntag and C. Zălinescu, Set convergences: a survey and a classification, Set-Valued Analysis 2 (1994), 339–356.
[72] T. Strömberg, The operation of infimal convolution, Dissert. Math. 352 (1996), 1–58.
[73] S. Villa, AW-convergence and well-posedness of non convex functions, J. Convex Anal. (2003), to appear.
[74] C. Zălinescu, On convex sets in general position, Linear Algebra Appl. 64 (1985), 191–198.
[75] C. Zălinescu, Stability for a class of nonlinear optimization problems and applications, in Nonsmooth Optimization and Related Fields, F.H. Clarke et al. eds., Plenum Press, London and New York (1989), 437–458.
[76] C. Zălinescu, Recession cones and asymptotically compact sets, J. Optim. Theory Appl., 77 (1993), 209–220.
[77] C. Zălinescu, Convex Analysis in General Vector Spaces, World Scientific, Singapore (2002).

CONTROL PROCESSES WITH DISTRIBUTED PARAMETERS IN UNBOUNDED SETS. APPROXIMATE CONTROLLABILITY WITH VARIABLE INITIAL LOCUS

G. Pulvirenti, G. Santagati and A. Villani
Dept. of Mathematics and Computer Sciences, University of Catania, Catania, Italy

Abstract: We consider the following distributed parameter linear control system

$$z_{xy} + A(x,y)z_x + B(x,y)z_y + C(x,y)z = F(x,y)U(x,y). \qquad \text{(E)}$$

Here (x,y) ranges over the unbounded set

$$L_{I,J} = \bigcup_{(u,v)\in I\times J} l(u,v),$$

where

$$l(u,v) = ([u,+\infty[\times\{v\}) \cup (\{u\}\times[v,+\infty[), \ (u,v)\in\mathbb{R}^2,$$

and I, J are two non-degenerate intervals of $\mathbb{R}$. The state vector function z belongs to the Sobolev type functional space

$$W^*_{p,\mathrm{loc}}(L_{I,J},\mathbb{R}^n) = \left\{z\in L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^n): z_x, z_y, z_{xy}\in L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^n)\right\}$$

and the control vector function U is in $L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^m)$. Moreover, for every $(u,v)\in I\times J$, the trace of z on $l(u,v)$ is taken as the system state corresponding to the values $x=u$, $y=v$ of the parameters. All these traces

belong to a functional space of Sobolev type, which does not depend on (u,v).

In this setting, given a point $(a,b) \in I \times J$, we study the controllability of system (E) from a given initial state, to be taken on the variable initial locus $l(a_0,b_0),(a_0,b_0) \in I \times J$, $a_0 \le a, b_0 \le b$, to an arbitrary final state, to be taken on the fixed final locus $l(a,b)$. We get a characterization of the approximate controllability when the set of the available controls is the unit ball of $L^\infty(L_{I,J},\mathbb{R}^m)$.

1. INTRODUCTION

We first introduce the notation for the unbounded subsets of $\mathbb{R}^2$ that we will use.

For every $(u,v) \in \mathbb{R}^2$ we put

$$l(u,v) = ([u,+\infty[\times\{v\}) \cup (\{u\} \times [v,+\infty[);$$

also, fixed any two non-degenerate intervals I, J of $\mathbb{R}$, we put

$$L_{I,J} = \bigcup_{(u,v)\in I\times J} l(u,v).$$

We consider the following distributed parameter linear hyperbolic control system:

$$z_{xy} + A(x,y)z_x + B(x,y)z_y + C(x,y)z = F(x,y)U(x,y) \quad \text{a.e. } (x,y) \in L_{I,J}. \tag{E}$$

Here, $A,B,C,A_x,B_y \in C^0(L_{I,J},\mathbb{R}^{n,n})$, $F \in L^\infty_{\mathrm{loc}}(L_{I,J},\mathbb{R}^{n,m})$, the *control vector function* U belongs to $L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^m)$ and the *state vector function* z is an element of the Sobolev type functional space

$$W^*_{p,\mathrm{loc}}(L_{I,J},\mathbb{R}^n) = \left\{ z \in L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^n) : z_x, z_y, z_{xy} \in L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^n) \right\}.$$

We take the trace of z on $l(u,v)$, $(u,v) \in I \times J$, as the system *state* corresponding to the values $x = u$, $y = v$ of the parameters. All these traces belong to the same functional space of Sobolev type

$$\Xi^{(n)}_{p,\mathrm{loc}} = \left\{(\varphi,\psi) \in W^{1,p}_{\mathrm{loc}}([0,+\infty[,\mathbb{R}^n) \times W^{1,p}_{\mathrm{loc}}([0,+\infty[,\mathbb{R}^n) : \varphi(0) = \psi(0)\right\},$$

which does not depend on (u,v).

In the previous papers A. Villani [11], G. Pulvirenti - G. Santagati [4] we allowed the control U to range over the entire space $L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^m)$ and assumed that the *initial locus* $l(a_0,b_0)$ and the *final locus* $l(a,b)$, where $(a_0,b_0),(a,b) \in I \times J$, $a_0 < a$, $b_0 < b$ – that are the loci where the system is required to take its *initial state* and *final state,* respectively – were both fixed. In that framework we provided conditions in order that for every initial state the corresponding set of final states (*attainable set*) be equal to the whole space $\Xi^{(n)}_{p,\mathrm{loc}}$ (*exact complete controllability problem*) or to a dense subspace of it (*approximate complete controllability problem*).

In this paper, taking an analogous controllability problem for a lumped parameter control process (see R. Conti [1], Sections VI. 4 and VI. 5) as a model, we no longer let the control U range over the whole space $L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^m)$, but assume that U is constrained within a proper subset $\mathcal{U}$ of that space. On the contrary, we do not consider a fixed initial locus $l(a_0,b_0)$, but we suppose that $l(a_0,b_0)$ may vary with

$$(a_0,b_0) \in I \times J,\ a_0 \le a,\ b_0 \le b.$$

We first introduce the functional spaces that will be used in the paper (Section 2) and study a linear hyperbolic system related with (E) (Section 3). Then, we give a representation formula for the solutions of (E) (Section 4) and show a characterization of the attainable set (Section 5).

Next (Section 6), assuming a given element (φ_0,ψ_0) of $\Xi^{(n)}_{p,\mathrm{loc}}$ as the initial state, to be taken on the variable initial locus $l(a_0,b_0)$, we consider the subset of $\Xi^{(n)}_{p,\mathrm{loc}}$ which is the union of all corresponding attainable sets on the fixed final locus $l(a,b)$. We focus about two main problems related to the above mentioned set: if such a set coincides with the whole space $\Xi^{(n)}_{p,\mathrm{loc}}$, or if such a set is a dense subset of $\Xi^{(n)}_{p,\mathrm{loc}}$. Regarding the second problem we establish (Section 7), by means of the *adjoint* map of a suitable functional transformation, a necessary and sufficient condition of solubility in the case that (φ_0,ψ_0) is the null element of $\Xi^{(n)}_{p,\mathrm{loc}}$ and $\mathcal{U}$ is the unit ball of $L^\infty(L_{I,J},\mathbb{R}^m)$. In this way we get a complete approximate controllability type result, which is of use when the whole space of controls is not available, but it is possible for the initial state to be taken on a variable initial locus. The above mentioned solubility condition is formulated by means of the integral of a functional constructed from the data. Also, this condition

presents some analogy with the necessary and sufficient condition for the complete approximate controllability, with fixed initial locus $l(a_0, b_0)$ and fixed final locus $l(a, b)$, established in the paper G. Pulvirenti - G. Santagati [4], already cited.

2. FUNCTIONAL SPACES

In this Section we introduce the functional spaces that we will make use of in the course of the paper, along with their main properties. The reader is referred to Section 2 of A. Villani [10] and Sections 2 and 3 of G. Pulvirenti - G. Santagati - A. Villani [5] for more information on these topics.

Henceforth, we shall assume that p and p' are conjugate exponents in $[1, +\infty]$ and that X is a measurable subset of $\mathbb{R}^d$, with positive measure.

Definition 2.1. $L^p_{\mathrm{loc}}(X, \mathbb{R}^s)$ is the complete l.t. space[1] of all [classes of] measurable functions $l : X \to \mathbb{R}^s$ whose restrictions to every compact set $K \subseteq X$ belong to $L^p(K, \mathbb{R}^s)$, endowed with the topology defined by the seminorms:

$$\pi_K(l) = \| l \|_{L^p(K, \mathbb{R}^s)} \qquad \forall l \in L^p_{\mathrm{loc}}(X, \mathbb{R}^s),$$

where K ranges in the collection of all compact subsets of X.

It is worth to remind that $L^p_{\mathrm{loc}}(X, \mathbb{R}^s)$ is a metrizable space (hence a Fréchet space) if and only if there exists a sequence $\{C_r\}$ of compact subsets of X having the following property: for each compact set $K \subseteq X$ there is some C_r for which $m(K \setminus C_r) = 0$. If the set X is such that $\overset{\circ}{X}$ is a dense subset of X, then the preceding metrizability condition notably simplifies, namely we have that $L^p_{\mathrm{loc}}(X, \mathbb{R}^s)$ is a metrizable space if and only if $\overline{X} \setminus X$ is a closed set. By means of this simplified condition the space $L^p_{\mathrm{loc}}(X, \mathbb{R}^s)$ is easily seen to be metrizable in all cases considered in this paper. The reader is referred to A. Villani [12] for the above mentioned metrizability conditions for $L^p_{\mathrm{loc}}(X, \mathbb{R}^s)$.

Definition 2.2. $L^{p'}_c(X, \mathbb{R}^s)$ is the l.t. space of all [classes of] measurable functions $\sigma : X \to \mathbb{R}^s$ which belong to $L^{p'}(X, \mathbb{R}^s)$ and vanish outside of

[1] By an l.t. space we mean a locally convex Hausdorf topological vector space.

some compact subset of X, endowed with the topology defined by the seminorms:

$$\pi_Y(\sigma) = \sup_{l \in Y} \left| \int_X \sigma^*(x) l(x) dx \right| \quad \forall \sigma \in L_c^{p'}(X, \mathbb{R}^s),$$

where Y ranges in the collection of all bounded[2] subsets of $L_{\text{loc}}^p(X, \mathbb{R}^s)$.

Consider the map $\sigma \to l'(\sigma)$, from $L_c^{p'}(X, \mathbb{R}^s)$ to $(L_{\text{loc}}^p(X, \mathbb{R}^s))'$, the strong dual space of $L_{\text{loc}}^p(X, \mathbb{R}^s)$, defined as follows:

$$< l, l'(\sigma) > = \int_X \sigma^*(x) l(x) dx \qquad \forall l \in L_{\text{loc}}^p(X, \mathbb{R}^s). \tag{2.1}$$

Then we have the following theorem.

Theorem 2.1. *The map* $\sigma \to l'(\sigma)$, *which assigns to each* $\sigma \in L_c^{p'}(X, \mathbb{R}^s)$ *the element* $l'(\sigma)$ *of* $(L_{\text{loc}}^p(X, \mathbb{R}^s))'$ *given by* (2.1), *is an algebraic and topological isomorphism between* $L_c^{p'}(X, \mathbb{R}^s)$ *and the linear subspace* $l'(L_c^{p'}(X, \mathbb{R}^s))$ *of* $(L_{\text{loc}}^p(X, \mathbb{R}^s))'$. *Moreover, for* $p \in [1, +\infty[$, *we have*

$$l'(L_c^{p'}(X, \mathbb{R}^s)) = (L_{\text{loc}}^p(X, \mathbb{R}^s))'.$$

Definition 2.3. Let Ω be an open subset of $\mathbb{R}^2$. Then $W_p^*(\Omega, \mathbb{R}^n)$ is the Banach space of all [classes of] measurable functions $w : \Omega \to \mathbb{R}^n$ which belong to $L^p(\Omega, \mathbb{R}^n)$ along with their weak derivatives w_x, w_y, w_{xy}, with the following norm:

$$\| w \|_{W_p^*(\Omega, \mathbb{R}^n)} = \left(\| w \|_{L^p(\Omega, \mathbb{R}^n)}^p + \| w_x \|_{L^p(\Omega, \mathbb{R}^n)}^p + \| w_y \|_{L^p(\Omega, \mathbb{R}^n)}^p + \| w_{xy} \|_{L^p(\Omega, \mathbb{R}^n)}^p \right)^{\frac{1}{p}},$$
$$p \in [1, +\infty[,$$

$$\| w \|_{W_\infty^*(\Omega, \mathbb{R}^n)} = \sup \left\{ \| w \|_{L^\infty(\Omega, \mathbb{R}^n)}, \| w_x \|_{L^\infty(\Omega, \mathbb{R}^n)}, \| w_y \|_{L^\infty(\Omega, \mathbb{R}^n)}, \| w_{xy} \|_{L^\infty(\Omega, \mathbb{R}^n)} \right\}.$$

The spaces $W_p^*(\Omega, \mathbb{R}^n)$, already studied in R. Di Vincenzo - A. Villani [2] and in M.B. Suryanarayana [9], play a role in the forthcoming definition, which is concerned with the space of the solutions to (E).

[2] Here, of course, boundedness has to be understood according to the theory of topological vector space.

Definition 2.4. Let I, J be any two non-degenerate intervals of $\mathbb{R}$.. Then $W^*_{p,\mathrm{loc}}(L_{I,J}, \mathbb{R}^n)$ is the Fréchet space of all [classes of] measurable functions $w : L_{I,J} \to \mathbb{R}^n$, whose restrictions to every open bounded set Ω such that $\overline{\Omega} \subseteq L_{I,J}$ belong to $W^*_p(\Omega, \mathbb{R}^n)$, endowed with the topology defined by the seminorms:

$$\pi_{\Omega}(w) = \| w \|_{W^*_p(\Omega,\mathbb{R}^n)} \qquad \forall w \in W^*_{p,\mathrm{loc}}(L_{I,J}, \mathbb{R}^n),$$

where Ω ranges in the collection of all bounded open set such that $\overline{\Omega} \subseteq L_{I,J}$.

It is apparent that the elements of $W^*_{p,\mathrm{loc}}(L_{I,J}, \mathbb{R}^n)$ are precisely those functions w such that $w, w_x, w_y, w_{xy} \in L^p_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^n)$, where w_x, w_y, w_{xy} are weak derivatives on $\overset{\circ}{\widehat{L_{I,J}}}$. Moreover, denoting by I^∞ [resp. J^∞] the union of the interval I [resp. J] and the set of its upper bounds (possibly empty), if we consider the product Fréchet space

$$S_{p,\mathrm{loc}}(L_{I,J}, \mathbb{R}^n) = L^p_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^n) \times L^p_{\mathrm{loc}}(I^\infty, \mathbb{R}^n) \times L^p_{\mathrm{loc}}(J^\infty, \mathbb{R}^n) \times \mathbb{R}^n,$$

we can show by similar arguments to those used in in A. Villani [10] (Theorem 2.1 and Proposition 2.2) that the following theorem holds.

Theorem 2.2. *Let $(\bar{a}, \bar{b})$ be any fixed point in $I \times J$. Then we have:*

1) *the elements of $W^*_{p,\mathrm{loc}}(L_{I,J}, \mathbb{R}^n)$ are precisely those functions w which can be written in the following form:*

$$w(x,y) = \int_{\bar{a}}^{x} \int_{\bar{b}}^{y} h(u,v)dudv + \int_{\bar{a}}^{x} h_1(u)du + \int_{\bar{b}}^{y} h_2(v)dv + \lambda, \qquad \forall (x,y) \in L_{I,J}, \tag{2.2}$$

with $(h, h_1, h_2, \lambda) \in S_{p,\mathrm{loc}}(L_{I,J}, \mathbb{R}^n)$;

2) *the linear map $(h, h_1, h_2, \lambda) \to w$, established by (2.2), is an algebraic and topological isomorphism between $S_{p,\mathrm{loc}}(L_{I,J}, \mathbb{R}^n)$ and $W^*_{p,\mathrm{loc}}(L_{I,J}, \mathbb{R}^n)$.*

Corollary 2.1. $W^{*}_{p,\mathrm{loc}}(L_{I,J},\mathbb{R}^n)$ *is embedded in* $C^0(L_{I,J},\mathbb{R}^n)$[3], *both algebraically and topologically.*

Since functions $w \in W^{*}_{p,\mathrm{loc}}(L_{I,J},\mathbb{R}^n)$ are continuous, it is possible to consider their traces on every set $l(u,v)$, $(u,v) \in I \times J$.

It is noteworthy that such traces can be regarded as elements of a unique functional space $\Xi^{(n)}_{p,\mathrm{loc}}$ (that is, the space $\tilde{\Xi}^{(n)}_{p}$ already considered in the previous papers; see, e.g., G. Pulvirenti - G. Santagati - A. Villani [6]) which does not depend on (u,v).

To the aim of reminding the definition of the space $\Xi^{(n)}_{p,\mathrm{loc}}$ and some useful properties of it, for the sake of completeness we start by setting the following definition.

Definition 2.5. Let G be any interval of $\mathbb{R}$. Then $W^{1,p}_{\mathrm{loc}}(\mathrm{G},\mathbb{R}^n)$ is the Fréchet space of all [classes of] measurable functions $\varphi : \mathrm{G} \to \mathbb{R}^n$ whose restrictions to every open bounded set A such that $\overline{\mathrm{A}} \subseteq \mathrm{G}$ belong to the Sobolev space $W^{1,p}(\mathrm{A},\mathbb{R}^n)$, endowed with the topology defined by the seminorms:

$$\pi_{\mathrm{A}}(\varphi) = \| \varphi \|_{W^{1,p}(\mathrm{A},\mathbb{R}^n)} \qquad \forall \varphi \in W^{1,p}_{\mathrm{loc}}(\mathrm{G},\mathbb{R}^n),$$

where A ranges in the collection of all bounded open set such that $\overline{\mathrm{A}} \subseteq \mathrm{G}$.

Similarly to what we noticed about $W^{*}_{p,\mathrm{loc}}(L_{I,J},\mathbb{R}^n)$, we have that the elements of $W^{1,p}_{\mathrm{loc}}(\mathrm{G},\mathbb{R}^n)$ are precisely those functions φ which belong to $L^{p}_{\mathrm{loc}}(\mathrm{G},\mathbb{R}^n)$ together with φ' (the weak derivative on $\overset{\circ}{\mathrm{G}}$). Also, the following theorem holds.

Theorem 2.3. *Let* $\bar{t}$ *be any fixed point in* G. *Then we have:*

1) *the elements of* $W^{1,p}_{\mathrm{loc}}(\mathrm{G},\mathbb{R}^n)$ *are precisely those functions* φ *which can be written in the form*

[3] $C^0(L_{I,J},\mathbb{R}^n)$ endowed with the topology of $L^{\infty}_{\mathrm{loc}}(L_{I,J},\mathbb{R}^n)$.

$$\varphi(t) = \int_{\bar{t}}^{t} k(s)ds + \delta \qquad \forall t \in \mathrm{G}, \tag{2.3}$$

with $(k,\delta) \in L^p_{\mathrm{loc}}(\mathrm{G},\mathbb{R}^n)\times\mathbb{R}^n$;

2) *the linear map* $(k,\delta) \to \varphi$, *established by* (2.3), *is an algebraic and topological isomorphism between the product Fréchet space* $L^p_{\mathrm{loc}}(\mathrm{G},\mathbb{R}^n)\times\mathbb{R}^n$ *and* $W^{1,p}_{\mathrm{loc}}(\mathrm{G},\mathbb{R}^n)$.

Theorems 2.2 and 2.3 imply, in an obvious way,

Corollary 2.2. *For every* $w \in W^{*}_{p,\mathrm{loc}}(L_{I,J},\mathbb{R}^n)$ *we have:*

$$w(\cdot,y) \in W^{1,p}_{\mathrm{loc}}(I^{\infty},\mathbb{R}^n)\ \forall y \in J, \quad w(x,\cdot) \in W^{1,p}_{\mathrm{loc}}(J^{\infty},\mathbb{R}^n)\ \ \forall x \in I.$$

Also, for each $y \in J$ [resp. $x \in I$], *we have that*

$$w \to w(\cdot,y)\ [\,\text{resp.}\, w \to w(x,\cdot)\,]$$

is a continuous linear map from the space $W^{*}_{p,\mathrm{loc}}(L_{I,J},\mathbb{R}^n)$ *onto the space* $W^{1,p}_{\mathrm{loc}}(I^{\infty},\mathbb{R}^n)$ [resp. $W^{1,p}_{\mathrm{loc}}(J^{\infty},\mathbb{R}^n)$].

Obviously, Theorem 2.3 ensures that every element of $W^{1,p}_{\mathrm{loc}}(\mathrm{G},\mathbb{R}^n)$ is a continuous function in G. Hence, the following definition is meaningful.

Definition 2.6. $\Xi^{(n)}_{p,\mathrm{loc}}$ is the Fréchet space

$$\left\{(\varphi,\psi) \in W^{1,p}_{\mathrm{loc}}([0,+\infty[,\mathbb{R}^n)\times W^{1,p}_{\mathrm{loc}}([0,+\infty[,\mathbb{R}^n) : \varphi(0) = \psi(0)\right\},$$

closed linear subspace of the product Fréchet space

$$W^{1,p}_{\mathrm{loc}}([0,+\infty[,\mathbb{R}^n)\times W^{1,p}_{\mathrm{loc}}([0,+\infty[,\mathbb{R}^n)$$

endowed with the topology given by the seminorms:

$$\pi_{\mathrm{A,B}}(\varphi,\psi) = \pi_{\mathrm{A}}(\varphi) + \pi_{\mathrm{B}}(\psi),$$
$$\forall(\varphi,\psi) \in W^{1,p}_{\mathrm{loc}}([0,+\infty[,\mathbb{R}^n)\times W^{1,p}_{\mathrm{loc}}([0,+\infty[,\mathbb{R}^n)$$

where A and B range in the collection of all bounded and open subsets of $]0,+\infty[$.

Another consequence of Theorem 2.3 is:

Theorem 2.4. *The elements of* $\Xi^{(n)}_{p,\mathrm{loc}}$ *are precisely those pairs of functions* (φ,ψ) *which can be represented as follows:*

$$\varphi(t)=\int_0^t k(s)ds+\delta,\ \ \psi(t)=\int_0^t l(s)ds+\delta\ \ \forall t\in[0,+\infty[, \tag{2.4}$$

where

$$(k,l,\delta)\in L^p_{\mathrm{loc}}([0,+\infty[,\mathbb{R}^n)\times L^p_{\mathrm{loc}}([0,+\infty[,\mathbb{R}^n)\times\mathbb{R}^n;$$

the linear map $(k,l,\delta)\to(\varphi,\psi)$, *established by* (2.4), *is an algebraic and topological isomorphism between the product Fréchet space*

$$L^p_{\mathrm{loc}}([0,+\infty[,\mathbb{R}^n)\times L^p_{\mathrm{loc}}([0,+\infty[,\mathbb{R}^n)\times\mathbb{R}^n$$

and $\Xi^{(n)}_{p,\mathrm{loc}}$.

Since the inverse isomorphism of $(k,l,\delta)\to(\varphi,\psi)$ is the map which associates

$$(\varphi',\psi',\varphi(0))\in L^p_{\mathrm{loc}}([0,+\infty[,\mathbb{R}^n)\times L^p_{\mathrm{loc}}([0,+\infty[,\mathbb{R}^n)\times\mathbb{R}^n$$

to each $(\varphi,\psi)\in\Xi^{(n)}_{p,\mathrm{loc}}$, taking into account Theorem 2.1 we obtain:

Theorem 2.5. *Let* $p\in[1,+\infty[$. *Then the map which transforms each*

$$(\mu,\nu,\xi)\in L^{p'}_c([0,+\infty[,\mathbb{R}^n)\times L^{p'}_c([0,+\infty[,\mathbb{R}^n)\times\mathbb{R}^n$$

into the element Q *of* $(\Xi^{(n)}_{p,\mathrm{loc}})'$, *the strong dual space of* $\Xi^{(n)}_{p,\mathrm{loc}}$, *given by*

$$<(\varphi,\psi),Q>=\int_0^{+\infty}\mu^*(t)\varphi'(t)dt+\int_0^{+\infty}\nu^*(t)\psi'(t)dt+\xi^*\varphi(0)\qquad\forall(\varphi,\psi)\in\Xi^{(n)}_{p,\mathrm{loc}}, \tag{2.5}$$

is an algebraic and topological isomorphism between the product l.t. space

$$L_c^{p'}([0,+\infty[,\mathbb{R}^n)\times L_c^{p'}([0,+\infty[,\mathbb{R}^n)\times\mathbb{R}^n$$

and $(\Xi_{p,\mathrm{loc}}^{(n)})'$.

The above theorem implies that (see G. Pulvirenti - G. Santagati - A. Villani [6]).

Theorem 2.6. *For* $p\in]1,+\infty[$ *the space* $\Xi_{p,\mathrm{loc}}^{(n)}$ *is reflexive.*

Now, in order to regard the traces of the functions $w\in W_{p,\mathrm{loc}}^{*}(L_{I,J},\mathbb{R}^n)$ on $l(u,v)$, $(u,v)\in I\times J$, as elements of the functional space $\Xi_{p,\mathrm{loc}}^{(n)}$, we notice that, given $(u,v)\in I\times J$, by Theorems 2.2 and 2.4 the restriction of each $w\in W_{p,\mathrm{loc}}^{*}(L_{I,J},\mathbb{R}^n)$ to $l(u,v)$ individualizes an element

$$\gamma_{(u,v)}w=(\varphi_{(u,v),w},\psi_{(u,v),w})$$

of $\Xi_{p,\mathrm{loc}}^{(n)}$ by means of the following equations:

$$\varphi_{(u,v),w}(t)=w(u+t,v),\ \psi_{(u,v),w}(t)=w(u,v+t)\ \ \forall t\in[0,+\infty[. \tag{2.6}$$

Consequently, the following definition is meaningful.

Definition 2.7. For each $w\in W_{p,\mathrm{loc}}^{*}(L_{I,J},\mathbb{R}^n)$ and each $(u,v)\in I\times J$, we call *trace* of w on $l(u,v)$ the element $\gamma_{(u,v)}w$ of $\Xi_{p,\mathrm{loc}}^{(n)}$ given by (2.6).

Furthermore, the following theorem is true.

Theorem 2.7. *For each* $(u,v)\in I\times J$, *the map* $w\to\gamma_{(u,v)}w$ *is a continuous linear map from* $W_{p,\mathrm{loc}}^{*}(L_{I,J},\mathbb{R}^n)$ *onto* $\Xi_{p,\mathrm{loc}}^{(n)}$.

3. EXISTENCE, UNIQUENESS, CONTINUOUS DEPENDENCE AND REPRESENTATION FORMULA FOR THE SOLUTIONS TO A LINEAR HYPERBOLIC SYSTEM

We henceforth suppose that I, J are two non-degenerate intervals of $\mathbb{R}$ and that the coefficients A, B and C of (E) satisfy the following assumption:

$$A, B, C, A_x, B_y \in C^0(L_{I,J}, \mathbb{R}^{n,n}). \tag{3.1}$$

We denote by P the continuous linear differential operator, from $W^*_{p,\mathrm{loc}}(L_{I,J}, \mathbb{R}^n)$ to $L^p_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^n)$, defined by putting:

$$Pw = w_{xy} + Aw_x + Bw_y + Cw \quad \forall w \in W^*_{p,\mathrm{loc}}(L_{I,J}, \mathbb{R}^n). \tag{3.2}$$

Fixed a point $(\bar{a}, \bar{b}) \in I \times J$, let us consider the problem:

$$\begin{cases} w \in W^*_{p,\mathrm{loc}}(L_{I,J}, \mathbb{R}^n), \\ Pw = f, \\ w(\cdot, \bar{b}) = \sigma, w(\bar{a}, \cdot) = \tau, \end{cases} \tag{3.3}$$

where $f \in L^p_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^n)$, $\sigma \in W^{1,p}_{\mathrm{loc}}(I^\infty, \mathbb{R}^n)$, $\tau \in W^{1,p}_{\mathrm{loc}}(J^\infty, \mathbb{R}^n)$, $\sigma(\bar{a}) = \tau(\bar{b})$.

By a similar argument to that used in A. Villani [10], Theorem 3.1, one proves that the following theorem holds true.

Theorem 3.1. *Let* $(\bar{a}, \bar{b}) \in I \times J$. *Then, for each fixed element* $((\sigma, \tau), f)$ *of the product Fréchet space*[4]

$$\left\{(\sigma, \tau) \in W^{1,p}_{\mathrm{loc}}(I^\infty, \mathbb{R}^n) \times W^{1,p}_{\mathrm{loc}}(J^\infty, \mathbb{R}^n) : \sigma(\bar{a}) = \tau(\bar{b})\right\} \times L^p_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^n), \tag{3.4}$$

Problem (3.3) *has a unique solution* $w_{(\sigma,\tau),f}$.

The map

[4] closed linear subspace of $W^{1,p}_{loc}(I^\infty, \mathbb{R}^n) \times W^{1,p}_{loc}(J^\infty, \mathbb{R}^n) \times L^p_{loc}(L_{I,J}, \mathbb{R}^n)$.

$$((\sigma,\tau),f) \to w_{(\sigma,\tau),f} \tag{3.5}$$

is an algebraic and topological isomorphism between the space (3.4) *and* $W^*_{p,\mathrm{loc}}(L_{I,J},\mathbb{R}^n)$.

Remark 3.1. In order to prove Theorem 3.1 the above assumption (3.1) may be weakened; indeed, the already mentioned argument used in A. Villani [10] simply requires that the coefficients A, B and C belong to $L^\infty_{\mathrm{loc}}(L_{I,J},\mathbb{R}^{n,n})$; moreover, it is well known that further generalizations are possible (see for instance G. Sturiale [8]). However, assumption (3.1) allows us to get also a representation formula for the solutions to Problem (3.3), through an *evolution matrix*, which is constructed by means of the coefficients of the operator P.

To define such an evolution matrix, we start by reminding the following result (see A. Villani [10], Theorem 4.1).

Theorem 3.2. *Let* $D=[u',u'']\times[v',v'']$ *be a closed rectangle of* $\mathbb{R}^2$ *such that* $D\subseteq L_{I,J}$.

Then, for each $(x,y)\in D$, *there exists a unique function* $(u,v)\to V^D(u,v;x.y)$, *from* D *to* $\mathbb{R}^{n,n}$, *continuous in* D *together with the derivatives* V^D_u, V^D_v, V^D_{uv}, *solution to the following problem:*

$$\begin{cases} V_{uv}-(VA)_u-(VB)_v+VC=O & \forall (u,v)\in D,\\ V_u-VB=O \quad v=y, & \forall u\in[u',u''],\\ V_v-VA=O \quad u=x, & \forall v\in[v',v''],\\ V(x,y;x,y)=I. \end{cases} \tag{3.6}$$

Also, we have that the function $(u,v;x,y)\to V^D(u,v;x,y)$, *from* $D\times D$ *to* $\mathbb{R}^{n,n}$, *is continuous in* $D\times D$ *together with the derivatives* $V^D_u, V^D_v, V^D_{uv}, V^D_x, V^D_y V^D_{xy}$.

Clearly, Theorem 3.2 also implies that if $(x,y)\in L_{I,J}$ and D_1, D_2 are any two closed rectangles such that

$$(x,y)\in D_1\cap D_2,\ D_1\cup D_2\subseteq L_{I,J},$$

then we have

$$V^{D_1}(u,v;x,y) = V^{D_2}(u,v;x,y), \qquad \forall (u,v) \in D_1 \cap D_2.$$

Denote by $T_{I,J}$ the subset of $\mathbb{R}^4$ consisting of all points $(u,v;x,y)$ for which there is some closed rectangle D of $\mathbb{R}^2$ such that $D \subseteq L_{I,J}$ and that $(u,v),(x,y) \in D$. Then, thanks to the previous remark, we are allowed to set the following definition.

Definition 3.1. We call *evolution matrix*, associated with the differential operator P, the function $(u,v;x,y) \to V(u,v;x,y)$, from $T_{I,J}$ to $\mathbb{R}^{n,n}$, that assigns to each $(u,v;x,y)$ the value $V^D(u,v;x,y)$, where D is any closed rectangle of $\mathbb{R}^2$ which is contained in $L_{I,J}$ and contains both points (u,v) and (x,y).

Then, by a similar argument to that in Section 4 of A. Villani [10], we get the following representation result for the solutions of Problem (3.3).

Theorem 3.3. *For each* $(\bar{a},\bar{b}) \in I \times J$, *each*

$$(\sigma,\tau) \in W^{1,p}_{\text{loc}}(I^\infty,\mathbb{R}^n) \times W^{1,p}_{\text{loc}}(J^\infty,\mathbb{R}^n),$$

such that $\sigma(\bar{a}) = \tau(\bar{b})$, *and each* $f \in L^p_{\text{loc}}(L_{I,J},\mathbb{R}^n)$, *we have that function* $w_{(\sigma,\tau),f}$, *the unique solution to Problem* (3.3), *is represented by means of the following formula:*

$$\begin{aligned} w_{(\sigma,\tau),f}(x,y) &= V(\bar{a},\bar{b};x,y)\sigma(\bar{a}) + \\ &+ \int_{\bar{a}}^{x} V(u,\bar{b};x,y)[\sigma'(u) + B(u,\bar{b})\sigma(u)]du + \\ &+ \int_{\bar{b}}^{y} V(\bar{a},v;x,y)[\tau'(v) + A(\bar{a},v)\tau(v)]dv + \\ &+ \int_{\bar{a}}^{x}\int_{\bar{b}}^{y} V(u,v;x,y)f(u,v))dudv, \qquad \forall (x,y) \in L_{I,J}. \end{aligned}$$

4. SOLUTIONS OF (E)

Keeping assumption (3.1), from now on we further suppose that

$$F \in L^{\infty}_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^{n,m}).$$

Let the initial locus $l(a_0, b_0)$, with $(a_0, b_0) \in I \times J$, the initial state $(\varphi_0, \psi_0) \in \Xi^{(n)}_{p,\mathrm{loc}}$ and the control $U \in L^{p}_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^m)$ be fixed. Then, it follows from Theorem 3.1 that there exist functions z, elements of $W^{*}_{p,\mathrm{loc}}(L_{I,J}, \mathbb{R}^n)$, which are solutions of (E) and, in addition, satisfy the following condition

$$\gamma_{(a_0,b_0)} z = (\varphi_0, \psi_0).$$

They are precisely all functions

$$w_{(\sigma,\tau),FU}$$

where (σ, τ) is any element of $W^{1,p}_{\mathrm{loc}}(I^{\infty}, \mathbb{R}^n) \times W^{1,p}_{\mathrm{loc}}(J^{\infty}, \mathbb{R}^n)$ such that

$$\sigma(x) = \varphi_0(x - a_0) \quad \forall x \geq a_0, \quad \tau(y) = \psi_0(y - b_0) \quad \forall y \geq b_0.$$

Thus, excluding the case $a_0 = \inf I$ and $b_0 = \inf J$, the set of such functions z is infinite.

Moreover, if $a_0 < \sup I$ and $b_0 < \sup J$, all the restrictions of the above mentioned functions z to the set

$$L_{I_{a_0}, J_{b_0}},$$

where $I_{a_0} = I \cap [a_0, +\infty[$, $J_{b_0} = J \cap [b_0, +\infty[$, coincide with a unique element

$$z(\cdot; (a_0, b_0), (\varphi_0, \psi_0), U) \tag{4.2}$$

of the space $W^{*}_{p,\mathrm{loc}}(L_{I_{a_0}, J_{b_0}}, \mathbb{R}^n)$ and the map

$$((\varphi_0, \psi_0), U) \to z(\cdot; (a_0, b_0), (\varphi_0, \psi_0), U), \tag{4.3}$$

from $\Xi^{(n)}_{p,\mathrm{loc}} \times L^p_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^m)$ to $W^*_{p,\mathrm{loc}}(L_{I_{a_0},J_{b_0}}, \mathbb{R}^n)$, is linear and continuous.

By Theorem 3.3, function (4.2) can be represented by the formula

$$z(x,y;(a_0,b_0),(\varphi_0,\psi_0),U) = \zeta(x,y;(a_0,b_0),(\varphi_0,\psi_0)) + \\ + \int_{a_0}^{x}\int_{b_0}^{y} V(u,v;x,y)F(u,v)U(u,v)dudv \qquad \forall (x,y) \in L_{I_{a_0},J_{b_0}}, \tag{4.4}$$

where V is the evolution matrix, associated with the differential operator P, and

$$\zeta(x,y;(a_0,b_0),(\varphi_0,\psi_0)) = V(a_0,b_0;x,y)\varphi_0(0) + \\ + \int_{a_0}^{x} V(u,b_0;x,y)[\varphi_0'(u-a_0) + B(u,b_0)\varphi_0(u-a_0)]du + \tag{4.5}$$

$$+ \int_{b_0}^{y} V(a_0,v;x,y)[\psi_0'(v-b_0) + A(a_0,v)\psi_0(v-b_0)]dv, \ \forall (x,y) \in L_{I_{a_0},J_{b_0}}.$$

Remark 4.1. It is apparent that the function

$$(x,y) \to \zeta(x,y;(a_0,b_0),(\varphi_0,\psi_0)) = z(x,y;(a_0,b_0),(\varphi_0,\psi_0),0), \quad (x,y) \in L_{I_{a_0},J_{b_0}},$$

is the unique solution to the problem below:

$$\begin{cases} \zeta \in W^*_{p,\mathrm{loc}}(L_{I_{a_0},J_{b_0}}, \mathbb{R}^n), \\ P\zeta = 0, \\ \gamma_{(a_0,b_0)}\zeta = (\varphi_0,\psi_0). \end{cases}$$

Likewise, the function

$$(x,y) \to \int_{a_0}^{x}\int_{b_0}^{y} V(u,v;x,y)F(u,v)U(u,v)dudv = z(x,y;(a_0,b_0),(0,0),U), \\ (x,y) \in L_{I_{a_0},J_{b_0}},$$

is the unique solution to the following problem :

$$\begin{cases} z \in W^*_{p,\mathrm{loc}}(L_{I_{a_0},J_{b_0}}, \mathbb{R}^n), \\ Pz = FU, \\ \gamma_{(a_0,b_0)} z = (0,0). \end{cases}$$

5. THE ATTAINABLE SET

Now we assume that, besides the initial locus $l(a_0,b_0)$, also the final locus $l(a,b)$, with $(a,b) \in I \times J$ such that $a_0 \le a$, $b_0 \le b$, and the set $\mathcal{U} \subseteq L^p_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^m)$ of the available controls are given. Moreover, just to simplify our exposition, we henceforth suppose that $a < \sup I$, $b < \sup J$[5].

Definition 5.1. Let $(a_0,b_0),(a,b) \in I \times J$, with[5]

$$a_0 \le a, \quad b_0 \le b, \quad (\varphi_0,\psi_0) \in \Xi^{(n)}_{p,\mathrm{loc}} \quad \text{and} \quad \mathcal{U} \subseteq L^p_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^m)$$

be given. We call *attainable set* on $l(a,b)$, from the initial state (φ_0,ψ_0) on $l(a_0,b_0)$, by means of the controls $U \in \mathcal{U}$, the set

$$\mathcal{A}((a_0,b_0),(a,b);(\varphi_0,\psi_0),\mathcal{U}) = \{\gamma_{(a,b)} z(\cdot;(a_0,b_0),(\varphi_0,\psi_0),U) : U \in \mathcal{U}\},$$

that is, the set consisting of all final states on $l(a,b)$, which are obtained as U ranges in $\mathcal{U}$.

From (4.4) and Remark 4.1 it follows

$$\mathcal{A}((a_0,b_0),(a,b);(\varphi_0,\psi_0),\mathcal{U}) = \gamma_{(a,b)}\zeta(\cdot;(a_0,b_0),(\varphi_0,\psi_0)) + \Lambda_{(a_0,b_0),(a,b)}\mathcal{U},$$

where $\Lambda_{(a_0,b_0),(a,b)}$ is the linear map specified in the definition below.

Definition 5.2. (*The map* $\Lambda_{(a_0,b_0),(a,b)}$.) Given $(a_0,b_0),(a,b) \in I \times J$, with $a_0 \le a$, $b_0 \le b$, we denote by

[5] We refer the reader to the forthcoming paper G. Pulvirenti – G. Santagati – A. Villani [7] to see how this assumption can be removed.

$$\Lambda_{(a_0,b_0),(a,b)}$$

the map, from $L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^m)$ to $\Xi^{(n)}_{p,\mathrm{loc}}$, which assigns to each $U \in L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^m)$ the following element of the space $\Xi^{(n)}_{p,\mathrm{loc}}$:

$$\Lambda_{(a_0,b_0),(a,b)}U = \gamma_{(a,b)}z(\cdot;(a_0,b_0),(0,0),U).$$

Having in mind that map (4.3) is linear and continuous, we have, as a particular case, that also

$$U \to z(\cdot;(a_0,b_0),(0,0),U)$$

is a continuous linear map from $L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^m)$ to $W^*_{p,\mathrm{loc}}(L_{I,J},\mathbb{R}^n)$; hence, by Theorem 2.7, we get the following proposition.

Proposition 5.1. *$\Lambda_{(a_0,b_0),(a,b)}$ is a continuous linear map from $L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^m)$ to $\Xi^{(n)}_{p,\mathrm{loc}}$.*

Let

$$\Lambda'_{(a_0,b_0),(a,b)} : \left(\Xi^{(n)}_{p,\mathrm{loc}}\right)' \to \left(L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^m)\right)',$$

be the adjoint map of $\Lambda_{(a_0,b_0),(a,b)}$. By a general result concerning the description of the closed convex hull of a set by means of its support function (see, for instance, G. Pulvirenti - G. Santagati - A. Villani [5], Proposition 5.1), we have the following characterization of the set $\overline{\mathrm{co}}(\Lambda_{(a_0,b_0),(a,b)}\mathcal{U})$.

Theorem 5.1. *Let $(a_0,b_0),(a,b) \in I \times J$, with $a_0 \le a$, $b_0 \le b$, and $\mathcal{U} \subseteq L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^m)$ be given. Then*

$$\overline{\mathrm{co}}(\Lambda_{(a_0,b_0),(a,b)}\mathcal{U}) =$$

$$= \{(\chi,\eta) \in \Xi^{(n)}_{p,\mathrm{loc}} : < (\chi,\eta), Q > \le \sup_{U\in\mathcal{U}} < \Lambda_{(a_0,b_0),(a,b)}U, Q > \forall Q \in (\Xi^{(n)}_{p,\mathrm{loc}})'\} =$$

$$= \{(\chi,\eta) \in \Xi^{(n)}_{p,\mathrm{loc}} : < (\chi,\eta), Q > \le \sup_{U\in\mathcal{U}} < U, \Lambda'_{(a_0,b_0),(a,b)}Q > \forall Q \in (\Xi^{(n)}_{p,\mathrm{loc}})'\}.$$

If $p \in [1,+\infty[$ it is possible to obtain an explicit representation of the adjoint map $\Lambda'_{(a_0,b_0),(a,b)}$ by identifying the dual space $\left(\Xi^{(n)}_{p,\mathrm{loc}}\right)'$ with the product space

$$L^{p'}_c([0,+\infty[,\mathbb{R}^n) \times L^{p'}_c([0,+\infty[,\mathbb{R}^n) \times \mathbb{R}^n$$

(Theorem 2.5) and the dual space $\left(L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^m)\right)'$ with $L^{p'}_c(L_{I,J},\mathbb{R}^m)$ (Theorem 2.1). Indeed, for each $(\mu,\nu,\xi) \in L^{p'}_c([0,+\infty[,\mathbb{R}^n) \times L^{p'}_c([0,+\infty[,\mathbb{R}^n) \times \mathbb{R}^n$ and each $U \in L^p_{\mathrm{loc}}(L_{I,J},\mathbb{R}^m)$, by a similar argument to that used in Section 6 of G. Pulvirenti - G. Santagati [4], one proves that

$$< U, \Lambda'_{(a_0,b_0),(a,b)}(\mu,\nu,\xi) > = < \Lambda_{(a_0,b_0),(a,b)}U, (\mu,\nu,\xi) > =$$
$$= \int\int_{L_{I_{a_0},J_{b_0}}} H^*_{(a,b)}(u,v;(\mu,\nu,\xi))U(u,v)dudv,$$

where

$$(u,v) \to H_{(a,b)}(u,v;\mu,\nu,\xi)$$

is the measurable function, from $L_{I,J}$ to $\mathbb{R}^m$, defined according to the rule:

$$H^*_{(a,b)}(u,v;(\mu,\nu,\xi)) =$$

$$= \begin{cases} \{\xi^* V(u,v;a,b) + \int_0^{+\infty} [\mu^*(t)V_x(u,v;a+t,b) + \\ \qquad +\nu^*(t)V_y(u,v;a,b+t)]dt\}F(u,v) \\ \qquad \text{a.e. } (u,v) \in L_{I,J} \cap]-\infty, a[\times]-\infty, b[, \\ \{\mu^*(u-a)V(u,v;u,b) + \int_{u-a}^{+\infty} \mu^*(t)V_x(u,v;a+t,b)dt\}F(u,v) \\ \qquad \text{a.e. } (u,v) \in L_{I,J} \cap ([a,+\infty[\times]-\infty, b[), \\ \{\nu^*(v-b)V(u,v;a,v) + \int_{v-b}^{+\infty} \nu^*(t)V_y(u,v;a,b+t)dt\}F(u,v) \\ \qquad \text{a.e. } (u,v) \in L_{I,J} \cap]-\infty, a[\times[b,+\infty[, \\ 0 \qquad \text{a.e. } (u,v) \in L_{I,J} \cap ([a,+\infty[\times[b,+\infty[). \end{cases} \tag{5.1}$$

It is easy to check that for every $(a_0,b_0) \in I \times J$, with $a_0 \le a$, $b_0 \le b$, the restriction of $H_{(a,b)}(\cdot;(\mu,\nu,\xi))$ to $L_{I_{a_0},J_{b_0}}$ is an element of $L_c^{p'}(L_{I_{a_0},J_{b_0}},\mathbb{R}^m)$, hence the function $\mathbf{1}_{L_{I_{a_0},J_{b_0}}} H_{(a,b)}(\cdot;(\mu,\nu,\xi))$ (where $\mathbf{1}_{L_{I_{a_0},J_{b_0}}}$ is the indicator of the set $L_{I_{a_0},J_{b_0}}$) belongs to $L_c^{p'}(L_{I,J},\mathbb{R}^m)$. Thus, we have the following theorem.

Theorem 5.2. *Let* $p \in [1,+\infty[$. *Given* $(a_0,b_0),(a,b) \in I \times J$, *with* $a_0 \le a$, $b_0 \le b$, *for each*

$(\mu,\nu,\xi) \in L_c^{p'}([0,+\infty[,\mathbb{R}^n) \times L_c^{p'}([0,+\infty[,\mathbb{R}^n) \times \mathbb{R}^n$

we have

$$\Lambda'_{(a_0,b_0),(a,b)}(\mu,\nu,\xi) = \mathbf{1}_{L_{I_{a_0},J_{b_0}}} H_{(a,b)}(\cdot;(\mu,\nu,\xi)).$$

Since for each $(\chi,\eta) \in \Xi_{p,\mathrm{loc}}^{(n)}$ and each

$(\mu,\nu,\xi) \in L_c^{p'}([0,+\infty[,\mathbb{R}^n) \times L_c^{p'}([0,+\infty[,\mathbb{R}^n) \times \mathbb{R}^n$

we have

$$<(\chi,\eta),(\mu,\nu,\xi)> \; = \int_0^{+\infty} \mu^*(t)\chi'(t)dt + \int_0^{+\infty} \nu^*(t)\eta'(t)dt + \xi^*\chi(0),$$

it follows that also the following theorem holds true.

Theorem 5.3. *Let* $p \in [1, +\infty[$. *Given* $(a_0, b_0), (a, b) \in I \times J$, *with* $a_0 \leq a$, $b_0 \leq b$, *and* $\mathcal{U} \subseteq L^p_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^m)$, *we have:*

$$\overline{\mathrm{co}}(\Lambda_{(a_0,b_0),(a,b)}\mathcal{U}) =$$
$$= \{(\chi, \eta) \in \Xi^{(n)}_{p,\mathrm{loc}} : \int_0^{+\infty} \mu^*(t)\chi'(t)dt + \int_0^{+\infty} \nu^*(t)\eta'(t)dt + \xi^*\chi(0) \leq$$
$$\leq \sup_{U \in \mathcal{U}} \iint_{L_{I_{a_0},J_{b_0}}} H^*_{(a,b)}(u, v; (\mu, \nu, \xi))U(u, v)dudv$$
$$\forall (\mu, \nu, \xi) \in L^{p'}_c([0, +\infty[, \mathbb{R}^n) \times L^{p'}_c([0, +\infty[, \mathbb{R}^n) \times \mathbb{R}^n\}.$$

6. CONTROLLABILITY PROBLEMS WITH VARIABLE INITIAL LOCUS

Let the final locus $l(a, b)$, $(a, b) \in I \times J$, the initial state $(\varphi_0, \psi_0) \in \Xi^{(n)}_{p,\mathrm{loc}}$ and the set $\mathcal{U} \subseteq L^p_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^m)$ of the available controls be assigned. In connection with the control process (E), we consider the following two controllability problems, where the initial locus $l(a_0, b_0)$ is allowed to vary.

Problem 6.1. (*Exact controllability with variable initial locus*).

Let $(a, b) \in I \times J$, $(\varphi_0, \psi_0) \in \Xi^{(n)}_{p,\mathrm{loc}}$ and $\mathcal{U} \subseteq L^p_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^m)$ be given.

For each $(\chi, \eta) \in \Xi^{(n)}_{p,\mathrm{loc}}$ find $(a_0, b_0) \in I \times J$, $a_0 \leq a$, $b_0 \leq b$, and $U \in \mathcal{U}$ such that

$$\gamma_{(a,b)}z(\cdot; (a_0, b_0), (\varphi_0, \psi_0), \mathcal{U}) = (\chi, \eta).$$

Of course, we have the following proposition.

Proposition 6.1. Problem 6.1 *is soluble if and only if*

$$\bigcup_{\substack{(a_0,b_0) \in I \times J \\ a_0 \leq a, b_0 \leq b}} \mathcal{A}((a_0, b_0), (a, b); (\varphi_0, \psi_0), \mathcal{U}) = \Xi^{(n)}_{p,\mathrm{loc}}. \tag{6.1}$$

Also, it is obvious that a sufficient condition in order that (6.1) hold true is the existence of some point, $(a_0, b_0) \in I \times J$, $a_0 \leq a$, $b_0 \leq b$, such that

$$\mathcal{A}((a_0, b_0), (a, b); (\varphi_0, \psi_0), \mathcal{U}) = \Xi^{(n)}_{p,\mathrm{loc}}, \tag{6.2}$$

while one expects that this condition is by no means necessary. The next example confirms this.

Example 6.1. Let $I = J = \mathbb{R}$; $n = m = 1$; $A = B = C = 0$ (hence $V(u, v; x, y) = 1$ in $\mathbb{R}^4$); $F = 1$. In other words, we are considering the scalar control process:

$$z_{xy} = U(x, y) \qquad \text{a.e. } (x, y) \in \mathbb{R}^2. \tag{6.3}$$

Also, let $(a, b) = (0, 0)$, $(\varphi_0, \psi_0) = (0, 0)$, the null element of $\Xi^{(1)}_{p,\mathrm{loc}}$, and

$$\mathcal{U} = \{U \in L^p_{\mathrm{loc}}(\mathbb{R}^2) : | U(x, y) | \leq 1 \text{ a.e. } (x, y) \in]-\infty, 0[\times]-\infty, 0[\}. \tag{6.4}$$

Let us show that Problem 6.1 has solution.

Given $(\overline{\chi}, \overline{\eta}) \in \Xi^{(n)}_{p,\mathrm{loc}}$, we choose $(a_0, b_0) \in \mathbb{R}^2$, $a_0 < 0$, $b_0 < 0$, in such a way that $a_0 b_0 \geq | \overline{\chi}(0) |$.

We remember that for each $U \in L^p_{\mathrm{loc}}(\mathbb{R}^2)$ we have

$$z(x, y; (a_0, b_0), (0, 0), U) = \int_{a_0}^{x} \int_{b_0}^{y} U(u, v) du dv,$$
$$\forall (x, y) \in [a_0, +\infty[\times[b_0, +\infty[.$$

Also, to be concise, we put

$$\gamma_{(0,0)} z(\cdot; (a_0, b_0), (0, 0), U) = (\varphi, \psi), \tag{6.5}$$

that is,

$$\varphi(t) = \int_{a_0}^{t} \int_{b_0}^{0} U(u, v) du dv \quad \forall t \in [0, +\infty[, \tag{6.6}$$

$$\psi(t) = \int_{a_0}^{0} \int_{b_0}^{t} U(u, v) du dv \quad \forall t \in [0, +\infty[. \tag{6.7}$$

Then, it is immediately checked that, if the control $U \in \mathcal{U}$ is chosen in such a way that

$$U(x,y) = \frac{\overline{\chi}(0)}{a_0 b_0} \quad \text{a.e. } (x,y) \in [a_0, 0[\times[b_0, 0[\tag{6.8}$$

(this is possible, since $\left|\frac{\overline{\chi}(0)}{a_0 b_0}\right| \leq 1$), we have $\varphi(0) = \overline{\chi}(0) = \psi(0) = \overline{\eta}(0)$.

It follows that, in order to get $\varphi = \overline{\chi}$, it is sufficient that $\varphi' = \overline{\chi}'$, that is

$$\int_{b_0}^{0} U(t,v)dv = \overline{\chi}'(t) \quad \text{a.e. } t \in [0, +\infty[. \tag{6.9}$$

The validity of (6.9) is ensured if the control $U \in \mathcal{U}$ is taken such that

$$U(x,y) = -\frac{\overline{\chi}'(x)}{b_0} \quad \text{a.e. } (x,y) \in [0, +\infty[\times[b_0, 0[. \tag{6.10}$$

Likewise, to obtain also $\psi' = \overline{\eta}'$, it is enough that $U \in \mathcal{U}$ is taken such that

$$U(x,y) = -\frac{\overline{\eta}'(y)}{a_0} \quad \text{a.e. } (x,y) \in [a_0, 0[\times[0, +\infty[. \tag{6.11}$$

In conclusion, if we choose $U \in \mathcal{U}$ in such a way that (6.8), (6.10) and (6.11) hold (this is possible), we get $(\varphi, \psi) = (\overline{\chi}, \overline{\eta})$.

Finally, let us prove that there is no point $(a_0, b_0) \in]-\infty, 0]\times]-\infty, 0]$ for which (6.2) is satisfied. To this aim, we notice that, if (a_0, b_0) is an arbitrary point of $]-\infty, 0]\times]-\infty, 0]$, then (6.4) and (6.6) imply

$$|\varphi(0)| \leq a_0 b_0 \qquad \forall (\varphi, \psi) \in \mathcal{A}((a_0, b_0), (0,0); (0,0), \mathcal{U})$$

and hence $\mathcal{A}((a_0, b_0), (0,0); (0,0), U) \neq \Xi^{(1)}_{p,\text{loc}}$.

As usual for a distributed parameter control process, in which the space of the states is infinite-dimensional, it is important to consider, besides the controllability problem of exact type, also the corresponding approximate problem.

Problem 6.2. (*Approximate controllability with variable initial locus*).

Let $(a,b) \in I \times J$, $(\varphi_0, \psi_0) \in \Xi^{(n)}_{p,\mathrm{loc}}$ and $\mathcal{U} \subseteq L^p_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^m)$ be given.

For each $(\chi, \eta) \in \Xi^{(n)}_{p,\mathrm{loc}}$, each seminorm $\pi_{\mathrm{A,B}}$ on $\Xi^{(n)}_{p,\mathrm{loc}}$ and each $\varepsilon > 0$ find $(a_0, b_0) \in I \times J$, $a_0 \leq a$, $b_0 \leq b$, and $U \in \mathcal{U}$ such that

$$\pi_{\mathrm{A,B}}\left(\gamma_{(a,b)} z(\cdot; (a_0, b_0), (\varphi_0, \psi_0), U) - (\chi, \eta)\right) < \varepsilon.$$

Owing to the nature of the seminorms $\pi_{\mathrm{A,B}}$, it is apparent that Problem 6.2 can be rephrased as follows: for each element (χ, η) of the space $\Xi^{(n)}_{p,\mathrm{loc}}$ and each neighbourhood $\mathcal{O}$ of (χ, η), find a point $(a_0, b_0) \in I \times J$, $a_0 \leq a$, $b_0 \leq b$, and a control $U \in \mathcal{U}$ such that

$$\gamma_{(a,b)} z(\cdot; (a_0, b_0), (\varphi_0, \psi_0), U) \in \mathcal{O}.$$

Hence, we have the following proposition.

Proposition 6.2. Problem 6.2 *is soluble if and only if the set*

$$\bigcup_{\substack{(a_0,b_0) \in I \times J \\ a_0 \leq a, b_0 \leq b}} \mathcal{A}((a_0, b_0), (a, b); (\varphi_0, \psi_0), \mathcal{U})$$

is dense in $\Xi^{(n)}_{p,\mathrm{loc}}$.

Of course, the solubility of Problem 6.1 implies that also Problem 6.2 is soluble. The next example shows that the vice versa is not true.

Example 6.2. Let us consider the same scalar control process (6.3) as in Example 6.1. Again, let $(a,b) = (0,0)$ and $(\varphi_0, \psi_0) = (0,0)$, while the set of the available controls now is

$$\mathcal{U} = \{U \in L^p_{\mathrm{loc}}(\mathbb{R}^2) : |\, U(x,y) \,| \leq 1 \ \text{ a.e. } (x,y) \in \mathbb{R}^2\}. \tag{6.12}$$

We first show that Problem 6.1 is not soluble. To this aim we notice that, for each $(a_0, b_0) \in]-\infty, 0] \times]-\infty, 0]$ and each $(\varphi, \psi) \in \mathcal{A}((a_0, b_0), (0,0); (0,0), \mathcal{U})$, from (6.6) and (6.12) we get

$$\left|\varphi'(t)\right| \le \int_{b_0}^{0} |\, U(t,v)\,|\, dv \le |\, b_0\,|\, t \qquad \text{a.e.} \quad t \in [0,+\infty[.$$

It follows that

$$\bigcup_{a_0 \le 0, b_0 \le 0} \mathcal{A}((a_0,b_0),(0,0);(0,0),\mathcal{U}) \neq \Xi^{(1)}_{p,\mathrm{loc}}.$$

Next, we prove that Problem 6.2 has solution. Let an element $(\overline{\chi},\overline{\eta}) \in \Xi^{(1)}_{p,\mathrm{loc}}$, a seminorm $\pi_{\mathrm{A,B}}$ on $\Xi^{(1)}_{p,\mathrm{loc}}$ and a number $\varepsilon > 0$ be given. We notice that we can assume, without loss of generality,

$$\mathrm{A} =]0,t_1[, \mathrm{B} =]0,t_2[,$$

for some $t_1, t_2 > 0$.

Given an arbitrary $\rho > 0$, we pick two functions $\alpha, \beta \in L^\infty(]0,+\infty[)$ such that

$$\left\|\alpha - \overline{\chi}'\right\|_{L^p(]0,t_1[)} < \rho \quad , \qquad \alpha(t) = 0 \text{ a.e. } t \ge t_1 \quad ,$$

$$\left\|\beta - \overline{\eta}'\right\|_{L^p(]0,t_2[)} < \rho \quad , \qquad \beta(t) = 0 \text{ a.e. } t \ge t_2 \quad ,$$

and denote by (χ_1,η_1) the element of $\Xi^{(1)}_{p,\mathrm{loc}}$ defined as follows:

$$\chi_1(t) = \overline{\chi}(0) + \int_0^t \alpha(s)ds \qquad \forall t \in [0,+\infty[,$$

$$\eta_1(t) = \overline{\chi}(0) + \int_0^t \beta(s)ds \qquad \forall t \in [0,+\infty[.$$

Now, we select $(a_0,b_0) \in]-\infty,0[\times]-\infty,0[$ such that

$$a_0 b_0 \ge |\, \overline{\chi}(0)\,|,$$

$$|\, a_0\,| \ge |\, \beta(t)\,| \quad \text{a.e. } t \in]0,+\infty[, \quad |\, b_0\,| \ge |\, \alpha(t)\,| \quad \text{a.e. } t \in]0,+\infty[$$

and choose a control $U \in \mathcal{U}$ satisfying the following conditions:

$$U(x,y) = \frac{\overline{\chi}(0)}{a_0 b_0} \qquad \text{a.e.} \quad (x,y) \in [a_0, 0[\times[b_0, 0[,$$

$$U(x,y) = -\frac{\alpha(x)}{b_0} \qquad \text{a.e.} \quad (x,y) \in [0, +\infty[\times[b_0, 0[,$$

$$U(x,y) = -\frac{\beta(y)}{a_0} \qquad \text{a.e.} \quad (x,y) \in [a_0, 0[\times[0, +\infty[.$$

Then, keeping the same notations (6.5), (6.6) and (6.7) as in Example 6.1, we have

$$(\varphi, \psi) = (\chi_1, \eta_1).$$

Let $k_{\mathrm{A,B}}$ be any positive constant such that

$$\pi_{\mathrm{A,B}}(\chi, \eta) \leq k_{\mathrm{A,B}} \left(|\chi(0)| + \|\chi'\|_{L^p(\mathrm{A})} + \|\eta'\|_{L^p(\mathrm{B})} \right), \quad \forall (\chi, \eta) \in \Xi^{(1)}_{p,\mathrm{loc}}.$$

Then, we have

$$\pi_{\mathrm{A,B}} \left(\gamma_{(0,0)} z(\cdot; (a_0, b_0), (0,0), U) - (\overline{\chi}, \overline{\eta}) \right) = \pi_{\mathrm{A,B}}((\chi_1, \eta_1) - (\overline{\chi}, \overline{\eta})) \leq$$

$$\leq k_{\mathrm{A,B}} \left(\|\alpha - \overline{\chi}'\|_{L^p(\mathrm{A})} + \|\beta - \overline{\eta}'\|_{L^p(\mathrm{B})} \right) < 2\rho k_{\mathrm{A,B}},$$

thus, choosing $\rho > 0$ in such a way that $2\rho k_{\mathrm{A,B}} < \varepsilon$, we have completed the argument.

Remark 6.1. Similarly to Problem 6.1, an obvious sufficient condition for the solubility of Problem 6.2 is the existence of $(a_0, b_0) \in I \times J$, $a_0 \leq a$, $b_0 \leq b$, such that the attainable set $\mathcal{A}((a_0, b_0), (a, b); (\varphi_0, \psi_0), \mathcal{U})$ be a dense subset of $\Xi^{(n)}_{p,\mathrm{loc}}$. This condition is not necessary. To see this, one can consider, for instance, the preceding Example 6.2 and take into account inequality (6.13).

7. ON THE SOLUBILITY OF PROBLEM 6.2

In this Section we establish a necessary and sufficient condition for the solubility of Problem 6.2 in the case that (φ_0, ψ_0) is the null element of $\Xi^{(n)}_{p,\mathrm{loc}}$ and $\mathcal{U}$ is the unit ball of $L^\infty(L_{I,J}, \mathbb{R}^m)$.

It is an interesting remark that the above mentioned condition is somewhat analogous to the necessary and sufficient condition for the complete approximate controllability with fixed initial locus $l(a_0, b_0)$ and fixed final locus $l(a, b)$, that has been established in G. Pulvirenti - G. Santagati [4], Theorem 6.1. This is similar to what happens in the case of a lumped parameter control process (compare Theorems II.2.1 and VI.5.1 of R. Conti [1]).

Theorem 7.1. *Let* $1 < p < +\infty$. *Also, let* $(a, b) \in I \times J$, $(\varphi_0, \psi_0) = (0, 0)$, *the null element of* $\Xi^{(n)}_{p,\mathrm{loc}}$, *and*

$$\mathcal{U} = \{U \in L^p_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^m) : |\, U(x, y)\,| \leq 1 \quad \text{a.e. } (x, y) \in L_{I,J}\}.$$

Then, in order that Problem 6.2 *be soluble, it is necessary and sufficient that the following condition be satisfied:*

$$\iint_{L_{I,J}} |\, H_{(a,b)}(u, v; (\mu, \nu, \xi))\,|\, du dv = +\infty$$

$$\forall (\mu, \nu, \xi) \in L^{p'}_c([0, +\infty[, \mathbb{R}^n) \times L^{p'}_c([0, +\infty[, \mathbb{R}^n) \times \mathbb{R}^n \setminus \{0\}. \tag{7.1}$$

Proof. We first show the necessity of the condition.

Assume that Problem 6.2 has solution and, arguing by contradiction, that there exists $(\overline{\mu}, \overline{\nu}, \overline{\xi}) \in L^{p'}_c([0, +\infty[, \mathbb{R}^n) \times L^{p'}_c([0, +\infty[, \mathbb{R}^n) \times \mathbb{R}^n \setminus \{0\}$ such that:

$$\iint_{L_{I,J}} |\, H_{(a,b)}(u, v; (\overline{\mu}, \overline{\nu}, \overline{\xi}))\,|\, du dv = k < +\infty.$$

Owing to the continuity of the linear functional $(\overline{\mu}, \overline{\nu}, \overline{\xi})$ and to the nature of the seminorms on $\Xi^{(n)}_{p,\mathrm{loc}}$, there exist a seminorm π on $\Xi^{(n)}_{p,\mathrm{loc}}$ and a positive constant c such that:

$$|<(\chi,\eta),(\overline{\mu},\overline{\nu},\overline{\xi})>| \leq c\pi(\chi,\eta) \qquad \forall(\chi,\eta) \in \Xi^{(n)}_{p,\mathrm{loc}}.$$

Let $r > 0$. Since the linear functional $(\overline{\mu},\overline{\nu},\overline{\xi})$ is not zero, there exists $(\overline{\chi},\overline{\eta}) \in \Xi^{(n)}_{p,\mathrm{loc}}$ such that

$$<(\overline{\chi},\overline{\eta}),(\overline{\mu},\overline{\nu},\overline{\xi})>> r + k. \tag{7.2}$$

Moreover, since Problem 6.2 has solution, in correspondence of the element $(\overline{\chi},\overline{\eta}) \in \Xi^{(n)}_{p,\mathrm{loc}}$, of the seminorm π and of the positive number $\frac{r}{c}$, there are a point $(a_0,b_0) \in I \times J$, $a_0 \leq a$, $b_0 \leq b$, and an element (χ,η) of the attainable set

$$\mathcal{A}((a_0,b_0),(a,b);(0,0),\mathcal{U}) = \Lambda_{(a_0,b_0),(a,b)}\,\mathcal{U}$$

such that

$$\pi((\chi,\eta)-(\overline{\chi},\overline{\eta})) < \frac{r}{c}.$$

It follows, by Theorem 5.3,

$$\begin{aligned}
&<(\overline{\chi},\overline{\eta}),(\overline{\mu},\overline{\nu},\overline{\xi})>=<(\overline{\chi},\overline{\eta})-(\chi,\eta),(\overline{\mu},\overline{\nu},\overline{\xi})>+<(\chi,\eta),(\overline{\mu},\overline{\nu},\overline{\xi})>\leq\\
&\quad\leq c\pi((\overline{\chi},\overline{\eta})-(\chi,\eta))+\sup_{U\in\mathcal{U}}\iint_{L_{I_{a_0},J_{b_0}}} H^*_{(a,b)}(u,v;(\overline{\mu},\overline{\nu},\overline{\xi}))U(u,v)dudv <\\
&\quad< r+\iint_{L_{I_{a_0},J_{b_0}}} |\,H^*_{(a,b)}(u,v;(\overline{\mu},\overline{\nu},\overline{\xi}))\,|\,dudv \leq\\
&\quad\leq r+\iint_{L_{I,J}} |\,H^*_{(a,b)}(u,v;(\overline{\mu},\overline{\nu},\overline{\xi}))\,|\,dudv = r+k,
\end{aligned}$$

but this contradicts (7.2).

Now, we prove that the condition is sufficient.

Let condition (7.1) be satisfied and, again by contradiction, let Problem 6.2 be not soluble, that is, there exists an element $(\overline{\chi},\overline{\eta}) \in \Xi^{(n)}_{p,\mathrm{loc}}$ such that

$$(\overline{\chi},\overline{\eta}) \notin \overline{\bigcup_{\substack{(a_0,b_0)\in I\times J\\ a_0\leq a,b_0\leq b}} \Lambda_{(a_0,b_0),(a,b)}\,\mathcal{U}}. \tag{7.3}$$

Since $\mathcal{U}$ is a convex set and $\Lambda_{(a_0,b_0),(a,b)}$ is a linear map, each set $\Lambda_{(a_0,b_0),(a,b)}\,\mathcal{U}$, $(a_0,b_0)\in I\times J$, $a_0\le a$, $b_0\le b$, is convex.

We also notice that $(a_0',b_0'),(a_0'',b_0'')\in I\times J$, $a_0'\le a_0''\le a$, $b_0'\le b_0''\le b$ imply

$$\Lambda_{(a_0',b_0'),(a,b)}\,\mathcal{U}\supseteq\Lambda_{(a_0'',b_0''),(a,b)}\,\mathcal{U}.$$

To see this, let $(\chi,\eta)\in\Lambda_{(a_0'',b_0''),(a,b)}\,\mathcal{U}$, that is,

$$(\chi,\eta)=\gamma_{(a,b)}z(\cdot;(a_0'',b_0''),(0,0),U'')$$

for some $U''\in\mathcal{U}$. Then, taking

$$U'(x,y)=\begin{cases}0 & \text{a.e. } (x,y)\in L_{I,J}\setminus L_{I_{a_0''},J_{b_0''}},\\ U''(x,y) & \text{a.e. } (x,y)\in L_{I_{a_0''},J_{b_0''}},\end{cases}$$

we have $U'\in\mathcal{U}$ and

$$z\big(x,y;(a_0',b_0'),(0,0),U'\big)=\begin{cases}0 & \forall(x,y)\in L_{I_{a_0'},J_{b_0'}}\setminus L_{I_{a_0''},J_{b_0''}},\\ z(x,y;(a_0'',b_0''), & \forall(\mathrm{x},\mathrm{y})\in L_{I_{a_0''},J_{b_0''}},\end{cases}$$

hence

$$(\chi,\eta)=\gamma_{(a,b)}z(\cdot;(a_0',b_0'),(0,0),U')\in\Lambda_{(a_0',b_0'),(a,b)}\,\mathcal{U}.$$

The above remarks easily imply that the union

$$\bigcup_{\substack{(a_0,b_0)\in I\times J\\ a_0\le a,b_0\le b}}\Lambda_{(a_0,b_0),(a,b)}\,\mathcal{U}$$

is a convex set, and the same is true for its closure.

From (7.3), by the strict separation theorem (N. Dunford - J.T. Schwartz [3], Theorem V.2.10), it follows the existence of some $(\overline{\mu},\overline{\nu},\overline{\xi}) \in L_c^{p'}([0,+\infty[,\mathbb{R}^n) \times L_c^{p'}([0,+\infty[,\mathbb{R}^n) \times \mathbb{R}^n \setminus \{0\}$ such that:

$$< (\chi,\eta),(\overline{\mu},\overline{\nu},\overline{\xi}) > \quad < \quad < (\overline{\chi},\overline{\eta}),(\overline{\mu},\overline{\nu},\overline{\xi}) >$$

$$\forall (\chi,\eta) \in \overline{\bigcup_{\substack{(a_0,b_0)\in I\times J \\ a_0\le a, b_0\le b}} \Lambda_{(a_0,b_0),(a,b)}\, \mathcal{U}}. \tag{7.4}$$

On the other hand, we shall verify that for each $(a_0,b_0) \in I \times J$, $a_0 \le a$, $b_0 \le b$, there exists $(\chi_0,\eta_0) \in \Lambda_{(a_0,b_0),(a,b)}\, \mathcal{U}$ such that

$$< (\chi_0,\eta_0),(\overline{\mu},\overline{\nu},\overline{\xi}) > \quad = \iint_{L_{I_{a_0},J_{b_0}}} | H_{(a,b)}(u,v;(\overline{\mu},\overline{\nu},\overline{\xi})) | \, du dv \tag{7.5}$$

and hence, by (7.4),

$$\iint_{L_{I_{a_0},J_{b_0}}} | H_{(a,b)}(u,v;(\overline{\mu},\overline{\nu},\overline{\xi})) | \, du dv << (\overline{\chi},\overline{\eta}),(\overline{\mu},\overline{\nu},\overline{\xi}) >;$$

since (a_0,b_0) is arbitrary, we get

$$\iint_{L_{I,J}} | H_{(a,b)}(u,v;(\overline{\mu},\overline{\nu},\overline{\xi})) | \, du dv \le < (\overline{\chi},\overline{\eta}),(\overline{\mu},\overline{\nu},\overline{\xi}) >,$$

but this contradicts condition (7.1).

To complete the proof, we are left to check (7.5). We have

$$\iint_{L_{I_{a_0},J_{b_0}}} | H_{(a,b)}(u,v;(\overline{\mu},\overline{\nu},\overline{\xi})) | \, du dv = \left\| H_{(a,b)}(\cdot;(\overline{\mu},\overline{\nu},\overline{\xi})) \right\|_{L^1(L_{I_{a_0},J_{b_0}},\mathbb{R}^m)} =$$

$$= \sup_{\substack{\tilde{U}\in L^\infty(L_{I_{a_0},J_{b_0}},\mathbb{R}^m) \\ \|\tilde{U}\|_{L^\infty(L_{I_{a_0},J_{b_0}},\mathbb{R}^m)}\le 1}} \iint_{L_{I_{a_0},J_{b_0}}} H^*_{(a,b)}(u,v;(\overline{\mu},\overline{\nu},\overline{\xi}))\tilde{U}(u,v)\, du dv =$$

$$= \sup_{U\in\mathcal{U}} \iint_{L_{I_{a_0},J_{b_0}}} H^*_{(a,b)}(u,v;(\overline{\mu},\overline{\nu},\overline{\xi}))U(u,v)\, du dv =$$

$$= \sup_{U \in \mathcal{U}} \iint_{L_{I,J}} \mathbf{1}_{L_{I_{a_0},J_{b_0}}}(u,v) H^*_{(a,b)}(u,v;(\overline{\mu},\overline{\nu},\overline{\xi}))U(u,v)\,dudv$$

and hence, by Theorem 5.2,

$$\iint_{L_{I_{a_0},J_{b_0}}} |\, H_{(a,b)}(u,v;(\overline{\mu},\overline{\nu},\overline{\xi}))\,|\,dudv =$$
$$= \sup_{U \in \mathcal{U}} < U, \Lambda'_{(a_0,b_0),(a,b)}(\overline{\mu},\overline{\nu},\overline{\xi}) > = \sup_{U \in \mathcal{U}} < \Lambda_{(a_0,b_0),(a,b)} U, (\overline{\mu},\overline{\nu},\overline{\xi}) > .$$

Moreover, since $\mathcal{U} \subseteq L^p_{\mathrm{loc}}(L_{I,J}, \mathbb{R}^m)$ is a bounded closed convex set (hence it is also weakly closed), we have that $\mathcal{U}$ is a weakly compact set (see G. Pulvirenti - G. Santagati - A. Villani [5], Proposition 2.6), hence also $\Lambda_{(a_0,b_0),(a,b)}\mathcal{U}$ is weakly compact. It follows that the linear and [weakly] continuous linear functional $(\overline{\mu},\overline{\nu},\overline{\xi})$ attains its maximum value in the set $\Lambda_{(a_0,b_0),(a,b)}\mathcal{U}$, that is, there exists $(\chi_0,\eta_0) \in \Lambda_{(a_0,b_0),(a,b)}\mathcal{U}$ such that

$$\sup_{U \in \mathcal{U}} < \Lambda_{(a_0,b_0),(a,b)} U, (\overline{\mu},\overline{\nu},\overline{\xi}) > = < (\chi_0,\eta_0), (\overline{\mu},\overline{\nu},\overline{\xi}) >,$$

so that (7.5) holds.

The above theorem can be applied, in particular, to the scalar control processes of the type

$$z_{xy} = F(x,y)U(x,y) \quad \text{a.e. } (x,y) \in \mathbb{R}^2,$$

that have been already considered in the previous papers (G. Pulvirenti - G. Santagati [4]; G. Pulvirenti - G. Santagati - A. Villani [5], [6]) and in Examples 6.1 and 6.2.

Corollary 7.1. *Let* $I = J = \mathbb{R}$; $n = m = 1$; $A = B = C = 0$. *Also, suppose that* $1 < p < +\infty$.

Given $(a,b) \in \mathbb{R}^2$, $(\varphi_0,\psi_0) = (0,0)$, *the null element of* $\Xi^{(1)}_{p,\mathrm{loc}}$, *and* (likewise in (6.12))

$$\mathcal{U} = \{U \in L^p_{\mathrm{loc}}(\mathbb{R}^2) : |\, U(x,y)\,| \leq 1 \quad \text{a.e. } (x,y) \in \mathbb{R}^2\},$$

we have that the following conditions (7.6) - (7.8) *are necessary and sufficient in order that* Problem 6.2 *be soluble:*

$$\int_{-\infty}^{a}\int_{-\infty}^{b} | F(u,v) | dudv = +\infty, \tag{7.6}$$

$$\int_{-\infty}^{b} | F(u,v) | dv = +\infty \quad \text{a.e.} \quad u \geq a, \tag{7.7}$$

$$\int_{-\infty}^{a} | F(u,v) | du = +\infty \quad \text{a.e.} \quad v \geq b. \tag{7.8}$$

Proof. We have $V(u,v;x,y) = 1$ in $\mathbb{R}^4$ and hence

$$H_{(a,b)}(u,v;(\mu,\nu,\xi)) = \begin{cases} \xi F(u,v) & \text{a.e. } (u,v) \in]-\infty,a[\times]-\infty,b[, \\ \mu(u-a)F(u,v) & \text{a.e. } (u,v) \in [a,+\infty[\times]-\infty,b[, \\ \nu(v-b)F(u,v) & \text{a.e. } (u,v) \in]-\infty,a[\times]b,+\infty[, \\ 0 & \text{a.e. } (u,v) \in [a,+\infty[\times[b,\infty[\end{cases}$$

for each $(\mu,\nu,\xi) \in L_c^{p'}([0,+\infty[) \times L_c^{p'}([0,+\infty[) \times \mathbb{R}$.

Assume that Problem 6.2 is soluble. Then, from (7.1), taking $(\mu,\nu,\xi) = (0,0,1)$, we immediately obtain that condition (7.6) is satisfied. Again from (7.1), taking $(\mu,\nu,\xi) = (\mathbf{1}_T,0,0)$, where T is an arbitrary bounded measurable subset of $[0,+\infty[$, with $m(T) > 0$, we get:

$$\int_T \int_{-\infty}^{b} | F(a+t,v) | dtdv = +\infty,$$

from which, owing to the arbitrariness of T, by means of an easy argument by contradiction it follows that also condition (7.7) is satisfied. The proof of (7.8) is quite analogous.

Vice versa, suppose now that conditions (7.6) - (7.8) hold and let (μ,ν,ξ) be an arbitrary non zero element of

$$L_c^{p'}([0,+\infty[) \times L_c^{p'}([0,+\infty[) \times \mathbb{R}.$$

If $\xi \neq 0$, from (7.6) we deduce:

$$\iint_{\mathbb{R}^2} | H_{(a,b)}(u,v;(\mu,\nu,\xi)) | dudv \geq \int_{-\infty}^{a} \int_{-\infty}^{b} | \xi F(u,v) | dudv = +\infty.$$

If $\mu \neq 0$, that is, $\mu(t) \neq 0$ in a set $T \subseteq [0,+\infty[$ of positive measure, from (7.7) we get:

$$\iint_{\mathbb{R}^2} | H_{(a,b)}(u,v;(\mu,\nu,\xi)) | dudv \geq \int_{a}^{+\infty} \int_{-\infty}^{b} | \mu(u-a)F(u,v) | dudv \geq$$

$$\geq \int_{a+T} [| \mu(u-a) | \int_{-\infty}^{b} | F(u,v) | dv] du = +\infty.$$

Similarly, if $\nu \neq 0$, condition (7.8) implies that

$$\iint_{\mathbb{R}2} | H_{(a,b)}(u,v;(\mu,\nu,\xi)) | dudv = +\infty.$$

In conclusion, we have that condition (7.1) is satisfied in every case, that is, Problem 6.2 is soluble.

REFERENCES

[1] R. Conti, Problemi di controllo e di controllo ottimale, UTET, Torino, 1974.

[2] R. Di Vincenzo - A. Villani, Sopra un problema ai limiti per una equazione lineare del terzo ordine di tipo iperbolico. Esistenza, unicità e rappresentazione della soluzione, Matematiche (Catania), 32 (1977), pp. 211-238.

[3] N. Dunford - J.T. Schwartz, Linear operators, Part I, Interscience Publishers, Inc., New York, 1958.

[4] G. Pulvirenti - G. Santagati, Processi di controllo con parametri distribuiti in insiemi non limitati. Controllabilità completa approssimata, Ann. Mat. Pura Appl., (4) 33 (1983), pp. 35-50.

[5] G. Pulvirenti - G. Santagati - A. Villani, Processi di controllo con parametri distribuiti in insiemi non limitati. Insieme raggiungibile, Boll. Un. Mat. Ital. B, (7) 4 (1990) , pp. 345-379.

[6] G. Pulvirenti - G. Santagati - A. Villani, Processi di controllo con parametri distribuiti in insiemi non limitati. Problemi di controllo, Matematiche (Catania), 51 (1996), pp. 413-443.

[7] G. Pulvirenti - G. Santagati - A. Villani, Processi di controllo con parametri distribuiti in insiemi non limitati. Controllabilità con luogo iniziale variabile, to appear.

[8] G. Sturiale, Un problema di Darboux in un insieme non limitato. Esistenza, unicità e dipendenza continua della soluzione, Matematiche (Catania), 53 (1998), pp. 359-373.

[9] M.B. Suryanarayana, A Sobolev space and a Darboux Problem, Pacific J. Math., 69 (1977), pp. 535-550.

[10] A.Villani, Un problema al contorno per un sistema lineare iperbolico su un insieme non limitato, Matematiche (Catania), 36 (1981), pp. 215-234.

[11] A.Villani, Processi di controllo con parametri distribuiti in insiemi non limitati. Controllabilità completa esatta, Ann. Mat. Pura Appl., (4) 33 (1983), pp. 19-33.

[12] A.Villani, On the metrizability of $L^p_{\text{loc}}(\Omega,\mu)$, Matematiche (Catania), 38 (1983), pp. 237-244.

WELL POSEDNESS AND OPTIMIZATION PROBLEMS

L. Pusillo
Dept. of Mathematics, University of Genoa, Genoa, Italy

Abstract: This contribution is in the field of Game Theory and Nash equilibria. The property of Tihkonov well posedness is analyzed in relation to other well posedeness properties which are ordinal, a very important property for games because it emphasizes the fact that players' decisions are expressed by preferences and not by a special choice of utility function.
Relations between *Twp* of an exact potential game and *Twp* of potential function as maximum problem are considered too.

Key words: well-posedness, approximate equilibria, non cooperative games.

1. INTRODUCTION

The development of Game Theory in the last decade has had a great interest in the economic theory, where mathematical tools are used very much.

At the beginning of '50s, John Nash, a young American mathematician, gave two important contributions to Game Theory: he developed the notion of equlibrium for non-cooperative games, so we call it Nash equilibrium (*NE* for short) and he studied a method to analyze bargaining games [13].

Today many economists have given an interesting contribution to Game Theory but this new theory is frequently taught not only in many economic courses, but also in operational research, engineering, mathematical economics and mathematical analysis courses.

So this contribution is in the field of mathematical analysis but with an eye to economic application.

The problem of well posedness was born for minimum problems. A minimum problem is said Tihkonov well posed if :
– there is a unique minimum point
– every minimizing sequence converges to the minimum point.

For more details see [2] and [5], [18]. This notion was generalized to other contexts: saddle points [1], Nash equilibria [5], [7], [15],[16], variational inequalities [4] .

In all the cases the idea is an extension of the idea of minimizing sequences seen as approximate solutions. Given a game $G=(X,Y,f,g)$ where X and Y are topological spaces, we shall say that G is Tikhonov well-posed ((Twp) for short) if there is a unique $NE(\overline{x},\overline{y})$ and every $aNE(x_n,y_n)$ converges to $(\overline{x},\overline{y})$, where (x_n,y_n) is a NE if

$$\sup_{x\in X} f(x,y_n)-f(x_n,y_n)\to 0, \sup_{y\in Y} g(x_n,y)-g(x_n,y_n)\to 0$$

For problems arising from economic theory it was investigated if Twp is an ordinal property [17] that is if the well posedness does not depend on the payoff of player but only on the total preorder represented by the function $f:X\to\mathbb{R}$ which induces a preorder $\preceq_f$ on X defined as $x\preceq_f y \Longleftrightarrow f(x)\le f(y)$. An interesting case is given when the preorder cannot be represented by any real valued function (as lexicographic order) and this problem was studied in [3].

Starting from the definition of (ε,k) equilibrium as the point where every player either guarantees at least k or he/she does not loose more then ε it has been selected a class of Twp games with the property of ordinality [10].

This $(\varepsilon.k)$ well posedness is studied in relations with the more known Twp. This class of (ε,k) well posedness games has the property of ordinality if the payoff functions are bounded from below. This study is considered in sections 3 and 4. In section 5 we speak about a new criterion of well posedness: Owp . This is an ordinal properties, it is better then $T^{\nu}wp$ of [7] because it seems to be the smallest ordinal extension of Twp , it does not pretend to taste the Nash equilibria and it has an interesting characterization of ordinal sequences.

In section 6 we speak about an interesting class of games introduced in [14]: exact potential games. We investigate about Twp of these games and relations with Twp of the potential function P as maximum problem; being a potential game the sum of a coordination game and a dummy one, we

prove that a foundamental role is given by dummy game... so the dummy game is not so dummy.

There are many open problems: is a metric characterization of *Owp* possible? The property of *Twp* as maximum problem of potential function has relations with $(\varepsilon,k)wp$ and with *Owp* ? Perhaps for ordinal potential function ([14]) there is a strict relation with *Owp* .

Works are in progress about these issues.

2. DEFINITIONS AND PRELIMINARIES

Let X,Y be Hausdorff topological spaces. By $G=(X,Y,f,g)$ we denote a game with two players, where X,Y are nonempty sets denoting the players' strategy spaces,
$f,g: X\times Y\to\mathbb{R}$ are real valued functions representing the utility functions of the players.

The most accreditate solution for non cooperative games is the Nash equilibrium (*NE* for short).

Definition 2.1. *Given a game* $G=(X,Y,f,g)$ *a Nash equilibrium (NE)for* G *is a pair* $(\overline{x},\overline{y})\in X\times Y$ *s.t.* $f(\overline{x},\overline{y})\geq f(x,\overline{y})\forall x\in X$, $g(\overline{x},\overline{y})\geq g(\overline{x},y)\forall y\in Y$.

In other words a *NE* is a couple of strategies such that each player's strategy is an optimal response to the other players' ones.

We now remind an alternative definition of equilibrium: (ε,k) equilibrium. We start from the definition of (ε,k) -equilibria introduced in [6] and define (ε,k) sequences as approximate equilibria that definitively guarantee to every player at least k or that he/she does not lose more than ε . We remind that

Definition 2.3. *Given* $\varepsilon>0, x\in X$ *is an* $\varepsilon-$*best reply to* y *if*

$$f(x,y)\geq\sup_{t\in X}f(t,y)-\varepsilon$$

Given $k\in\mathbb{R},\mathrm{x}\in\mathrm{X}$ *is a* $k-$*guaranteeing reply to* $y\in Y$ *if*

$$f(x,y)\geq k$$

If $x\in X$ *is either an* $\varepsilon-$*best reply or a* $k-$*guaranteeing reply (or both) to* y *then* x *is called*

(ε,k) *best reply (to* y*)*

Furthermore, we say that $(\overline{x},\overline{y}) \in X \times Y$ *is an* (ε,k) *equilibrium if* $\overline{x}$ *is an* (ε,k) *best reply to* $\overline{y}$ *and conversely.*

A minimum problem is said well posed if there is a unique minimum point and every minimizing sequence is converging to minimum point. So in [1],[5] were introduced the asymptotic Nash sequences to generalize the criterion of well posedness to Nah equilibria:

Definition 2.4. *Given a game* $G=(X,Y,f,g)$ *we shall say that a sequence* $(x_n,y_n) \in X \times Y$ *is an asymptotically Nash equilibrium (aNE for short) if* $\sup_{x\in X} f(x,y_n) - f(x_n,y_n) \to 0$, $\sup_{y\in Y} g(x_n,y) - g(x_n,y_n) \to 0$.

Notice that this implies that $\exists \overline{n} \in \mathbb{N}$ s.t.

$$\sup_{x\in X} f(x,y_n) < +\infty \text{ and } \sup_{y\in Y} g(x_n,y) < +\infty \forall n > \overline{n}$$

So a sequence $(x_n,y_n) \in X \times Y$ is an asymptotically Nash equilibrium if $\forall \varepsilon > 0, (x_n,y_n) \in \Omega_\varepsilon$ for n sufficiently large.

Definition 2.5 *We shall say that a property* (P) *is ordinal if:* (P) *is true for* G *implies that* (P) *is true for* $\tilde{G}$ *too. Where* $G=(X,Y,f,g)$ *and* $\tilde{G}=(X,Y,\tilde{f},\tilde{g})$ *with* $\tilde{f}=\phi f$, $\tilde{g}=\psi g$ *for all* ϕ,ψ *continuous and strictly increasing functions s.t.*

$\phi : I \to \mathbb{R}$, $\mathrm{I} \supset f(\mathrm{X}\times\mathrm{Y})$, I interval

$\psi : J \to \mathbb{R}$, $\mathrm{J} \supset g(\mathrm{X}\times\mathrm{Y})$, J interval

G *and* $\tilde{G}$ *are called ordinally equivalent.*

Definition 2.6 (x_n,y_n) *is called ordinal asymptotic Nash equilibrium if it is aNE for* G *then it is the same for* $\tilde{G}$.

Definition 2.7. *Given* $\varepsilon > 0$ *and* $k \in \mathbb{R}$ *we denote by* Ω_ε *the set of* ε*-equilibria (*ε *– NE for short).*

$$\Omega_\varepsilon = \{(\overline{x},\overline{y}) \in X \times Y : \sup_{x\in X} f(x,\overline{y}) - f(\overline{x},\overline{y}) \le \varepsilon,$$

$\sup_{y\in Y} g(\overline{x},y)-g(\overline{x},\overline{y})\le\varepsilon\}$

We denote by Ω^k *the set of* k *-equilibria*

$$\Omega^k=\{(\overline{x},\overline{y})\in X\times Y: f(\overline{x},\overline{y})\ge k, g(\overline{x},\overline{y})\ge k\}.$$

By Ω_ε^k *we denote the set of* (ε,k) *equilibria, that is:*

$$\Omega_\varepsilon^k=\{(\overline{x},\overline{y})\in X\times Y:(\overline{x},\overline{y})\textit{verifies}\,[a)\,\textit{or}\,b)]\ \textit{and}\,[c)\,\textit{or}\,d)]\}\ \textit{where}$$

a) $\sup_{x\in X} f(x,\overline{y})-f(\overline{x},\overline{y})\le\varepsilon$
b) $f(\overline{x},\overline{y})\ge k$

and

c) $\sup_{y\in Y} g(\overline{x},y)-g(\overline{x},\overline{y})\le\varepsilon$
d) $g(\overline{x},\overline{y})\ge k.$

Remark 2.8 We note that Ω_0 is the set of Nash equilibria and a sequence (x_n,y_n) is an ε -sequence or asymptotically Nash equilibrium, if $\forall\varepsilon>0$, $(x_n,y_n)\in\Omega_\varepsilon$ definitively,
(x_n,y_n) is a k -sequence if $\forall k\in\mathbb{R}(x_n,y_n)\in\Omega^k$ definitively,
(x_n,y_n) is an (ε,k) -sequence if $\forall\varepsilon>0,\forall k>0,(x_n,y_n)\in\Omega_\varepsilon^k$ definitively.

Definition 2.9 *Given a game* $G=(X,Y,f,g)$, *we shall say that* G *is*

(i) Tikhonov well-posed (Twp) if there is a unique NE $(\overline{x},\overline{y})$ *and every a NE* (x_n,y_n) *converges to* $(\overline{x},\overline{y})$ (see [2], [1], [16]).

(ii) Tikhonov well-posed in value (T^vwp) *if there is at least one NE and every* a^vNE (x_n,y_n) *converges to a NE*

where a sequence (x_n,y_n) *is called* a^vNE *if it is aNE and it converges in value to a Nash equilibrium that is* $f(x_n,y_n)\to f(\tilde{x},\tilde{y})$ *and analogously for* g , *with* $(\tilde{x},\tilde{y})$ *Nash equilibrium* ([8], [9]).

3. (ε,k) WELL-POSEDNESS

We have proven in previous papers [7], [17] that Tihkonov well posedness is not an ordinal property, that is if a game $G(X,Y,f,g)$ is *Twp* then it is not so for the game $\tilde{G}(X,Y,\tilde{f},\tilde{g})$, where $\tilde{f}$ and $\tilde{g}$ are the

composite of f and g with increasing functions as in definition 2.5 . This is a bad thing for games.

For example let us consider $G(X,Y,f,g)$ with $f(x,y)=g(x,y)=xy$ and $\tilde{G}(X,Y,\tilde{f},\tilde{g})$ with $\tilde{f}=\tilde{g}=\arctan xy$. G is *Twp* but $\tilde{G}$ is not in fact the sequence (n,n) is an *aNE* for $\tilde{G}$ but not converging to the unique $NE=\{(0,0)\}$.

We have so located a class of *Twp* games which has this important property introducing the following definition:

Definition 3.1. *We say that a game G is (ε,k) well-posed ($(\varepsilon,k)wp$ for short) if:*
– there exist at least one Nash equilibrium (NE);
– every (ε,k) sequence converges to a NE.

This definition guarantees the uniqueness of NE, otherwise the sequence which alternates two NE is (ε,k) sequence non converging.

We remark that the definitions given of ε -sequence, k sequence, (ε,k) -sequence and (ε,k) well-posedness (definitions 2.3, 3.1) clearly do not depend on ε and k, but we choose this terminology since it is more expressive.

Proposition 3.2. *If $G=(X,Y,f,g)$ is (ε,k) well posed then G is Twp. Moreover if the payoff functions f,g are bounded from above, the converse is true as well.*

Proof. If G is (ε,k) well posed then it is *Twp*. The converse, in general, does not hold: it is sufficient to consider example 3.3, where the game is *Twp* but not $(\varepsilon,k)wp$.

On the other hand, if f,g are bounded from above, then $Twp\equiv(\varepsilon,k)wp$; in fact if (x_n,y_n) is a (ε,k) sequence and we consider $k>\max\{\sup f,\sup g\}$, then only conditions a) and c) of definition 2.7 can be satisfied, thus (x_n,y_n) is a $\varepsilon-$sequence. □

Let us show by two examples that $Twp\neq(\varepsilon,k)wp\neq\varnothing$

Examples 3.3. $G_1=(\mathbb{R}^2,\mathbb{R}^2,\mathrm{f}_1,\mathrm{g}_1)$, $f_1(x,y)=g_1(x,y)=xy$ is *Twp* but it is not $(\varepsilon,k)wp$.

$$G_2=(\mathbb{R}^2,\mathbb{R}^2,\mathrm{f}_2,\mathrm{g}_2),\quad f_2(x,y)=xy,\ g_2(x,y)=-xy \text{ is } (\varepsilon,k)wp.$$

There is a nice metric property of $(\varepsilon,k)wp$ as shown in the following

Theorem 3.4. *Let X,Y be metric spaces. G is (ε,k) well posed if and only if there is a Nash equilibrium and*

$$\lim_{\varepsilon\to 0,k\to+\infty} diam\Omega_\varepsilon^k = 0$$

(for details of proof see [10]).

Example 3.5.

Let $G=(X,Y,f,g)$ be defined as follows: $X=Y=[0,1]$,

$$f(x,y)=g(x,y)=\begin{cases}1/(x^2+y^2) & \text{if } x\neq 0, y\neq 0\\ 0 & \text{if } x=0 \text{ or } y=0)\end{cases}$$

then G is (ε,k) well posed.

There is a unique $NE:(0,0)$; the (ε,k) sequences are those converging to $(0,0)$.

$$\Omega_\varepsilon(f)=\left\{(x,y)\in\mathbb{R}^2 \text{s.t.x}\le\sqrt{\varepsilon y^4/(1-\varepsilon y^2)}\right\}.$$

If $\varepsilon<1/2$ then $\Omega_\varepsilon=\{(0,0)\}$ so G is Twp. $\Omega^k=\{(x,y)\in\mathbb{R}^2 \text{s.t.x}^2+y^2\le 1/k\}=\Omega_k(f)=\Omega_k(g)$; $\Omega_\varepsilon^k\neq\Omega_\varepsilon$ but $\text{diam}\,\Omega_\varepsilon^k\to 0$ so G is (ε,k) well posed.

4. ORDINALITY PROPERTY OF (ε,k) WELL-POSEDNESS

Our intention is to study ordinality of well-posedness for games and, as already said, this property is very important: in fact the problems' data are the preferences of the players, not a special choice of the utility function.

If f,g are functions bounded from below , $(\varepsilon,k)wp$ is an ordinal property as follows from the theorem 4.4:

Lemma 4.1. *Let (x_n,y_n) be a sequence in $X\times Y$. (x_n,y_n) is an (ε,k) sequence if and only if the following are satisfied definitively:*

a) $\sup_{x\in X} f(x,y_n)-f(x_n,y_n)\le\varepsilon$ *or b')* $1-\frac{f(x_n,y_n)}{1+|f(x_n,y_n)|}\le\varepsilon$

and

c) $\sup_{y\in Y} g(x_n,y)-g(x_n,y_n)\le\varepsilon$ *or d')* $1-\frac{g(x_n,y_n)}{1+|g(x_n,y_n)|}\le\varepsilon$

Proof. $\Rightarrow$ Let us choose $0<\varepsilon<1$, $k=1/\varepsilon-1$. Then $f(x_n,y_n)\geq 1/\varepsilon-1$ implies $1-\frac{f(x_n,y_n)}{1+|f(x_n,y_n)|}\leq\varepsilon$.

$\Leftarrow$ Given $\varepsilon,k>0$, let us choose $\varepsilon'<1$ such that $1/\varepsilon'-1>k$. Then $f(x_n,y_n)\geq 1/\varepsilon'-1>k$ definitively. □

Proposition 4.2. *Let $\alpha=(x_n,y_n)$ be a sequence in $X\times Y$. Then it is an (ε,k) sequence for the game G, if and only if there exist four disjoint indices sets: S_1,S_2,S_3,S_4 such that their union is $\mathbb{N}$ and such that*

- *S_1 is finite or determines a subsequence α_1 of α (made by all the terms with indices in S_1), α_1 is an ε -sequence*
- *S_2 is finite or determines a subsequence α_2 of α (made by all the terms with indices in S_2), α_2 is a k sequence.*
- *S_3 is finite or determines a subsequence α_3 of α (made by all the terms with indices in S_3), α_3 is an ε sequence for f and a k sequence for g*
- *S_4 is finite or determines a subsequence α_4 of α (made by all the terms with indices in S_4), α_4 is a k sequence for f and an ε sequence for g.*

Proof. $\Leftarrow$: trivial.

$\Rightarrow$: let (x_n,y_n) be an (ε,k) sequence for G. Then for the previous lemma, a) or b) and c) or d) hold. This is equivalent to

$$\min\left(\sup_{x\in X} f(x,y_n)-f(x_n,y_n),1-\frac{f(x_n,y_n)}{1+|f(x_n,y_n)|}\right)\leq\varepsilon$$

and analogously for g. This is true if and only if $\max\{\phi_1,\phi_2\}\leq\varepsilon$ definitively, where

$$\phi_1(x_n,y_n)=\min\{\sup_{x\in X} f(x,y_n)-f(x_n,y_n),1-f(x_n,y_n)/[1+|f(x_n,y_n)|]\}$$

and

$$\phi_2(x_n,y_n)=\min\{\sup_{y\in Y} g(x_n,y)-g(x_n,y_n),1-g(x_n,y_n)/[1+|g(x_n,y_n)|]\}$$

if and only if

$$\lim_n \max\{\phi_1(x_n,y_n),\phi_2(x_n,y_n)\}=0 \text{ if and only if}$$

$\lim_n \phi_1(x_n, y_n) = 0$ and $\lim_n \phi_2(x_n, y_n) = 0$.

Let us define $S_i, i = 1,2,3,4$ in the following way:

$$S_1 = \{n \in \mathbb{N} : \sup_{x\in X} f(x, y_n) - f(x_n, y_n) \le 1 - \tfrac{f(x_n,y_n)}{1+|f(x_n,y_n)|},$$
$$\sup_{y\in Y} g(x_n, y) - g(x_n, y_n) \le 1 - \tfrac{g(x_n,y_n)}{1+|g(x_n,y_n)|}\}$$

then

$$\phi_1(x_n, y_n) = \sup_{x\in X} f(x, y_n) - f(x_n, y_n)$$

$$\phi_2(x_n, y_n) = \sup_{y\in Y} g(x_n, y) - g(x_n, y_n)$$

S_1 is finite or S_1 determines a subsequence α_1 of α and it is an ε sequence.

Analogously:

$$S_2 = \{n \in \mathbb{N} : \sup_{x\in X} f(x, y_n) - f(x_n, y_n) > 1 - \tfrac{f(x_n,y_n)}{1+|f(x_n,y_n)|},$$
$$\sup_{y\in Y} g(x_n, y) - g(x_n, y_n) > 1 - \tfrac{g(x_n,y_n)}{1+|g(x_n,y_n)|}\}$$

$$S_3 = \{n \in \mathbb{N} : \sup_{x\in X} f(x, y_n) - f(x_n, y_n) > 1 - \tfrac{f(x_n,y_n)}{1+|f(x_n,y_n)|},$$
$$\sup_{y\in Y} g(x_n, y) - g(x_n, y_n) \le 1 - \tfrac{g(x_n,y_n)}{1+|g(x_n,y_n)|}\}$$

then

$$S_4 = \{n \in \mathbb{N} : \sup_{x\in X} f(x, y_n) - f(x_n, y_n) \le 1 - \tfrac{f(x_n,y_n)}{1+|f(x_n,y_n)|},$$
$$\sup_{y\in Y} g(x_n, y) - g(x_n, y_n) > 1 - \tfrac{g(x_n,y_n)}{1+|g(x_n,y_n)|}\}$$

Summarizing, each sets S_i is either finite or determines a subsequence α_i of α (i=1,2,3,4) which turn out an ε sequence, a k sequence, a k sequence for f and ε sequence for g, ε sequence for f and k sequence for g respectively. □

Proposition 4.3. *If (x_n, y_n) is an ε sequence for f then it is an (ε, k) sequence for ϕf. If (x_n, y_n) is a k -sequence for f, then it is also an (ε, k) sequence for ϕf.*

Proof. We must consider the cases: ϕ bounded, and ϕ unbounded with the subcases f bounded and unbounded. □

Finally we have the so much waited result:

Theorem 4.4. *Let* $G=(X,Y,f,g)$, $\tilde{G}=(X,Y,\tilde{f},\tilde{g})$ *be two games s.t.* $G\sim\tilde{G}$ $f,\tilde{f},g,\tilde{g}$ *are bounded from below and* $\tilde{f}=\phi f$, $\tilde{g}=\psi g$. *If* (x_n,y_n) *is an* (ε,k) *sequence for* G *then it is an* (ε,k)*-sequence for* $\cup\tilde{G}$ *as well. So if* G *is* $(\varepsilon,k)wp$ *then* $\tilde{G}$ *is too.*

Example 4.5 The duopoly Cournot model with the hypothesis considered in [9] is an (ε,k) well posed game.

5. VALUE BOUNDED WELL-POSEDNESS

In this section we introduce a new criterion of well-posedness, it is a variation of Tihkonov well-posedness: *Owp* (for more details see [11]) To introduce it, let us define the sequences which must approximate the eqilibrium.

Definition 5.1 *A sequence* $(x_n,y_n)\in X\times Y$ *is value bounded if there are four numbers:* $a,b,c,d\in\mathbb{R}$ *s.t.*

$$a\le f(x_n,y_n)\le b,\ \ c\le g(x_n,y_n)\le d,\forall n\in\mathbb{N}$$

$$a,b\in f(X\times Y),\ \ c,d\in g(X\times Y)$$

So we call (x_n,y_n) *a value bounded asynptotically Nash if it is aNE and it is value bounded*

A sequence (x_n,y_n) *is "definitively NE" if its elements are all Nash equilibria for n large enough.*

The following theorem gives an interesting property of value bounded sequences:

Theorem 5.2 *Let* (x_n,y_n) *be an aNE. Then* (x_n,y_n) *is ordinal if it is definitively NE or its elements, which are not NE, form a value bounded aNE.*

Proof: Since two ordinally equivalent games G and $\tilde{G}$ have the same NE, it is sufficient to prove that a value bounded aNE (whose elements are not NE) is ordinal too.

Let (x_n, y_n) be a value bounded *aNE*. Let $(\overline{x}, \overline{y}), (\tilde{x}, \tilde{y})$ be s.t.

$$f(\overline{x}, \overline{y}) \le f(x_n, y_n) \le f(\tilde{x}, \tilde{y})$$

and further

$$\forall \varepsilon > 0 \ \exists \overline{n}_\varepsilon \ \ s.t. \ f(x, y_n) - f(x_n, y_n) \le \varepsilon \ \ \forall n > \overline{n}_\varepsilon, \ \ \forall x \in X \ \text{(analogously for } g\text{)}. \qquad (5.1.1)$$

Fixing ε_1, we choose ε according to the following cases:

i) if $f(\tilde{x}, \tilde{y}) = \sup f$ then ε is the modulus of uniform continuity of $\phi(.)$ in the interval $[f(\overline{x}, \overline{y}), f(\tilde{x}, \tilde{y})]$

ii) if $f(\tilde{x}, \tilde{y}) < \sup f$, let (x^*, y^*) be s.t. $f(\tilde{x}, \tilde{y}) < f(x^*, y^*)$, so we choose $\varepsilon = \min \{f(x^*, y^*) - f(\tilde{x}, \tilde{y}), \delta^*\}$ where δ^* is the modulus of uniform continuity of $\phi(.)$ in the interval $[f(\overline{x}, \overline{y}), f(x^*, y^*)]$. After the choice of ε, from hypotheses we must distinguish two cases:

(1) $f(x, y_n) \le f(x_n, y_n)$, and so

$$\phi f(x, y_n) \le \phi f(x_n, y_n) \le \phi f(x_n, y_n) + \varepsilon_1 \quad \text{or}$$

(2) $f(x_n, y_n) \le f(x, y_n) \le f(x_n, y_n) + \varepsilon$

if $f(\tilde{x}, \tilde{y}) = \sup f$ then

$$f(\overline{x}, \overline{y}) \le f(x_n, y_n) \le f(x, y_n) \le f(\tilde{x}, \tilde{y})$$

So $f(x_n, y_n)$ and $f(x, y_n)$ are in the interval of uniform continuity and their difference is less than ε, so after applying ϕ, their difference is less than ε_1.

Instead if $f(\tilde{x}, \tilde{y}) < \sup f$ then

$$f(\overline{x}, \overline{y}) \le f(x_n, y_n) \le f(x, y_n) \le f(x_n, y_n) + \varepsilon \le f(\tilde{x}, \tilde{y}) + \varepsilon \le f(x^*, y^*)$$

so we are again in the interval of uniform continuity of ϕ and we can conclude about this sufficient condition. □

Now we can introduce a new criterion of well-posedness:

Definition 5.3 *A game G is ordinally well-posed (Owp for short) if:*

1) there is a NE $(\overline{x}, \overline{y})$

2) every value bounded aNE converges to a NE.

It follows that the NE is unique (as in Definition 3.1).

Proposition 5.4

The followings properties are true:

1) $\mathcal{O}wp \supset Twp$

2) $\mathcal{O}wp = Twp$ *if the payoff functions* f, g *attain maximum and minimum points*

3) $\mathcal{O}wp$ *is ordinal*

Considering the previous well-posedness properties the following inclusions are true:

$$(\varepsilon, k)wp \subset Twp \subset \mathcal{O}wp \subset T^{v}wp$$

By the following example and example 3.2 we see that the previous inclusions are proper.

Example 5.5 Let $G = (X, Y, f, g)$ be a game. $X = Y = [0, +\infty)$

$$f(x,y) = g(x,y) = \begin{cases} \arctan xy \text{ if } x \in \mathbb{Z} \text{ or } y \in \mathbb{Z} \\ xy \text{ otherwise} \end{cases}$$

The unique NE is $(0,0)$ and (x_n, y_n) is a aNE if it is equal to $(0,0)$ definitively, so G is $T^{v}wp$. G is not Owp because (n,n) is value bounded aNE but non convergent.

6. POTENTIAL GAMES AND WELL POSEDNESS

In general it is not trivial to find Nash equilibria for a strategic non cooperative game but the class of potential games introduced by Monderer and Shapley is a special one, because the problem of equilibria is reduced to study a unique function called Potential function.

We shall call $G = (X, Y, f, g)$ a game with exact potential P, if it exists a function $P : X \times Y \to \mathbb{R}$ s.t. the increase of P along x is equal to the increase of f and the increase along y is equal to that of g ([14],[20]) that is:

Definition 6.1 *A game* $G = (X, Y, f, g)$ *is said an exact potential game if it exists a function* P *s.t.*

$$f(x_1, y) - f(x_2, y) = P(x_1, y) - P(x_2, y)$$

$$g(x,y_1)-g(x,y_2)=P(x,y_1)-P(x,y_2)$$

$\forall x,x_1,x_2 \in X e \forall y,y_1,y_2 \in Y$. *The function* P *is called an exact potential function for* G. (see [14],[20])

Example 6.2

	C	D
A	5 7	2 3
B	4 3	4 4

all potential functions are $P=$

	C	D
A	k	k-4
B	k-1	k

$k \in \mathbb{R}$

Remark 6.3 If G is an exact potential game then it has the same Nash equilibria of $G^P=(X,Y,P,P)$.

Each finite game with exact potential function has at least a *NE* and it is the maximum point of potential. If X or Y are infinite this fact is not true, it is sufficient consider

$$X=Y=\mathbb{R}, \mathrm{f}=\mathrm{g}=\mathrm{x}+\mathrm{y}$$

Definition 6.4 *A game* $G=(X,Y,f,g)$ *is a:*

- *coordination game if* $f(x,y)=g(x,y)=P(x,y)$
- *dummy game if there are two functions* $h:Y\to\mathbb{R}$, $k:X\to\mathbb{R}$ *s.t.* $f(x,y)=h(y)$ *and* $g(x,y)=k(x)$. In a dummy game the payoff of one player depends only by the strategies of the other .

Theorem 6.5 *Let* G *be a strategic game.* G *is an exact potential game if and only if* $G=G_c+G_d$ *where* G_c *is a pure coordination game (sometimes we call it* $G^P=(X,Y,P,P)$*) and* G_d *is a dummy game.*

It is very interesting the identification between an exact potential game and a congestion game (a special class of games which consider utilities or machines used by players). These games are important for traffic problems, but this argument is fast from our goal, so we invite the interested reader to Rosenthal's paper [19]and Voorneveld's book [20].

Theorem 6.6 *Let* $G=(X,Y,f,g)$ *be an exact potential game.* P *potential function has a maximum point. If* $G^P=(X,Y,P,P)$ *is* $T^v wp$ *then* P *is Twp . Further if the NE is unique,* P *is Twp as maximum problem if and only if* G^P *is* $T^v wp$

Proof. At first, we remark that P has only one maximum point, in fact, if by absurd, there were two maximum points, these would be two repeated NE then G were not $T^{v}wp$. Let $(\overline{x},\overline{y})$ be a maximum point for P and let us suppose, by contradiction that P is not Twp (as maximum problem) so there is (x_n,y_n) maximizing sequence that is $P(x_n,y_n)\to P(\overline{x},\overline{y})$ but (x_n,y_n) does not converges to $(\overline{x},\overline{y})$.

So we can choose:

$$(a_n,b_n):=\begin{cases}(x_n,y_n) & \text{for } n \text{ even}\\ (\overline{x},\overline{y}) & \text{for } n \text{ odd}\end{cases}$$

So (a_n,b_n) is $a^{v}NE$ but it is not converging. This is absurd.

For the second part of theorem, we note that the $a^{v}NE$ coincide with maximizing sequences of P. □

Generally there is no relation between Twp of a game and Twp of a potential function P, as we learn by the following examples:

1) $f(x,y)=g(x,y)=xy=P(x,y)$, $G(\mathbb{R},\mathbb{R},\mathrm{f},\mathrm{g})$ is Twp but P has no maximum point.

2)

0	1
0	0

P has maximum point and it is Twp as maximum problem.

$G=G^{P}$ is not Twp because there are two NE. But

Theorem 6.7 *Let G be an exact potential game. G is Twp and P has maximum point, then P is Twp as maximum problem.*

Proof. for details see [12].

Example 6.8 Let D be the dummy game:

D :

0 0	1 0
0 2	1 2

P :

0	0
0	0

P is a potential function for D which is $T^{v}wp$ but D^{P} is not $T^{v}wp$ because there are repeated Nash equilibria. So dummy game preserve Twp instead of $T^{v}wp$, so dummy game is not so dummy...

REFERENCES

[1] E. Cavazzuti, J. Morgan J, *Well-Posed Saddle Point Problems* in "Optimization, theory and algorithms" (J.B. Hirriart-Urruty, W. Oettli and J. Stoer eds.), Proc. Conf. Confolant/France 1981,(1983), 61-76.

[2] A. Dontchev and T. Zolezzi , *Well-posed Optimization Problems.* Lecture Notes in Mathematics, 1543, Springer, Berlin,(1993).

[3] V. Fragnelli, F. Patrone and A. Torre, *The Nucleolus is Well Posed* preprint 2003

[4] M.B.Lignola and J. Morgan *Well-posedness for Optimization Problems with Constraints Defined by Variational Inequalities having a Unique Solution* J. Global Optimization 16 (2000),57-67.

[5] R. Lucchetti, F. Patrone: *Hadamard and Tyhonov Well-Posedness of Certain Class of Convex Functions*, Journal of Mathematical Analysis and Applications 88, (1982), 204-215

[6] R. Lucchetti, F. Patrone, S. Tijs, *Determinateness of two-person games.* Bollettino U.M.I. 6: (1986), 907-924.

[7] M. Margiocco, F. Patrone, L. Pusillo Chicco: *A New Approach to Tikhonov Well–Posedness for Nash Equilibria,* Optimization 40, (1997), 385–400

[8] M. Margiocco, F. Patrone, and L. Pusillo Chicco: *Metric Characterizations of Tikhonov Well-Posedness in Value,* Journal of Optimization Theory Application, 100 n.2,(1999), 377–387.

[9] Margiocco M., Patrone F., Pusillo Chicco L.: *On Tikhonov well-posedness of concave games and Cournot oligopoly,* Journal of Optimization Theory Application 112(2002) 361-369

[10] M. Margiocco and L. Pusillo (ε,k) *Equilibria and Well Posedness* preprint (2003)

[11] M. MARGIOCCO and L. PUSILLO: *Value Bounded Approximations for Nash Equilibria* preprint (2003)

[12] M. Margiocco and L. Pusillo: *Potential Games and Well-Posedness: Static and Dynamic Cases* preprint (2004)

[13] R.B. Myerson: *Game Theory: Analysis of Conflict*, Harvard University Press, Cambridge(MA) (1991)

[14] P. Monderer, L. Shapley, *Potential games*, Games and Economic Behaviour 14, (1996), 124-143;

[15] F. Patrone, *Well-Posedness as an Ordinal Property.* Rivista di Matematica pura ed applicata 1, (1987), 95-104.

[16] F. Patrone, *Well-Posedness for Nash Equilibria and Related Topics* in "Recent Developments in Well-Posed Variational Problems" Lucchetti R. and Revalski J. eds. Kluwer, Dordrecht, (1995).

[17] F. Patrone and L. Pusillo Chicco, *Antagonism for two person games:taxonomy and applications to Tihkonov well- posedness.* preprint DIMA (1995)

[18] L. Pusillo Chicco: *Approximate solutions and Tikhonov well-posedness for Nash equilibria*, in "Equilibrium Problems: Nonsmooth Optimization and variational Inequality Models", F. Giannessi, A. Maugeri and Panos M. Pardalos eds., Kluwer Academic Publishers (2001) 231-244

[19] R.W. Rosenthal *A class of games possessing pure strategy Nash equilibria* International Journal of Game Theory 2(1973)65-67

[20] M. Voorneveld, *Potential Games and Interactive Decisions with Multiple Criteria.* Center Dissertation Series. Tilburg University (1999).

SEMISMOOTH NEWTON METHODS FOR SHAPE-PRESERVING INTERPOLATION, OPTION PRICE AND SEMI-INFINITE PROGRAMS

L. Qi
Dept. of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong

Abstract: In this paper, we survey the development of semismooth Newton methods for solving the shape-preserving interpolation problem, the option price problem, and the semi-infinite programming problem.

Key words: Smoothness, Semismoothness, shape-preserving interpolation, option price problem, semi-infinite programming

1. INTRODUCTION

The semismooth Newton method was initiated in 1993 by Qi [31], Qi and Sun [37]. It has found applications in nonlinear complementarity problems [27,7,22,14,45,8,26], variational inequality problems [13,33], civil engineering problems [5] and data mining problems [16]. In [16], a semismooth Newton method was used for solving a 60 million variable support vector machine problem successfully. A survey on the semismooth Newton method can be found in [23]. Some developments after [23] can be found in [17,33,18]. The applications of semismooth Newton methods on nonlinear complementarity problems and variational inequality problems can be found in the recent book of Facchinei and Pang [15], where several hundreds of references on semismooth Newton methods were given.

Recently, semismooth Newton methods have been applied to the shape-preserving interpolation problem, the option price problem, and the semi-infinite programming problem. We survey these appplications in this paper.

2. SEMISMOOTH NEWTON METHODS

Let $G:\Re^n \to \Re^m$ be a locally Lipschitz function. By the Radamacher theorem, G is differentiable almost every where. Denote the set on which G is differentiable as D_G. Clarke [6] defined the **generalized Jacobians** of G at x as

$$\partial G(x) = \text{conv}\{\lim_{x^j \to x, x^j \in D_G} \nabla G(x^j)\},$$

which is a nonempty compact convex set.

Suppose that $F:\Re^n \to \Re^n$ is a locally Lipschitz function. We aim to solve

$$F(x) = 0. \tag{1}$$

A generalized Newton method is naturally available, i.e., given x^k, if it is not a solution of (1), solve

$$F(x^k) + V_k d = 0, \tag{2}$$

where $V_k \in \partial F(x^k)$. Let d^k be a solution of (2). Then we find x^{k+1} by:

$$x^{k+1} = x^k + d^k. \tag{3}$$

The subproblem (2) is a system of linear equations. So we may expect it is efficient. But counterexamples are available to show that the generalized Newton method (2-3) may be divergent if F is merely locally Lipschitz.

The superlinear and quadratic convergence of the generalized Newton method (2-3) can be established under the condition of semismoothness and strong semismoothness.

The concept of semismoothness for vector valued functions [37] is as follows.

Definition 2.1. *Let $G:\mathbf{R}^n \to \Re^m$ be a locally Lipschitz continuous function. Then G is said to be* ***semismooth*** *at $x \in \mathbf{R}^n$ if for any $h \in \mathbf{R}^n$ the limit*

$$\lim_{h' \to h, \tau \downarrow 0} \{Vh' \mid V \in \partial G(x + \tau h')\}$$

exists.

It was proved in [37] that a vector function G is semismooth if and only if all its components are semismooth. Also, if G is semismooth at x, then the directional derivative of G at x along a direction h, denoted $G'(x;h)$, exists for any $h \in \mathbf{R}^n$. A function G is said to be semismooth if it is semismooth at each point of its domain.

A function $G: \mathbf{R}^n \to \mathbf{R}^m$ is said **strongly semismooth** at x [31,34] if it is locally Lipschitz and directionally differentiable at x, and for all $h \to 0$ and $V \in \partial G(x+h)$ one has

$$G(x+h) - G(x) - Vh = O(\|h\|^2). \tag{4}$$

The following local convergence result was established in [37].

Theorem 2.2 *Let x^* be a solution of the equation (1), $F(x) = 0$, and let F be a locally Lipschitz function which is semismooth at x^*. Assume that all $V \in \partial F(x^*)$ are nonsingular matrices. Then the generalized Newton method (2-3) is well defined and converges superlinearly to x^* if the initial point x^0 is sufficiently close to x^*. If furthermore F is strongly semismooth at x^*, then the convergence rate is quadratic.*

Another method which is closely related to the semismooth Newton method is the smoothing Newton method [3,19,4,17,18,36].

3. SHAPE-PRESERVING INTERPOLATION

The constrained approximation problem comes from practical applications in computer aided geometric design where one has not only to approximate data points but also to achieve a desired shape of a curve or a surface. This is also called **shape preserving approximation**.

Examples of a desired shape property include convexity and monotonicity. A special case of shape preserving approximation is **shape preserving interpolation**. That is, to find a function, whose graph has a desired shape, to interpolate given points.

Consider the following **convex best interpolation** problem:

$$\text{minimize } \|f''\|_2 \tag{5}$$

$$\text{subject to} \quad f(t_i) = y_i, \quad i = 1,2,\cdots,N+2,$$
$$f \text{ is convex on } [a,b], \quad f \in W^{2,2}[a,b],$$

where $a = t_1 < t_2 < \ldots < t_{N+2} = b$ and y_i, $i = 1,\ldots,N+2$ are given numbers, $\|\cdot\|_2$ is the Lebesgue $L^2[a,b]$ norm, and $W^{2,2}[a,b]$ denotes the Sobolev space of functions with absolutely continuous first derivatives and second derivatives in $L^2[a,b]$, and equipped with the norm being the sum of the $L^2[a,b]$ norms of the function, its first, and its second derivatives. Employing the normalized B-splines B_i of order two associated with (t_i, y_i) and the corresponding second divided differences d_i, the interpolation conditions can be equivalently written in terms of the second derivative of f,

$$\int_a^b B_i(t) f''(t) dt = d_i, \quad i = 1,2,\cdots,N,$$

and then the problem (5) becomes a problem of projection of the origin in $L^2[a,b]$ on the intersection of finitely many hyperplanes and the cone of nonnegative functions:

$$\begin{aligned} &\text{minimize } \|u\|_2 \\ &\text{subject to} \quad \int_a^b B_i(t) u(t) dt = d_i, \quad i = 1,2,\cdots,N, \\ &\qquad u \geq 0 \text{ a.e. } [a,b], \quad u \in L^2[a,b]. \end{aligned} \tag{6}$$

This problem is a well-known optimization problem; it can be viewed as a version of the moment problem. In order to derive optimality conditions, e.g. from the Lagrange duality theory, we need certain regularity of the constraints. For this purpose, we take the positivity of d_i as a blanket assumption.

According to the Lagrange multiplier rule, u^* is the unique solution of (6) if and only if there exist numbers λ_i^*, $i = 1,2,\cdots,N$ such that u^* is a solution of the problem

$$\begin{aligned} &\text{minimize } \int_a^b \left(\frac{1}{2} | u(t) |^2 - \sum_{i=1}^N \lambda_i^* B_i(t) u(t)\right) dt + \sum_{i=1}^N \lambda_i^* d_i \\ &\text{subject to } u \in L^2[a,b], \quad u \geq 0 \text{ a.e. } [a,b]. \end{aligned} \tag{7}$$

The minimum is attained at a function whose value at t gives the minimum at t of the integrand, hence

$$u^*(t) = \left(\sum_{i=1}^{N} \lambda_i^* B_i(t) \right)_+, \tag{8}$$

where $a_+ := \max\{0, a\}$.

By duality, substituting the solution (8) in (7), we obtain that the value of the Lagrange multiplier vector $\lambda^* = (\lambda_1, \cdots, \lambda_N) \in \Re^N$ is a solution of the dual problem which in our case is an unconstrained finite-dimensional concave program of the form

$$\max_{\lambda \in \Re^N} \left| -\frac{1}{2} \int_a^b \left(\sum_{i=1}^{N} \lambda_i B_i(t) \right)_+^2 dt + \sum_{i=1}^{N} \lambda_i d_i \right|. \tag{9}$$

By the first-order optimality condition and concavity, the latter problem is equivalent to the system of **nonsmooth equations**

$$F(x) = d, \tag{10}$$

where $d = (d_1, \cdots, d_N)$ and the i th component of F is defined by

$$F_i(x) = \int_a^b \left(\sum_{l=1}^{n} x_l B_l(s) \right)_+ B_i(s) ds. \tag{11}$$

Irvine, Marin and Smith [20] proposed in 1986 a Newton-type method for solving the equation (10). By monitoring the decrease of the norm of the residual $F(\lambda) - d$, they observed fast convergence in their numerical experiments and raised the question of theoretically estimating the rate of convergence. They wrote: **"Although we have not established rigorous convergence results for Newton's method we have been very encouraged by numerical experiments...."**.

The conjecture of Irvine, Marin and Smith remained unproved for 15 years. In 2001, Dontchev, Qi and Qi [9] proved this conjecture by viewing the method of Irving, Marin and Smith as a semismooth Newton method. They added line search to the method and established global convergence.

To apply the superlinear (quadratic) convergence theory of the semismooth Newton method to the system of **nonsmooth equations**

$$F(x) = d, \tag{12}$$

where $d=(d_1,\cdots,d_N)$ and the i th component of F is defined by

$$F_i(x)=\int_a^b\left(\sum_{l=1}^{n}x_lB_l(s)\right)_+B_i(s)ds, \tag{13}$$

Dontchev, Qi and Qi [9] proved that an integral functional of a semismooth function is semismooth; in particular, the function F in (12) is semismooth:

Proposition 3.1. *Suppose that $\partial f_t(\lambda)$, viewed as a joint mapping of t and λ, is upper semicontinuous, i.e., for every $\varepsilon>0$, there exists $\delta>0$ such that*

$$\partial f_{t'}(\lambda')\subseteq\partial f_t(\lambda)+\varepsilon U,\qquad \text{for all } \lambda'\in U(\lambda,\delta),t'\in U_1(t,\delta),$$

where

$$U_1(t,\delta)=\{t'\,|\,|t'-t|\le\delta\}\cap[a,b]$$

and

$$U(\lambda,\delta)=\{\lambda'\,|\,\|\lambda'-\lambda\|\le\delta\}.$$

Then φ is semismooth at λ if $f_t(\cdot)$ is semismooth at λ for every $t\in[a,b]$.

Theorem 3.2 *The function F in (12) is semismooth.*

Dontchev, Qi and Qi [9] proved that near the solution the elements of the generalized Jacobian of F are positive definite. Based upon these, they established superlinear convergence of the nonsmooth Newton method applied to F. As a special case, they obtained that the Newton-type method proposed by Irvine, Marin and Smith [20] is locally superlinearly convergent. They further globalized the nonsmooth Newton method by employing the dual problem (9).

However, there are two questions from here:

(1) Is F **piecewise smooth**? If so, the semismoothness theory may be not necessary.

(2) Is F strongly semismooth? If so, we may get quadratic convergence of the Newton-type method of Irvine, Marine and Smith.

These two questions are answered in Dontchev, Qi and Qi [10]. The answer to the first question is "no". The answer to the second question is "yes".

In [10], the following results were established:

Proposition 3. 3. *The function F is not differentiable at λ if and only if λ belongs to the set*

$$\Omega = \left\{\lambda \in \mathbf{R}^N \mid \lambda_1 = 0 \text{ or} \lambda_N = 0 \text{or} \lambda_i = \lambda_{i+1} = 0 \text{ for some } i \in \{1,\ldots,N-1\}\right\}.$$

Proposition 3.4 *F is LC^1 (smooth with a Lipschitz gradient) in the complement of Ω, and C^∞ in the interior of each orthant of $\mathbf{R}^N$.*

It was observed by Pang and Ralph [29] that if G is piecewise smooth then the B-subdifferential of G at x in the sense of Qi [31] contains finitely many elements. Dontchev, Qi and Qi [10] showed that the B-subdifferential of F_1 at the origin contains infinitely many elements. Therefore, by the above mentioned observation of Pang and Ralph, F_1 is not piecewise smooth. This answered the first question. For the second question, recall that the B-spline B_i is given by

$$B_i(t) = \begin{cases} \alpha_i(t - t_i) & \text{for } t \in [t_i, t_{i+1}] \\ \bar{\alpha}_i(t_{i+2} - t) & \text{for } t \in [t_{i+1}, t_{i+2}] \\ 0 & \text{otherwise,} \end{cases}$$

where we denote

$$\alpha_i = 2/((t_{i+2} - t_i)(t_{i+1} - t_i)), \quad \bar{\alpha}_i = 2/((t_{i+2} - t_i)(t_{i+2} - t_{i+1})).$$

In the sequel we study the following functions:

$$\Phi_1(\lambda_1) = \int_{t_1}^{t_2} \left(\lambda_1 B_1(t)\right)_+ B_1(t)dt,$$

$$\Phi_2(\lambda_N) = \int_{t_{N+1}}^{t_{N+2}} \left(\lambda_N B_N(t)\right)_+ B_N(t)dt,$$

$$\Gamma_i(\lambda_{i-1}, \lambda_i) = \int_{t_i}^{t_{i+1}} \left(\lambda_{i-1} B_{i-1}(t) + \lambda_i B_i(t)\right)_+ B_i(t)dt, \quad i = 2, \ldots, N,$$

$$\Psi_i(\lambda_i, \lambda_{i+1}) = \int_{t_{i+1}}^{t_{i+2}} \left(\lambda_i B_i(t) + \lambda_{i+1} B_{i+1}(t)\right)_+ B_i(t), \quad i = 1, \ldots, N-1.$$

Then

$$F_1(\lambda) = \Phi_1(\lambda_1) + \Psi_1(\lambda_1, \lambda_2),$$

$$F_i(\lambda) = \Gamma_i(\lambda_{i-1}, \lambda_i) + \Psi_i(\lambda_i, \lambda_{i+1}), \quad i = 2, \ldots, N-1,$$

$$F_N(\lambda) = \Gamma_N(\lambda_{N-1}, \lambda_N) + \Phi_2(\lambda_N).$$

Dontchev, Qi and Qi [10] proved the following theorem.

Theorem 3.5

(a) Φ_1 *and* Φ_2 *are piecewise linear;*

(b) Γ_i *and* Ψ_i *are* LC^1 *away from the origin, they are not piecewise smooth, but are strongly smooth;*

(c) F *is not piecewise smooth, but strongly semismooth.*

Based upon this result, Dontchev, Qi and Qi [10] established quadratic convergence of the Newton-type method for convex best interpolation.

The one-dimensional shape-preserving spline interpolation problem preserves the function to be convex for subintervals between nodes if the successive second-order divided differences are positive, and to be concave for subintervals between nodes if the successive second-order divided differences are negative. Irvine, Marin and Smith [20] also proposed a Newton-type method for nonsmooth equation reformulation of this shape-preserving interpolation problem. Again, there were no convergence results for the method. Dontchev, Qi, Qi and Yin [11] proved the nonsmooth equation reformulation is strongly semismooth and established quadratic convergence of the generalized Newton method for solving this problem. They also established global convergence for a modification of the algorithm.

4. THE OPTION PRICE PROBLEM

Recently, Wang, Yin and Qi [44] developed an interpolation method to preserve the shape of the **option price** function. The interpolation is optimal in terms of minimizing the distance between the implied risk-neutral density and a prior approximation function in L^2-norm, which is very important when only a few observations are available.

Since the seminal paper of Black-Scholes [2], numerous theoretical and empirical studies have been done on the no-arbitrage pricing theory, see Duffie [12] and the references therein. If the uncertainty of nature can be described by a stochastic process q_t, then the absence of arbitrage opportunities implies that there exists a state-price density (SPD) or risk-neutral density, which is denoted by $p(q_{t_2} \mid F_{t_1})$, where t_2 is any time after time t_1, F_{t_1} is all the information available at time t_1. The price of any financial security can be expressed as the expected net present value of future payoffs, where the expectation is taken with respect to the risk-neutral density. In the call option pricing case, the underlying asset price S_t can be used as the state variable, the risk-free rate is considered as constant. So the price at time t is

$$C(S_t, s, \tau, r_{t,\tau}) = e^{-r_{t,\tau}\tau} \int_0^\infty (S_T - s)_+ p(S_T \mid S_t, \tau, r_{t,\tau}) dS_T, \tag{14}$$

where S_t is the underlying asset price at time t, s is the strike price of the option contract, τ is the time-to-expiration, $T = t + \tau$ is the expiration time, $r_{t,\tau}$ is the risk free rate from time t to $T = t + \tau$. No matter what kind of process of the underlying asset price S_t is, and whether the market is complete or not, the equation above always holds.

By taking the first derivative with respect to s for equation (14), we have:

$$\frac{\partial C(S_t, s, \tau, r_{t,\tau})}{\partial s} = -e^{-r_{t,\tau}\tau} \int_s^\infty p(S_T \mid S_t, \tau, r_{t,\tau}) dS_T. \tag{15}$$

It is obvious that the right-hand side of above equation is negative. Thus the call option price must be a decreasing function of strike price x.

For any strike price values s_1 and s_2 with $s_1 \le s_2$, we have:

$$\frac{\partial C(S_t, s_2, \tau, r_{t,\tau})}{\partial s} - \frac{\partial C(S_t, s_1, \tau, r_{t,\tau})}{\partial s} = e^{-r_{t,\tau}\tau} \int_{s_1}^{s_2} p(S_T \mid S_t, \tau, r_{t,\tau}) dS_T.$$

The density function p is non-negative. This shows that $\frac{\partial C}{\partial s}$ is nondecreasing. Hence, the option price function is convex with respect to the strike price s.

Without loss of the generality, we assume that $0 < a = s_0 < s_1 < \cdots < s_{n+2} = b < +\infty$ and consider the following constrained interpolation problem:

$$\begin{array}{ll} \min & \|f''(s) - h(s)\|_2 \\ \text{s.t.} & f(s_i) = y_i, \ i = 1,2,\cdots,n+2, \\ & f''(s) \geq 0 a.e.[a,b], f \in W_2^2[a,b], \end{array} \tag{16}$$

where

$$h(s) = e^{-r_{t,\tau}\tau} \frac{1}{x\sigma\sqrt{2\pi\tau}} exp\{-\frac{(\log s - \log S_t - r_{t,\tau}\tau + \sigma^2\tau/2)^2}{2\sigma^2\tau}\}. \tag{17}$$

Similar to Section 3, by using the duality theory and Lagrange multipliers, Wang, Yin and Qi [44] converted the minimization problem (16) to a system of **nonsmooth equations**

$$F(x) = d, \tag{18}$$

where $d = (d_1, d_2, \cdots, d_n)^T$, $F = (F_1, F_2, \cdots, F_n)^T : \mathbf{R}^n \to \mathbf{R}^n$ and the $i-$th component of F is defined by

$$F_i(x) = \int_a^b (\sum_{l=1}^{n} x_l B_l(s) + h(s))_+ B_i(s) ds. \tag{19}$$

Based upon Proposition 1, Wang, Yin and Qi [44] obtained the following theorem.

Theorem 4.1 *Suppose h is continuously differentiable on $[a,b](a > 0)$. Then the function F defined by (19) is semismooth for any $x \in \mathbf{R}^n$.*

With this theorem, Wang, Yin and Qi [44] established superlinear convergence of the Newton-type method for solving (18).

May the Newton-type method for solving (18) have quadratic convergence? This needs to show the function F defined by (19) is strongly semismooth for any $x \in \mathbf{R}^n$. The answer is negative in general.

Qi and Yin [39] extended the strong semismoothness result of F defined by (10) to a more general integral function class.

Theorem 4.2 *Suppose that p is a continuous function on $[a,b]$ $(-\infty < a < b < \infty)$ and u,v are two strongly semismooth functions on $\mathbf{R}^n$. Then the integral function $G:\mathbf{R}^n \to \mathbf{R}$ defined by*

$$G(x) = \int_a^b \left(su(x) + v(x)\right)_+ p(s)ds \tag{20}$$

is strongly semismooth on $\mathbf{R}^n$.

Andersson, Elfving, Iliev and Vlaxhkova [1] converted the edge convex minimum norm network interpolation to a system of *nonsmooth equations*, and proposed a Newton-type method for solving the system. They could not prove the convergence of their method. By using Theorem 2, Qi and Yin [39] established quadratic convergence of the Newton-type method of Andersson, Elfving, Iliev and Vlaxhkova.

Qi and Yin [39] also gave an example that G, defined by

$$G(x) = \int_a^b [g(x,s)]_+ p(s)ds, \tag{21}$$

where $\alpha_+ := \max\{0,\alpha\}$, may not be a strongly semismooth function, even if $p(s) \equiv 1$ and g is a quadratic polynomial with respect to t and infinitely many times smooth with respect to x. The example is as follows:

In (21), let $n = 2, a = -1, b = 1, p(s) \equiv 1$ and $g(x,s) = s^2 + x_1 s + x_2$. Let

$$u = x_1^2 - 4x_2 \quad \text{and} \quad w = \sqrt{u_+}.$$

Assume that $\|x\| \le \frac{1}{8}$. Then

$$f(x) = \frac{2}{3} + 2x_2 + \frac{1}{6}q(x),$$

where

$$q(x) = w^3.$$

Then

$$\nabla q(x) = 3w\begin{pmatrix} x_1 \\ -2 \end{pmatrix},$$

i.e., q is smooth. This implies that f is smooth. However,

$$\nabla q(x)^T x - q'(0;x) = 3w(x_1^2 - 2x_2) = O(\|x\|^{1.5}).$$

By (4), q is not strongly semismooth. This implies that f is not strongly semismooth.

Hence, in general, the function F defined by (19) is not strongly semismooth, and the Newton-type method for solving (18) is not quadratically convergent.

Recently, Ling and Qi [25] showed that the function F defined by (19) is at least $\frac{1}{3}$-order semismooth in the sense of Qi and Sun [37], and the Newton-type method for solving (18) has at least $\frac{4}{3}$-order convergence rate.

5. SEMI-INFINITE PROGRAMS

Consider the following semi-infinite programming (SIP) problem:

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } \quad h_j(x) \le 0, \quad j = 1,2,\cdots,p, \\ &\qquad\qquad\quad g_j(x,s) \le 0, \quad s \in [a,b] \quad j = 1,2,\cdots,m, \end{aligned} \tag{22}$$

where $h_j(x) \le 0, \quad j = 1,2,\cdots,p$ are conventional inequality constraints, while $g_j(x,s) \le 0, \quad s \in [a,b] \quad j = 1,2,\cdots,m$ are infinite functional constraints, g is continuously differentiable (smooth) in x and s.

Such a SIP problem has wide applications [30,40]. Recently, Qi, Wu and Zhou [38] and Li, Qi, Tam and Wu [24] proposed a semismooth Newton method and a smoothing method to find a KKT pont of (22) respectively. However, It is possible to have another semismooth approach to solve (22).

In 1989-1993, Teo and his collaborators [21,41,42,43] proposed to aggregate the functional constraints to

$$G_j(x) := \int_a^b (g(x,s))_+ ds = 0, \quad j = 1,2,\cdots,m.$$

Then the SIP problem (22) is converted to a nonlinear programming problem:

$$\text{minimize } f(x) \tag{23}$$

$$\text{subject to} \quad h_j(x) \le 0, \quad j = 1,2,\cdots,p,$$
$$G_j(x) \le 0, \quad j = 1,2,\cdots,m,$$

where G_j may be nonsmooth. Teo and his collaborators [21,41,42,43] proposed a smoothing method to solve (23).

A function is called an SC^1 function if it is smooth and its gradient function is semismooth [32,28]. Recently, Qi and Shapiro [35] gave conditions under which G defined by

$$G(x) := \int_a^b (g(x,s))_+ ds,$$

is SC^1. Therefore, by [32,28], we may establish superlinear convergence of the SQP method for solving (23). This gives another semismooth approach to solve (22).

REFERENCES

[1] L.-E. Andersson, T. Elfving, G. Iliev AND K. Vlachkova, *Interpolation of convex scattered data in $\Re^3$ based upon an edged convex minimum norm network*, J. Approx. Theory. 80 (1995), pp. 299–320.

[2] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, J. Political Economy, 81 (1973), pp.637-659.

[3] C. Chen and O.L. Mangasarian, *A class of smoothing functions for nonlinear and mixed complementarity problems*, Comput. Optim. Appl. 5 (1996), pp. 97–138.

[4] X. Chen, L. Qi and D. Sun, *Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities*, Math. Comput., 67 (1998), pp. 519-540.

[5] P.W. Christensen and J.S. Pang, *Frictional contact algorithms based on semismooth Newton methods*, in: M. Fukushima and L. Qi (eds.): *Reformulation – Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods*, Kluwer Academic Publisher, Nowell, 1999, pp. 81-116.

[6] F. H. Clarke, *Optimization and Nonsmooth Analysis,* John Wiley & Sons, New York, 1983. (Reprinted by SIAM, Philadelphia, 1990.)

[7] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Prog. 75 (1996), pp. 407–439.

[8] T. De Luca, F. Facchinei, and C. Kanzow, *A theoretical and numerical comparison of some semismooth algorithms for complementarity problems,* Comput. Optim. Appl. 16 (2000), pp. 173–205.

[9] A. L. Dontchev, H.-D. Qi and L. Qi, *Convergence of Newton's method for convex best interpolation*, Numer. Math. 87 (2001), pp. 435–456.

[10] A. L. Dontchev, H.-D. Qi and L. Qi, *Quadratic convergence of Newton's method for convex interpolation and smoothing*, Constr. Approx. 19 (2003) 123-143.

[11] A. L. Dontchev, H.-D. Qi, L. Qi and H. Yin, *A Newton method for shape-preserving spline interpolation*, SIAM J. Optim. 13 (2003) 588-602.

[12] D. Duffie, *Dynamic Asset Pricing Theory*, Princeton University Press, Princeton, 1996.
[13] F. Facchinei, A. Fischer, and C. Kanzow, *Inexact Newton methods for semismooth equations with applications to variational inequality problems*, in: G. Di Pillo and F. Giannessi (eds.): *Nonlinear Optimization and Applications*, Plenum Press, New York, 1996, pp. 125–139.
[14] F. Facchinei and C. Kanzow, *A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems*, Math. Prog. 76 (1997), pp. 493–512.
[15] F. Facchinei and J.S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, 2003.
[16] M.C. Ferris and T.S. Munson, *Semismooth support vector machines*, Data Mining Institute Technical Report 00-09, Computer Sciences Department, University of Wisconsin, 2000.
[17] M. Fukushima and L. Qi, *Reformulation: Nonsmooth, piecewise smooth, semismooth and smoothing methods*, Applied Optimization 22, Kluwer Academic Publishers, Nowell, 1999.
[18] M. Fukushima and L. Qi, *Nonsmooth and smoothing methods*, Special Issue of Comput. Optim. Appl. 17, No 2/3, Kluwer Academic Publishers, Nowell, 2000.
[19] S.A. Gabrial and J.J. Moré, *Smoothing of mixed complementarity problems*, in: M.C. FERRIS AND J.S. PANG (eds.): *Complementarity and Variational Problems: State of the Art*, SIAM, Philadelphia, 1996, pp. 105–116.
[20] L. D. Irvine, S. P. Marin and P. W. Smith, *Constrained interpolation and smoothing*, Constr. Approx. 2 (1986), pp. 129–151.
[21] L.S. Jennings and K.L. Teo, *A computational algorithm for functional inequality constrained optimization problems*, Automatica. 26 (1990), pp. 371-375.
[22] H. Jiang and L. Qi, *A new nonsmooth equations approach to nonlinear complementarity problems*, SIAM J. Control & Optim. 35 (1997), PP. 178-193.
[23] H. Jiang, L. Qi, X. Chen and D. Sun, *Semismoothness and superlinear convergence in nonsmooth optimization and nonsmooth equations*, in: G. DI PILLO AND F. GIANNESSI (eds.): *Nonlinear Optimization and Applications*, Plenum Press, New York, 1996, pp. 197–212.
[24] D. Li, L. Qi, J. Tam and S. Y. Wu, *Smoothing Newton methods for semi-infinite programming*, to appear in: J. Global Optim.
[25] C. Ling and L. Qi, $\frac{4}{3}$*-order convergence of the generalized Newton method for solving the no-arbitrage option price interpolation problem*, Department of Applied Mathematics, The Hong Kong Polytechnic University, 2003.
[26] T.S. Munson, F. Facchinei, M.C. Ferris, A. Fischer and C. Kanzow, *The semismooth algorithm for large scale complementarity problems*, INFORMS J. Comput. 13 (2001), pp. 294–311.
[27] J.S. Pang and L. Qi, *Nonsmooth equations: Motivation and algorithms*, SIAM J. Optim. 3 (1993), pp. 443–465.
[28] J.S. Pang and L. Qi, *A globally convergent Newton method for convex* SC^1 *minimization problems*, J. Optim. Theory Appl. 85 (1995), pp. 633-648.
[29] J.-S. Pang and D. Ralph, *Piecewise smoothness, local invertibility, and parametric analysis of normal maps*, Math. Oper. Res., 21 (1996), pp. 401–426.
[30] E. Polak, *Optimization: Algorithms and Consistent Approximation*, Springer-Verlag, New York, 1997.
[31] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., 18 (1993), pp. 227–253.
[32] L. Qi, *Superlinearly convergent approximate Newton methods for* LC^1 *optimization problems*, Math. Prog., 64 (1994), pp. 277-294.

[33] L. Qi, *Regular pseudo-smooth NCP and BVIP functions and globally and quadratically convergent generalized Newton methods for complementarity and variational inequality problems*, Math. Oper. Res. 24 (1999), pp. 440–471.

[34] L. Qi and H. Jiang, *Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations*, Math. Oper. Res. 22 (1997), pp. 301–325.

[35] L. Qi and A. Shapiro, *Differentiability properties of integral functions and their applications*, Department of Applied Mathematics, The Hong Kong Polytechnic University, 2003.

[36] L. Qi, D. Sun and G. Zhou, *A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities*, Math. Prog., 87 (2000), pp. 1–35.

[37] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Prog. 58 (1993), pp. 353–367.

[38] L. Qi, S. Y. Wu and G. Zhou, *Semismooth Newton methods for solving semi-infinite programming problems*, J. Global Optim., 47 (2003), pp. 215-232.

[39] L. Qi and H. Yin, *A strongly semismooth integral function and its application*, Comput. Optim. Appl. 25 (2003) 223-246.

[40] R. Reemsten and J. Rückmann, *Semi-Infinite Programming*, Kluwer, Boston, 1998.

[41] K.L. Teo and L.S. Jennings, *Nonlinear optimal control problems with continuous state inequality constraints*, J. Optim. Theory Appl. 63 (1989), pp. 1-22.

[42] K.L. Teo, C.J. Goh and K.H. Wong, *A Unified Computational Approach to Optimal Control Problems*, Longman Scientific and Technical, 1991.

[43] K.L. Teo, V. Rehbock and L.S. Jennings, *A new computational algorithm for functional inequality constrained optimization problems*, Automatica, 29 (1993), pp. 789-792.

[44] Y. Wang, H. Yin and L. Qi, *No-Arbitrage interpolation of the option price function and its reformulation*, J. Optim. Theory Appl., 120 (2004)629-649.

[45] N. Yamashita and M. Fukushima, *Modified Newton methods for solving semismooth reformulations of monotone complementarity problems*, Math. Prog. 76 (1997), pp. 469–491.

HÖLDER REGULARITY RESULTS FOR SOLUTIONS OF PARABOLIC EQUATIONS

M.A. Ragusa
Dept. of Mathematics and Computer Sciences, University of Catania, Catania, Italy

Key words: Second order liner parabolic operator, divergence from equations

Mathematics Subject Classification (2000): Primary 31B10, 43A15,35K20. Secondary 32A37, 46E35

1. Introduction

In this paper we consider a linear parabolic operator of second order with coefficients belonging to the closure, in the parabolic BMO norm, of uniformly continuous functions. Aim of this note is to study some properties of the solution of the parabolic equation and extend the regularity results contained in [12] in order to allow operators to have lower order terms.

If the coefficients are discontinuous neither in elliptic case nor in the parabolic case there is a general theory, thus we wish to mention the study made by Di Fazio in [7] where the well posedness of a Dirichlet problem for divergence form elliptic equations is obtained in the case that the coefficients of the principal part belong to the Sarason class.

The technique used in this paper is inspired to that one used in [5], [6], where the authors consider an elliptic second order equation in nondivergence form and study the well posedness of the associated Dirichlet problem.

Let $\Omega \subset \mathbb{R}^n$ be an open bounded set with $\partial\Omega \in C^{1,1}$ and $T > 0$.

We set in the sequel

$$\mathcal{M}u \equiv u_t - (a_{ij}(X)u_{x_i})_{x_j}$$

and the linear parabolic operator

$$\mathcal{L}u = \mathcal{M}u + b_i u_{x_i} + cu - (d_i u)_{x_i}$$

in the cylinder $Q_T = \Omega \times (-T, 0)$, where $X = (x,t) = (x_1, .., x_n, t) \in \mathbb{R}^{n+1}$.

We are interested in the study of the Cauchy-Dirichlet problem

$$\begin{cases} \mathcal{M}u + b_i u_{x_i} + cu - (d_i u)_{x_i} = \operatorname{div} f \text{ in } Q_T \\ u = 0 \qquad\qquad\qquad\qquad\qquad \text{on } \partial\Omega \times (-T, 0) \\ u(x, -T) = 0 \qquad\qquad\qquad\qquad \text{in } \Omega. \end{cases}$$

We consider $\mathbb{R}^{n+1}$ $(n \geq 3)$ with the following parabolic metric that was first defined by Fabes and Riviére in [8] $d(X,Y) = \rho(X - Y)$, where

$$\rho(X) = \sqrt{\frac{|x|^2 + \sqrt{|x|^4 + 4t^2}}{2}}.$$

We define parabolic cube centered at $X = (x,t)$ with radius r the set

$$Q = Q_{r,X}(Y) = \left\{ Y = (y,\tau) \in \mathbb{R}^{n+1} : |x - y| < r; \quad |t - \tau| < r^2 \right\}.$$

Let us assume the coefficients a_{ij} such that

$$a_{ij}(X) = a_{ji}(X) \qquad \forall X \in Q_T \qquad \forall i, j = 1.., n;$$
$$\exists s > 0 : s^{-1} |\xi|^2 \leq a_{ij}(X)\xi_i\xi_j \leq s |\xi|^2, \qquad \forall \xi \in \mathbb{R}^n \qquad a.\ e.\ X \in Q_T.$$

We also suppose that a_{ij} belongs to the following class of vanishing mean oscillations functions.

Let us first define the more general class of functions of bounded mean oscillations.

Definition 1.1 *Let f be a locally integrable function defined in $\mathbb{R}^{n+1}$ The function f is in the Parabolic $BMO(\mathbb{R}^{n+1})$ (see [9]) if*

$$\sup_{Q\subset\mathbb{R}^{n+1}} \frac{1}{|Q|}\int_Q |f(y)-f_Q|\,dy < \infty$$

where Q run over the class of all parabolic cubes of $\mathbb{R}^{n+1}$ and f_Q is the integral average $f_Q = \frac{1}{Q}\int_Q f(y)dy$.

Let us consider a function $f \in BMO(\mathbb{R}^{n+1})$ and $r>0$. We set the Parabolic VMO modulus of f

$$\eta(r) = \sup_{\tau\le r} \frac{1}{|Q_\tau|}\int_{Q_\tau} |f(y) - f_{Q_\rho}|\,dy$$

where Q_τ is a parabolic cube with radius $\tau, \tau \le r$.

BMO is a Banach space with the following norm $\|f\|_* = \sup_{r>0}\eta(r)$.

Definition 1.2 *We say that a function $f \in BMO$ is in the Sarason class $VMO(\mathbb{R}^{n+1})$ (see [13]) if*

$$\lim_{r\to 0^+} \eta(r) = 0.$$

In the sequel we denote by η_{ij} the VMO modulus of a_{ij}, $i,j = 1,\dots,n$, and we define $\eta(r) = \left(\sum_{i,j=1}^n \eta_{ij}^2(r)\right)^{1/2}$.

Let us assume that the known term f belongs to $[L^p(Q_T)]^{n+1}, 1 < p < +\infty$.

The lower order terms satisfy the following hypothesis:

$b_i, d_i \in L^q(Q_T), c \in L^{\frac{q}{2}}(Q_T)$ where

$$q = \begin{cases} q = n+1, & 1 < q < n+1 \\ q > n+1, & q = n+1 \\ q = p, & p > n+1. \end{cases} \tag{1.1}$$

Definition 1.3 *A weak solution of the equation*

$$Mu + b_i u_{x_i} + cu - (d_i u)_{x_i} = \operatorname{div} f$$

is a function $u: Q_T \to \mathbb{R}$ *such that* $u, u_{x_j} \in L^2_{loc}(Q_T), \forall_j = 1,\ldots,n$ *and is true the equality*

$$\int_{Q_T} \Big(a_{ij}(x,t)(u_{x_j}\phi_{x_j})(x,t) - b_i(x,t)\big(u_{x_i}\phi\big)(x,t) - c(x,t)(u\phi)(x,t)\Big)dxdt - \\ - \int_{Q_T} u(x,t)\frac{\partial\phi(x,t)}{\partial t}dxdt = \\ = -\int_{Q_T}\Big(f_i(x,t)\phi_{x_i}(x,t)dxdt + d_i(x,t)u(x,t)\phi_{x_i}(x,t)\Big)dxdt, \quad \forall\phi \in C_0^\infty(Q_T).$$

Definition 1.4 *(see [11]). We say that a function* k *is a Parabolic Calderón-Zygmund kernel(PCZ kernel) on* $\mathbb{R}^{n+1}$ *we respect to the above defined parabolic metric* ρ *if:*

1. k *is smooth on* $\mathbb{R}^{n+1} \setminus \{0\}$;
2. $k(rx, r^2t) = r^{-(n+2)}k(x,t), \quad \forall r > 0$, *(homogeneity condition);*
3. $\int_{\rho(X)=r} k(X)d\sigma(X) = 0, \quad \forall > 0$, *(cancellation property on ellipsoids).*

Let us consider the fundamental solution of the parabolic operator.

Definition 1.5 *(see [1]). We denote by* $\Gamma^0(X) = \Gamma(X_0, \xi)$ *the fundamental solution of the constant coefficient operator* $\mathcal{M}_0$ *obtained from* $\mathcal{M}$ *by freezing the coefficients at a fixed point* $X_0 \in Q_T$

$$\Gamma(X_0,\xi) = \begin{cases} \dfrac{1}{(4\pi(t-T))^{(\frac{n}{2})}\sqrt{\det\{a_{ij}(X_0)\}}}\, exp\left(-\dfrac{\sum_{i,j=1}^n A_{ij}(X_0)\xi_i\xi_j}{4(t-T)}\right), & t-T>0 \\ & (1.2) \\ 0, & t-T<0 \end{cases}$$

where A_{ij} *are the entries of the inverse matrix of* $\{a_{ij}\}_{i,j=1,\ldots,n}$.

The first derivatives of the function will be denoted in general by

$$\Gamma_i(X,\xi) = \frac{\partial}{\partial\xi_i}\Gamma(X,\xi),$$

and the second derivatives by

$$\Gamma_{ij}(X,\xi) = \frac{\partial}{\partial\xi_i\partial\xi_j}\Gamma(X,\xi).$$

Definition 1.6 *(see [11]). Let us denote by* $a_n(x,t)=(a_{in}(x,t))_{i=1,\dots,n}$ *the last column (row) of the matrix* $\{a_{ij}\}_{i,j=1,\dots,n}$ *and define the following operator*

$$T(x,t;y,t) = x - 2x_n\frac{a_n(y,t)}{a_{nn}(y,t)} \qquad \forall x,y \in \mathbb{R}^n_+$$

and any fixed $t \in \mathbb{R}_+$. *Let us define*

$$T(X) = T(x,t;x,t) \qquad \forall x \in \mathbb{R}^n_+$$

for any fixed t *in* $\mathbb{R}_+$.

We point out that if $k(X,\cdot)$ *is a variable PCZ kernel,* $k(X,T(X)-Y)$ *is a nonsingular kernel for every point* X *and* Y.

ACKNOWLEDGEMENTS

The author takes this opportunity to thank prof. A. Maugeri for useful suggestions.

2. MAIN RESULTS

The main goal of this note is to prove the following theorem.

Theorem 2.1 *(Main Result). Let* $a_{ij} \in VMO \cap L^\infty(\mathbb{R}^{n+1}), i,j\dots,n,$ *be symmetric and uniformly elliptic,* $b_i, d_i \in L^q(Q_T), c \in L^2(Q_T)$ *where* q *is defined in (1.1) and the known term* $f \in L^p(Q_T), \quad 1 < p < \infty.$
Then the Cauchy-Dirichlet problem

$$\begin{cases} Lu = \operatorname{div} f & \text{in } Q_T \\ u = 0 & \text{on } \partial\Omega\times(-T,0) \\ u(x,-T) = 0 & \text{in } \Omega \end{cases} \tag{2.3}$$

has a unique solution u such that if $p > n+1$, belong to $C^{0,\alpha}(\overline{Q_T}), \alpha = 1 - \frac{n+1}{p}$, and also exists a costant $C > 0$ indipendent of f such that

$$\|u\|_{C^{0,\alpha}(\overline{Q_T})} \leq C \|f\|_{L^p(Q_T)}. \tag{2.4}$$

We need some properties before proving the above theorem.

Theorem 2.2 *(see [1]). Let us consider k a variable PCZ kernel and*

$$Kf(X) = P.V. \int_{\mathbb{R}^{n+1}} k(X, X-Y) f(Y) dY.$$

Then, for every $f \in L^p(Q_T), 1 < p < \infty$, there exists $C = C(p,k)$ independent of f such that

$$\|Kf\|_{L^p(Q_T)} \leq C \|f\|_{L^p(Q_T)}.$$

Theorem 2.3 *(see [1]). Let k be a variable PCZ kernel and K as in the previous theorem. For $a \in VMO \cap L^\infty(\mathbb{R}^{n+1})$, we consider the commutator*

$$\begin{aligned} C[a,f](X) &= P.V. \int_{\mathbb{R}^{n+1}} k(X, X-Y)[a(X) - a(Y)] f(Y) dY = \\ &= a(X)(K(f))(X) - K(af)(X). \end{aligned}$$

Then, for every $\varepsilon > 0 \exists c > 0$ and also $\exists r_0 > 0$ depending only on ε and the VMO modulus η_a of a such that

$$\|C[a,f]\|_{L^p(Q_r)} \leq c \in \|f\|_{L^p(Q_r)}, \qquad \forall f \in L^p(Q_r), \quad with\ r \leq r_0.$$

The following two results are established in [1], respectively they are Theorem 3.1 and Corollary 3.8.

Theorem 2.4 *Let k be a PCZ variable kernel and T the operator considered in Definition 1.6. We consider the operator*

$$\tilde{K}f(X) = \int_{\mathbb{R}^{n+1}} k(X, T(X) - Y) f(Y) d(Y).$$

Then, for every $1 < p < \infty \exists C = C(p, \mu, k)$ such that for $f \in L^p\left(\mathbb{R}^{n+1}_+\right)$ we have

$$\left\|\tilde{K}f\right\|_{L^p\left(\mathbb{R}_+^{n+1}\right)} \leq C\left\|f\right\|_{L^p\left(\mathbb{R}_+^{n+1}\right)}.$$

Theorem 2.5 *Let $k(X,Y)$ be a variable PCZ kernel, T as in Theorem 2.4, $a \in VMO \cap L^\infty(\mathbb{R}_+^{n+1})$ and*

$$\tilde{C}[a,f] = \int_{\mathbb{R}_+^{n+1}} k(X, T(X) - Y)[a(X) - a(Y)]f(Y)dY.$$

Then, for every $\varepsilon > 0 \exists r_0$, depending only on ε and the VMO modulus η_a of a, such that for every $f \in L^p(Q_r^+), 1 < p < \infty, Q_r^+ = Q_r \cap \mathbb{R}_+^{n+1},$, with $r < r_0$, yields

$$\left\|\tilde{C}[a,f]\right\|_{L^p(Q_r^+)} \leq c(p,\mu,k) \in \left\|f\right\|_{L^p(Q_r^+)}.$$

Theorem 2.6 *(see [14]). Let $\beta \in]0, n+1[$ and let $V \in C^\infty(\mathbb{R}^{n+1} \setminus 0)$ be a homogeneous function of degree $-\beta$. If $g \in L^\mu(\mathbb{R}^{n+1})$ then the operator*

$$Jg(X) = \int_{\mathbb{R}^{n+1}} V(X - Y)g(Y)dY$$

is defined almost every where and $\exists c = c(p,\mu) > 0$ such that

$$\|Jg\|_{L^p\left(\mathbb{R}_+^{n+1}\right)} \leq c \max_{\|X\|=1} |V(X)| \cdot \|g\|_{L^\mu\left(\mathbb{R}^{n+1}\right)}, \qquad \tfrac{1}{p} + 1 = \tfrac{1}{\mu} + \tfrac{\beta}{n+1}.$$

Let Q_σ be some parabolic cube $Q_\sigma \subset\subset Q_T$, $\Theta \in C^\infty(Q_T)$ a standard cut-off function, $\Theta(x) = 1$ in $Q_{\gamma\sigma}, 0 < \gamma < 1$.

If u is a solution of $Lu = divf$ on Q_T with zero boundary data, we may consider u as a solution, with support in Q_σ, of the equation $\mathcal{M}(\Theta u) = \mathcal{S} + \operatorname{div}\mathcal{F}$, where

$$\mathcal{F} = -(a_{ij}\Theta_{x_i}u - (\Theta f_j + d_j u)),$$

and

$$\mathcal{S} = -(a_{ij}\Theta_{x_j}u_{x_i} + \Theta_{x_j}(f_j + d_j u) + \Theta b_i u_{x_i} + c\Theta u) + \Theta_t u.$$

Theorem 2.7 *(Interior Representation Formula). Let the hypotheses of symmetry and ellipticity for $a_{ij} \in C^\infty(\mathbb{R}^{n+1}) \cap L^\infty(\mathbb{R}^{n+1})$ hold, let*

$\mathcal{F} \in [C_0^\infty(Q)]^{n+1}$, $\mathcal{S} \in C_0^\infty(Q)$ *and* $v = \Theta u \in C_0^\infty(Q)$,, *for some* $Q \subset\subset Q_T$, *as a solution of* $\mathcal{M}(\Theta u) = \mathcal{S} + \operatorname{div}\mathcal{F}$. *Then*

$$v_{xi} = (\Theta u)_{xi} = P.V.\int_Q \Gamma_{ij}(X, X-Y)\{[a_{hj}(X) - a_{hj}(Y)](\Theta u)_{x_h}(Y) - \\ -\mathcal{F}_j(Y)\}dY + \int_Q \mathcal{S}(Y)\Gamma_i(X, X-Y)dY + \mathcal{F}_j(X)\int_{\sum_{n+1}} \Gamma_i(x,t)\eta_j d\sigma_t \tag{2.5}$$

where η_j *stands for the* j*-th component of the outer normal to the surface* Σ_{n+1}.

Proof. Let $X_0 \in Q$. We have

$$\mathcal{M}_0(\Theta u)(X) \equiv (\Theta u)_t - (a_{ij}(X_0)(\Theta u)_{xi}(X))_{xj} = \\ = \mathcal{M}_0(\Theta u) - \Theta\mathcal{M}u + \Theta\{(f_j + d_j u)_{xi} - (b_i u_{xi} + cu)\} = \\ = -[(a_{ij}(X_0) - a_{ij}(X))(\Theta u)_{xi}(X) - \mathcal{F}_j(X)]_{xj} + \mathcal{S}(X) = \\ = -(\omega_j^{X_0}(X))_{xj} + \mathcal{S}(X).$$

Since (Θu), $\mathcal{F}$ and $\mathcal{S}$ are compactly supported in Q, we have

$$(\Theta u)(X) = \int_Q \Gamma_j(X_0, X-Y)\omega_j^{X_0}(Y)dY + \int_Q \Gamma(X_0, X-Y)\mathcal{S}(Y)dY, \quad \forall X \in Q.$$

Differentiating twice yields

$$(\Theta u)_{xi}(X) = P.V.\int_Q \Gamma_{ij}(X_0, X-Y)\{[a_{hj}(X_0) - a_{hj}(Y)](\Theta u)_{x_h}(Y) - \mathcal{F}_j(Y)\}dY + \\ + \int_Q \Gamma_i(X_0, X-Y)\mathcal{S}(Y)dY + c_{ij}\omega_j^{X_0}(X); \qquad c_{ij}(X) = \int_{\Sigma_{n+1}} \Gamma_i(x,t)t_j d\sigma_t.$$

This relation is true for every $X_0 \in Q$ Letting $X = X$ we obtain the interior representation formula for v_{xi}.

In the sequel we will consider the L^p estimates only for $p > 2$ because the case $p = 2$ is a classical result by Campanato and the case $1 < p < 2$ will be recovered by duality.

Theorem 2.8 *(Interior Estimate). Let* a_{ij} *satisfy the hypotheses of Theorem 2.7, and let* $u \in C^\infty(Q_T)$ *be a solution of* $Lu = \operatorname{div} f$, *with* $f \in [C^\infty(Q_T)]^{n+1}$. *Then* $\exists \sigma, c$ *depending on* n, p, η_{ij}, μ *and* $\operatorname{dist}(Q_\sigma, \partial Q_T)$ *such that*

$$\|\nabla u\|_{L^p(Q_{\frac{\sigma}{2}})} \leq c\Big(\|\nabla u\|_{L^2(Q_\sigma)} + \|f\|_{L^p(Q_\sigma)} + \|u\|_{L^p(Q_\sigma)}\Big), \qquad \forall Q_\sigma \subset\subset Q_T. \tag{2.6}$$

Proof. Let us first examine the assertion if $2 < p \leq 2^*$. Using the above obtained interior representation formula, Theorems 2.2, 2.3, 2.6, and taking the p-norms of $\nabla(\Theta u)$ over the cylinders Q_σ, we find that

$$\|\nabla(\Theta u)\|_{L^p(Q_\sigma)} \leq c\Big(\| a \|_* \| \nabla(\Theta u) \|_{L^p(Q_\sigma)} + \|\mathcal{F}\|_{L^p(Q_\sigma)} + \|\mathcal{S}\|_{L^{p_*}(Q_\sigma)}\Big),$$

where $\frac{1}{p_*} = \frac{1}{p} + \frac{1}{n+1}$. If σ is chosen for which $\| a \|_* < \frac{1}{2}$ then $\exists c$ independent of $\Theta, u, \mathcal{F}$ and $\mathcal{S}$ such that

$$\| \nabla(\Theta u) \|_{L^p(Q_\sigma)} \leq c\Big(\|\mathcal{F}\|_{L^p(Q_\sigma)} + \|\mathcal{S}\|_{L^{p_*}(Q_\sigma)}\Big).$$

From the definition of $\mathcal{F}$ and $\mathcal{S}$ we have

$$\|\mathcal{S}\|_{L^{p_*}(Q_\sigma)} \leq \|\nabla u\|_{L^{p_*}(Q_\sigma)} + \|f\|_{L^{p_*}(Q_\sigma)} + \| u \|_{L^{p_*}(Q_\sigma)} \leq$$

$$\leq \|\nabla u\|_{L^{p_*}(Q_\sigma)} + \|f\|_{L^p(Q_\sigma)} + \| u \|_{L^p(Q_\sigma)},$$

and

$$\|\mathcal{F}\|_{L^{p_*}(Q_\sigma)} \leq c\Big(\|f\|_{L^p(Q_\sigma)} + \| u \|_{L^p(Q_\sigma)}\Big).$$

Let us observe that, if $2 < p \leq 2^*$, $\frac{1}{p_*} \geq \frac{1}{2}$. It follows that $p_* \leq 2$ and we have:

$$\| \nabla u \|_{L^{p_*}(Q_\sigma)} \leq c \| \nabla u \|_{L^2(Q_\sigma)}.$$

Hence

$$\| \nabla u \|_{L^p(Q_{\gamma\sigma})} \leq c\Big(\| u \|_{L^p(Q_\sigma)} + \| f \|_{L^p(Q_\sigma)} + \| \nabla u \|_{L^2(Q_\sigma)}\Big). \tag{2.7}$$

If $2^* < p \leq 2^{**}$ (with 2^{**} such that $\frac{1}{2^{**}} = \frac{1}{2^*} - \frac{1}{n+1}$), it follows that $p_* \leq 2^*$. Letting Θ be a cut-off function identically equal to 1 in $Q_{\gamma^2\sigma}$ and supported in $Q_{\gamma\sigma}$, such that $0 \leq \Theta \leq 1$ and $| \nabla\Theta | \leq \frac{c}{\gamma\sigma(1-\gamma)}$, we obtain

$$\| \nabla u \|_{L^p(Q_{\gamma^2\sigma})} \le c\Big(\| u \|_{L^p(Q_{\gamma\sigma})} + \| f \|_{L^p(Q_{\gamma\sigma})} + \| \nabla u \|_{L^{p_*}(Q_{\gamma\sigma})}\Big) \le$$

$$\le c\Big(\| u \|_{L^p(Q_{\gamma\sigma})} + \| f \|_{L^p(Q_{\gamma\sigma})} + \| \nabla u \|_{L^{2^*}(Q_{\gamma\sigma})}\Big). \tag{2.8}$$

Using (2.7) with $p = 2^*$ we see that (2.8) reduces to the inequality

$$\| \nabla u \|_{L^p(Q_{\gamma^2\sigma})} \le c\Big(\| u \|_{L^p(Q_\sigma)} + \| f \|_{L^p(Q_\sigma)} + \| \nabla u \|_{L^2(Q_\sigma)}\Big).$$

If $\gamma^2 = \frac{1}{2}$ we obtain (2.6).

Finally, iterating this method with various exponents $p > 2$ we will find $h \in \mathbb{N}$, an exponent $2^{\overbrace{***}^{h-1}} < p \le 2^{\overbrace{***}^{h}}$ so that $\gamma^h = \frac{1}{2}$ concluding with (2.6).

We set $Q_\sigma^+ = \{(x_1..x_n, t) \in Q_\sigma : x_n > 0, t > 0\}$, where Q_σ is a parabolic cube.

Theorem 2.9 *(Boundary Representation Formula). Let us assume the above hypotheses about the coefficients* a_{ij} *and, in addition, that* $a_{ij} \in C^\infty\left(\mathbb{R}^{n+1}\right), b_i, d_i, c \in C^\infty\left(Q_\sigma^+\right)$. *Let* $\mathcal{F} \in [C^\infty(\overline{Q}_\sigma^+)]^{n+1}$ *and* $\mathcal{S} \in C^\infty(\overline{Q}_\sigma^+)$, *vanish in a neighbourhood of* $\mathbb{R}_+^{n+1} \cap \partial Q_\sigma$.

If u *is a restriction to* Q_σ^+ *of some function in* $C_0^\infty(Q_\sigma)$ *vanishing in* $\{\{x_n = 0\}\times]0,T[\} \cap \overline{Q}_\sigma^+$ *and satisfies the equation* $\mathcal{L}u = \mathcal{S} + \operatorname{div}\mathcal{F}$ *in* $\overline{Q}_\sigma^+$, *then*

$$u_{x_i} = P.V.\int_{Q_\sigma^+} \Gamma_{ij}(X, X-Y)\Big\{[a_{hj}(X) - a_{hj}(Y)]u_{x_h}(Y) - \mathcal{F}_j(Y)\Big\}dY +$$

$$+c_{ij}(X)\mathcal{F}_j(X) + \int_{Q_\sigma^+} \Gamma_i(X, X-Y)\mathcal{S}(Y)dy + I_i(X), \qquad \forall X \in Q_\sigma^+,$$

where $c_{ij}(X) = \int_{\sum_{n+1}} \Gamma_i(x,t)\eta_j d\sigma_t$ *are bounded functions arising from (2.5),*

$$I_i(X) = \int_{Q_\sigma^+} \Gamma_{ij}(X, T(X) - Y)\Big\{\big[a_{hj}(X) - a_{hj}(Y)\big]u_{x_h}(Y) - F_j(Y)\Big\}dY -$$

$$-\int_{Q_\sigma^+} \Gamma_i(X, T(X) - Y)\mathcal{S}(Y)dY, \qquad i = 1,..,n-1$$

and

$$I_n(X) = \int_{Q_\sigma^+} B_h(Y)\Gamma_{hj}(X,T(X)-Y)\{[a_{hj}(X)-a_{hj}(Y)]u_{x_h}(Y)-\mathcal{F}_j(Y)\}dY$$

where B_h are bounded functions having L^∞ norm estimated in terms of the ellipticity constant s.

Proof. Let us consider the Green's function $G_0(X,Y)$ for the halfspace $\mathbb{R}^{n+1}_+$ and u as in the above hypothesis, then

$$u(X) = \int_{Q_\sigma^+} G_0(X,Y)\mathcal{M}_0 u(Y)dY.$$

We claim that for fixed $X_0 \in \mathbb{R}^{n+1}$

$$G_0(X,Y) = \Gamma^0(X-Y) - \Gamma^0(T(X_0,X)-Y)$$

thus

$$u(X) = \int_{Q_\sigma^+} \Gamma^0(X_0,X-Y)\mathcal{M}_0 u(Y)dY - \int_{Q_\sigma^+} \Gamma^0(T(X_0,X)-Y)\mathcal{M}_0 u(Y)dY =$$
$$= -\int_{Q_\sigma^+} \Gamma^0(X_0,X-Y)(\omega_j^{X_0}(Y))_{x_j}dY + \int_{Q_\sigma^+} \Gamma^0(X_0,X-Y)\mathcal{S}(Y)dY -$$
$$-\left\{-\int_{Q_\sigma^+} \Gamma^0(X,T(X_0,X)-Y)(\omega_j^{X_0}(Y))_{x_j}dY + \int_{Q_\sigma^+} \Gamma^0(X_0,T(X_0,X)-Y)\mathcal{S}(Y)dY\right\};$$

this means that

$$u(X) = \int_{Q_\sigma^+} \Gamma_j^0(X_0,X-Y)(\omega_j^{X_0}(Y))dY + \int_{Q_\sigma^+} \Gamma^0(X_0,X-Y)\mathcal{S}(Y)dY -$$
$$-\int_{Q_\sigma^+} \Gamma_j^0(X_0,T(X_0,X_0-Y)(\omega_j^{X_0}(Y))dY - \int_{Q_\sigma^+} \Gamma^0(X_0,T(X_0,X-Y)\mathcal{S}(Y)dY$$

and it is easy to see that

$$u_{x_i}(X) = \int_{Q_\sigma^+} \Gamma_{ij}^0(X_0,X-Y)(\omega_j^{X_0}(Y))dY + c_{ij}(X)\omega_j^{X_0}(X) + \int_{Q_\sigma^+} \Gamma_i^0(X_0,X-Y)\mathcal{S}(Y)dY + I_i(X) \tag{2.9}$$

where, for $i = 1,..,n-1$,

$$I_i(X) = \int_{Q_\sigma^+} \Gamma_{ij}^0(X_0,T(X_0,X-Y)(\omega_j^{X_0}(Y))dY,$$

and

$$I_n(X) = \int_{Q_\sigma^+} B_h(Y)\Gamma_{hj}(X_0, T(X_0, X) - Y)(\omega_j^{X_0}(Y))dY.$$

By our smoothess assumption, (2.9) equality is true for any fixed X_0 and, in particular, for $X = X_0$.

Then we have proved the boundary representation formula.

Theorem 2.10 *(Boundary Estimate). Let* $a_{ij} \in C^\infty(\mathbb{R}^{n+1}) \cap L^\infty(\mathbb{R}^{n+1})$ *be symmetric and uniformly elliptic. Let also be* $f \in [C_0^\infty(\overline{Q}_\sigma^+)]^{n+1}, b_i, d_i, c \in C^\infty(Q_\sigma^+), \forall i = 1, \ldots, n.$

Then there exists $\sigma > 0$ *such that for every* $u \in C^\infty(Q_\sigma^+)$ *such that* $Lu = \operatorname{div} f$ *in* Q_σ^+, *which vanishes on* $\{\{x_n = 0\} \times]0, T[\} \cap \overline{Q}_\sigma^+$, *we obtain*

$$\| \nabla u \|_{L^r(Q_{\sigma/2}^+)} \leq c\left(\| \nabla u \|_{L^2(Q_\sigma^+)} + \| f \|_{L^r(Q_\sigma^+)} + \| u \|_{L^r(Q_\sigma^+)} \right).$$

for a suitable constant c *independent of* u *and* f.

Proof. Using the boundary representation formula (Theorem 2.9) and applying Theorems 2.4 and 2.5 we have the requested inequality. We point out that the terms B_h are not relevant because are bounded functions.

We are now ready to prove the *main result,* following the lines of [12].

First we obtain that $\exists c > 0$ independent of u and f such that $\| \nabla u \|_{L^r(Q_T)} \leq c \| f \|_{L^r(Q_T)}$ with restrictive hypotheses that $a_{ij} \in C^\infty(\mathbb{R}^{n+1}) \cap L^\infty(\mathbb{R}^{n+1})$, $f \in [C^\infty(Q_T)]^{n+1}$, $b_i, d_i, c \in C^\infty(Q_T)$.

Interior and boundary estimates allow us to have

$$\| \nabla u \|_{L^r(Q_T)} \leq c\left(\| \nabla u \|_{L^2(Q_T)} + \| f \|_{L^r(Q_T)} + \| u \|_{L^r(Q_T)} \right), \qquad \forall p > 2. \tag{2.10}$$

Let us suppose $2 < p \leq 2^*$, using the Sobolev theorem we obtain

$$\| u \|_{L^r(Q_T)} \leq c \| \nabla u \|_{L^2(Q_T)}. \tag{2.11}$$

Making use of the last inequality, (2.10) and the L^2 theory we write

$$\| \nabla u \|_{L^r(Q_T)} \leq c\left(\| \nabla u \|_{L^2(Q_T)} + \| f \|_{L^r(Q_T)} \right) \leq c \| f \|_{L^r(Q_T)} \tag{2.12}$$

which gives

$$\| \nabla u \|_{L^p(Q_T)} \leq c \| f \|_{L^p(Q_T)} .$$

From this relation we have

$$\| \nabla u \|_{L^{2^*}(Q_T)} \leq c \| f \|_{L^{2^*}(Q_T)} . \tag{2.13}$$

Further, is $2^* < p \leq 2^{**}$, the Sobolev lemma ensures

$$\| u \|_{L^p(Q_T)} \leq c \| \nabla u \|_{L^{p^*}(Q_T)} \leq c \| f \|_{L^{2^*}(Q_T)} \leq c \| f \|_{L^p(Q_T)} . \tag{2.14}$$

Making use of (2.10) and (2.14) we conclude that

$$\| \nabla u \|_{L^p(Q_T)} \leq c \left(\| \nabla u \|_{L^2(Q_T)} + \| f \|_{L^p(Q_T)} \right) \leq c \| f \|_{L^p(Q_T)}, \qquad \forall 2 < p \leq 2^{**}.$$

After a finite number of steps the required estimate $\forall p > 2$ is true.

Let us set $a_{ij} \in VMO \cap L^\infty(\mathbb{R}^{n+1})$, $\{a_{ij}^k\}_{k\in\mathbb{N}}$ a sequence of smooth functions converging to a_{ij} in the BMO norm, $\{f_k\}_{k\in\mathbb{N}}$ a sequence of functions in $[C^\infty(Q_T)]^{n+1}$ converging to $f \in [L^p(Q_T)]^{n+1}$, $\{b_i^k\}, \{d_i^k\}$, and $\{c^k\}$ sequences of functions in $C^\infty(Q_T)$ converging respectively to b_i, d_i, c and by u_k the solution of the Dirichlet problem

$$\begin{cases} u_t - (a_{ij}^k u_{x_i})_{x_j} + b_i^k u_{x_i} + c^k u - (d_i^k u)_{x_i} = \operatorname{div} f^k, & Q_T \\ u = 0 & \partial\Omega \times (-T, 0) \\ u(x, -T) = 0 & \text{in } \Omega. \end{cases}$$

Therefore, we have

$$\| \nabla u_k \|_{L^p(Q_T)} \leq c \| f_k \|_{L^p(Q_T)}, \qquad \forall k \in \mathbb{N}$$

where c is indipendent of $k \in \mathbb{N}$. From the above inequality there exists a function u such that

$$\| \nabla u \|_{L^p(Q_T)} \leq c \| f \|_{L^p(Q_T)} . \tag{2.15}$$

and moreover u verifies the original Dirichlet problem.

Using this L^p estimate and the Sobolev imbedding theorem we have that, if $p > n+1$, $u \in C^{0,\alpha}(\overline{Q_T})$, with $\alpha = 1 - \frac{n+1}{p}$.

REFERENCES

[1] M.Bramanti, M.C. Cerutti, $W_p^{1,2}$ Solvability for the Cauchy-Dirichlet Problem for Parabolic Equations with VMO coefficients, *Comm. in Partial Differential Equations* **18 (9 and 10)** (1993), 1735–1763.

[2] F. Chiarenza, M. Frasca, P. Longo, Interior $W^{2,p}$ estimates for nondivergence elliptic equations with discontinuous coefficients, *Ricerche di Mat.* **40** (1991), 149–168.

[3] F. Chiarenza, M. Frasca, P. Longo, $W^{2,p}$-solvability of the Dirichlet problem for non divergence elliptic equations with *VMO* coefficients, *Trans. Amer. Math. Soc.* **336** (1993), 841–853.

[4] G. Di Fazio, L^p estimates for divergence form elliptic equations with discontinuous coefficients, *Bollettino U.M.I.* **(7) 10-A** (1996), 409–420.

[5] E. Fabes, N. Riviére, Singular integrals with mixed homogeneity, *Studia Math.* **27** (1966), 19–38.

[6] F. John, L. Nirenberg, On functions of bounded mean oscillation, *Commun. Pure Appl. Math.* **14** (1961), 415–426.

[7] D. Palagachev, M.A. Ragusa, L. Softova, Cauchy-Dirichlet problem in Morrey spaces for parabolic equations with discontinuous coefficients, *B.U.M.I.* (8) 6-B (2003), 667-675

[8] M.A.Ragusa, Dirichlet problem associated to divergence form parabolic equations, *preprint.*

[9] D. Sarason, On functions of vanishing mean oscillation, *Trans. Amer. Math. Soc.* **207** (1975), 391–405.

[10] E.M. Stein Armonic Analysis: Real Variable methods, orthogonality and oscillatory integrals, *Princeton Univ. Press, Princeton, New Jersey*, 1993.

SURVEY ON THE FENCHEL PROBLEM OF LEVEL SETS*

T. Rapcsák
Computer and Automation Research Institute, Hungarian Academy of Sciences, Budapest, Hungary

Abstract: The Fenchel problem of level sets was formulated by Roberts and Varberg in their book titled "Convex functions" (1973, p. 271) is as follows: "What "nice" conditions on a nested family of convex sets will ensure that it is the family of level sets of a convex function?" The aim of the paper is to draw attention to this structural question of convex analysis and to survey some results in different directions.

1. INTRODUCTION

The problem of level sets, formulated and discussed first by Fenchel in 1951, is as follows: Under what conditions is a nested family of closed convex sets the family of the level sets of a convex function ?

Fenchel (1951, 1956) gave necessary and sufficient conditions for the existence of a convex function with the prescribed level sets, furthermore, the existence of a smooth convex function under the additional assumption that the given subsets are the level sets of a twice continuously differentiable function. In the first case, seven conditions were deduced, and while the first six are simple and intuitive, the seventh is rather complicated. This fact and the additional assumption in the smooth case, according to which the given

* This research was supported in part by the Hungarian National Research Fund, Grant No. OTKA-TO43241 and CNR.

subsets are the level sets of a twice continuously differentiable function, seem to be the motivation for Roberts and Varberg (1973, p. 271) to draw up anew the following problem of level sets among some unsolved problems: "What "nice" conditions on a nested family of convex sets will ensure that it is the family of level sets of a convex function?"

The aim of the paper is to draw attention to a structural question of convex analysis raised by Fenchel (1951) which contains several open subproblems, and to survey some results in different directions. The second part is devoted to the original Fenchel result (1951) in the continuous case, because it cannot be found in textbooks and research monographs. The third part is only a short summary of some results in the smooth case, because a detailed review can be found in Avriel et al. (1988) (Section 8.2, about 40 pages), and the concluding remarks contain some open problems.

2. THE CONTINUOUS CASE

Level sets of a function

Let $A \subseteq R^n$ be a subset and $f : A \to R$ an arbitrary, not necessarily convex function. Then, the level sets of the function f are

$$\mathrm{lev}_{\leq\alpha} f = \{\mathbf{x} \in A \mid f(\mathbf{x}) \leq \alpha\}, \qquad \alpha \in R. \tag{2.1}$$

Clearly, $\mathrm{lev}_{\leq\alpha} f$ is empty for $\alpha < \mathbf{x} \in A\inf f(\mathbf{x})$. Therefore, α will be restricted to the smallest interval J containing the whole range $Im_f(A)$ of f. This interval may be finite or infinite, open, half open or closed. To exclude the trivial case of a constant function we assume that J has interior points. In the following space, all numbers $\alpha, \beta, \ldots$ are supposed to belong to J. On observing that $f(\mathbf{x}) \leq \alpha, \mathbf{x} \in A$, is equivalent to $f(\mathbf{x}) \leq \beta$ for all $\beta > \alpha$, it is immediately seen that the family of level sets $\mathrm{lev}_{\leq\alpha} f$ has the following properties:

$$\alpha \in J \cup \mathrm{lev}_{\leq\alpha} f = A, \tag{2.2}$$

$$\textit{if } \alpha, \beta \in J \textit{ and } \alpha \leq \beta, \textit{ then } \mathrm{lev}_{\leq\alpha} f \subseteq \mathrm{lev}_{\leq\beta} f, \tag{2.3}$$

$$\textit{if } \alpha, \beta \in J, \textit{ then } \beta > \alpha \cap \mathrm{lev}_{\leq\beta} f = \mathrm{lev}_{\leq\alpha} f, \tag{2.4}$$

if J does not contain a lower bound, then $\alpha \in J \cap \operatorname{lev}_{\leq\alpha} f = \varnothing$. (2.5)

So far, our steps have been reversible, i.e., given a set A in R^n and a family $\{L_\alpha\}$ of sets indexed by the real numbers of some interval J and satisfying (2.2)–(2.5), we may construct a function $f : A \to R$ having L_α, $\alpha \in J$, as its level sets. If f is defined by

$$f(\mathbf{x}) = \inf\{\alpha \in J \mid \mathbf{x} \in L_\alpha\}, \tag{2.6}$$

then, by Fenchel (1951, p. 116),

$$L_\alpha = \{\mathbf{x} \in A \mid f(\mathbf{x}) \leq \alpha\}. \tag{2.7}$$

If follows that the function f is finite for all $\mathbf{x} \in A$, because for every $\mathbf{x} \in A$, property (2.2) ensures that some L_α contains $\mathbf{x}$, while (2.5) ensures that if J is unbounded below, there is some L_α which does not contain $\mathbf{x}$.

The level set corresponding to α of this function consists of all $\mathbf{x}$ such that $\operatorname{int}_{\mathbf{x}\in L_\beta} \beta \leq \alpha$. Thus, $\mathbf{x}$ is in this level set iff f for every $\varepsilon > 0$, there is a $\beta < \alpha + \varepsilon$ such that $\mathbf{x} \in L_\beta$. Because of (2.3), this means that $\mathbf{x} \in L_{\alpha+\varepsilon}$ for all $\varepsilon > 0$, and hence, by (2.4), $\mathbf{x} \in L_\alpha$. A further consequence of (2.4) is that $f(\mathbf{x}) = \min_{\mathbf{x}\in L_\alpha} \alpha$.. This equation establishes a one-to-one correspondence between the function f defined over A and the indexed families of subsets satisfying (2.2)–(2.5).

It is known well that a function $f : A \to R$ with level sets L_α, $\alpha \in J$, is lower semicontinuous iff

L_α is closed relative to A for every $\alpha \in J$. (2.8)

The condition for upper semicontinuity is

$\bigcup_{\beta<\alpha} L_\beta$ is open relative to A for every $\alpha \in J$

will not be used explicitly.

Let $t = \varphi(\alpha)$ be a strictly increasing continuous function defined for $\alpha \in J$. Denote the range $\varphi(\alpha)$, $\alpha \in J$, by J_φ, and let $\alpha = \varphi^{-1}(t)$,

$t \in J_\varphi$, be the inverse of φ. Then, the family of the sets $L_{\varphi^{-1}(t)}$, $t \in J_\varphi$, is the family of the level sets of the function $\varphi(f(\mathbf{x}))$, $\mathbf{x} \in A$, and satisfies conditions (2.2)-(2.5) and (2.8) if L_α, $\alpha \in J$, does so. For the sake of brevity, two families such as $\{L_\alpha\}$ and $\left\{L_{\varphi^{-1}(t)}\right\}$ obtained from each other by a strictly increasing and continuous index transformation $t = \varphi(\alpha)$, $\alpha \in J$, is said to be transformable into each other.

Figure 1 shows some of the level sets for a function $f : R^2 \to R$ that has a surface of revolution as its graph.

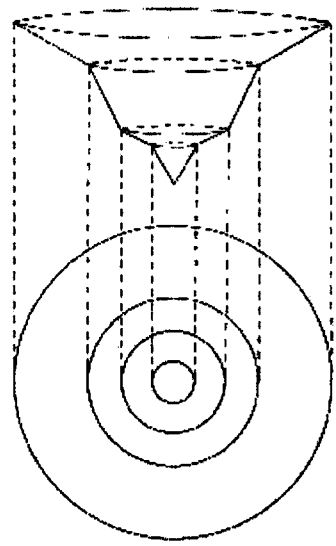


Figure 1.

Level sets of a convex function

With f now constructed from the sets $\{L_\alpha\}$ satisfying $(2.2)-(2.5)$ and (2.8), it is natural to wonder what additional knowledge about the sets $\{L_\alpha\}$ will enable us to draw conclusions about f. By (2.8), if the sets $\{L_\alpha\}$ are all closed, then f is lower semicontinuous.

The Fenchel problem of level sets was formulated on page 117 as follows:

"*Under what conditions is a family of sets* $\{L_\alpha\}$ *satisfying* $(2.2)-(2.5)$ *and* (2.8) *transformable into the family of level sets of a convex function? To avoid inessential difficulties the domain* A *will henceforth be assumed to be convex and open.*"

It is well-known that if f is convex, then $\operatorname{lev}_{\leq\alpha} f$ is convex for each $\alpha \in J$. De Finetti (1949) asked the obvious question: What about the converse? The question of de Finetti, an inverse problem, is of the above type, and it led him to study the class of functions that are now called quasiconvex. If we add the obvious necessary assumption

$$L_\alpha \text{ is convex for each } \alpha \in J, \tag{2.9}$$

what can we say about the corresponding function? It is clear from Figure 1 where all the level sets of a nonconvex function are concentric circles that we cannot conclude the convexity of f. It is well-known that the level sets of a function defined on a convex set are convex iff the function is quasiconvex. Both de Finetti (1949) and Fenchel (1951, 1956) gave further restrictions on the family $\{L_\alpha\}$ which, together with (2.9), guarantee the convexity of f.

The class of the quasiconvex functions seems to be considerably larger than the class of the convex functions, because it contains, in addition to the convex functions, all functions illustrated in Figure 1, all monoton functions $f: I \to R$, and many other ones. However, there remains an open question, namely, how big the difference is between pseudoconvex and convex functions.

A family of subsets $\{L_\alpha\}$ of A with the range J satisfying $(2.2)-(2.5)$ and (2.8), (2.9), i.e., the family of the level sets of a lower semicontinuous, quasiconvex function f defined on A with the range J, is briefly called a quasiconvex family. Suppose now $\{L_\alpha\}$ is transformable into the family of level sets $L_{\varphi^{-1}(t)}$, $t \in J_\varphi$, of a convex function $\varphi(f(\mathbf{x}))$, $\mathbf{x} \in A$, briefly called a convex family. Then, both f and φf are continuous. The interval J_φ, the image of J by φ, is open to the right, since a convex function in an open domain has no maximum. Hence, J must have the same property. This implies that all sets $L_\alpha = L_{\varphi^{-1}(t)}$ are proper subsets of A. If J_φ is closed to the left, then J is closed to the left.

Let the interior of J be denoted by *int* J. A rather obvious necessary condition which a quasiconvex family $\{L_\alpha\}$ must satisfy in order that it be transformable into a convex family is

$$\overline{\bigcup_{\beta<\alpha} L_\beta} = L_\alpha, \qquad \forall \alpha \in intJ. \tag{2.10}$$

This property expresses that a convex function cannot assume a constant value except, possibly, its minimum on a relatively open subset of its domain. This condition will not, however, be used explicitly.

The further discussion of the problem stated above will be based on the following characterization of a convex family:

Lemma 2.1. *A quasiconvex family* L_α, $\alpha \in J$, *is a convex family iff*

$$(1-\theta)L_{\alpha_0} + \theta L_{\alpha_1} \subseteq L_{\alpha_\theta}, \tag{2.11}$$

where $0 \leq \theta \leq 1$, $\alpha_0 \in J$, $\alpha_1 \in J$, $\alpha_\theta = (1-\theta)\alpha_0 + \theta\alpha_1$.

Proof. Suppose that L_α, $\alpha \in J$, are the level sets of the convex function $f(\mathbf{x})$, $\mathbf{x} \in A$. Let $\mathbf{x}_\theta = (1-\theta)\mathbf{x}_0 + \theta\mathbf{x}_1$, where $\mathbf{x}_0 \in L_{\alpha_0}$, $\mathbf{x}_1 \in L_{\alpha_1}$, be an arbitrary point of $(1-\theta)L_{\alpha_0} + \theta L_{\alpha_1}$. Then,

$$f(\mathbf{x}_\theta) \leq (1-\theta)f(\mathbf{x}_0) + \theta f(\mathbf{x}_1) \leq (1-\theta)\alpha_0 + \theta\alpha_1 = \alpha_\theta.$$

Hence, $\mathbf{x}_\theta \in L_\theta$.

Conversely, let (2.11) be satisfied and define $f(\mathbf{x}) = \mathbf{x} \in L_\alpha \min \alpha$, $\alpha \in J$. As mentioned above, this function has the level sets L_α, $\alpha \in J$. Let $\mathbf{x}_0$ and $\mathbf{x}_1$ be arbitrary points of A, and put $f(\mathbf{x}_0) = \alpha_0$, $f(\mathbf{x}_1) = \alpha_1$, and $\mathbf{x}_\theta = (1-\theta)\mathbf{x}_0 + \theta\mathbf{x}_1$. Then, $\mathbf{x}_0 \in L_{\alpha_0}$, $\mathbf{x}_1 \in L_{\alpha_1}$, and $\mathbf{x}_\theta \in L_{\alpha_\theta}$ because of (2.11). Hence,

$$f(\mathbf{x}_\theta) = \mathbf{x}_\theta \in L_\alpha \min\alpha \leq \alpha_\theta = (1-\theta)f(\mathbf{x}_0) + \theta f(\mathbf{x}_1),$$

which proves the statement.

Let A be a point set. The cone with vertex at the origin consisting of all directions in which A is bounded will be denoted by $B(A)$. The following, rather obvious properties of cones B will be used:

for any two point sets A_1, A_2

$$\begin{aligned} B(A_1) &\supseteq B(A_2) \qquad \text{if} \quad A_1 \subseteq A_2, \\ B(\lambda A) &= B(A) \qquad \text{for} \quad \lambda > 0, \\ B(A_1 + A_2) &= B(A_1) \cap B(A_2). \end{aligned} \tag{2.12}$$

Lemma 2.2. *For a quasiconvex family* L_α, $\alpha \in J$, *transformable into a convex family, all sets* L_α, $\alpha \in J$, *are bounded in the same directions, i.e.,* $B = B(L_\alpha)$, $\alpha \in intJ$, *is independent of* α. *If* J *is closed to the left, then*

$$B \subseteq B\left(L_{\min\limits_{\alpha \in J} \alpha}\right) \subseteq \overline{B}. \tag{2.13}$$

Proof. Since this statement is invariant under index transformations, it is sufficient to prove it for a family $L_{\varphi^{-1}(t)}$, $t \in J_\varphi$, satisfying (2.11). Let $t \in intJ_\varphi$, $t_1 \in intJ_\varphi$, $t_1 > t$, be given and choose $t_0 < t$ in J_φ. With $\theta = (t - t_0)/(t_1 - t_0)$, relation (2.11) yields

$$(1-\theta) L_{\varphi^{-1}(t_0)} + \theta L_{\varphi^{-1}(t_1)} \subseteq L_{\varphi^{-1}}(t). \tag{2.14}$$

Hence, because $L_{\varphi^{-1}(t_0)} \subseteq L_{\varphi^{-1}(t)} \subseteq L_{\varphi^{-1}(t_1)}$,

$$B\left(L_{\varphi^{-1}(t_1)}\right) \subseteq B\left(L_{\varphi^{-1}(t)}\right) \subseteq \left[B\left(L_{\varphi^{-1}(t_0)}\right) \cap B\left(L_{\varphi^{-1}(t_1)}\right)\right] = B\left(L_{\varphi^{-1}(t_1)}\right). \tag{2.15}$$

Thus, $B\left(L_{\varphi^{-1}(t)}\right) = B\left(L_{\varphi^{-1}(t_1)}\right)$, which proves the first part of the statement.

If J is closed to the left, then $B \subseteq B\left(L_{\min\limits_{\alpha \in J} \alpha}\right)$, because $L_{\min\limits_{\alpha \in J} \alpha} \subseteq L_\alpha$, $\alpha \in intJ$. It only remains to prove that $B\left(L_{\min\limits_{\alpha \in J} \alpha}\right) \subseteq \overline{B}$.

Let $\eta \neq 0$, $\|\eta\| = 1$, be in $B(L_\alpha)$ and let H be the supporting hyperplane of L_α with the normal direction η. In L_α, there is some point $\hat{\mathbf{x}}$ whose distance from H is less than a given $\varepsilon > 0$. Let H_ε denote the hyperplane parallel to H at distance ε which is separated from $\hat{\mathbf{x}}$ by H. In H_ε, consider the $(n-1)$-dimensional closed (solid) unit sphere whose center is the orthogonal projection of $\hat{x}$ on H_ε. Because the compact unit sphere has

a positive distance from L_α, by (2.4), there is some $L_{\varphi^{-1}(t)}$ such that $L_{\varphi^{-1}(t)}$ and the unit sphere are disjoint. By a separation argument, there is a hyperplane H' separating $L_{\varphi^{-1}(t)}$ and the unit sphere. The normal vector η', $\|\eta'\|=1$, of H' which is directed towards H_ε belongs to B because $L_{\varphi^{-1}(t)}$ is bounded in this direction. The tangent of the angle formed by η and η is less than 2ε, since H separates $\hat{\mathbf{x}}$ from the unit sphere. Hence, the ray η is a limit ray of rays $\eta' \in B$, which proves that $B(L_\alpha) \subseteq \overline{B}$. □

Let us introduce the assumption

all sets L_α, $\alpha \in J$, *are bounded in the same directions*. (2.16)

The recession cone of the set L_α for every $\alpha \in J$ is the set of vectors $\xi \in R^n$ such that $L_\alpha + \xi \subseteq L_\alpha$.

Corollary 2.1. *All level sets of a convex function have the same recession cone.*

The function $h(L_\alpha, \eta): R^n \to R \cup \{+\infty\}$ defined for every $\alpha \in J$ by

$$h(L_\alpha, \eta) = \mathbf{x} \in L_\alpha \sup \eta^T \mathbf{x}, \qquad \eta \in R^n,$$

is called the support function of L_α for every $\alpha \in J$ and its effective domain

$$\{\eta \in R^n \mid h(L_\alpha, \eta) < +\infty\}$$

is called the barrier cone of L_α for every $\alpha \in J$ The support function of L_α, $\alpha \in J$, describes all the closed half-spaces which contain L_α, $\alpha \in J$. From Lemma 2.2, it follows that $h(L_\alpha, \eta)$, $\alpha \in J$, with a finite value is defined over the cone B and nowhere else.

Lemma 2.3. *Let* L_α, $\alpha \in J$, *be the family of the level sets of a lower semicontinuous, quasiconvex function* $f: A \to R$ *such that the cone* $B(L_\alpha) = B$ *is independent of* α *for* $\alpha \in intJ$. *Let* $h(L_\alpha, \eta)$, $\eta \in B$, $\|\eta\| = 1$, *be the support function of* L_α, $\alpha \in J$,. *Further, let* $t = \varphi(\alpha)$, $\alpha \in J$, *be a strictly increasing continuous function and* $\alpha = \varphi^{-1}(t)$, $t \in J_\varphi$, *its inverse. Then,* $\varphi(f(\mathbf{x}))$, $\mathbf{x} \in A$, *is convex if* $f\ h\left(L_{\varphi^{-1}}(t), \eta\right)$ *is a concave function of* $t \in J_\varphi$ *for every fixed* $\eta \in B$.

Proof. The family of level sets L_α, $\alpha \in J$, satisfies conditions (2.2)–(2.5) and (2.8), (2.9). By (2.3), $h(L_\alpha, \eta)$ is an increasing function of $\alpha \in J$.

I. Suppose that there is a strictly increasing continuous function $t = \varphi(\alpha)$, $\alpha \in J$, such that $\varphi(f(\mathbf{x}))$, $\mathbf{x} \in A$, is convex. Then, the sets $L_{\varphi^{-1}(t)}$, $t \in J_\varphi$, satisfy (2.11), and hence, by the properties of support functions, we obtain that

$$h\left(L_{\varphi^{-1}(t_\theta)}, \eta\right) \geq (1-\theta)\, h\left(L_{\varphi^{-1}(t_0)}, \eta\right) + \theta\, h\left(L_{\varphi^{-1}(t_1)}, \eta\right), \tag{2.17}$$

where $t_\theta = (1-\theta)\, t_0 + \theta t_1$, which means that $h\left(L_{\varphi^{-1}(t)}, \eta\right)$, $t \in J_\varphi$, is a concave function of t for every fixed $\eta \in B$.

II. Conversely, suppose there exists a strictly increasing continuous function $t = \varphi(\alpha)$, $\alpha \in J$, such that for a family L_α, $\alpha \in J$, the function $h\left(L_{\varphi^{-1}(t)}, \eta\right)$, $t \in J_\varphi$, is a concave function of t for every fixed $\eta \in B$. It follows from this hypothesis that $\varphi(f(\mathbf{x}))$, $\mathbf{x} \in A$, is a convex function. To prove this it is sufficient to show (2.11). Now, (2.17) is valid, and for two point sets A_1 and A_2,, $h(A_1, \eta) \leq h(A_2, \eta)$ implies $\overline{A_1} \subseteq \overline{A_2}$. Hence,

$$\overline{L_{\varphi^{-1}(t_\theta)}} \supseteq \overline{(1-\theta)\, L_{\varphi^{-1}(t_0)} + \theta L_{\varphi^{-1}(t_1)}} \supseteq (1-\theta)\, L_{\varphi^{-1}(t_0)} + \theta L_{\varphi^{-1}(t_1)}. \tag{2.18}$$

Condition (2.8) implies that

$$\overline{L_{\varphi^{-1}(t)}} \cap A = L_{\varphi^{-1}(t)}, \qquad t \in J_\varphi,$$

thus,

$$L_{\varphi^{-1}(t_\theta)} \supseteq A \cap \left[(1-\theta)\, L_{\varphi^{-1}(t_0)} + \theta L_{\varphi^{-1}(t_1)}\right] = (1-\theta)\, L_{\varphi^{-1}(t_0)} + \theta L_{\varphi^{-1}(t_1)}. \tag{2.19}$$

The latter equality follows from the inclusions $L_{\varphi^{-1}(t_0)} \subseteq A$, $L_{\varphi^{-1}(t_1)} \subseteq A$, and the convexity of A, which completes the proof of the statement. □

In terms of slopes, the concavity of $h\left(L_{\varphi^{-1}(t)}, \eta\right)$, $t \in J_\varphi$, for every fixed $\eta \in \beta$ is equivalent to

$$\frac{h\left(L_{\varphi^{-1}(t_2)},\eta\right)-h\left(L_{\varphi^{-1}(t_1)},\eta\right)}{t_2-t_1}\geq\frac{h\left(L_{\varphi^{-1}(t_3)},\eta\right)-h\left(L_{\varphi^{-1}(t_2)},\eta\right)}{t_3-t_2},\qquad(2.20)$$
$$\forall t_1<t_2<t_3,\qquad t_1,t_2,t_3\in J_\varphi.$$

This condition may be given in the different form of

$$\frac{\varphi(\alpha_3)-\varphi(\alpha_2)}{\varphi(\alpha_2)-\varphi(\alpha_1)}\geq\frac{h\left(L_{\alpha_3},\eta\right)-h\left(L_{\alpha_2},\eta\right)}{h\left(L_{\alpha_2},\eta\right)-h\left(L_{\alpha_1},\eta\right)},\quad\forall\alpha_1<\alpha_2<\alpha_3,\quad\alpha_1,\alpha_2,\alpha_3\in J,$$

where the right-hand side is interpreted as 0 whenever the denominator vanishes. Let

$$\psi(\alpha_1,\alpha_2,\alpha_3)=\sup_{\eta\in B}\frac{h\left(L_{\alpha_3},\eta\right)-h\left(L_{\alpha_2},\eta\right)}{h\left(L_{\alpha_2},\eta\right)-h\left(L_{\alpha_1},\eta\right)},\quad\forall\alpha_1<\alpha_2<\alpha_3,\quad\alpha_1,\alpha_2,\alpha_3\in J.\qquad(2.21)$$

The function $\psi:J^3\to R$ only depends on the family L_α, $\alpha\in J$, thus it is used for stating the necessary condition as follows:

There is a strictly increasing continuous function $\varphi(\alpha)$, $\alpha\in J$, *such that*

$$\varphi(\alpha_3)-\varphi(\alpha_2)\geq\left(\varphi(\alpha_2)-\varphi(\alpha_1)\right)\psi(\alpha_1,\alpha_2,\alpha_3),\qquad(2.22)$$
$$\forall\alpha_1<\alpha_2<\alpha_3,\qquad\alpha_1,\alpha_2,\alpha_3\in J.$$

From the preceding lemmas and reasoning, the following statement follows:

Fenchel theorem (1951). *A family of subsets of an open convex set* $A\subseteq R^n$ *suitably indexed by real numbers forms the family of the level sets of a convex function defined on* A *iff* (2.2)–(2.5), (2.8), (2.9), (2.16), (2.22) *hold.*

Fenchel (1953) remarked that while (2.2)–(2.5), (2.8), (2.9), (2.16) are simple and intuitive, condition (2.22) is rather complicated. He added that there was no simple test to decide whether the function $\psi:J^3\to R$ is such that can admit a strictly increasing continuous solution of the functional inequality (2.22), and both local and global properties of ψ entered decisively. When he compared it with the original problem, there seemed to be no progress, however, condition (2.22) had the advantage of leading to a nice construction of the required function $\varphi:J\to R$.

In the theory of economics, Debreu (1954) proved his famous theorem on the representation of a continuous and complete preference ordering by a utility function. It is obvious that the utility function, whose existence is given by the Debreu theorem, is quasiconcave if the preference ordering is convex. Crouzeix (1977) and Kannai (1977, 1981) studied the problem of the concavifiability of convex preference orderings, i.e., the problem of the existence of a concave function having the same level sets as a given continuous quasiconcave one, and they improved the Fenchel results. This problem can be important in several economic and bargaining situations. The conditions provided for the cases of continuous, differentiable and twice differentiable quasiconcave functions are intimately related to constructions of special (least concave) utility representations. (By Debreu (1976), a utility function is said to be least concave on a convex set if every concave utility function defined on the same set can be represented by a concave transformation of the given utility function.) Crouzeix (1977) and Kannai (1981) introduced auxiliary functions, observing that the concavifiability of a quasiconcave function is essentially a one-dimensional phenomenon, and that if the convex preference ordering is concavifiable, then a suitably constructed auxiliary quasiconcave utility function has to possess finite and non-vanishing one-sided directional derivatives.

An unusual feature of concavifiability theory, as presented in Kannai (1977), was the use of Perron's integral in expressing concavifiability in terms of second-order (one-point) conditions involving a twice differentiable quasiconcave utility function. It turns out that in case a function like this exists at all, the auxiliary function is also twice differentiable, and the associated function, whose Perron integrability is equivalent to concavifiability, has a constant sign, hence Perron integrability is equivalent (in the term of auxiliary functions) to Lebesque integrability (Crouzeix, 1977 and Kannai, 1981).

3. THE SMOOTH CASE

In the smooth case, the original problem is divided into two parts. The first one is to give conditions for the existence of a smooth pseudoconvex function with the prescribed level sets, while the second one is to characterize the smooth convex image transformable functions.

Existence of a smooth pseudoconvex function with the prescribed level sets

By formula (2.4), the existence of a smooth pseudoconvex function with the prescribed level sets can be studied, subject to the prescribed equality level sets, thus in the smooth case, differential geometric tools can be applied. Based on this idea, Rapcsák (1991) gave an explicit formulation of the gradient of the class of the smooth pseudolinear functions, which results in the solution of the first part of the Fenchel problem in the case of a nested family of convex sets whose boundaries are of hyperplanes defining an open convex set. This result was generalized by Rapcsák (1997) for the case where the boundaries of the nested family of convex sets in R^{n+1} are given by n-dimensional differentiable manifolds of class C^3 and the convex sets determine an open or closed convex set in R^{n+1}. Here, the first results is recalled.

Theorem (Rapcsák, 1991) *Let a C^3 function f be defined on an open convex set $A \subseteq R^n$ and assume that $\nabla f \neq 0$ on A. Then, f is pseudolinear on A if f there exist C^2 functions $l(\mathbf{x})$, $\eta_i(f(\mathbf{x}))$, $i = 1,\ldots,n$, $\mathbf{x} \in A$, such that*

$$\partial f(\mathbf{x}) \partial x_i = l(\mathbf{x}) \eta_i(f(\mathbf{x})), \qquad i = 1,\ldots,n, \qquad \mathbf{x} \in A. \tag{3.1}$$

Komlósi (1993) proved the statement without the C^2 property of the functions $l, \eta_i(f), i = 1,\ldots,n$, on the set A under the continuous differentiability of the function f. Pseudolinear or pseudoaffine maps were characterized in the n-dimensional Euclidean space by Bianchi et al. (2000), and the general form of the pseudolinear maps defined on the whole n-dimensional Euclidean space was represented by Bianchi et al. (2003).

Characterization of the smooth convex image transformable functions

In Fenchel (1951, 1956), the problem discussed in the preceding section was studied and solved under the additional assumption that the prescribed subsets of A are the level sets of a twice continuously differentiable function $f(\mathbf{x})$, $\mathbf{x} \in A$. The second part of the Fenchel problem of level sets in the smooth case is to find necessary and sufficient conditions for the convex image transformability of a twice continuously differentiable function f over the open convex set $A \subseteq R^n$, i.e., the problem is the existence of a twice continuously differentiable strictly increasing function

$\varphi(\alpha)$, $\alpha \in J$, such that the function $\varphi(f(\mathbf{x}))$, $\mathbf{x} \in A$, is convex. A first complete set of necessary and sufficient conditions for the convexifiability of C^2 functions was derived by Fenchel (1951, 1956). Following Fenchel and Avriel et al. (1988), it can be formulated in two steps via a hierarchy of four conditions. In the first step, two local conditions are discussed that must be satisfied at every point of A. The second step consists of global conditions that must hold, taking the entire domain $A \subseteq R^n$ into account.

Definition 3.1 *A nonconvex function is convex image transformable if it can be transformed into a convex function by a one-to-one increasing transformation of its image.*

Let $A \subseteq R^n$ be an open convex set and the augmented Hessian matrix of the function $f \in C^2(A, R)$ be given by

$$Hf(\mathbf{x}; r) = Hf(\mathbf{x}) + r\nabla f(\mathbf{x})^T \nabla f(\mathbf{x}), \qquad \mathbf{x} \in A, \qquad r \in R. \tag{3.2}$$

Definition 3.2 *Let H, H_c and H_{IL} be the family of C^2 functions for which a positive semidefinite augmented Hessian matrix with a function $\Psi : A \to R$, a continuous function $\Psi : A \to R$ and a locally Lipschitz function $\Psi : A \to R$ exists at every $x \in A$, respectively.*

The family H of C^2 functions was introduced by Fenchel (1951) as a necessary condition for the convexifiability of f on a convex set A. In his original work, this condition consists of two properties the characterization of which can be found in Avriel et al. (1988). The family of the functions H_c was introduced by Avriel and Schaible (1978) and characterized by Schaible and Zhang (1980), see, Avriel et al. (1988). The family of the functions H_{IL} was introduced as a new pseudoconvex subclass originated from analytical mechanics and characterized by Rapcsák (2003).
Alternatively, if the function ρ_0 defined by

$$\rho_0(\mathbf{x}) = \inf \left\{ \mathbf{y}^{\mathrm{T}} Hf(\mathbf{x})\mathbf{y} / (\nabla f(\mathbf{x})\,\mathbf{y})^2 \;\middle|\; \|\mathbf{y}\| = 1,\ \nabla f(\mathbf{x})\mathbf{y} \neq 0,\ \mathbf{y} \in R^n \right\} \tag{3.3}$$

satisfies $\rho_0(\mathbf{x}) > -\infty$ for every $\mathbf{x} \in A$, then the matrix function $H(\mathbf{x}; \Psi(\mathbf{x}))$ is positive semidefinite for every function $\Psi : A \to R$ satisfying $\Psi \geq -\rho_0$ on A.

Let us introduce the following conditions:

$$\nabla f(\mathbf{x}) \neq 0 \text{ for all } \mathbf{x} \in A, \text{ except for the global minimum points in } A \text{ if such points exist;} \tag{3.4}$$

f *is pseudoconvex over* A; (3.5)

$f \in H$; (3.6)

the function g *defined on the open interval int* J *by*
$g(t) = \inf\{\rho_0(\mathbf{x}) \mid \mathbf{x} \in A, f(\mathbf{x}) = t\}$ (3.7)
is finite for every $t \in int\ J$;

there exists a differentiable positive function h *on* $t \in int\ J$ *satisfying*
$$\frac{d}{dt} ln(h(t)) \leq g(t), \qquad t \in int\, J. \tag{3.8}$$

Fenchel-Avriel-Diewert-Schaible-Zang theorem (1988). *If* $A \subseteq R^n$ *is an open convex set and* $f \in C^2(A,R)$, *then,* f *is convex image transformable on* A *if* f *conditions* (3.4), (3.6), (3.7), (3.8) *hold.*

A new geometric necessary and sufficient condition was obtained for the existence of a smooth convex function with the level sets of a given smooth pseudoconvex function by Rapcsák (2003), which is a new solution for the second part of the Fenchel problem of level sets in the smooth case. This approach provides ageometric characterization of the new subclass of pseudoconvex functions H_{IL} originated from analytical mechanics, an extension of the local-global property of nonlinear optimization to nonconvex open sets, and a new view on the convexlike and generalized convexlike mappings in the image analysis (see, e.g., Giannessi, 1984; Mastroeni et al., 2000).

The main statements are as follows:

Theorem 3.1. *Let* $A \subseteq R^n$ *be an open convex set,* $f \in C^2(A,R)$ *and* $\psi : A \to R$ *a locally Lipschitz function. Then,* $f \in H_{IL}$ *if* f *for every* $\mathbf{x} \in A$ *there exists a convex neighbourhood* $U(\mathbf{x}) \subseteq A$ *such that for every pair* $(\mathbf{x}, \mathbf{y} = \mathbf{z} - \mathbf{x})$, $\mathbf{z} \in A$, *the single variable function*

$$f\left(\mathbf{x} + \varphi_{(\mathbf{x},\mathbf{y})}(t)\,\mathbf{y}\right), \qquad \mathbf{x} + \varphi_{(\mathbf{x},\mathbf{y})}(t)\,\mathbf{y} \in U(\mathbf{x}), \qquad t \in [0,1], \tag{3.9}$$

is convex where $\varphi_{(\mathbf{x},\mathbf{y})} : [0,1] \to R$, $\varphi_{(\mathbf{x},\mathbf{y})}(0) = 0$, $\varphi_{(\mathbf{x},\mathbf{y})}(0) = 1$, *is a strictly increasing function given by the following differential equation*:

$$-\left(\frac{1}{\varphi_{(\mathbf{x},\mathbf{y})}(t)}\right) = \psi\left(\mathbf{x}+\varphi_{(\mathbf{x},\mathbf{y})}(t)\,\mathbf{y}\right)\nabla f\left(\mathbf{x}+\varphi_{(\mathbf{x},\mathbf{y})}(t)\,\mathbf{y}\right)\mathbf{y}, \qquad t\in[0,1].$$

Moreover, if $\psi : A \to R_+$, *and*

$$\nabla f(\mathbf{x})\,\mathbf{y} > 0, \tag{3.10}$$

then, $\varphi_{(\mathbf{x},\mathbf{y})}$ *is strictly convex.*

Theorem 3.2 *Let* $f \in H_{iL}$ *be a real-valued function defined on an open convex set* $A \subseteq R^n$. *Then,* f *is convex image transformable by a one-to-one increasing function* $\phi \in C^2\left(Im_f(A), R\right)$ *iff for every* $\mathbf{x} \in A$, *there exists a convex neigbourhood* $U(\mathbf{x}) \subseteq A$ *such that for every pair* $(\mathbf{x}, \mathbf{y} = \mathbf{z} - \mathbf{x})$, $\mathbf{z} \in A$, *the single variable function*

$$f\left(\mathbf{x}+\varphi_{(\mathbf{x},\mathbf{y})}(t)\,\mathbf{y}\right), \qquad \mathbf{x}+\varphi_{(\mathbf{x},\mathbf{y})}(t)\,\mathbf{y}\in U(\mathbf{x}), \qquad t\in[0,1], \tag{3.11}$$

is convex where $\varphi_{(\mathbf{x},\mathbf{y})} : [0,1] \to R$, $\varphi_{(\mathbf{x},\mathbf{y})}(0) = 0$, $\varphi'_{(\mathbf{x},\mathbf{y})}(0) = 1$, *is a strictly increasing function given by the following differential equation:*

$$\varphi'_{(\mathbf{x},\mathbf{y})}(t) = \frac{1}{\phi'(f(\mathbf{x}))}\phi'\left(f\left(\mathbf{x}+\varphi_{(\mathbf{x},\mathbf{y})}(t)\,\mathbf{y}\right)\right), \qquad t\in[0,1]. \tag{3.12}$$

Moreover, if $\phi : A \to R_+$, *and*

$$\nabla f(\mathbf{x})\mathbf{y} > 0, \tag{3.13}$$

then, $\varphi_{(\mathbf{x},\mathbf{y})}$ *is strictly convex.*

4. CONCLUDING REMARKS

In the paper, a survey is given on some results of the Fenchel problem of level sets. It is emphasized that this problem is an important structural question of convex analysis which is related to economics and analytical mechanics. Some open questions are as follows:

1. How large is the difference between pseudoconvex and convex functions?

2. How to solve the Fenchel problem of level sets in the case of a nor open neither closed convex set?
3. How to solve the Fenchel problem of level sets in the case of C^1 functions?
4. Whether the necessary and sufficient conditions obtained in the smooth case can be preserved under milder conditions?

REFERENCES

[1] Avriel, M., Diewert, W.E., Schaible, S. and Zang, I., Generalized concavity, Plenum Press, 1988.

[2] Bianchi, M. and Schaible, S., An extension of pseudolinear functions and variational inequality problems, Journal of Optimization Theory and Applications 104 (2000) 59-71.

[3] Bianchi, M., Hadjisavvas, N. and Schaible, S., On pseudomonotone maps T for which -T is also pseudomonotone, Journal of Convex Analysis 10 (2003) 149–168.

[4] Crouzeix, J.P, Contributions à l'étude des fonctions quasiconvexes, Thèse, Université de Clermont-Ferrand, 1977.

[5] Debreu, G., Representation of a preference ordering by a numerical function, in: Decision processes, Thrall, Coombs and Davis (eds.), John-Wiley and Sons, 1954.

[6] Debreu, G., Least concave utility functions, Journal of Mathematical Economics 3 (1976) 121-129.

[7] De Finetti, B., Sulle stratificazioni convesse, Annali di Matematica Pura ed Applicata 30 (1949) 173-183.

[8] Fenchel, W., Convex cones, sets and functions, Mimeographed lecture notes, Princeton University Press, Princeton, New Jersey, 1951.german

[9] Fenchel, W., Über konvexe Funktionen mit vorgeschriebenen Niveaumannigfaltigkeiten, Mathematische Zeitschrift 63 (1956) 496-506.english

[10] Giannessi, F., Theorems of the alternative and optimality conditions, Journal of Optimization Theory and Applications 42 (1984) 331-365.

[11] Kannai, Y, Concavifiability and constructions of concave utility functions, Journal of Mathematical Economics 4 (1977) 1-56.

[12] Kannai, Y., Concave utility functions - existence, constructions and cardinality, in: Generalized concavity in optimization and economics, (eds.): S. Schaible and W.T. Ziemba, Academic Press, New York (1981) 543-611.

[13] Komlósi, S., First and second order characterizations of pseudolinear functions, European Journal of Operational Research 67 (1993) 278-286.

[14] Mastroeni, G. and Rapcsák, T., On convex generalized systems, Journal of Optimization Theory and Applications 3 (2000) 605-627.

[15] Rapcsák, T., On pseudolinear functions, European Journal of Operational Research 50 (1991) 353-360.

[16] Rapcsák, T., An unsolved problem of Fenchel, Journal of Global Optimization 11 (1997) 207-217.

[17] Rapcsák, T., Smooth nonlinear optimization in R^n, Kluwer Academic Publishers, 1997.

[18] Rapcsák, T., Geometry of paths and the Fenchel problem of level sets, LORDS WP 2003-9.

[19] Roberts, A. W. and Varberg, D. E., Convex functions, Academic Press, New York, London, 1973.

INTEGRAL FUNCTIONALS ON SOBOLEV SPACES HAVING MULTIPLE LOCAL MINIMA

B. Ricceri
Dept. of Mathematics, University of Catania, Catania, Italy

If (X,τ) is a topological space, for any $\Psi : X \to]-\infty,+\infty]$, I denote by τ_Ψ the smallest topology on X which contains both τ and the family of sets $\{\Psi^{-1}(]-\infty,r[)\}_{r\in\mathbb{R}}$.

In [2], I have established the following general result:

Theorem A. *Let (X,τ) be a Hausdorff topological space and $\Psi : X \to]-\infty,+\infty]$, $\Phi : X \to \mathbb{R}$ two functions. Assume that there is $r > \inf_X \Psi$ such that the set $\Psi^{-1}(]-\infty,r])$ is compact and first-countable. Moreover, suppose that the function Φ is bounded below in $\Psi^{-1}(]-\infty,r])$ and that the function $\Psi + \lambda\Phi$ is sequentially lower semicontinuous for each $\lambda \geq 0$ small enough. Finally, assume that the set of all global minima of Ψ has at least k connected components.*

Then, there exists $\lambda^* > 0$ such that, for each $\lambda \in]0,\lambda^*[$, the function $\Psi + \lambda\Phi$ has at least k τ_Ψ-local minima lying in $\Psi^{-1}(]-\infty,r[)$.

In the context of a systematic series of applications of Theorem A, I intend to present here two multiplicity results about local minima of integrals of the calculus of variations.

In the sequel, Ω will denote a bounded, open and connected subset of $\mathbb{R}^n$ with sufficiently smooth boundary.

Recall that a function $f : \Omega \times \mathbb{R}^m \to \mathbb{R}$ is said to be sup-measurable if, for each measurable function $u : \Omega \to \mathbb{R}^m$, the composite function $x \to f(x,u(x))$ is measurable. Following [3], f is said to be a normal integrand if it is $\mathcal{L}(\Omega) \otimes \mathcal{B}(\mathbb{R}^m)$-measurable and $f(x,\cdot)$ is lower

semicontinuous for a.e. $x \in \Omega$. Here $\mathcal{L}(\Omega)$ and $\mathcal{B}(\mathbb{R}^m)$ denote the Lebesgue and the Borel σ-algebras of subsets of Ω and $\mathbb{R}^m$, respectively. Also, f is said to be a Carathéodory function if $f(x,\cdot)$ is continuous for a.e. $x \in \Omega$ and $f(\cdot,y)$ is measurable for every $y \in \mathbb{R}^m$. Note that any Carathéodory function is a normal integrand and that any normal integrand is sup-measurable ([3], pp. 174-175).

The aim of the present paper is to establish the following two results, in the conclusions of which the space $W^{1,p}(\Omega)$ is considered with the topology induced by the usual norm $\|u\|_{W^{1,p}(\Omega)} = \left(\int_\Omega (|\nabla u(x)|^p + |u(x)|^p)dx\right)^{\frac{1}{p}}$:

Theorem 1. *Let $1 < p < n$. Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ and $\psi : \mathbb{R}^n \to [0,+\infty[$ be two functions, with $\psi(0) = 0$ and $\psi(\eta) > 0$ for all $\eta \in \mathbb{R}^n \setminus \{0\}$, such that, for every $\lambda \geq 0$ small enough, the function $\psi + \lambda\varphi$ is convex in $\mathbb{R}^n$. Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function such that the set $g^{-1}(\inf_{\mathbb{R}} g)$ has at least k connected components. Furthermore, let $\beta : \Omega \times \mathbb{R} \to \mathbb{R}$ be a normal integrand. Assume that there are $c > 1$, $q \in [p, \frac{pn}{n-p}[$ and $\delta \in L^1(\Omega)$ such that, for a.e. $x \in \Omega$ and for every $(\xi,\eta) \in \mathbb{R} \times \mathbb{R}^n$, one has*

$$\frac{1}{c}(|\eta|^p + |\xi|^p - c^2) \leq \psi(\eta) + g(\xi) \leq c(|\eta|^p + |\xi|^{\frac{pn}{n-p}} + 1) \tag{1}$$

and

$$-c(|\eta|^p + |\xi|^q + \delta(x)) \leq \varphi(\eta) + \beta(x,\xi) \leq c(|\eta|^p + |\xi|^{\frac{pn}{n-p}} + \delta(x)). \tag{2}$$

Then, for every $\alpha \in L^\infty(\Omega)$, with $\operatorname{ess\,inf}_\Omega \alpha > 0$, for every sequentially weakly closed set $X \subseteq W^{1,p}(\Omega)$ containing all the constant functions and for every $r > \inf_{\mathbb{R}} g \|\alpha\|_{L^1(\Omega)}$, there exists $\lambda^* > 0$ such that, for each $\lambda \in]0,\lambda^*[$, the restriction to X of the functional

$$u \to \int_\Omega (\psi(\nabla u(x)) + \alpha(x) g(u(x)))dx + \lambda \int_\Omega (\varphi(\nabla u(x)) + \beta(x,u(x)))dx$$

has at least k local minima lying in the set

$$\left\{u \in X : \int_\Omega (\psi(\nabla u(x)) + \alpha(x) g(u(x)))dx < r\right\}.$$

Theorem 2. *Let $2 \le n < p$. Let X be the space of all $u \in W^{1,p}(\Omega)$ which are harmonic in Ω. Let $\psi : \Omega \times \mathbb{R} \times \mathbb{R}^n \to [0, +\infty[$ be a Carathéodory function, with $\psi(x,\xi,0) = 0$ and $\psi(x,\xi,\eta) > 0$ for a.e. $x \in \Omega$ and for every $(\xi,\eta) \in \mathbb{R} \times (\mathbb{R}^n \setminus \{0\})$. Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function such that the set $g^{-1}(\inf_{\mathbb{R}} g)$ has at least k connected components. Furthermore, let $\beta : \Omega \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ be a normal integrand. Assume that, for some $c > 1$, one has*

$$\frac{1}{c}(|\eta|^p + |\xi|^p - c^2) \le \min\{\psi(x,\xi,\eta), g(\xi)\} \tag{3}$$

for a.e. $x \in \Omega$ and for every $(\xi,\eta) \in \mathbb{R} \times \mathbb{R}^n$, and that, for each $s > 0$, there exist $c_s > 0$ and $M_s \in L^1(\Omega)$, such that

$$\begin{aligned} -M_s(x) &\le \min\{\psi(x,\xi,\eta), \beta(x,\xi,\eta)\} \le \max\{\psi(x,\xi,\eta), \beta(x,\xi,\eta)\} \\ &\le M_s(x) + c_s |\eta|^p \end{aligned} \tag{4}$$

for a. e. $x \in \Omega$, for every $\eta \in \mathbb{R}^n$ and for every $\xi \in \mathbb{R}$ satisfying $|\xi| \le s$.

Then, for every $\alpha \in L^1(\Omega)$, with $\operatorname{ess\,inf}_{\Omega} \alpha > 0$, and for every $r > \inf_{\mathbb{R}} g \|\alpha\|_{L^1(\Omega)}$, there exists $\lambda^* > 0$ such that, for each $\lambda \in]0, \lambda^*[$, the functional

$$u \to \int_{\Omega} (\psi(x,u(x),\nabla u(x)) + \alpha(x) g(u(x)))dx + \lambda \int_{\Omega} \beta(x,u(x),\nabla u(x))dx, \; u \in X$$

has at least k local minima lying in the set

$$\left\{u \in X : \int_{\Omega} (\psi(x,u(x),\nabla u(x)) + \alpha(x) g(u(x)))dx < r\right\}.$$

Let us start proving the following

Proposition 1. *Let $\psi : \Omega \times \mathbb{R} \times \mathbb{R}^n \to [0, +\infty[$ be a sup-measurable function with $\psi(x,\xi,0) = 0$ and $\psi(x,\xi,\eta) > 0$ for a.e. $x \in \Omega$ and for every $(\xi,\eta) \in \mathbb{R} \times (\mathbb{R}^n \setminus \{0\})$. Let $g : \mathbb{R} \to \mathbb{R}$ be a Borel function such that the set $g^{-1}(\inf_{\mathbb{R}} g)$ has at least k connected components. Let $\alpha \in L^1(\Omega)$ be a non-negative function. For each $u \in W^{1,1}(\Omega)$, put*

$$\Psi(u) = \int_{\Omega} (\psi(x,u(x),\nabla u(x)) + \alpha(x) g(u(x)))dx.$$

Then, for every set $X \subseteq W^{1,1}(\Omega)$ *which contains the set* R *of all (equivalence classes of) constant functions, one has*

$$\inf_X \Psi = \inf_{\mathbb{R}} g \|\alpha\|_{L^1(\Omega)}$$

and the set

$$\left\{u \in X : \Psi(u) = \inf_{\mathbb{R}} g \|\alpha\|_{L^1(\Omega)}\right\}$$

is contained in R *and has at least* k *connected components in the Euclidean topology of* R.

Proof. For each $u \in X$, we clearly have

$$\Psi(u) \geq \inf_{\mathbb{R}} g \|\alpha\|_{L^1(\Omega)}$$

and that equality holds if u is almost everywhere equal to a constant c such that $g(c) = \inf_{\mathbb{R}} g$. On the other hand, if $u \in W^{1,1}(\Omega)$ is not almost everywhere equal to a constant, then $|\nabla u| > 0$ in some set of positive measure (recall that Ω is connected), and so $\int_\Omega \psi(x, u(x), \nabla u(x))dx > 0$. From this, it clearly follows that

$$\left\{u \in X : \Psi(u) = \inf_{\mathbb{R}} g \|\alpha\|_{L^1(\Omega)}\right\} = \gamma(g^{-1}(\inf_{\mathbb{R}} g))$$

where γ denotes the mapping that to each $r \in \mathbb{R}$ associates the equivalence class of functions almost everywhere equal in Ω to r. If one considers on R the Euclidean topology, the mapping γ is a homeomorphism between $\mathbb{R}$ and R, and from this the conclusion follows. □

Proof of Theorem 1. Fix $a > 0$ such that the function $\psi + \lambda\varphi$ is convex in $\mathbb{R}^n$ (and hence continuous) for all $\lambda \in [0, a[$. This, of course, implies that φ is continuous too. Now, for each $u \in W^{1,p}(\Omega)$, put

$$\Psi(u) = \int_\Omega (\psi(\nabla u(x)) + \alpha(x) g(u(x)))dx$$

and

$$\Phi(u) = \int_\Omega (\varphi(\nabla u(x)) + \beta(x,u(x)))dx.$$

From (1) and (2), since $W^{1,p}(\Omega)$ is continuously embedded in $L^{\frac{pn}{n-p}}(\Omega)$, it follows that Ψ and Φ are well defined, with finite values, and that Ψ is $\|\cdot\|_{W^{1,p}(\Omega)}$-continuous, since g is continuous. Fix $\lambda \in [0,a[$. We show that the functional $\Psi + \lambda\Phi$ is sequentially weakly lower semicontinuous. Since the functional $u \to \int_\Omega (\psi(\nabla u(x)) + \lambda\varphi(\nabla u(x)))dx$ is weakly lower semicontinuous, being convex and continuous, it is enough to prove that the functional $u \to \int_\Omega (\alpha(x)g(u(x)) + \lambda\beta(x,u(x)))dx$ is sequentially weakly lower semicontinuous. To this end, let $u \in W^{1,p}(\Omega)$ and let $\{u_k\}$ be a any sequence in $W^{1,p}(\Omega)$ weakly converging to u. Since $q < \frac{pn}{n-p}$, by the Rellich-Kondrachov theorem, there is a subsequence $\{u_{k_h}\}$ strongly converging to u in $L^q(\Omega)$. Of course, we may assume that $\lim_{h\to+\infty} u_{k_h}(x) = u(x)$ and that $\sup_{h\in\mathbb{N}} |u_{k_h}(x)|^q \le \omega(x)$ for a.e. $x \in \Omega$, for a suitable $\omega \in L^1(\Omega)$. Clearly, we have

$$\alpha(x)g(u(x)) + \lambda\beta(x,u(x)) \le \liminf_{h\to+\infty} (\alpha(x)g(u_{k_h}(x)) + \lambda\beta(x,u_{k_h}(x)))$$

for a.e. $x \in \Omega$. On the other hand, by (1) and (2), there is a suitable $b > 0$ such that

$$-b(\omega(x) + \delta(x) + 1) \le \alpha(x)g(u_{k_h}(x)) + \lambda\beta(x,u_{k_h}(x))$$

for a.e. $x \in \Omega$ and for every $h \in \mathbb{N}$. So, by Fatou's lemma, we get

$$\int_\Omega (\alpha(x)g(u(x)) + \lambda\beta(x,u(x)))dx \le \int_\Omega \liminf_{h\to+\infty} (\alpha(x)g(u_{k_h}(x)) + \lambda\beta(x,u_{k_h}(x)))dx$$
$$\le \liminf_{h\to+\infty} \int_\Omega (\alpha(x)g(u_{k_h}(x)) + \lambda\beta(x,u_{k_h}(x)))dx,$$

as desired. Again by (1) and (2), there is a suitable $\theta > 0$ such that, for every $u \in W^{1,p}(\Omega)$, one has

$$\Psi(u) \ge \frac{\min\{1, \operatorname{ess\,inf}_\Omega \alpha\}}{c} \|u\|^p_{W^{1,p}(\Omega)} - \theta$$

and

$$|\Phi(u)| \leq \theta\left(\int_{\Omega}(|\nabla u(x)|^{p} + |u(x)|^{\frac{pn}{n-p}})dx + 1\right).$$

So, Φ is bounded in each bounded subset of $W^{1,p}(\Omega)$, and the set $\{u \in X : \Psi(u) \leq r\}$ is weakly compact and metrizable, being a bounded and sequentially weakly closed subset of the reflexive and separable space $W^{1,p}(\Omega)$. Finally, by Proposition 1, $r > \inf_X \Psi$ and the set of all global minima of the functional $\Psi_{|X}$ has at least k connected components in the weak topology, since the relativization of this to R is the Euclidean topology. So, if τ is the relativization to X of the weak topology, we realize that $\Phi_{|X}$ and $\Psi_{|X}$ satisfy all the assumtpions of Theorem A. Therefore, there exists $\lambda^{*} > 0$ such that, for each $\lambda \in]0, \lambda^{*}[$, the functional $\Psi_{|X} + \lambda\Phi_{|X}$ has at least k τ_{Ψ}-local minima lying in $\Psi^{-1}(]-\infty, r[) \cap X$. But, since Ψ is $\|\,.\,\|_{W^{1,p}(\Omega)}$-continuous, the topology τ_{Ψ} is weaker than the relative $\|\,.\,\|_{W^{1,p}(\Omega)}$- topology, and so the above mentioned τ_{Ψ}-local minima of $\Psi_{|X} + \lambda\Phi_{|X}$ are local minima of this functional in the latter topology, as claimed. □

Proof of Theorem 2. For each $u \in X$, put

$$\Psi(u) = \int_{\Omega}(\psi(x, u(x), \nabla u(x)) + \alpha(x)g(u(x)))dx$$

and

$$\Phi(u) = \int_{\Omega}\beta(x, u(x), \nabla u(x)))dx.$$

Since $p > n$, $W^{1,p}(\Omega)$ is compactly embedded in $C^{0}(\overline{\Omega})$. From this and from (4), it follows that Ψ and Φ are well defined, with finite values. We are now going to apply Theorem A taking as τ the topology induced by the norm $\|u\|_{C^{0}(\overline{\Omega})} = \max_{\overline{\Omega}} |u|$. We prove that Ψ and Φ are sequentially lower semicontinuous. We do that for Φ only, the other case being analogous. So, let $u \in X$ and let $\{u_k\}$ be a sequence in X converging to u. By a classical property of harmonic functions ([1], p. 16), the sequence $\{\nabla u_k(x)\}$

converges to $\nabla u(x)$ for all $x \in \Omega$. Moreover, one has $\sup_{k\in\mathbb{N}} \|u_k\|_{C^0(\overline{\Omega})} < +\infty$. Thus, if we apply (4) taking $s = \sup_{k\in\mathbb{N}} \|u_k\|_{C^0(\overline{\Omega})}$, we get

$$-M_s(x) \le \beta(x, u_k(x), \nabla u_k(x))$$

for a.e. $x \in \Omega$ and for every $k \in \mathbb{N}$. Thus, we can apply Fatou's lemma, obtaining

$$\Phi(u) \le \int_\Omega \liminf_{k\to+\infty} \beta(x, u_k(x), \nabla u_k(x)) \le \liminf_{k\to+\infty} \Phi(u_k),$$

as desired. Let us also prove that Ψ is $\|\,.\,\|_{W^{1,p}(\Omega)}$-continuous. So, let $w \in X$ and let $\{w_k\}$ be a sequence in X with $\lim_{k\to+\infty} \|w_k - w\|_{W^{1,p}(\Omega)} = 0$. Hence, $\lim_{k\to+\infty} \|w_k - w\|_{C^0(\overline{\Omega})} = 0$ and there are $\omega \in L^1(\Omega)$ and a subsequence $\{w_{k_h}\}$ such that $\{\nabla w_{k_h}(x)\}$ converges to $\nabla w(x)$ and $\sup_{h\in\mathbb{N}} |w_{k_h}(x)|^p \le \omega(x)$ for a.e. $x \in \Omega$. By continuity, we get

$$\lim_{h\to+\infty} (\psi(x, w_{k_h}(x), \nabla w_{k_h}(x)) + \alpha(x) g(w_{k_h}(x))) = \psi(x, w(x), \nabla w(x)) + \alpha(x) g(w(x))$$

for a.e. $x \in \Omega$. On the other hand, applying (4) with $s = \sup_{h\in\mathbb{N}} \|w_{k_h}\|_{C^0(\overline{\Omega})}$, we get

$$|\psi(x, w_{k_h}(x), \nabla w_{k_h}(x)) + \alpha(x) g(w_{k_h}(x))| \le M_s(x) + c_s \omega(x) + \alpha(x) \sup_{|\xi|\le s} |g(\xi)|$$

for a.e. $x \in \Omega$ and for every $h \in \mathbb{N}$. Hence, we can apply the dominated converge theorem, obtaining $\lim_{h\to+\infty} \Psi(w_{k_h}) = \Psi(w)$, as desired. Now, we prove that $\Psi^{-1}(]-\infty, r])$ is compact. Since we are in a metric setting, this is equivalent to prove that $\Psi^{-1}(]-\infty, r])$ is sequentially compact. Thus, let $\{v_k\}$ be any sequence in $\Psi^{-1}(]-\infty, r])$. By (3), we get a suitable $\nu > 0$ such that

$$\nu \|u\|_{W^{1,p}(\Omega)} - \frac{1}{\nu} \le \Psi(u)$$

for all $u \in X$. So, the sequence $\{v_k\}$ is bounded in $W^{1,p}(\Omega)$. This implies that there is a subsequence $\{v_{k_h}\}$ weakly converging in $W^{1,p}(\Omega)$ to some v. Consequently, by compact embedding, the sequence $\{v_{k_h}\}$ converges strongly to v in $C^0(\overline{\Omega})$. By another classical property of harmonic functions ([1], p. 16), the function v turns out to be harmonic in Ω, and hence $v \in X$. On the other hand, by the lower semicontinuity of Ψ, we have $\Psi(v) \le \liminf_{h \to +\infty} \Psi(v_{k_h}) \le r$, and so $v \in \Psi^{-1}(]-\infty, r])$, as desired. Also, note that Φ is bounded below in $\Psi^{-1}(]-\infty, r])$ as it is lower semicontinuous. Finally, by Proposition 1, $r > \inf_X \Psi$ and the set of all global minima of the functional Ψ has at least k connected components, since the relativization of τ to R is the Euclidean topology. At this point, all the assumptions of Theorem A are satisfied, and hence there exists $\lambda^* > 0$ such that, for each $\lambda \in]0, \lambda^*[$, the functional $\Psi + \lambda\Phi$ has at least k τ_Ψ-local minima lying in $\Psi^{-1}(]-\infty, r[)$. But, since Ψ is $\| . \|_{W^{1,p}(\Omega)}$-continuous, the topology τ_Ψ is weaker than the $\| . \|_{W^{1,p}(\Omega)}$- topology, and so the above mentioned τ_Ψ-local minima of $\Psi + \lambda\Phi$ are local minima of this functional in the latter topology, as claimed. □

Remark. In both Theorems 1 and 2 the key assumption is that the set of all global minima of g has at least k connected components. Knowing simply that this set is infinite is not useful in order to the multiplicity of local minima of the considered functionals. In this connection, for $p > 1$, consider the function g defined by

$$g(\xi) = \begin{cases} |\xi|^p & \text{if } \xi < 0 \\ 0 & \text{if } \xi \in [0,1] \\ (\xi - 1)^p & \text{if } \xi > 1 \end{cases}$$

So, $g^{-1}(\inf_{\mathbf{R}} g) = [0,1]$ and $\lim_{|\xi| \to +\infty} \frac{g(\xi)}{|\xi|^p} > 0$. Nevertheless, for each $\lambda > 0$, the functional

$$u \to \int_\Omega (|\nabla u(x)|^p + g(u(x)))dx + \lambda \int_\Omega |u(x)|^p \, dx$$

is strictly convex, and so its restriction to any convex subset of $W^{1,p}(\Omega)$ has at most one local minimum.

REFERENCES

[1] S. Axler, P. Bourdon and W. Ramey, *Harmonic Function Theory*, Springer, 2001.

[2] B. Ricceri, *Sublevel sets and global minima of coercive functionals and local minima of their perturbations*, J. Nonlinear Convex Anal., to appear.

[3] R.T. Rockafellar, *Integral functionals, normal integrands and measurable selections*, in *Nonlinear Operators and the Calculus of Variations*, 157-207, Lecture Notes in Math., vol. 543, Springer, 1976.

ASPECTS OF THE PROJECTOR ON PROX-REGULAR SETS[1]

S.M. Robinson
Dept. of Industrial Engineering, University of Wisconsin–Madison, Madison, USA

Abstract: This paper deals with results about projectors on the important class of closed, *prox-regular* sets in $\mathbb{R}^n$. These sets, which include all closed convex sets but also many nonconvex sets, have the property that their associated projection mappings are very well behaved, being locally single-valued and continuous among other good properties. We give elementary proofs of these properties of the projector, and for the case in which the projection is made onto a perturbed set we show that under suitable conditions the projector is jointly continuous in the perturbation variable and the variable expressing the point that is projected. We briefly describe an application to the extension of a normal-map construction from variational inequalities posed over polyhedral convex sets to variational conditions posed over sets that satisfy prox-regularity.

Key words: Projector, variational condition, variational inequality, closest-point mapping, prox-regular, weakly convex, weak convexity

[1] The research reported here was sponsored in part by the National Science Foundation under Grant DMS-0305930, in part by the U. S. Army Research Office under Grant DAAG19-01-1-0502, and in part by the Air Force Research Laboratory under agreement number F49620-01-1-0040. The U. S. Government has certain rights in this material, and is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the sponsoring agencies or the U. S. Government.

INTRODUCTION

This paper deals with results about projectors on the important class of closed, *prox-regular* sets in $\mathbb{R}^n$. These sets include all closed convex sets but also many nonconvex sets. However, they preserve some very important properties of convex sets, and in so doing they substantially extend the usefulness of those properties.

One of the best known results about closed convex sets in $\mathbb{R}^n$ is that for such a set, say C, the *projector* on C (that is, the function taking a point $x \in \mathbb{R}^n$ to the (necessarily unique) point $\Pi_C(x)$ in C that is closest to x) is single-valued and Lipschitz continuous with modulus 1. For a discussion see [5, Example 2.25]; the Lipschitz property is an easy extension of this.

However, for closed sets that may not be convex these properties may well fail. An obvious example is a circle S in $\mathbb{R}^2$; this is a closed set, but if x is the center of the circle then $\Pi_S(x) = S$: that is, Π_S is multivalued at x and indeed every point of S is closest to x. A less trivial example, also in $\mathbb{R}^2$, is provided by the parabola $S = \{(x_1, x_2) | x_2 = (1/2)x_1^2\}$. For points $x(\xi)$ of the form $(0,\xi)$ with $\xi > 1$, the two points $(\pm[2(\xi-1)]^{1/2}, \xi - 1)$ are closest to x in S, so that the projection of $x(\xi)$ on S does not consist of a single point. Moreover, if we consider the point $x(\xi) = (0,1)$ corresponding to the value $\xi = 1$, for which these two points coalesce into one at the origin, and then increase ξ slightly, the points of $x(\xi)$ move away from the origin at a rate faster than linear. Thus, Π_S fails to display not only Lipschitz continuity but even upper Lipschitz continuity (sometimes called calmness) at that point.

However, this unpleasant example also furnishes us with a clue to what we might do to remedy the problem. The critical value of ξ in this simple example was $\xi = 1$, and it is easy to verify that for $\xi \in [0,1]$ the value of $\Pi_S(x(\xi))$ consists of one point only. Thus, we might conjecture that if we constrained the point x to be close enough to the set S, Π_S would retain the good properties that we want.

Even this is not quite true, though, as we can see by modifying this example slightly. Let p be a real number in $(0,1)$ and for small nonnegative α define $\xi(\alpha) = (1+p)^{-1}\alpha^{1+p} + \alpha^{1-p}$, $x_+(\alpha) = (\alpha, (1+p)^{-1}\alpha^{1+p})$, and $x_-(\alpha) = (-\alpha, (1+p)^{-1}\alpha^{1+p})$. One can then verify that $x_+(\alpha)$ and $x_-(\alpha)$ are projections of $\xi(\alpha)$ on $S = \{(x_1, x_2) | x_2 = (1+p)^{-1} | x_1 |^{1+p}\}$. This example and other like it show that the properties we are looking for require that the boundary of the set in question display a certain amount of good behavior. We shall give a precise definition later, in the form that we will need for the work of this paper, but first we discuss some previous work that has identified the condition and some of its properties.

As far as the author is aware, Vial [7] first identified the property expressing this good behavior. He called it weak convexity, and in addition to technical descriptions he gave an excellent geometric depiction of this property as the requirement that one could roll a suitably small ball along the boundary of the set while maintaining contact with the set at one point only. This geometric view makes it easy to see why outward corners (such as one finds on many convex sets) make no difficulty, but inward corners do, as do boundaries that are concave outward and not sufficiently smooth, like the one in the last example above. Vial analyzed several properties of such sets.

In 1994 Shapiro [6] gave a definition of this property in terms of what he called "$O(2)$-convexity." He proved the important fact that the property implied that, near a point $x_0 \in S$ where this property held, the projector Π_S was single-valued and Lipschitzian [6, Theorem 2.2]. Clarke, Stern, and Wolenski [1] developed a similar idea under the name of "proximal smoothness," and obtained results about its consequences for projection. Subsequently Poliquin and Rockafellar [3] developed a number of properties of this basic geometric idea, in substantial generality, under the name of "prox-regularity." More such work has since taken place, for example that of [2], of which we make substantial use later in this paper. Much of this work is summarized and extended in [5, Section 13.F]. In this latter work, the basic prox-regularity property is defined for a *function,* and the application to sets then follows by letting this function be the indicator of the set in question.

The present paper was motivated by the need to establish some results about projectors as a foundation for the work carried out in [4]. That paper introduced a localized version of *normal maps,* which are single-valued functions that encapsulate important properties of the solutions of variational conditions. Through this localization it was possible to extend the normal-map concept to deal with variational conditions posed over nonconvex sets, and thereby to draw conclusions about the existence and continuity of solutions of the underlying variational conditions.

As part of the work of [4] we investigated solutions of variational conditions in which the underlying set, as well as the function appearing in the condition, may vary continuously. In order to deal with these varying sets it was necessary to establish some basic results about projectors on perturbed sets, and the class of sets for which this could be done turned out to be the prox-regular sets. However, the exact results required seemed not to be available in the current literature, and the aim of this paper is therefore to establish these results and, moreover, to do so using the most elementary methods possible.

In the following sections we first define precisely the class of sets we will use, employing a definition due to Levy, Poliquin, and Rockafeller [2], and

then establish the main result about single-valuedness and continuity of the projector. These are done in Section 1. Then, in Section 2, we establish a quantitative bound that we subsequently use to show that the projector is jointly continuous when considered as a function of the point being projected as well as a perturbation parameter defining the set on which the projection is done.

1. PROX-REGULARITY AND THE PROJECTOR

This section establishes the notation and underlying hypotheses we use, then proves the main theorem about properties of the projector. We employ a variable set, given by a multifunction $S:\mathbb{R}^m \to \mathbb{R}^n$. For a parameter $u \in \mathbb{R}^m$, we will project points x in $\mathbb{R}^n$ onto the set $S(u)$. Our interest will be in properties of the projection when considered as a function of x and u together.

The property we require of the set $S(u)$ is prox-regularity in x with compatible parametrization by u, in the terminology of Levy, Poliquin, and Rockafellar. This property is stated for functions in the following definition, adapted from [2, Definition 2.1].

Definition 1. *Let f be a lower semicontinuous extended real-valued function on $\mathbb{R}^n \times \mathbb{R}^m$, and let $(x_0,u_0) \in \mathbb{R}^n \times \mathbb{R}^m$ with $v_0 \in \partial_x f(x_0,u_0)$. We say that the function f is* prox-regular in x at x_0 for v_0 with compatible parametrization by u *at u_0 if there exist neighborhoods U, V, and X of u_0, v_0, and x_0 respectively, with $\varepsilon > 0$ and $\rho \geq 0$, such that whenever (x,u,v) belongs to the intersection of $X \times U \times V$ with the graph of $\partial_x f$ and $f(x,u) \leq f(x_0,u_0) + \varepsilon$, one has for each $x' \in X$ the inequality*

$$f(x',u) \geq f(x,u) + \langle v, x' - x \rangle - (\rho/2)\|x' - x\|^2 . \tag{1}$$

For our application we want the function $f(x,u)$ to be the indicator of $S(u)$ evaluated at x. We will require the multifunction S to be continuous at each point of U, which implies in particular that f is lower semicontinuous. As the indicator takes only the values 0 and $+\infty$, the statement in (1) becomes the assertion that

$$\langle v, x' - x \rangle - (\rho/2)\|x' - x\|^2 \leq 0, \tag{2}$$

whenever $(x,u,v) \in X \times U \times V$, $x \in S(u)$, $x' \in X \cap S(u)$, and $v \in N_{S(u)}(x)$. The situation where a prox-regular function f is an indicator (that is, where

we are dealing with geometric properties of a set) has been studied previously.

The next theorem provides information on the properties of the projector on a prox-regular set that also satisfies a continuity condition. In dealing with the projector we employ the notation $N_{S(u)}(x)$ for the normal cone to $S(u)$ at x in the sense of [5, Chapter 6].

Theorem 2. *Let X and U be open subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, and let S be a multifunction from U to $\mathbb{R}^n$ that is continuous on U. Let x_0 be a point of X and u_0 a point of U, such that $x_0 \in S(u_0)$. Let $v_0 = 0$, and suppose that the indicator of S is prox-regular in x at x_0 for v_0 with compatible parametrization by u at u_0. Then for each real number $\beta > 1$ there exist an open neighborhood U_0 of u_0 and closed neighborhoods X_0 of x_0 and Z_0 of the point $z_0 := x_0$, such that*

a. The localization to $(U_0 \times Z_0) \times X_0$ of the multifunction taking $(u,z) \in U \times Z$ to $(I + N_{S(u)})^{-1}(z) \subset \mathbb{R}^n$ is a single-valued function π that coincides with the localization to $(U_0 \times Z_0) \times X_0$ of the multifunction taking $(u,z) \in U \times Z$ to the projection $\Pi_{S(u)}(z)$ (the set of points in $S(u)$ closest to z).

b. For each $u \in U_0$ the function $\pi(u,\cdot)$ is Lipschitzian on Z_0 with modulus β.

Note that in the statement of Theorem 2 we require a special choice of z_0: namely, the point x_0, so that $z_0 - x_0 = 0$. For the application to variational conditions this presents no difficulty, as we can always multiply the function f by some positive scalar Φ to ensure that $f(u_0, x_0)$ is as close to the origin as we wish.

Proof. The proof employs two arguments. In the first, we use the machinery of prox-regularity to show that for each $\beta > 1$ there are neighborhoods X_0 of x_0, U_1 of u_0, and Z_1 of z_0, with X_0 closed, such that the localization to $(U_1 \times Z_1) \times X_0$ of the multifunction, say Q, taking $(u,z) \in U \times Z$ to $(I + N_{S(u)})^{-1}(z)$, has desirable continuity properties. We will use the symbol π_1 for this localization, and will show that for each $u \in U_1$, $\pi_1(u,\cdot)$ is Lipschitz continuous on Z_1 with modulus β (hence, in particular, its values contain no more than one point). Then, in the second argument we employ the continuity of S to show that by further shrinking the neighborhoods U_1 and Z_1 to an open neighborhood U_0 of u_0 and a closed neighborhood Z_0 of z_0, we can ensure that for each $(u,z) \in U_0 \times Z_0$ the set X_0 contains a point of $\Pi_{S(u)}(z)$. We then combine the two arguments to prove claims (a) and (b).

For the first argument we recall that because of the prox-regularity, if we take $v_0 = 0 \in N_{S(u_0)}(x_0)$ then there exist neighborhoods U_1, V', and X' of

u_0, v_0, and x_0 respectively, with $\varepsilon > 0$ and $\rho > 0$, such that whenever $x \in S(u)$ and (x,u,v) belongs to $X' \times U_1 \times V'$ and satisfies $v \in N_{S(u)}(x)$, one has for each $x' \in X' \cap S(u)$ the inequality

$$0 \geq \langle v, x' - x \rangle - (\rho/2)\|x' - x\|^2. \tag{3}$$

By making X' and U_1 smaller if necessary, we may suppose them to be contained in X and U respectively.

Choose a real number $\beta > 1$. The interiors of the neighborhoods V' and X' contain closed balls of radius $\sigma > 0$ around v_0 and x_0 respectively. Let $X_0 = B(x_0, \gamma)$, where $0 < \gamma \max\{1, 2\rho(1-\beta^{-1})^{-1}\} \leq \sigma$, and take $Z_1 = X_0$. Fix any $u \in U_1$ and suppose that for $i = 1,2$ we have $z_i \in Z_1$ and $x_i \in X_0$ with $z_i \in N_{S(u)}(x_i)$ for $i = 1,2$. Define $v_i = z_i - x_i$. If we suppose for the moment that $\rho > 0$, then we have

$$\|v_i\| \leq \|(z_i - x_0) + (x_0 - x_i)\| \leq 2\gamma,$$

so that $\|\rho(1-\beta^{-1})^{-1} v_i\| \leq \sigma$ and accordingly $[\rho(1-\beta^{-1})^{-1}]v_i \in V'$. As the points x_i belong to $X' \cap S(u)$ we conclude from (3) that for $i = 1,2$ and for any $x' \in X' \cap S(u)$ we have

$$0 \geq [\rho(1-\beta^{-1})^{-1}]\langle v_i, x' - x_i \rangle - (\rho/2)\|x' - x_i\|^2,$$

and therefore

$$0 \geq \langle v_i, x' - x_i \rangle - (1/2)(1-\beta^{-1})\|x' - x_i\|^2. \tag{4}$$

On the other hand, if $\rho = 0$ then (4) follows *a fortiori* from (3), so that in fact (4) holds for any $\rho \geq 0$.

For each i replace x' by the other point x_j ($j \neq i$) in (4) and add the two inequalities to obtain

$$\langle v_1 - v_2, x_1 - x_2 \rangle \geq -(1-\beta^{-1})\|x_1 - x_2\|^2.$$

Recall that $v_i = z_i - x_i$, so that we can rewrite this inequality as

$$\langle z_1 - z_2, x_1 - x_2 \rangle \geq \beta^{-1}\|x_1 - x_2\|^2, \tag{5}$$

which shows that for fixed $u \in U_1$ the function $Z_1 \cap (I + N_{S(u)})(\cdot)$ is strongly monotone on X_0 with modulus β^{-1}. By applying the Schwarz inequality to the left side of (5) we can derive the bound

$$\|x_1 - x_2\| \le \beta \|z_1 - z_2\|,$$

which shows that for each $u \in U_1$, $\pi_1(u,\cdot) := X_0 \cap (I + N_{S(u)})^{-1}(\cdot)$ is in fact Lipschitzian on its domain Z_1, with a modulus β that does not depend on $u \in U_1$. In particular, then, each value $\pi_1(u,z)$ contains no more than one point, and this concludes the first argument.

For the second argument, choose positive numbers ε and δ small enough so that X_0 contains the ball $B(x_0, 2\delta + \varepsilon)$. We have assumed that $x_0 \in S(u_0)$, so $S(u_0)$ meets the open ball of radius ε about x_0. Use the inner semicontinuity of S at u_0 to choose an open neighborhood U_0 of u_0 contained in U_1 and small enough so that for each $u \in U_0$ the set $S(u)$ meets that ball. Let $Z_0 := B(z_0, \min\{\gamma, \delta\})$, so that $Z_0 \subset Z_1$, and choose any $z \in Z_0$. Then if $u \in U_0$ the distance from z to $S(u)$ is less than

$$\|z - z_0\| + \|z_0 - x_0\| + \varepsilon = \|z - z_0\| + \varepsilon \le \delta + \varepsilon,$$

where we used the fact that $z_0 = x_0$. The prox-regularity assumption requires the indicator of S to be lower semicontinuous, so S has closed values. Thus there is a point x of $\Pi_{S(u)}(z)$ in the ball $B(z, \delta + \varepsilon)$. But this ball is contained in $B(z_0, 2\delta + \varepsilon) = B(x_0, 2\delta + \varepsilon)$, which by hypothesis is contained in X_0. This concludes the second argument.

Now we put the two arguments together. For any $(u,z) \in U_0 \times Z_0$ the first argument shows that $\pi(u,z) = X_0 \cap Q(u,z)$ is at most a singleton. The second argument shows that X_0 contains a point $x \in \Pi_{S(u)}(z)$. But we always have $\Pi_{S(u)}(z) \subset Q(u,z)$, so in fact $x = \pi(u,z)$, and we have

$$X_0 \cap \Pi_{S(u)}(z) = \{x\} = X_0 \cap Q(u,z).$$

This shows that π is a single-valued function on $U_0 \times Z_0$, which is simultaneously the localization to $(U_0 \times Z_0) \times X_0$ of the multifunction taking $(u,z) \in U \times Z$ to $(I + N_{S(u)})^{-1}(z) \subset \mathbb{R}^n$ and of the multifunction taking $(u,z) \in U \times Z$ to the projection $\Pi_{S(u)}(z)$, and it proves the claim in (a). The claim in (b) follows from what we showed in the first argument about π_1, of which π is a restriction. □

Theorem 2 showed that for fixed u the function $\pi(u,z)$ was continuous in z, but we actually need joint continuity. Therefore in the next section we

first establish in Theorem 3 a quantitative bound on the difference between $\pi(u',z)$ and $\pi(u,z)$ for two points u and u' close to u_0. Then we apply that bound in Corollary 4 to establish the required joint continuity.

2. QUANTITATIVE BOUNDS

Here we first prove that for points u and u' close to u_0 we can bound the distance between the projections of a point z on $S(u)$ and on $S(u')$ respectively by a quantitative expression involving the Pompeiu-Hausdorff distance between the intersections of these two sets with a neighborhood X_0. We use the symbol $d[P,Q]$ for the Pompeiu-Hausdorff distance between subsets P and Q of $\mathbb{R}^n$.

Theorem 3. *Assume the notation and hypotheses of Theorem 2, let* $\beta > 1$, *and determine the neighborhoods* U_0, X_0, *and* Z_0 *whose existence that theorem guarantees. For each* $u \in U_0$ *define* $T(u) := S(u) \cap X_0$. *Let* $z \in Z_0$ *and* $u \in U_0$. *Then for each* $u' \in U_0$ *one has*

$$\begin{aligned} \|\pi(u',z) - \pi(u,z)\| \;&\le\; \delta(\beta/2) + \delta^{1/2}\{2\beta\|z - \pi(u,z)\| \\ &+ \delta[\beta^2/4 + \beta - 1]\}^{1/2}, \end{aligned} \tag{6}$$

where $\delta = d[T(u'),T(u)]$.

Proof. Write $x = \pi(u,z)$, $x' = \pi(u',z)$, $\delta = d[T(u),T(u')]$, and $\nu := \|z - \pi(u,z)\|$. We have $x \in T(u)$ and $x' \in T(u')$. Moreover, $z - x \in N_{S(u)}(x)$ and $z - x' \in N_{S(u')}(x')$. The multifunction S is continuous on U and hence has closed values; as X_0 is closed the values of $T(u)$ are also closed. Therefore there exist points $y \in T(u)$ and $y' \in T(u')$ with

$$y' = x + r, \quad \|r\| \le \delta, \qquad y = x' + r', \quad \|r'\| \le \delta.$$

Applying (4) to the triples $(z - x, x, y)$ and $(z - x', x', y')$ we obtain the two inequalities

$$\begin{aligned} 0 \;&\ge\; \langle z - x, y - x\rangle - (1/2)(1 - \beta^{-1})\|y - x\|^2, \\ 0 \;&\ge\; \langle z - x', y' - x'\rangle - (1/2)(1 - \beta^{-1})\|y' - x'\|^2. \end{aligned} \tag{7}$$

We have $y - x = r' + (x' - x)$, so

$$\|y - x\|^2 = \|r' + (x' - x)\|^2 = \|r'\|^2 + 2\langle r', x' - x\rangle + \|x' - x\|^2.$$

Writing similar equations for $y' - x' = r + (x - x')$, substituting all of these into (7), and adding yields

$$\begin{aligned} 0 \geq\ & \langle z - x, r'\rangle + \langle z - x', r\rangle + \|x' - x\|^2 \\ & - (1/2)(1 - \beta^{-1})(\|r\|^2 + \|r'\|^2) \\ & - (1 - \beta^{-1})(\langle r' - r, x' - x\rangle + \|x' - x\|^2). \end{aligned} \tag{8}$$

By observing that

$$\begin{aligned} \langle z - x, r'\rangle + \langle z - x', r\rangle &= \langle z - x, r' + r\rangle - \langle x' - x, r\rangle \\ &\geq -2\delta\nu - \langle x' - x, r\rangle, \end{aligned} \tag{9}$$

and that

$$-(1/2)(1 - \beta^{-1})(\|r\|^2 + \|r'\|^2) \geq -\delta^2(1 - \beta^{-1}),$$

and by combining some terms, we can obtain from (8)

$$\begin{aligned} 0 \geq\ & -2\delta\nu - \langle x' - x, r\rangle + \|x' - x\|^2 - \delta^2(1 - \beta^{-1}) \\ & - (1 - \beta^{-1})(\langle r' - r, x' - x\rangle + \|x' - x\|^2) \\ =\ & -\delta[2\nu + \delta(1 - \beta^{-1})] - \langle \beta^{-1} r + (1 - \beta^{-1}) r', x' - x\rangle \\ & + \beta^{-1}\|x' - x\|^2 \\ \geq\ & -\delta[2\nu + \delta(1 - \beta^{-1})] - \delta\|x' - x\| + \beta^{-1}\|x' - x\|^2. \end{aligned} \tag{10}$$

By applying the quadratic formula to (10) we obtain (6). □

Observe that in the special case $z \in S(u)$ (*i.e.*, $\nu = 0$), (6) provides a bound of the form

$$\|\pi(u', z) - \pi(u, z)\| \leq \kappa d[T(u'), T(u)],$$

so that as we would expect $\pi(\cdot,z)$ obeys a Lipschitz condition in the Pompeiu-Hausdorff metric applied to $T(\cdot)$. In the general case we only have a Hölder condition with exponent 1/2, as the form of the bound is

$$\|\pi(u',z)-\pi(u,z)\| \le \lambda d[T(u'),T(u)]^{1/2},$$

as long as u' remains near u.

With this result it is now easy to show the joint continuity of the function π in the arguments (u,z). We do this in the following corollary.

Corollary 4. *Assume the notation and hypotheses of Theorem 2. Fix* $\beta > 1$ *and determine the neighborhoods* U_0, X_0, *and* Z_0 *whose existence is guaranteed by that theorem. Then the function* π *is continuous at each* $(u,z) \in U_0 \times Z_0$.

Proof. Choose any $(u,z) \in U_0 \times Z_0$ and any $\varepsilon > 0$. Theorem 3 shows that $\pi(\cdot,z)$ is continuous at u, so we can find a neighborhood V of u, contained in the open set U_0, such that if $u' \in V$ then $\|\pi(u,z)-\pi(u',z)\| < \varepsilon/2$. Now choose any $(u',z') \in V \times Z_0$ such that $\|z'-z\| < (2\beta)^{-1}\varepsilon$. By using the Lipschitz continuity of $\pi(u',\cdot)$ established in Theorem 2, we obtain

$$\begin{aligned}\|\pi(u,z)-\pi(u',z')\| &\le \|\pi(u,z)-\pi(u',z)\| + \|\pi(u',z)-\pi(u',z')\| \\ &\le \varepsilon/2 + \varepsilon/2 = \varepsilon,\end{aligned}$$

from which we see that π is continuous at (u,z). □

ACKNOWLEDGMENT

The author thanks the Ettore Majorana Foundation and Centre for Scientific Culture, Erice, Sicily, for its hospitality and for its very stimulating intellectual atmosphere, both of which greatly assisted the research for this paper.

REFERENCES

[1] Francis H. Clarke, R.J. Stern, and P.R. Wolenski. Proximal smoothness and the lower-C^2 property. *Journal of Convex Analysis*, 1–2:117–144, 1995.

[2] Adam B. Levy, René A. Poliquin, and R. Tyrrell Rockafellar. Stability of locally optimal solutions. *SIAM Journal on Optimization*, 10:580–604, 2000.

[3] R.A. Poliquin and R.T. Rockafellar. Prox-regular functions in variational analysis. *Transactions of the American Mathematical Society*, 348:1805–1838, 1996.

[4] Stephen M. Robinson. Localized normal maps and the stability of variational conditions. Accepted by *Set-Valued Analysis*.

[5] R. Tyrrell Rockafellar and Roger J–B Wets. *Variational Analysis*. Number 317 in Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin, 1998.

[6] Alexander Shapiro. Existence and differentiability of metric projections in Hilbert spaces. *SIAM Journal on Optimization*, 4:130–141, 1994.

[7] Jean-Philippe Vial. Strong and weak convexity of sets and functions. *Mathematics of Operations Research*, 8:231–259, 1983.

APPLICATION OF OPTIMAL CONTROL THEORY TO DYNAMIC SOARING OF SEABIRDS

G. Sachs[1] and P. Bussotti[2]
Institute of Flight Mechanics and Flight Control, Technical University of Munich, Garching, Germany;[1] Humboldt Foundation, Institute of Science History, Ludwig-Maximilians University of Munich, Munich,, Germany[2]

Abstract: Optimal control theory is applied as a method for determining the minimum wind strength required for dynamic soaring of seabirds. Dynamic soaring is a flight technique by which seabirds extract energy from shear wind existing in an altitude layer close to the water surface. Mathematical models for describing the soaring motion of a bird and for the shear wind are presented. Optimality conditions are formulated using the minimum principle. Switching conditions are introduced to deal with a state constraint. Numerical results of high accuracy are generated using an efficient computational procedure based on the method of the multiple shooting for an albatross as a representative for seabirds performing dynamic soaring.

NOMENCLATURE

a_{ki}	abbreviation factor
C_D	drag coefficient
C_L	lift coefficient
D	drag
g	acceleration due to gravity
H	Hamiltonian

h	altitude
J	performance criterion
k	drag factor for describing lift effect
L	lift
m	mass of bird
S	reference area
t	time
u_{Kg}, v_{Kg}, w_{Kg}	speed components
V	airspeed
V_K	inertial speed
V_W	wind speed
x_K, y_K, z_K	geodetic coordinate system
χ_a	flight azimuth angle
γ_a	flight path wind angle
λ_i	Lagrange multiplier
μ_a	flight bank wind angle

1. INTRODUCTION

Dynamic soaring is a flight method by which an object in gliding flight (bird, sailplane) extracts energy from horizontally moving air. The possibility of extracting energy for continuous dynamic soaring requires that the horizontally moving air is non-uniform. This means that the horizontal wind speed changes with altitude. Such a type of wind is called shear wind or shear flow.

Dynamic soaring is observed with seabirds which utilize the shear flow in an altitude region close to the water surface. Here, the wind speed rapidly increases in a small altitude interval termed boundary layer, from zero to the value of the free air flow. There are several seabirds that perform dynamic soaring. Among these, the albatross is the most famous and considered the master of dynamic soaring (Ref. 1).

The possibility of utilizing shear wind for soaring flight and the basic mechanism of the energy transfer from the moving air to the bird have been early considered and clarified, Refs. 1-4. Since then, dynamic soaring is an issue of continuous interest and the knowledge has been continually increased, Ref. 5. Investigations on energy estimations and numerical simulations were performed, yielding an improved understanding of dynamic soaring. The research includes papers which are concerned with

mathematical treatments of dynamic soaring, Refs. 6-11. Other papers are more dealing with ornithological aspects, Refs. 12-14. Modern optimization techniques have been applied to the dynamic soaring problem, yielding results on the minimum required wind strength, Refs. 15-17. Recent experimental research is concerned with tracking of albatrosses, providing results on their enormous flight performance, Refs. 18-23.

It is purpose of this paper to present a rigorous mathematical treatment concerned with dynamic soaring of seabirds. This relates to the mathematical model for describing the motion of the birds and to the optimization method to generate solutions for the minimum wind strength required for dynamic soaring. Numerical results are presented showing the form of the optimal dynamic soaring trajectory requiring minimum wind strength and the properties of state and control variables for achieving this goal.

2. BASIC CONSIDERATIONS ON DYNAMIC SOARING

Dynamic soaring comprises a rather complex and a highly dynamic flight maneuver involving a complicated control structure and a corresponding behavior of the state variables. Consequently, no simple and direct access to this problem is possible to yield closed-form solutions. By contrast, there are other soaring techniques which consist of rather simple flight maneuvers to enable an energy gain for the bird. These problems can be solved with direct approaches. Such soaring flights are thermalling and hang gliding in up-wind fields which are utilized by birds or sailplanes. An illustration is given in Fig. 1 which shows flight conditions in up-wind fields. Basically, the flying object is moved upwards by the air, resulting in a corresponding energy increase. This energy source can be utilized with rather simple maneuvers, like circling or even straight motions representing basically steady-state flight conditions. Correspondingly, comparatively simple solutions are known for such problems, Ref. 24.

By contrast, dynamic soaring consists of a complex and unsteady flight maneuver for transferring energy to the flying object from the moving air which is not moving upwards, but horizontally. A dynamic soaring flight maneuver is illustrated in Fig. 2. There is a horizontal wind which shows a shear flow characteristic. The wind rapidly increases from very small values immediately above the sea surface to the free air flow speed. For transferring energy from the moving air to the bird from such a shear wind, a complex

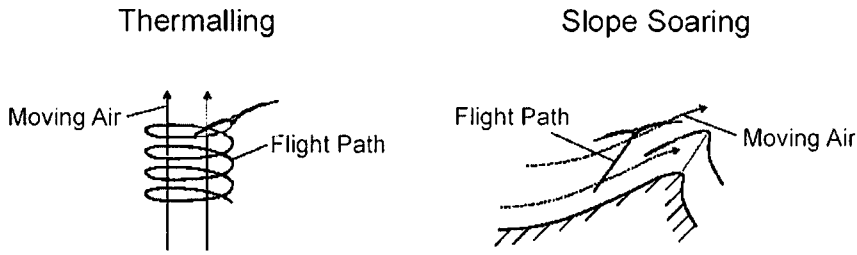


Figure 1. Flight paths and energy transfer in up-wind regions

flight maneuver as illustrated in Fig. 2 is necessary. It basically consists of four phases 1 to 4, yielding:

1) Phase 1
   Lower curve close to sea surface (change of flight direction from leeward to windward)
2) Phase 2
   Climb (windward flight)
3) Phase 3
   Upper curve (change of flight direction from windward to leeward)
4) Phase 4
   Descent (leeward flight)

Phases 1 to 4 form a cycle which is the basic constituent of dynamic soaring. By periodically repeating the dynamic soaring cycle, the bird can perform a range flight without flapping its wings.

With a dynamic soaring flight maneuver as shown in Fig. 2, it is possible for the bird to attain an energy gain from the moving air. This enables it to achieve an enormous flight performance. An example for the achievable performance is given in Fig. 3 which presents the trajectory of an albatross tracked by satellite measurements. The length of the trajectory shown in Fig. 3 is 6479 km which the bird has traveled in 8.15 days.

Focus of this paper is on the minimum shear wind strength required for dynamic soaring. There are dynamic soaring trajectories which show the same energy state of the bird at the end of a dynamic soaring cycle as at its beginning. These trajectories can be designated as energy-neutral. The designation "energy-neutral" means that the energy gain from the moving air is just sufficient to compensate for the energy loss due to drag after completing a dynamic soaring cycle. There is a great variety of energy-neutral trajectories. One of these is of particular concern: It is the one which requires the minimum shear wind strength. Having knowledge of this energy-neutral trajectory, it is possible to judge whether or not the shear

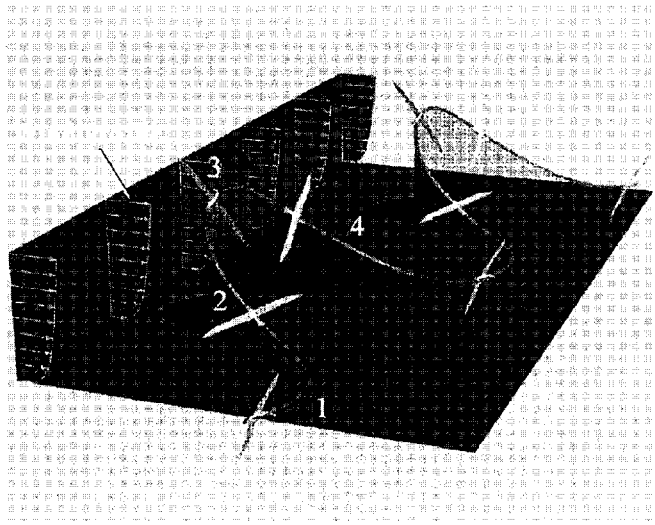


Figure 2. Dynamic soaring trajectory and energy transfer

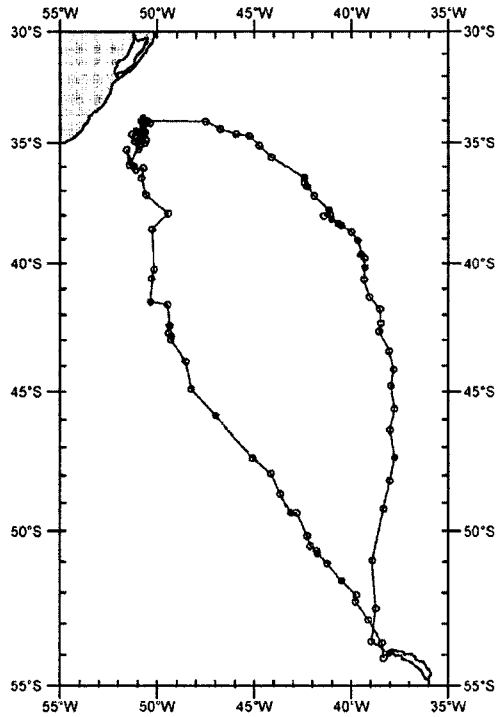


Figure 3. Flight path of an albatross (from Ref. 19)

wind strength in the areas of the seabirds in mind is sufficient for dynamic soaring.

3. MATHEMATICAL MODEL FOR DESCRIBING THE MOTION OF THE BIRD

The dynamics of birds in soaring flight can be described using a point mass model. Reference is made to an earth fixed coordinate system, and the moving air is appropriately accounted for. Fig. 4 shows the earth fixed reference system and the speed vectors describing the moving air and the motion of the bird. The equations of motion may be expressed as

$$\begin{aligned}
\frac{\mathrm{d}u_{Kg}}{\mathrm{d}t} &= -a_{u1}\frac{D}{m} - a_{u2}\frac{L}{m} \\
\frac{\mathrm{d}v_{Kg}}{\mathrm{d}t} &= -a_{v1}\frac{D}{m} - a_{v2}\frac{L}{m} \\
\frac{\mathrm{d}w_{Kg}}{\mathrm{d}t} &= -a_{w1}\frac{D}{m} - a_{w2}\frac{L}{m} + g \\
\frac{\mathrm{d}x_{Kg}}{\mathrm{d}t} &= u_{Kg} \\
\frac{\mathrm{d}y_{Kg}}{\mathrm{d}t} &= v_{Kg} \\
\frac{\mathrm{d}h}{\mathrm{d}t} &= -w_{Kg}
\end{aligned} \tag{1}$$

where the coefficients a_{ij} denote functions of the path angles χ_a, γ_a and μ_a, given by the following relations

$$\begin{aligned}
a_{u1} &= \cos\gamma_a \cos\chi_a \\
a_{u2} &= \cos\mu_a \sin\gamma_a \cos\chi_a + \sin\mu_a \sin\chi_a \\
a_{v1} &= \cos\gamma_a \sin\chi_a \\
a_{v2} &= \cos\mu_a \sin\gamma_a \sin\chi_a - \sin\mu_a \cos\chi_a \\
a_{w1} &= -\sin\gamma_a \\
a_{w2} &= \cos\mu_a \cos\gamma_a
\end{aligned} \tag{2}$$

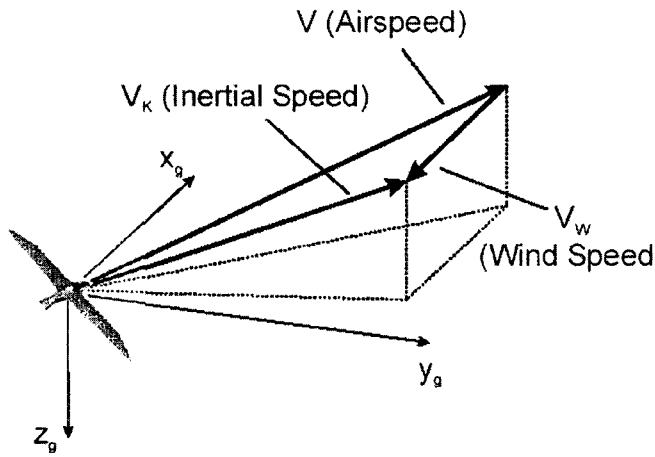


Figure 4. Coordinate system x_g, y_g, z_g and speed vectors V_K, V, V_w for describing the flight in a horizontal shear wind (V_K: inertial speed measured relative to the ground, V: airspeed of the bird relative to the moving air, V_w: wind speed). The x_g axis is aligned with the wind speed vector V_w, the z_g axis points vertically downward.

The aerodynamic forces are drag and lift which read

$$\begin{aligned} L &= C_L(\rho/2)V^2S \\ D &= C_D(\rho/2)V^2S \end{aligned} \tag{3}$$

The drag characteristics may be modeled as

$$C_D = C_{D0} + kC_L^2 \tag{4}$$

where the lift coefficient C_L is a control which is determined by optimality conditions described in a subsequent section.

The aerodynamic forces are dependent on airspeed $\vec{V}$, while the motion of the bird with regard to the earth is described by the inertial speed $\vec{V}_K = (u_{Kg}, v_{Kg}, w_{Kg})^T$. They are related to each other by the following expression

$$\vec{V} = \vec{V}_K - \vec{V}_W \tag{5}$$

With the use of $\vec{V}_W = (-V_W, 0, 0)^T$, Eq. (5) yields

$$\vec{V} = (u_{Kg} + V_W, v_{Kg}, w_{Kg})^{\mathrm{T}}$$
$$V = \sqrt{(u_{Kg} + V_W)^2 + v^2{}_{Kg} + u_{Kg}^2} \tag{6}$$

Two angles of the aerodynamic coordinate system used in Eqs. (1) and (2) are given by

$$\sin \gamma_a = -\frac{w_{Kg}}{V}$$
$$\tan \chi_a = \frac{v_{Kg}}{u_{Kg} + V_W} \tag{7}$$

The remaining angle μ_a which describes the banking of the lift vector is a control. It is determined by optimality conditions described in a subsequent section.

For a cycle of an energy-neutral trajectory, the following boundary conditions implying periodicity hold

$$u_{Kg}(0) = u_{Kg}(t_{cyc}), \quad v_{Kg}(0) = v_{Kg}(t_{cyc}), \quad w_{Kg}(0) = w_{Kg}(t_{cyc}), \quad h(0) = h(t_{cyc}) \tag{8}$$

where t_{cyc} describes the time at the end of a cycle. With an appropriate choice of the coordinate system, the boundary values of the longitudinal and lateral coordinates read

$$x_g(0) = 0, \quad x_g(t_{cyc}) : \text{free}$$
$$y_g(0) = 0, \quad y_g(t_{cyc}) : \text{free} \tag{9}$$

Control variables are the lift coefficient C_L and the bank angle μ_a. The lift coefficient is subject to the following constraint relation

$$C_{L\min} \le C_L \le C_{L\max} \tag{10}$$

4. MATHEMATICAL SHEAR WIND MODEL

The dynamic soaring of seabirds is possible because there is a shear wind at the sea. The shear wind is due to boundary layer effects of the moving air. This is illustrated in Fig. 5 which shows the shear wind profile (wind speed vs. altitude) for the altitude region of concern for the dynamic soaring of sea birds. From zero or very small values immediately above the sea surface, the wind speed rapidly increases and approaches the value of the free air flow.

There are various models for describing the shear wind characteristics. For the shear wind above the water surface, logarithmic or exponential models are used (Refs. 6,7,13,25). They may be expressed as

$$V_W = V_{W\,ref}\left(\frac{h}{h_{ref}}\right)^p$$
$$V_W = \overline{V}_W \frac{\ln(h/h_0)}{\ln(\overline{h}/h_0)} \tag{11}$$

The quantities $V_{W\,ref}$, h_{ref} and p as well as $\overline{V}_W$, h_0 and $\overline{h}$, denote reference values which are used for indicating the strength of the shear wind and for taking properties of the surface into account. The exponential model is applied in this paper, with $p = 0.143$.

5. OPTIMALITY CONSIDERATIONS

For the performance criterion, designated by J, the quantity $V_{W\,ref}$ can be used. This is because $V_{W\,ref}$ is a measure for the shear wind strength, yielding

$$J = V_{W\,ref} \tag{12}$$

The optimal control problem can then be formulated as to determine the controls, the initial conditions $\vec{V}_K(0) = (u_{Kg}(0), v_{Kg}(0), w_{Kg}(0))^T$ and $h(0)$ and the optimal cycle time t_{cyc} which minimize the performance criterion $J = V_{W\,ref}$ subject to the dynamic system Eq. (1), the boundary conditions Eqs. (8), (9) and the control constraints Eq. (10).

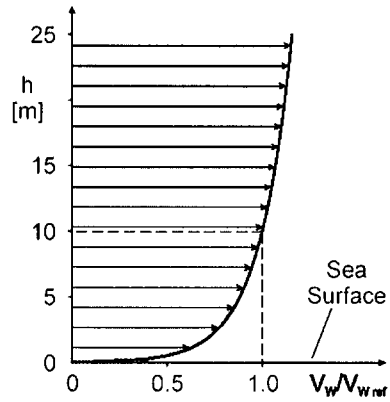


Figure 5. Model for describing shear wind in boundary layer above sea surface. The quantity V_W ref denotes the wind speed at the reference altitude for which a value of h = 10 m is chosen.

The optimal control problem described is solved with use of the minimum principle. For this purpose, the Hamiltonian is introduced

$$H = -\frac{a_{u1}D + a_{u2}L}{m}\lambda_u - \frac{a_{v1}D + a_{v2}L}{m}\lambda_v + \left(g - \frac{a_{w1}D + a_{w2}L}{m}\right)\lambda_w + u_{Kg}\lambda_x + v_{Kg}\lambda_y - w_{Kg}\lambda_h \tag{13}$$

with Lagrange multipliers

$$\lambda = (\lambda_u, \lambda_v, \lambda_w, \lambda_x, \lambda_y, \lambda_h)^{\mathrm{T}} \tag{14}$$

adjoined to the system Eq. (1). The following relations hold for the Lagrange multipliers

$$\frac{d\lambda_u}{dt} = \frac{\partial}{\partial u_{Kg}}(a_{u1}D + a_{u2}L)\frac{\lambda_u}{m} + \frac{\partial}{\partial u_{Kg}}(a_{v1}D + a_{v2}L)\frac{\lambda_v}{m} + \frac{\partial}{\partial u_{Kg}}(a_{w1}D + a_{w2}L)\frac{\lambda_w}{m} - \lambda_x$$

$$\frac{d\lambda_v}{dt} = \frac{\partial}{\partial v_{Kg}}(a_{u1}D + a_{u2}L)\frac{\lambda_u}{m} + \frac{\partial}{\partial v_{Kg}}(a_{v1}D + a_{v2}L)\frac{\lambda_v}{m} + \frac{\partial}{\partial v_{Kg}}(a_{w1}D + a_{w2}L)\frac{\lambda_w}{m} - \lambda_y$$

$$\frac{d\lambda_w}{dt} = \frac{\partial}{\partial w_{Kg}}(a_{u1}D + a_{u2}L)\frac{\lambda_u}{m} + \frac{\partial}{\partial w_{Kg}}(a_{v1}D + a_{v2}L)\frac{\lambda_v}{m} + \frac{\partial}{\partial w_{Kg}}(a_{w1}D + a_{w2}L)\frac{\lambda_w}{m} - \lambda_h \tag{15}$$

$$\frac{d\lambda_x}{dt} = 0$$

$$\frac{d\lambda_y}{dt} = 0$$

$$\frac{d\lambda_h}{dt} = \left[\frac{\partial}{\partial V_W}(a_{u1}D + a_{u2}L)\frac{\lambda_u}{m} + \frac{\partial}{\partial V_W}(a_{v1}D + a_{v2}L)\frac{\lambda_v}{m} + \frac{\partial}{\partial V_W}(a_{w1}D + a_{w2}L)\frac{\lambda_w}{m}\right]\frac{dV_W}{dh}$$

Because of periodicity properties, the following boundary conditions hold

$$\lambda_u(0) = \lambda_u(t_{cyc}), \quad \lambda_v(0) = \lambda_v(t_{cyc}), \quad \lambda_w(0) = \lambda_w(t_{cyc}) \tag{16a}$$

Further to the boundary conditions

$$\lambda_x(t_{cyc}) = 0, \quad \lambda_y(t_{cyc}) = 0, \quad \lambda_h(t_{cyc}) = -1 \tag{16b}$$

The Hamiltonian is constant because the system described by Eq. (1) is autonomous (i.e., not explicitly dependent on t). Furthermore, the time at the end of a cycle, t_{cyc}, is treated as free, yielding

$$H = 0 \tag{17}$$

The optimal controls are such that the Hamiltonian is minimized. From $\partial H / \partial \mu_a = 0$, it follows for the optimal bank angle control

$$\tan(\mu_a)_{opt} = \frac{\lambda_u \sin\chi_a - \lambda_v \cos\chi_a}{\lambda_u \sin\gamma_a \cos\chi_a + \lambda_v \sin\gamma_a \sin\chi_a + \lambda_w \cos\gamma_a} \tag{18}$$

Similarly from $\partial H / \partial C_L = 0$ for the optimal lift coefficient

$$(C_L)_{opt} = -\frac{1}{2k}\left[\frac{\lambda_u(\cos\mu_a \sin\gamma_a \sin\chi_a - \sin\mu_a \cos\chi_a)}{\lambda_u \cos\gamma_a \cos\chi_a + \lambda_v \cos\gamma_a \sin\chi_a - \lambda_w \sin\gamma_a} + \frac{\lambda_v(\cos\mu_a \sin\gamma_a \sin\chi_a - \sin\mu_a \cos\chi_a) + \lambda_w \cos\mu_a \cos\gamma_a}{\lambda_u \cos\gamma_a \cos\chi_a + \lambda_v \cos\gamma_a \sin\chi_a - \lambda_w \sin\gamma_a}\right] \tag{19}$$

Otherwise, the constraining bounds given by $C_{L\min}$ and $C_{L\max}$, Eq. (10), become active.

6. ALTITUDE CONSTRAINT AND SWITCHING CONDITIONS

The altitude range for the dynamic soaring maneuver has a lower limit given by the water surface. Therefore, it is necessary to introduce an altitude limit described as $h_{\min}$. This results in an altitude constraint given by

$$h \geq h_{\min} \tag{20}$$

As a consequence, there are additional optimization conditions which are presented in the following according to Ref. 26.

Basically, the altitude constraint can become active in two ways. First, a point of contact can exist at which the dynamic soaring trajectory touches the altitude limit $h_{\min}$. Second, there can be an arc on which the dynamic soaring trajectory stays for a finite interval. The second possibility occurred in the computational treatment of the optimization problem so that it will be considered in the following.

For treating the altitude constraint problem, the relation

$$G(h) = h_{\min} - h \leq 0 \tag{21}$$

is introduced. Calculating successively the time derivatives of G until an expression is obtained which is explicitly dependent on the control variables, the following result is obtained

$$\begin{aligned} G^{(1)}(w_{Kg}) &= w_{Kg} \\ G^{(2)}(u_{Kg}, v_{Kg}, w_{Kg}, h, C_L, \mu_a) &= -a_{w1}\frac{D}{m} - a_{w2}\frac{L}{m} + g \end{aligned} \tag{22}$$

Accordingly, the altitude constraint is of second order.

For incorporating the altitude constraint in the optimization, the Hamiltonian is changed to yield

$$H^* = H + \mu(t)G^{(2)}(u_{Kg}, v_{Kg}, w_{Kg}, h, C_L, \mu_a) \tag{23}$$

with addition of an Lagrange multiplier denoted by $\mu(t)$.

As a consequence, the following relations result

$$\begin{aligned} \frac{\mathrm{d}\lambda_u}{\mathrm{d}t} &= -\frac{\partial H}{\partial u_{Kg}} - \mu(t)\frac{\partial}{\partial u_{Kg}}G^{(2)} \\ \frac{\mathrm{d}\lambda_v}{\mathrm{d}t} &= -\frac{\partial H}{\partial v_{Kg}} - \mu(t)\frac{\partial}{\partial v_{Kg}}G^{(2)} \\ \frac{\mathrm{d}\lambda_w}{\mathrm{d}t} &= -\frac{\partial H}{\partial w_{Kg}} - \mu(t)\frac{\partial}{\partial w_{Kg}}G^{(2)} \\ \frac{\mathrm{d}\lambda_h}{\mathrm{d}t} &= -\frac{\partial H}{\partial h} - \mu(t)\frac{\partial}{\partial h}G^{(2)} \end{aligned} \tag{24}$$

The expressions $\mathrm{d}\lambda_x/\mathrm{d}t = 0$ und $\mathrm{d}\lambda_y/\mathrm{d}t = 0$ remain unchanged, because $G^{(2)}$ is independent of x_g and y_g.

For $\mu(t)$, the following relations hold:

a) $\mu(t) = 0$ on the unconstrained arc.

b) $\mu(t) \geq 0$ on the constrained arc. From $\partial H/\partial C_L = 0$ and $\gamma_a = 0$ it follows that

$$\begin{aligned} \mu(t) = &-\frac{2kC_L}{\cos\mu_a}(\lambda_u \cos\chi_a + \lambda_v \sin\chi_a) \\ &- \lambda_u \tan\mu_a \sin\chi_a + \lambda_v \tan\mu_a \cos\chi_a - \lambda_w \end{aligned} \tag{25}$$

c) μ can be discontinuous at the entry point of the constrained arc. At the exit point of the constrained arc, μ is continuous. Using a), it is given by $\mu(t_2) = 0$ where t_2 denotes the exit point.

From Eq. (22), the following relation is obtained for the optimal control of the lift coefficient on the constrained arc ($G^{(2)} = 0$ with $\gamma_a = 0$)

$$(C_L)_{opt} = \frac{1}{\cos \mu_a} \frac{mg}{(\rho/2) V^2 S} \tag{26}$$

This relation describes the aerodynamic lift required for an (unsteady) horizontal turn, i.e. $L = mg / \cos \mu_a$.

The optimal bank angle on the constrained arc can be determined, using $\partial H / \partial \mu_a = 0$ with Eqs. (25) and (26), to yield

$$\tan(\mu_a)_{opt} = -\frac{(\rho/2)\ V^2 S}{2kmg} \frac{\lambda_u \sin \chi_a - \lambda_v \cos \chi_a}{\lambda_u \cos \chi_a + \lambda_v \sin \chi_a} \tag{27}$$

In regard to the entry point of the constrained arc, denoted by t_1, the following relations apply:

a) Eqs. (21) and (22) can be used to yield

$$\begin{aligned} h(t_1) &= h_{\min} \\ w_{Kg}(t_1) &= 0 \end{aligned} \tag{28}$$

b) Some of the multipliers are discontinuous. Denoting by t_1^- the time just before the entry point and by t_1^+ immediately after, the following relations hold

$$\begin{aligned} \lambda_h(t_1^+) &= \lambda_h(t_1^-) + \nu_0 \\ \lambda_w(t_1^+) &= \lambda_w(t_1^-) - \nu_1 \end{aligned} \tag{29}$$

where $\nu_0 \geq 0$ und $\nu_1 \geq 0$ are additional unknowns. The other multipliers are continuous at the entry point.

c) The controls are continuous at the entry point. This results from $\nu_1 = \mu(t_1)$ and $H = \text{const}$.

The introduction of switching functions is appropriate for the numerical treatment of constrained arcs. These functions may be specified for the entry and exit points as

$$S_1 := h(t_1) - h_{\min} = 0$$
$$S_2 := \mu(t_2) = 0 \tag{30}$$

In case that the control and state constraints, Eq. (10) and (21), become simultaneously active, the optimal bank angle can be determined with the use of Eq. (26) to yield

$$\cos(\mu_a)_{opt} = \frac{1}{C_{L\max}} \frac{mg}{(\rho/2)^2 S} \tag{31}$$

For the Lagrange multiplier μ, Eqs. (25) remains valid with $C_L = C_{L\max}$ and $\mu_a = (\mu_a)_{opt}$ from Eq. (31).

7. NUMERICAL RESULTS

Numerical results have been achieved treating the optimization of dynamic soaring of seabirds as a boundary value problem. The numerical difficulties in determining optimal dynamic soaring trajectories require powerful computational procedures and efficient computer programs. These problems include the precise treatment of switching conditions, internal point and jump conditions, etc. The computer code applied is based on the method of multiple shooting and provides results of high accuracy (Refs. 27 and 28).

In the numerical treatment, data of an albatross is used as a representative for the seabirds performing dynamic soaring. Reference is made to albatross data given in Refs. 1,12,13. The model data applied in the present investigation are given in Table 1.

	Model Data
m [kg]	9.0
S [m^2]	0.65
b [m]	3.47
C_{D0}	0.033
k	0.019
$(L/D)_{\max}$	20

Table 1 Albatross data

The optimal dynamic soaring cycle which requires minimum wind strength is presented in Fig. 6 which provides a perspective view on its form and shows its extensions in the three dimensions. The dotted arrows denote the

begin and end of the optimal cycle, referring to flight conditions of the same energy state and the same direction corresponding to an energy-neutral

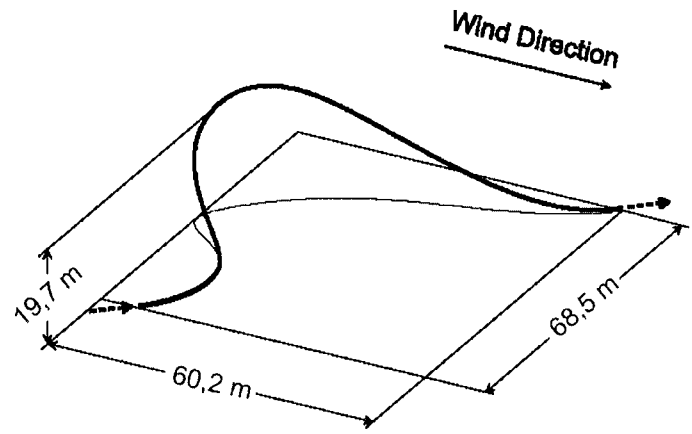


Figure 6. Optimal dynamic soaring cycle requiring minimum wind strength
Optimal cycle time: $(t_{cyc})_{opt} = 7.2$ sec
Wind speed: 9.4 m/s at highest point (h =19.7 m)
5.6 m/s at lowest point ($h_{min} = 0.5$ m)

cycle. As a basic issue, the results concerning the required wind speed and the altitude region compare well with observations and empirical data.

Figs. 7 and 8 present the time histories of state variables, providing more quantitative information about the motion. The altitude range of an optimal cycle, depicted in Fig. 7, extends to about 20 m. The lower altitude limit becomes active for quite a part of the optimal cycle. It is related to the lower curve close to the water surface. The speed behavior is shown in Fig. 8. During the windward climb, the airspeed is larger than the speed relative to the earth while the opposite holds for the leeward descent. The highest speed level is attained in the lower curve.

Results for the optimal controls are presented in Figs. 9 and 10. Fig. 9 shows that the lifting capability is utilized to a large extent. In both the upper and lower curve, the lift coefficient reaches its maximum limit. Correspondingly, the constraint in the lift coefficient becomes active (straight line segments). The unsteady character of dynamic soaring also manifests in the behavior of the bank angle the time history of which is shown in Fig. 10. The greatest bank angle amounts to about 75 deg.

8. CONCLUSIONS

Dynamic soaring which is a flight technique of seabirds for extracting energy from horizontally moving air in a shear flow is treated as an optimal

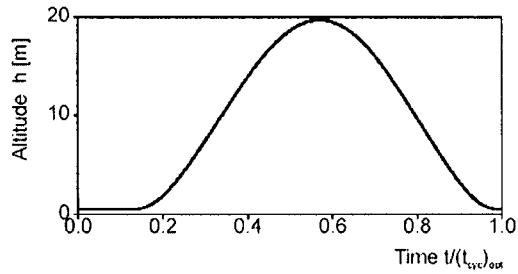


Figure 7. Altitude

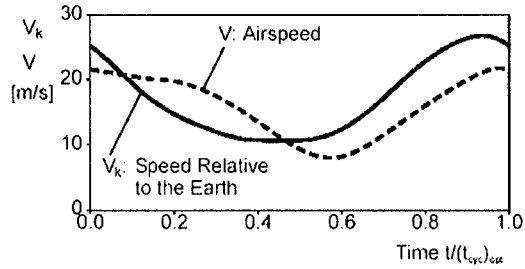


Figure 8. Speeds

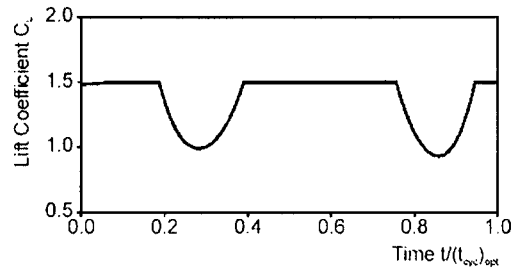


Figure 9. Lift coefficient

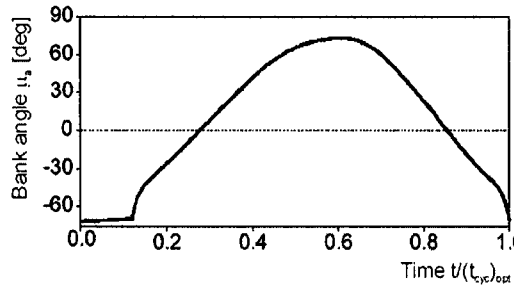


Figure 10. Bank Angle

control problem. It involves a rather complex and highly dynamic maneuver consisting of a sequence of climbing, turning and descending flight phases in order to achieve an energy gain from the shear wind which exists in an altitude layer above the water surface. The basic objective of this paper is to determine the minimum wind strength required for dynamic soaring of seabirds, using optimal control theory. Mathematical models for the motion of a bird in horizontally moving air and for the shear wind are presented. Optimization considerations are applied based on the minimum principle, yielding necessary optimality conditions. Furthermore, switching conditions are introduced in order to deal with an altitude constraint of the dynamic soaring trajectory. Numerical results of high accuracy are generated using an efficient computational procedure based on the method of multiple shooting. The results which concern an albatross as a representative for seabirds performing dynamic soaring concern the minimum shear wind gradient and properties of the related optimal trajectory.

REFERENCES

[1] Idrac, P.: Exprimentelle Untersuchungen über den Segelflug. Berlin: Oldenburg-Verlag 1932.

[2] Rayleigh, J. W. S.: The Soaring of Birds. Nature 27, pp. 534-535, 1883.

[3] Idrac, M. P.: Étude théoretique des manoeuvres des albatros par vent croissant avec l'alitude. C. r. hebd. Séanc. Acad. Sci., Paris 179, pp. 1136-1139, 1924.

[4] Prandtl, L.: Beobachtungen über den dynamischen Segelflug. Zeitschrift für Flugtechnik und Motorluftschiffahrt, Vol. 21, pp. 116, 1930.

[5] Tickell, W.L.N.: Albatrosses. Pica Press, Robertsbridge, 2000.

[6] Cone, C.D., Jr.: A Mathematical Analysis of the Dynamic Soaring Flight of the Albatross with Ecological Interpretations. Virginia Institute of Marine Science, Gloucester Point, Virgina, Special Scientific Report 50, 1964.

[7] Wood, C.J.: The Flight of Albatrosses (A Computer Simulation). IBIS 115, pp. 244-256, 1973.

[8] Lighthill, J.: Aerodynamics Aspects of Animal Flight. Bull. Inst. Math. Applics., 10, pp. 369-393, 1974.
[9] Hendriks, F.: Dynamic Soaring in Shear Slow. AIAA Paper No. 74-1003, 1974.
[10] Lighthill, J.: Animal Flight. In: Mathematical Biofluiddynamics, SIAM, Philadelphia, Chapter 8, pp. 151-178, 1975.
[11] Wilson, J.A.: Sweeping flight and soaring by albatrosses, Nature, Vol. 257, pp. 307-308, 1975.
[12] Berger, M., Göhde, W.: Zur Theorie des Segelfluges von Vögeln über dem Meere. Zool. Jb. Physiol., Vol. 71, pp. 217-224, 1965.
[13] Pennycuick, C.J.: The Flight of Petrels and Albatrosses (Procellaiiformes), Observed in South Georgia and its Vicinity. Phil. Trans. R. Soc. Lond. B 300, pp. 75-106, 1982.
[14] Pennycuick, C. J.: Gust soaring as a basis for the flight of petrels and albatrosses (Procellariiformes). Avian Science 2, pp. 1-12, 2002.
[15] Sachs, G.: Minimalbedingungen für den dynamischen Segelflug. Zeitschrift für Flugwissenschaften und Weltraumforschung, Band 13 (1989), S. 188-198.
[16] Sachs, G., Lesch, K.: Optimal Periodic Trajectorries of Aircraft with Singular Control. 11th IFAC World Congress Proceedings, 1990.
[17] Sachs, G.: Minimaler Windbedarf für den dynamischen Segelflug der Albatrosse. Journal für Ornithologie 134, pp. 435-445, 1993.
[18] Jouventin, Pp., Weimerskirch, H. 1990. Satellite Tracking of Wandering Albatrosses. Nature 343, pp. 746-748.
[19] Prince, P. A., Wood, A. G., Barton, T., Croxall, J. P.: Satellite tracking of wandering albatrosses (Diomedea exulans) in the South Atlantic. Antarctic Science 4 (1), pp. 31-36, 1992.
[20] Alerstam, T., Gudmundsson, G.A., Larsson, B.: Flight tracks and speeds of Antarctic and Atlantic seabirds: radar and optical measurements. Phil. Trans. R. Soc. Lond. 340, pp. 55-67, 1993.
[21] Tuck, G.N., Polacheck, T., Croxall, J.P., Weimerskirch, H., Prince, P.A., Wotherspoon, S.: The Potential of Archival Tags to Provide Long-term Movement and Behaviour Data for Seabirds: First Results from Wandering Albatross Diomedea exulans of South Georgia and the Crozet Islands. EMU Vol. 99, pp. 60-68, 1999.
[22] Weimerskirch, H., Wilson, R.P.: Oceanic respite for wandering albatrosses. Nature, Vol. 406. pp. 955-956, 2000.
[23] Weimerskirch, H., Bonadonna, F., Billeul, F., Mabille, G., Dell'omo, G., Lipp, H-P.: GPS Tracking of Foraging Albatrosses. Science, Vol. 295, pp. 1259, 2002.
[24] Thomas, F.: Grundlagen für den Entwurf von Segelflugzeugen. Motorbuch Verlag, Stuttgart, 1979.
[25] Swolinsky, M.: Beiträge zur Modellierung von Scherwind für Gefährdungsuntersuchungen. Dissertation, Technische Universität Carolo-Wilhelminia zu Braunschweig, 1986.
[26] Bock, H.G.: Numerische Behandlung von zustandsbeschränkten und Chebychef-Steuerungsproblemen. Course R 1.06 of the Carl-Cranz-Gesellschaft, Oberpfaffenhofen, Germany, 1983.
[27] Bulirsch, R.: Die Mehrzielmethode zur numerischen Lösung von nichtlinearen Randwertproblemen und Aufgaben der optimalen Steuerung. Bericht der Carl-Cranz-Gesellschaft, Oberpfaffenhofen, 1971; Nachdruck: Report des Sonderforschungsbereiches 255, Lehrstuhl für Höhere Mathematik und Numerische Mathematik, TU München, 1993.

[28] Oberle, H.J.: Numerische Berechnung optimaler Steuerungen von Heizung und Kühlung für ein realistisches Sonnenhausmodell. Institut für Mathematik der Technischen Universität München, TUM-M8310, 1983.

ON THE CONVERGENCE OF THE MATRICES ASSOCIATED TO THE ADJUGATE JACOBIANS

C. Sbordone
Dept. of Mathematics and Applications "R. Caccioppoli" Napoli, Italy

Abstract: For any $n \times n$ matrix $D \in \mathbb{R}^{n \times n}$ let adj$D$ denote the transpose of its cofactors. If $\det D > 0$ then there exists a symmetric matrix $\mathcal{A} = \mathcal{A}(D)$ with $\det \mathcal{A} = 1$ such that

$$\text{adj}D = (\det D)^{\frac{n-2}{n}} \mathcal{A} D'$$

Where D' is the transpose of D.
For $n = 2$, $K \geq 1$, the set of K-quasiconformal matrices is denoted by

$$\mathcal{Q}_2(K) = \left\{ D \in \mathbb{R}^{2\times 2} : \|D\|^2 \leq K \det D \right\}$$

Furthermore define

$$\mathcal{E}_2(K) = \left\{ \mathcal{A} \in \mathbb{R}^{2\times 2} : \frac{I}{K} \leq \mathcal{A}' = \mathcal{A} \leq KI, \det \mathcal{A} = 1 \right\}$$

For variable matrices $D = D(x) \in \mathcal{E}_2(K)$ for a.e. $x \in \Omega \subset \mathbb{R}^{2\times 2}$, Ω a simply connected and bounded domain, a natural question is to see how does $\mathcal{A} = \mathcal{A}(x, D)$ change with $D(x)$.

Theorem 0.1 *Let* D_j, $D \in L^2(\Omega, R^{2\times 2})$, $D_j(x) \in \mathcal{Q}_2(K)$ *a.e.* $x \in \Omega$. *Assume* $\text{Curl} D_j = 0$ *(resp.* $\text{Div} D_j = 0$*) and* $D_j \rightharpoonup D \neq 0$ *weakly in* $L^2(\Omega, \mathbb{R}^{2\times 2})$. *Then* $D(x) \in \mathcal{Q}_2(K)$ *for a.e.* $x \in \Omega$ *and*

$$\mathcal{A}(x, D_j) \xrightarrow{G} \mathcal{A}(x, D) \qquad (resp.\ \mathcal{A}^{-1}(x, D_j) \xrightarrow{G} \mathcal{A}^{-1}(x, D))$$

1. INTRODUCTION

For any $n \times n$ matrix $D \in \mathbb{R}^{n \times n}$ let adj$D$ denote the transpose of its cofactors, defined by the algebraic identity

$$D \operatorname{adj} D = (\det D)\mathbf{I}$$

where $\mathbf{I}$ is the unit matrix. If D is invertible then

$$\operatorname{adj} D = (\det D)D^{-1}.$$

However, if $D \in \mathbb{R}^{n \times n}$ is an arbitrary matrix with $\det D > 0$ we point out the following representation:

$$\operatorname{adj} D = (\det D)^{\frac{n-2}{n}} \mathcal{A} D^t \tag{1.1}$$

by means of a symmetric matrix $\mathcal{A} = \mathcal{A}(D)$ with $\det \mathcal{A} = 1$.

In the following we supply $\mathbb{R}^{n \times n}$ with the operator norm

$$\|D\| = \max_{|\xi|=1} |D\xi|$$

Proposition 1.1 *For any matrix $D \in \mathbb{R}^{n \times n}$ with $\det D > 0$ there exists a symmetric matrix $\mathcal{A} = \mathcal{A}(D)$, with $\det \mathcal{A} = 1$ such that* (1.1) *holds. Moreover, the sharp ellipticity bounds*

$$\frac{(\det D)^{\frac{2}{n}}}{\|D\|^2} |\xi|^2 \le \langle \mathcal{A}\xi, \xi \rangle \le \frac{\|adjD\|^2}{(\det D)^{\frac{2(n-1)}{n}}} |\xi|^2 \tag{1.2}$$

for any $\xi \in \mathbb{R}^n$, hold.

Remark 1.1 Notice that $\mathcal{A}(D) = \mathbf{I}$ if and only if D is a conformal matrix in $\mathbb{CO}_+(n)$. Our next objective is to show that, in general, $\mathcal{A}(D)$ measures how far D is from being conformal.

For $n = 2$, we can make precise statements. For $K \ge 1$, let us introduce the set $Q_2(K)$ of K *-quasiconformal matrices*, i.e.

$$Q_2(K) = \left\{ D \in \mathbb{R}^{2 \times 2} : \|D\|^2 \le K \det D \right\}$$

and the set

$$\mathcal{E}_2(K)=\left\{\mathcal{A}\in\mathbb{R}^{2\times 2}:\frac{\mathbf{I}}{K}\le\mathcal{A}^t=\mathcal{A}\le K\mathbf{I},\ \det\mathcal{A}=1\right\}$$

We have the following

Theorem 1.1 *A matrix* $D\in\mathbb{R}^{2\times 2}$ *belongs to* $Q_2(K)$, *if and only if the matrix*

$$\mathcal{A}(D)=\begin{cases}\left[\dfrac{D^tD}{\det D}\right]^{-1} & \text{if } \det D>0\\ \mathbf{I} & \text{if } \det D=0\end{cases}\tag{1.3}$$

belongs to $\mathcal{E}_2(K)$. *For* $D\in Q_2(K)$, $\det D>0$, $\mathcal{A}(D)$ *is the unique matrix in* $\mathcal{E}_2(K)$ *such that*

$$\operatorname{adj}D=\mathcal{A}D^t\tag{1.4}$$

We refer to $\mathcal{A}(D)$ as the *inverse distortion tensor* of D. See [1], [3] for analogous results in different settings.

Now we are interested in variable matrices $D=D(x)\in Q_2(K)$, for a. e. $x\in\Omega$ where $\Omega\subset\mathbb{R}^2$ is a simply connected bounded domain. If $D(x)$ is measurable then the pointwise distortion tensor $\mathcal{A}(x)=\mathcal{A}(x,D(x))$, associated with $\operatorname{adj}D(x)$, i.e. satisfying

$$\operatorname{adj}D(x)=\mathcal{A}(x)D(x)^t\tag{1.5}$$

is a measurable matrix field which is uniformly elliptic with $\det\mathcal{A}(x)=1$ a.e.. An important point here is that a converse statement is also true. By the so-called measurable Riemann mapping theorem, given any measurable symmetric matrix field $\mathcal{A}(x)$ in $\Omega\subset\mathbb{R}^2$ such that $\mathcal{A}(x)\in\mathcal{E}_2(K)$ a.e. $x\in\Omega$ we can find $D\in L^2(\Omega,\mathbb{R}^{2\times 2})$ such that $D(x)\in Q_2(K)$ a.e., $\operatorname{Curl}D(x)=0$, for which (1.5) holds. A natural question is to see how does the *pointwise inverse distortion tensor* $\mathcal{A}=\mathcal{A}(x,D)$ change with $D(x)$.

We are particularly concerned with the continuity properties of the operator

$$D\in L^2(\Omega,\mathbb{R}^2)\longrightarrow\mathcal{A}(x,D)\in L^\infty(\Omega,\mathbb{R}^2)$$

when we supply $L^2(\Omega,\mathbb{R}^2)$ with the weak topology. Weak convergence of D_j to D, does not guarantee the convergence of matrices $\mathcal{A}(x,D_j)$ to $\mathcal{A}(x,D)$ in any familiar sense. Note that the condition $\det\mathcal{A}(x,D_j)=1$ is not necessarily preserved under the weak *convergence of $\mathcal{A}(x,D_j)$. The

right notion of convergence to be considered here is the G-convergence, at least in the case $\mathrm{Curl} D_j = 0$ a. e. in Ω (see also related ideas in [13]) . Let $A_j = A_j(x)$ be a sequence of measurable matrix valued functions

$$A_j : \Omega \longrightarrow \mathbb{R}^{2\times 2}$$

satisfying the ellipticity condition

$$\frac{|\xi|^2}{K} \le \langle A_j(x)\xi, \xi \rangle \le K |\xi|^2 \tag{1.6}$$

for a.e. $x \in \Omega$ and $\forall \xi \in \mathbb{R}^2$, with $K \ge 1$. Assume that

$$\det A_j(x) = 1 \text{ a.e. } x \in \Omega \tag{1.7}$$

We are ready for the definition of G-convergence of A_j to a matrix valued function $A = A(x)$ satisfying (1.6) and (1.7).

Definition 1.1 The sequence $\{A_j\}$ G-converges to A if and only if, for $D_j \in L^2_{loc}(\Omega, \mathbb{R}^{2\times 2})$ satisfying

$$\begin{cases} \mathrm{Div}(A_j(x)D_j^t(x)) &= 0 \\ \mathrm{Curl}\, D_j(x) &= 0 \end{cases}$$

the conditions

(*i*) $D_j(x) \rightharpoonup D(x)$

(*ii*) $A_j(x)D_j^t(x) \rightharpoonup A(x)D^t(x)$

are equivalent to each other.

Here, the Div operator is defined as

$$\left(DivM(x)\right)_i = \sum_{k=1}^{2} \frac{\partial M_{ki}(x)}{\partial x_k} \qquad i = 1,2$$

$$M \in L^2_{loc}(\Omega, \mathbb{R}^{2\times 2}) .$$

We will prove the following

Theorem 1.2 *Let Ω be a simply connected bounded open set in $\mathbb{R}^2$. Let D_j belong to $L^2(\Omega, \mathbb{R}^2)$ and $D_j(x) \in Q_2(K)$ a.e. ($K \ge 1$).*

Assume

$$D_j \rightharpoonup D \neq 0 \quad \textit{weakly in } L^2(\Omega, \mathbb{R}^{2\times 2})$$

Then (*i*) *and* (*ii*) *hold true :*

(*i*) *if* Curl $D_j = 0$, *then* $D(x) \in Q_2(K)$ *a.e. and*

$$\mathcal{A}(x, D_j) \xrightarrow{G} \mathcal{A}(x, D)$$

(*ii*) *if* Div $D_j = 0$, then $D(x) \in Q_2(K)$ a.e. and

$$\mathcal{A}(x, D_j)^{-1} \xrightarrow{G} \mathcal{A}(x, D)^{-1}$$

We can prove that such a result does not hold in all dimensions $n > 2$ (See [13]).

2. QUASICONFORMAL MATRICES

In the following we supply the space $\mathbb{R}^{n\times n}$ of real $n \times n$ matrices with the operator norm

$$\|D\| = \max_{|\xi|=1} |D\xi|$$

Sometimes $\mathbb{R}^{n\times n}$ will be equipped with the Hilbert-Schmidt norm

$$|D|^2 = \mathrm{Tr} D^t D$$

where D^t is the transpose of D and $\mathrm{Tr}C$ denotes the trace of matrix $C \in \mathbb{R}^{n\times n}$, $C = (c_{ij})$, that is $\mathrm{Tr}C = \sum_{i=1}^{n} c_{ii}$.

The adjugate matrix of D is the transpose of its cofactors and is denoted by $\mathrm{adj}D$. Therefore we have a non linear multiplicative mapping

$$\mathrm{adj} : \mathbb{R}^{n\times n} \longrightarrow \mathbb{R}^{n\times n}$$

which is a matrix valued homogeneous polynomial of degree $n-1$ (linear if $n = 2$). The adjugate matrix satisfies the rule

$$D\text{adj}\,D = (\det D)\mathbf{I} \tag{2.1}$$

where $\mathbf{I} = (\delta_{ij})$ is the unit matrix.

To any $D \in \mathbb{R}^{n\times n}$ with $\det D > 0$ we associate the positive definite and symmetric matrix

$$\mathcal{A}(D) = \left[\frac{D'D}{(\det D)^{\frac{2}{n}}}\right]^{-1} \tag{2.2}$$

called the *inverse distortion tensor* of D which clearly satisfies $\det \mathcal{A}(D) = 1$.

Proposition 2.1 [11] *Let* D_1, $D_2 \in \mathbb{R}^{n\times n}$ *with* $\det D_i > 0$. *Then*

$$A(D_1) = A(D_2) \tag{2.3}$$

if and only if there exist an orthogonal matrix O *and* $\gamma \in \mathbb{R}$ *such that*

$$D_2 = \gamma O D_1 \tag{2.4}$$

We will give now the

Proof (of Proposition 1.1). By (2.2) we obtain immediately

$$\begin{aligned}(\det D)^{\frac{n-2}{n}}\mathcal{A}D' &= \\ &= (\det D)^{\frac{n-2}{n}}(\det D)^{\frac{2}{n}}D^{-1}[D']^{-1}D' = \\ &= (\det D)D^{-1} = \text{adj}D\end{aligned}$$

For the (sharp) ellipticity bounds (1.2) see [7] p. 112.

The set of two-dimensional K-quasiconformal matrices ($K \geq 1$) is defined as follows

$$Q_2(K) = \left\{D \in \mathbb{R}^{2\times 2} : \|D\|^2 \leq K \det D\right\} \tag{2.5}$$

Let us introduce now the set

$$\mathcal{E}_2(K) = \left\{A \in \mathbb{R}^{2\times 2} : \frac{\mathbf{I}}{K} \leq A' = A \leq K\mathbf{I}, \det A = 1\right\} \tag{2.6}$$

If $\mathcal{A} \in E_2(K)$ and $D \in \mathbb{R}^{2\times 2}$ with positive determinant are related by the identity

$$\operatorname{adj} D = \mathcal{A} D^t \tag{2.7}$$

then it is easy to check that $D \in Q_2(K)$.

To show this, notice that D belongs to $Q_2(K)$ if and only if

$$\operatorname{Tr} D^t D \le (K + \frac{1}{K}) \det D,$$

(see the forthcoming proof)

Hence by (2.7) we have

$$\begin{aligned} \operatorname{Tr} D^t D &= \operatorname{Tr}(\mathcal{A}^{-1}(\operatorname{adj} D) D) = \\ &= \operatorname{Tr}(\mathcal{A}^{-1}(\det D)\mathbf{I}) = \\ &= \operatorname{Tr}(\mathcal{A}^{-1}) \det D \end{aligned}$$

Now, for any $\mathcal{A} \in \mathcal{E}_2(K)$ the inequality

$$\operatorname{Tr} \mathcal{A}^{-1} \le K + \frac{1}{K}$$

holds true.

Let us now pass to the

Proof (of Theorem 1). First of all let us prove that a matrix D belongs to $Q_2(K)$ if and only if

$$|D|^2 \le \left(K + \frac{1}{K} \right) \det D \tag{2.8}$$

If $D = (d_{ij})$, the conformal and anticonformal part are represented by the matrices

$$D^{\pm} = \frac{1}{2} \begin{pmatrix} d_{11} \pm d_{22} & d_{12} \mp d_{21} \\ d_{21} \mp d_{12} & d_{22} \pm d_{11} \end{pmatrix}$$

It is then immediate that

$$\|D\| = |D^+| + |D^-|$$

$$|D|^2 = \mathrm{Tr}(D'D) = 2(|D^+|^2 + |D^-|^2)$$

$$\det D = |D^+|^2 - |D^-|^2$$

Hence the distortion inequality

$$\|D\|^2 \le K \det D$$

is easily seen to be equivalent to

$$|D^-| \le \frac{K-1}{K+1}|D^+|$$

This, in turn, is equivalent to (2.8). Now let $D \in Q_2(K)$. If $\det D = 0$ then $\mathcal{A}(D) = \mathbf{I}$ and therefore $\mathcal{A}(D)$ belongs to $\mathcal{E}_2(K)$. If $\det D > 0$, consider the inverse matrix of $\mathcal{A}$

$$\mathcal{G} = \frac{D'D}{\det D}$$

Then, obviously $\det \mathcal{G} = 1$ and the distortion inequality (2.8) is equivalent to

$$\mathrm{Tr}(\mathcal{G}) \le K + \frac{1}{K}$$

Let λ and $\frac{1}{\lambda}$ be the eigenvalues of $\mathcal{G}$. Then the last inequality means that

$$\lambda + \frac{1}{\lambda} \le K + \frac{1}{K}$$

hence $\frac{1}{K} \le \lambda \le K$ and the first statement of the theorem is proven. From previous considerations the identity (1.4) follows immediately. The uniqueness of the matrix $\mathcal{A}$ in the class $\mathcal{E}_2(K)$ satisfying (1.4) is obvious.

Remark 2.1 If $D' = D$ then it is easy to check that the following conditions

(i) $D \in Q_2(K)$

(ii) $\left(\sqrt{K}+1/\sqrt{K}\right)^2 \mathrm{Tr}\left(D^t D\right) \le \left(K+1/K\right)\left(\mathrm{Tr}\, D\right)^2$

(iii) $|\xi|^2 / K \le \langle D\xi, \xi\rangle \le \sqrt{K}\,|\xi|^2$

are equivalent to each other.

3. THE *G*-CONVERGENCE OF THE ASSOCIATED MATRICES FOR $n = 2$

We begin the present section with a simple example. Let $f_j = f_j(x_1, x_2)$ be of the special form

$$f_j(x_1, x_2) = (f_j^{(1)}(x_1), f_j^{(2)}(x_2))$$

where $f_j^{(i)}(x_i) : \mathbb{R} \to \mathbb{R}$ verify $(K > 1)$

$$\frac{1}{\sqrt{K}} \le \frac{d}{dt} f_j^{(i)}(t) \le \sqrt{K}$$

for any $t \in \mathbb{R}$. The matrix associated with $\mathrm{adj}Df_j$ is diagonal

$$\mathcal{A}_j(x) = \begin{pmatrix} \dfrac{\dfrac{\partial f_j^{(2)}}{\partial x_2}(x_2)}{\dfrac{\partial f_j^{(1)}}{\partial x_1}(x_1)} & 0 \\ 0 & \dfrac{\dfrac{\partial f_j^{(1)}}{\partial x_1}(x_1)}{\dfrac{\partial f_j^{(2)}}{\partial x_2}(x_2)} \end{pmatrix}$$

If we assume $Df_j \rightharpoonup Df$ weakly in $L^2(\mathbb{R}^{2\times 2})$, the sequences $\mathcal{A}_j$ does not converge weakly in $L^2(\Omega, \mathbb{R}^{2\times 2})$ to the matrix associated $\mathcal{A}(x)$ to $Df(x)$

$$\mathcal{A}(x)=\begin{pmatrix} \dfrac{\dfrac{\partial f^{(2)}}{\partial x_2}(x_2)}{\dfrac{\partial f^{(1)}}{\partial x_1}(x_1)} & 0 \\ 0 & \dfrac{\dfrac{\partial f^{(1)}}{\partial x_1}(x_1)}{\dfrac{\partial f^{(2)}}{\partial x_2}(x_2)} \end{pmatrix}$$

unless $Df_j \to Df$ strongly in L^2. On the other hand, by means of a characterization of G-convergence of diagonal matrices whose entries are products of functions of one variable (due to L. Tartar [15]) we deduce that

$$\mathcal{A}_j \xrightarrow{G} \mathcal{A}$$

This is consistent with Theorem 1.2 whose proof we present below, which relies on a deep result of G-compactness.

Proof of Theorem 1.2. (*i*) Let $f_j, f \in W^{1,2}(\Omega, \mathbb{R}^2)$ satisfy $Df_j = D_j$, $Df = D$. By our assumption :

$$Df_j \rightharpoonup Df \tag{3.1}$$

we obtain, via a classical result of Reshetnyak [11],[7]

$$\det Df_j \rightharpoonup \det Df \qquad \text{weakly in } L^1_{loc}(\Omega)$$

and so $Df(x) \in \mathcal{Q}_2(K)$, a.e. in Ω in virtue of the lower semicontinuity of the norm. By the G-compactness theorem [13] (see also [4], [8] for more general cases of degenerate elliptic equations) we may assume

$$\mathcal{A}(x, Df_j) \xrightarrow{G} A_0(x)$$

Since

$$\mathrm{Div}(\mathcal{A}(x, Df_j)Df_j^t) = \mathrm{Div}(\mathrm{adj}Df_j) = 0$$

by definition of G-convergence we have

$$\mathcal{A}(x, Df_j)Df_j' \rightharpoonup A_0(x)Df'$$

But (3.1) and the definition of $\mathcal{A}(x, Df_j)$ imply

$$\mathcal{A}(x, Df_j)Df_j' \rightharpoonup \mathcal{A}(x, Df)Df'$$

and so

$$\mathcal{A}(x, Df)Df' = A_0(x)Df'$$

Since $Df' \neq 0$ a.e., we deduce

$$\mathcal{A}(x, Df) = A_0(x).$$

(*ii*) Taking into account that Ω is a simply connected open set in $\mathbb{R}^2$,the condition $\operatorname{Div} D_j = 0$ implies $D_j = \operatorname{adj} Dg_j$ for some $g_j \in W^{1,2}(\Omega; \mathbb{R}^2)$. Hence, by the definition of $\mathcal{A}(x, Dg_j)$

$$D_j = \mathcal{A}(x, Dg_j)Dg_j'$$

which, of course can be rewritten as

$$Dg_j = \mathcal{A}(x, Dg_j)^{-1} D_j'$$

Now, the hypothesis $D_j \rightharpoonup D$ in $L^2(\Omega, \mathbb{R}^{2\times 2})$ is equivalent to $Dg_j \rightharpoonup Dg$ and so, by part (*i*) we have $D(x) \in Q_2(K)$ a.e. in Ω, and

$$\mathcal{A}(x, Dg_j) \xrightarrow{G} \mathcal{A}(x, Dg).$$

Note that

$$\mathcal{A}(x, Dg_j) = \mathcal{A}(x, D_j)^{-1} \tag{3.2}$$

Actually (3.2) is a consequence of the equivalence

$$C = \operatorname{adj} D \iff D = \operatorname{adj} C$$

which holds for all 2×2 matrices $C, D \in \mathbb{R}^{2\times 2}$. □

The results of this paper have been partially announced in [12].

ACKNOWLEDGMENTS

Research supported by MIUR and GNAMPA-INdAM.

REFERENCES

[1] C C. Capone. Quasiharmonic fields and Beltrami operators. *Comment. Math. Univ. Carolinae* **43** (2) (2002), 363-377.

[2] R. De Arcangelis and P. Donato. On the convergence of Laplace-Beltrami operators associated to quasiregular mappings. *Studia Math.* **86**, (3) (1987), 189-204.

[3] L. D'onofrio and L. Greco. A counterexample in *G*-convergence of nondivergence elliptic operators. Proc. Royal Soc. Edinburgh, **133A**, (2003), 1299-1310 .

[4] M.R. Formica. On the Γ-convergence of Laplace-Beltrami operators in the plane. *Annales Academiæ Scientiarium Fennicæ. Mathematica* **25** (2000), 423-438.

[5] G.A. Francfort and F. Murat. Optimal bounds for conduction in two dimensional, two phase, anisotropic media. *Non Classical Continuum Mechanics.* London Math. Soc. Lecture Notes Ser. **122**, Cambridge Univ. Press (1987), 197-212.

[6] T. Iwaniec, P. Koskela, G. Martin and C. Sbordone. Mappings of finite distortion: $L^n \log^{\alpha} L$ -integrability. *J. London Math. Soc.* (2) **67** (2003), no. 1, 123–136.

[7] T. Iwaniec and G. Martin. *Geometric function theory and non-linear analysis.* Oxford Mathematical Monographs (2001).

[8] F. Giannetti, T. Iwaniec, L.Kovalev, G. Moscariello and C. Sbordone. *On G-compactness of the Beltrami Operators,* NATO Adv. Res. Workshop on Nonlinear Homogenization, Kazimierz Dolny, June 2003.

[9] P. Marcellini and C. Sbordone. An approach to the asymptotic behaviour of elliptic-parabolic operators. *J. Math. Pures Appl.* (9) **56** (1977), no. 2, 157–182.

[10] F. Murat and L. Tartar. *H*-convergence. *Topics in the mathematical modelling of composite materials.* Progr. Nonlinear Differential Equations Appl. Birkhäser Boston. **31**. (1997) 21–43.

[11] Yu G. Reshetnyak. Mappings of bounded deformations as extremals of Dirichlet type integrals *Sibirsk. Mat. Ž* **9** (1968) 625-666.

[12] C. Sbordone, On the Γ-convergence of matrix fields related to the adjugate Jacobian, Comptes Rendus, Ac. Sci. Paris Ser I 337 (2003) 165-170.

[13] C. Sbordone, On the Convergence of the Associated Matrix to the Adjugate Jacobian, (2004) to appear.

[14] S. Spagnolo. Some convergence problems. *Symposia Mathematica, Vol. XVIII (Convegno sulle Transformazioni Quasiconformi e Questioni Connesse, INDAM, Rome, 1974).* (1976) 391–398.

[15] L. Tartar. Convergence d'operators differentials. *Analisi Convessa Appl.* Roma (1974) 101-104.

QUASI-VARIATONAL INEQUALITIES APPLIED TO RETARDED EQUILIBRIA IN TIME-DEPENDENT TRAFFIC PROBLEMS

L. Scrimali
Dept. of Mathematics and Computer Sciences, University of Catania, Catania, Italy

Abstract: We present a time-dependent and elastic model of transportation networks. We also take into consideration the presence of delay effects and propose a variational approach to the corresponding traffic equilibrium problem.

Key words: Quasi-variational inequalities, delay, dynamics of flows.

1. INTRODUCTION

In this paper we deal with equilibrium problems in time-dependent and elastic traffic networks. The presented model explicitly depends on time and, in order to allow for possible congestion effects, some capacity restrictions on flows are imposed. Moreover, we adopt the assumption of elastic travel demands, in the sense that they are affected by the equilibrium pattern.

We are mainly interested in studying delay effects in the distribution of flows through the network. In fact, the finite speed of the information which travel through the network and the time that the users spend to choose the best route slow down the propagation of flows. Hence the conservation of flows condition is required at a certain instant, but it is satisfied only later; consequently we have to cope with a *retarded* equilibrium pattern (Raciti, 2001). In addition, we suggest a variational approach to the problem, showing how the retarded equilibrium flow solves a quasi-variational

inequality, for which we are able to ensure the existence of solutions. In particular, we adopt an integral formulation of the problem (Daniele et al., 1999; Friesz et al.,1993; Gwinner, 2003; Raciti, 2001; Raciti and Scrimali, 2003), which is the most suitable since we want to focus our attention only on a particular interval of time.

Moreover, we are concerned with considering the dynamics of flows. It results indeed that, under some regularity conditions, the distribution of flows follows the first-in-first-out queue discipline (Friesz et al., 1993).

2. THE VARIATIONAL FORMULATION

Let us consider a time-dependent and elastic traffic network model where:

W is the set of Origin Destination (O/D) pairs w_j, $j=1,2,...,l$;

R_j is the set of routes R_r, $r=1,2,...,m$, which connect the pair w_j;

$\Phi=(\varphi_{jr})_{j=1,...,l\, r=1,...,m}$ is the incidence matrix, where $\varphi_{jr}=1$ if $R_r \in R_j$ and $\varphi_{jr}=0$ otherwise.

Let us suppose that the functional space for the trajectories of route flows is $L^2(I;\mathbb{R}^m_+)$, $I \subseteq \mathbb{R}$. Thus, the vector of flows is given by $F(u)=(F_1(u),...,F_m(u)) \in L^2(I;\mathbb{R}^m_+)$. We also suppose that the flow $F_r(u), r=1,...,m$ is nondecreasing in I and bounded by some capacity restrictions, denoted by $\lambda(u)=(\lambda_1(u),...,\lambda_m(u))$ and $\mu(u)=(\mu_1(u),...,\mu_m(u))$, where $0 \le \lambda_r(u) \le \mu_r(u)$, $r=1,...,m$. Moreover, let us assume that $C:[0,T]\times\mathbb{R}^m_+ \to \mathbb{R}^m_+$ is the cost function on routes.

Since we want to take into consideration the propagation of flows through the network, the presence of delay effects can not be neglected. In fact, the information travel through the network at a finite speed, hence it is reasonable to suppose that users take a certain time before evaluating the best route and consequently adjusting their route choices. Therefore, it is plausible to expect that demand requirements imposed at time t are satisfied after a delay, namely after that the distribution of flows is complete. Now, let us introduce the delay vector $d(t,l)=(d_1(t,l_1),...,d_m(t,l_m)) \in \mathbb{R}^m_+$, where $t \in [0,T]$ is the departure time and l_r is the length of the route R_r, $r=1,...,m$.

In our model, we assume that all the components of the delay are nonnegative, thus we do not deal with the case of early arrivals. We also suppose that $d_r(t,l_r), r=1,...,m$ is a nondecreasing linear function with respect to t. Hence we want to cope with the most unfavorable case of a bottleneck in $[0,T]$ where the delay increases.

Now, let us define the function $\psi(t)$ as $\psi(t)=t+d(t,l)$, $t \in [0,T]$, with the image covered by I, *i.e.* $\psi([0,T]) \subseteq I$. We are then entitled to consider the

retarded flow $F(t+d(t,l))=(F_1(t+d_1(t,l_1)),...,F_m(t+d_m(t,l_m))):[0,T]\to\mathbb{R}_+^m$ and, analogously, the retarded capacity constraints $\lambda(t+d(t,l))=(\lambda_1(t+d_1(t,l_1)),...,\lambda_m(t+d_m(t,l_m)))$ and $\mu(t+d(t,l))=(\mu_1(t+d_1(t,l_1)),...,\mu_m(t+d_m(t,l_m)))$.

Remark 1. *We want to highlight that* $d(t,l)\in L^2(0,T;\mathbb{R}_+^m)$, *it is indeed a nonnegative, nondecreasing with respect to time and bounded function and hence it is measurable and Lebesgue-integrable. Therefore, since* $F(t+d(t,l))$ *is in turn nonnegative, nondecreasing with respect to time and bounded, it results that* $F(t+d(t,l))\in L^2(0,T;\mathbb{R}_+^m)$.

Thus, we are able to introduce the following definition of retarded equilibrium flow (Raciti, 2001).

Definition 1. *A flow* $H(t+d(t,l))\in L^2(0,T;\mathbb{R}_+^m)$ *is said to be a retarded equilibrium flow if and only if* $\forall w_j\in W,\forall R_q,R_s\in\mathcal{R}_j$ *and a.e. in* $[0,T]$*:*

$$\begin{aligned}&C_q(t,H(t+d(t,l)))<C_s(t,H(t+d(t,l)))\Rightarrow\\&H_q(t+d_q(t,l_q))=\mu_q(t+d_q(t,l_q))\,or\\&H_s(t+d_s(t,l_s))=\lambda_s(t+d_s(t,l_s)).\end{aligned}\tag{1}$$

To describe better the behavior of flows, we assume that each flow $F(t)$ fulfills the following uniform integral continuity condition:

$$\lim_{|h|\to 0}\int_0^T |F(t+h+d(t+h,l))-F(t+d(t,l))|^2\,dt=0\tag{2}$$

uniformly in F, namely $\forall\varepsilon>0\ \ \exists\delta>0$ such that $\forall h\in\mathbb{R}$, $|h|<\delta$ and $\forall F$

$$\int_0^T |F(t+h+d(t+h,l))-F(t+d(t,l))|^2\,dt<\varepsilon,$$

provided that $F(t+d(t,l))=0$ if $t+d(t,l)\notin[0,T]$. Let us introduce (Friesz et al., 1993) the flow rate $v(t)=(v_1(t),...,v_m(t))$, which represents the derivative of the route flow which enters the first link of the route at time t: $\frac{d}{dt}F_r(t+d(t,l_r))=v_r(t), r=1,...,m$, a.e. in $[0,T]$. Condition (2) is obviously satisfied if, for instance, we require that $\exists\eta\in\mathbb{R}_+:\|v(t)\|_{L^2}\le\eta\ \ \forall v(t)$ or if we assume that flows verify an integral Hölder condition: $\int_0^T |F(t+h+d(t+h,l))-F(t+d(t,l))|^2\,dt\le L|h|^\alpha$, $0<\alpha\le 1, L\in\mathbb{R}_+$. It is worth noting the importance of the uniform integral continuity condition which allows us to take under control the flow rates.

Thus we can introduce the following set:

$$E = \Big\{F(t+d(t,l)) \in L^2(0,T;R_+^m) : \lambda_r(t+d_r(t,l_r)) \le F_r(t+d_r(t,l_r)) \le$$
$$\le \mu_r(t+d_r(t,l_r)) \text{ a.e. in } [0,T],\ \ r=1,2,\ldots,m;$$
$$F_r(t_1+d_r(t_1,l_r)) \le F_r(t_2+d_r(t_2,l_r))\ \forall t_1,t_2 \text{ a.e. in}[0,T],\ r=1,2,\ldots,m,$$
$$\lim_{|h|\to 0}\int_0^T \mid F(t+h+d(t+h,l)) - F(t+d(t,l)) \mid^2 dt = 0$$
$$\text{uniformly in } F,\ F(t+d(t,l)) = 0 \text{if} t+d(t,l) \notin [0,T]\ \Big\}.$$

The set of feasible flows is then the set-valued function $K : E \to 2^{\mathbb{R}_+^m}$ given by:

$$K_d(H) = \{F(t+d(t,l)) \in E : \sum_{r=1}^{m} \varphi_{jr} F_r(t+d_r(t,l_r)) = \frac{1}{T}\int_0^T \rho_j(t,H(\tau))d\tau$$
$$\text{a.e.in}[0,T], j=1,2,\ldots,l\},$$

where $\rho(t,H)$, defined in $[0,T]\times E \to \mathbb{R}_+^l$, is the elastic demand depending on the equilibrium pattern. We also suppose that the condition $\Phi\lambda(t+d(t,l)) \le \Phi F(t+d(t,l)) \le \Phi\mu(t+d(t,l))$ is satisfied a.e. in $[0,T]$, so that the set $K_d(H)$ is nonempty.

Now, we present the following theorem, which gives a complete characterization of the retarded equilibrium flow (Maugeri, 1998; Raciti, 2001).

Theorem 1. *A feasible flow is a retarded equilibrium flow if and only if it solves the following retarded quasi-variational inequality (R.Q.V.I.):*

$$H(t+d(t,l)) \in K_d(H)$$

$$\int_0^T C(t,H(t+d(t,l)))(F(t+d(t,l)) - H(t+d(t,l)))dt \ge 0, \qquad (3)$$

$$\forall F(t+d(t,l)) \in K_d(H).$$

Proof. We argue by reductio ad absurdum and suppose that (3) does not hold, so that there exist $w_j \in W$, $R_q\ R_s \in \mathcal{R}_j$ and a set $G \subset [0,T]$ with positive measure such that a.e. in G it results that:

$$C_q(t,H(t+d(t,l))) < C_s(t,H(t+d(t,l))) \Rightarrow$$
$$H_q(t+d_q(t,l_q)) < \mu_q(t+d_q(t,l_q)) \text{or}$$
$$H_s(t+d_s(t,l_s)) > \lambda_s(t+d_s(t,l_s)).$$

Let us set:

$$\delta(t+\bar{d}) = \min_{t\in[0,T]}\{\mu_q(t+d_q(t,l_q)) - H_q(t+d_q(t,l_q)), H_s(t+d_s(t,l_s)) - \lambda_s(t+d_s(t,l_s))\},$$

with $\delta(t+\bar{d}) > 0$ a.e. in G. We construct the following flow $F \in K_d(H)$:

$$F_q(t+d_q(t,l_q)) = \begin{cases} H_q(t+d_q(t,l_q)) + \delta(t+\bar{d}) & \forall t \in G \\ H_q(t+d_q(t,l_q)) & \forall t \in [0,T] \setminus G \end{cases}$$

$$F_s(t+d_s(t,l_s)) = \begin{cases} H_s(t+d_s(t,l_s)) - \delta(t+\bar{d}) & \forall t \in G \\ H_s(t+d_s(t,l_s)) & \forall t \in [0,T] \setminus G \end{cases}$$

$$F_r(t+d(t,l_r)) = H_r(t+d(t,l_r))\ r \neq q,s\ \forall t \in [0,T].$$

$F \in K_d(H)$, therefore we can write:

$$\int_0^T C(t, H(t+d(t,l)))(F(t+d(t,l)) - H(t+d(t,l)))dt =$$
$$\int_0^T \delta(t+\bar{d})(C_q(t, H(t+d(t,l))) - C_s(t, H(t+d(t,l))))dt < 0.$$

Now we suppose that H is a retarded equilibrium flow and prove that it solves the (R.Q.V.I.) It results that $C_q(t, H(t+d(t,l))) < C_s(t, H(t+d(t,l)))$ implies that $H_q(t+d_q(t,l_q)) = \mu_q(t+d_q(t,l_q))$ or
$H_s(t+d_s(t,l_s)) = \lambda_s(t+d_s(t,l_s))$.
$\forall w_j \in W$ let us set:

$$A = \{R_q \in \mathcal{R}_j : H_q(t+d_q(t,l_q)) < \mu_q(t+d_q(t,l_q))\}$$
$$B = \{R_s \in \mathcal{R}_j : H_s(t+d_s(t,l_s)) > \lambda_s(t+d_s(t,l_s))\}.$$

It follows that:
$C_s(t, H(t+d(t,l))) \leq C_q(t, H(t+d(t,l)))\ \ \forall R_q \in A, \forall R_s \in B$.
Thus, there exists $\gamma_{w_j,t} \in \mathbb{R}$ such that:
$\sup_B C_s(t, H(t+d(t,l))) \leq \gamma_{w_j,t} \leq \inf_A C_q(t, H(t+d(t,l)))$. Let us consider $F(t+d(t,l)) \in K_d(H)$, $\forall R_r \in \mathcal{R}_j$, we have that if:

$$C_r(t, H(t+d(t,l))) < \gamma_{w_j,t}$$

then $R_r \notin A$, hence:

$$H_r(t+d_r(t,l_r)) = \mu_q(t+d_q(t,l_q)); \quad F_r(t+d_r(t,l_r)) - H_r(t+d_r(t,l_r)) \le 0.$$

Therefore, we obtain that:

$$(C_r(t,H(t+d(t,l))) - \gamma_{w_j,t})(F_r(t+d_r(t,l_r)) - H_r(t+d_r(t,l_r))) \ge 0.$$

If

$$C_r(t,H(t+d(t,l))) > \gamma_{w_j,t}$$

then

$$(C_r(t,H(t+d(t,l))) - \gamma_{w_j,t})(F_r(t+d_r(t,l_r)) - H_r(t+d_r(t,l_r))) \ge 0.$$

We conclude that:

$$\sum_{R_r \in \mathcal{R}_j} C_r(t,H(t+d(t,l)))(F_r(t+d_r(t,l_r)) - H_r(t+d_r(t,l_r))) \ge 0;$$

summing up $\forall w_j \in W$ and integrating, we find that:

$$\int_0^T C(t,H(t+d(t,l)))(F(t+d(t,l)) - H(t+d(t,l)))dt \ge 0.$$

3. AN EXISTENCE RESULT

In this section, we provide a theorem for the existence of solutions to the retarded model, generalizing previous results (De Luca, 1997; De Luca and Maugeri, 1992 A; Raciti and Scrimali, 2003). First, let us recall the following result adapted to our case (Tan, 1985):

Theorem 2. *Let X be a locally convex, Hausdorff topological vector space, E a nonempty compact, convex subset of X, $C: E \to X^*$ a continuous function, $K: E \to 2^E$ a closed lower semicontinuous multifunction with $K(H) \subset E$ nonempty, compact, convex $\forall H \in E$. Then, there exists a solution for the quasi-variational inequality:*

$$H \in K(H),\ \langle F - H, C(H) \rangle \geq 0 \quad \forall F \in K(H).$$

Now, let us consider the L^2-version of Ascoli's theorem, due to Riesz, Fréchet and Kolmogorov (Brexis, 1983) adapted to our case:

Theorem 3. *Let F be a bounded set in $L^2([0,T])$. Let us suppose that*

$$\lim_{|h| \to 0} \left\| F(t+h+d(t+h,l)) - F(t+d(t,l)) \right\|_{L^2} = 0 \quad \text{uniformly in } F \in \mathcal{F},$$

provided that $F(t+d(t,l)) = 0$ if $t + d(t,l) \notin [0,T]$. Then F has compact closure in $L^2([0,T])$.

Now, we are able to prove the following result.

Theorem 4. *Let us assume that the functions*

$$C : [0,T] \times \mathbb{R}_+^m \to R_+^m \text{ and } \rho : [0,T] \times \mathbb{R}_+^m \to \mathbb{R}_+^l$$

satisfy the following conditions:

a) $C(t,v)$ *is measurable in* t $\forall v \in \mathbb{R}_+^m$, *continuous in* v *for* t *a.e. in* $[0,T]$,

$$\exists \gamma \in L^2(0,T) : |C(t,v)| \leq \gamma(t) + |v|;$$

b) $\rho(t,v)$ *is measurable in* t $\forall v \in \mathbb{R}_+^m$, *continuous in* v *for* t *a.e. in* $[0,T]$,

$$\exists \psi \in L^1(0,T) : |\rho(t,v)| \leq \psi(t) + |v|^2;$$

c) $\exists h(t) \geq 0$ *a.e. in* $[0,T]$, $h \in L^2(0,T)$:

$$\forall v_1, v_2 \in \mathbb{R}_+^m,\ |\rho(t,v_1) - \rho(t,v_2)| \leq h(t) |v_1 - v_2|;$$

Then the R.Q.V.I. admits a solution.

Proof. At first we can observe that under the hypotheses a), b) and since $H(t+d(t,l)) \in L^2(0,T;\mathbb{R}_+^m)$, it results that

$$C(t, H(t+d(t,l))) \in L^2(0,T;\mathbb{R}_+^m) \text{ and } \rho(t, H(t)) \in L^1(0,T;\mathbb{R}_+^l).$$

Moreover, by a) and b) it follows that C and ρ belong to the class of Nemytskii operators, therefore if $\{H^n\} \longrightarrow^{L^2} H$ then

$$\left\| C(t,H^n(t+d(t,l))) - C(t,H(t+d(t,l))) \right\|_{L^2} \to 0, \ \left\| \rho(t,H^n(t)) - \rho(t,H(t)) \right\|_{L^1} \to 0$$

and the functions C, ρ are L^2 and L^1-continuous respectively.

Now we prove that $K_d(H)$ is a closed multifunction, for this aim we show that:

$$\forall \{H^n\} \longrightarrow^{L^2} H, \ \forall \{F^n\} \longrightarrow^{L^2} F \text{ with } F^n \in K_d(H^n), \ \forall n \in \mathbb{N},$$

then $F \in K_d(H)$.

Let $\{H^n\}, \{F^n\} \in L^2$ be two arbitrary convergent sequences. Since $F^n \in K_d(H^n)$ we have that:

$$\lambda_r(t + d_r(t,l_r)) \le F_r^n(t + d_r(t,l_r)) \le \mu_r(t + d_r(t,l_r)) \text{ a.e. in } [0,T], \ r = 1,2,...,m,$$

and the convergence of the sequence $\{F^n\}$ in L^2 implies that even F satisfies capacity constraints. It can be easily proved that F verifies (2), since F^n satisfies the above assumption and L^2-converges to F.

Moreover, the following relationship holds:

$$\sum_{r=1}^{m} \varphi_{jr} F_r^n(t + d_r(t,l_r)) = \frac{1}{T} \int_0^T \rho_j(t, H^n(\tau)) d\tau \text{ a.e. in } [0,T], \ j = 1,2,\ldots,l.$$

The left-hand side converges almost everywhere to $\sum_{r=1}^{m} \varphi_{jr} F_r^n(t + d_r(t,l_r))$; the right-hand side, meanwhile, results in:

$$\begin{aligned} &\left| \int_0^T \rho_j(t, H^n(\tau)) d\tau - \int_0^T \rho_j(t, H(\tau)) d\tau \right| \\ &\le \int_0^T \left| \rho_j(t, H^n(\tau)) - \rho_j(t, H(\tau)) \right| d\tau \\ &\le h(t) \int_0^T \left| H^n(\tau) - H(\tau) \right| d\tau. \end{aligned}$$

By applying c) and considering that the convergence of $\{H^n\}$ in L^2 implies the convergence also in L^1, we achieve the assertion.

In order to show the lower semi-continuity of $K_d(H)$, we prove that $\forall \{H^n\} \longrightarrow^{L^2} H, \forall F \in K_d(H)$ there exists $\{F^n\}$ such that:

$$\{F^n\} \longrightarrow^{L^2} F \text{ with } F^n \in K_d(H^n) \ \forall n \in \mathbb{N}.$$

Let us consider an arbitrary $\{H^n\} \longrightarrow^{L^2} H, F \in K_d(H)$ and fix $n \in \mathbb{N}, t \in [0,T]$. We introduce the following sets:

$$A_j = \{r \in \{1,2,...,m\} : \varphi_{jr} = 1\}$$
$$B_j(n,t) = \{r \in A_j : \overline{\rho}_j(t) - \overline{\rho}_j^n(t) \le 0\}$$
$$C_j(n,t) = \{r \in A_j : 0 < \overline{\rho}_j(t) - \overline{\rho}_j^n(t) < F_r(t + d_r(t,l_r)) - \lambda_r(t + d_r(t,l_r))\}$$
$$Dj(n,t) = \{r \in A_j : F_r(t + d_r(t,l_r)) - \lambda_r(t + d_r(t,l_r)) \le \overline{\rho}_j(t) - \overline{\rho}_j^n(t)\}$$

where $j \in \{1,2,...,l\}$ and

$$\overline{\rho}_j(t) = \frac{1}{T}\int_0^T \rho_j(t, H(\tau))d\tau, \ \overline{\rho}_j^n(t) = \frac{1}{T}\int_0^T \rho_j(t, H^n(\tau))d\tau.$$

Let us construct the following sequence $\{F^n\}$:

$$F_r^n(t + d_r(t,l_r)) = \begin{cases} F_r(t + d_r(t,l_r)) & \text{if } \in B_j \cup D_j, t \in [0,T] \\ F_r(t + d_r(t,l_r)) - \dfrac{\overline{\rho}_j(t) - \overline{\rho}_j^n(t)}{\sum\limits_{s \in C_i} \varphi_{js}} & r \in C_j, t \in [0,T]. \end{cases}$$

If $r \in B_j \cup D_j$, then $F_r^n(t + d(t,l)) = F_r(t + d_r(t,l_r))$ a.e. in $[0,T]$ and, since $F \in K_d(H)$, it results that:

$$\lambda_r(t + d_r(t,l_r)) \le F_r^n(t + d_r(t,l_r)) \le \mu_r(t + d_r(t,l_r)) \text{ a.e. in } [0,T].$$

If $r \in C_j$, then it is easy to show that a.e. in $[0,T]$:

$$\lambda_r(t + d_r(t,l_r)) < F_r^n(t + d_r(t,l_r)) = F_r(t + d_r(t,l_r)) - \frac{\overline{\rho}_j(t) - \overline{\rho}_j^n(t)}{\sum\limits_{s \in C_j} \varphi_{js}} < \mu_r(t + d_r(t,l_r)).$$

Therefore, F_r^n satisfies the capacity restrictions $\forall r = 1,2,...,m, \forall n \in \mathbb{N}$. Moreover,

$$\sum_{r=1}^{m} \varphi_{jr} F_r^n(t + d_r(t,l_r)) = \sum_{r \in A_j} \varphi_{jr} F_r^n(t + d_r(t,l_r)) = \sum_{r \in B_j \cup D_j} \varphi_{jr} F_r(t + d_r(t,l_r)) +$$

$$+ \sum_{r \in C_j} \varphi_{jr} (F_r(t + d_r(t,l_r)) - \frac{\overline{\rho}_j(t) - \overline{\rho}_j^n(t)}{\sum_{s \in C_j} \varphi_{js}})$$

$$= \sum_{r \in A_j} \varphi_{jr} F_r(t + d_r(t,l_r)) - (\overline{\rho}_j(t) - \overline{\rho}_j^n(t)) = \overline{\rho}_j^n(t).$$

As demand requirements are verified and assumption (2) holds, we deduce that F^n belongs to $K_d(H^n)\ \forall n \in \mathbb{N}$. To show that $\{F^n\}$ converges to F in L^2, we proceed as follows. Let t be in $[0,T]$

$$\int_0^T (\sum_{r=1}^{m} \varphi_{jr}(F_r^n(t + d_r(t,l_r)) - F_r(t + d_r(t,l_r))))^2 dt = \int_0^T (\overline{\rho}_j(t) - \overline{\rho}_j^n(t))^2 dt =$$

$$= \frac{1}{T^2} \int_0^T (\int_0^T [\rho_j(t, H(\tau)) - \rho_j(t, H^n(\tau))] d\tau)^2 dt \leq$$

$$\leq \frac{1}{T^2} T \int_0^T | \rho_j(t, H(\tau)) - \rho_j(t, H^n(\tau)) |^2 \, d\tau dt \leq$$

$$\leq \frac{1}{T} (\int_0^T h^2(t) dt)(\int_0^T | H(\tau) - H^n(\tau) |^2 \, d\tau).$$

Additionally, it results that:

$$(\sum_{r=1}^{m} \varphi_{jr}(F_r^n(t+d_r(t,l_r)) - F_r(t+d_r(t,l_r))))^2 = (\sum_{r\in A_j} \varphi_{jr}(F_r^n(t+d_r(t,l_r)) +$$

$$-F_r(t+d_r(t,l))))^2 = (\sum_{r\in B_j\cup D_j} (F_r^n(t+d_r(t,l_r)) - F_r(t+d_r(t,l_r))) +$$

$$+\sum_{r\in C_j}(F_r^n(t+d_r(t,l_r)) - F_r(t+d_r(t,l_r))))^2 =$$

$$=(\sum_{r\in C_j}(-\frac{\overline{\rho}_j(t)-\overline{\rho}_j^n(t)}{\sum_{s\in C_j}\varphi_{js}}))^2 = (\sum_{r\in C_j}(\frac{\overline{\rho}_j(t)-\overline{\rho}_j^n(t)}{\sum_{s\in C_j}\varphi_{js}}))^2 \geq$$

$$\geq \sum_{r\in C_j}(\frac{\overline{\rho}_j(t)-\overline{\rho}_j^n(t)}{\sum_{s\in C_j}\varphi_{js}})^2 \geq \frac{1}{m^2}\sum_{r\in C_j}(F_r^n(t+d_r(t,l_r)) - F_r(t+d_r(t,l_r)))^2 =$$

$$=\frac{1}{m^2}\sum_{r\in A_j}(F_r^n(t+d_r(t,l_r)) - F_r(t+d_r(t,l_r)))^2 =$$

$$\frac{1}{m^2}|F_r^n(t+d_r(t,l_r)) - F_r(t+d_r(t,l_r))|^2 .$$

Therefore,

$$0 \leq \frac{1}{m^2}\int_0^T |F^n(t+d(t,l)) - F(t+d(t,l))|^2\, dt$$

$$\leq \int_0^T (\sum_{r=1}^{m} \varphi_{jr}(F_r^n(t+d_r(t,l_r)) - F_r(t+d_r(t,l_r))))^2 dt$$

$$\leq \frac{1}{T}(\int_0^T h^2(t)dt)(\int_0^T |H(\tau) - H^n(\tau)|^2\, d\tau)$$

and, due to the fact that $\{H^n\} \longrightarrow^{L^2} H$, we obtain the convergence of the sequence $\{F^n\}$ to F. It is easy to show that $K_d(H)$ is a closed, bounded and convex set. By assumption (2) and Theorem 3, it follows that $K_d(H)$ is compact. Thus, all the hypotheses of Theorem 2 are satisfied and the existence of at least one solution is ensured.

4. RETARDED EQUILIBRIUM MODELS AND FIFO APPROACH

In this section, we want to apply the FIFO queue discipline to our retarded model. In fact, it is reasonable to suppose that the distribution of flows through the network have a first-in-first-out behavior. The FIFO discipline requires that, on average, the traffic which enters the first link of a route will exit first, or equivalently that vehicles do not pass each other. It has recently been proposed (Friesz et al., 1993) a dynamic equilibrium model which fulfills Wardrop's user equilibrium principle and establishes that users can choose their own routes as well as departure times. Thus it is possible to cope with more realistic dynamic models and study different behaviors of the users. In the above mentioned paper, the authors have also shown that the no overtaking requirement is equivalent to the invertibility of exit time functions.

Now, let us denote by $D_r(t) = \alpha_r t + \beta_r d_r(t,l)$, $\alpha, \beta \in \mathbb{R}_+^m$ the traversal time for the route R_r, assuming that the departure time from the origin occurs at time $t = 0$. Let also $\tau_r(t) = t + D_r(t)$ be the exit time function for the route R_r. It results that for any linear traversal time function the resulting exit time function is strictly increasing and hence invertible (Friesz el al., 1993). Since we are considering linear delay functions, we are entitled to deduce that the invertibility of exit time functions is ensured and FIFO requirements are satisfied.

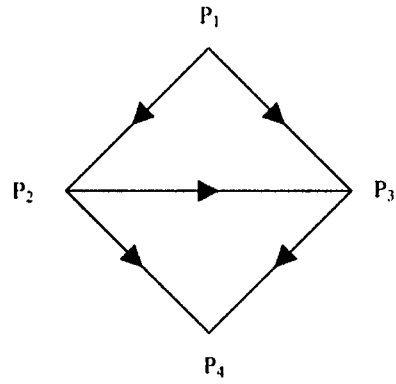
Throughout our paper we deal with route-flow variables, they are indeed the most suitable tools to examine problems with multiple destinations or in case of non-additive cost functions. But, in order to discuss the dynamic of flows, we should take into consideration the network topology and the distribution of flows through the links. Nevertheless, the route traversal functions can be expressed in terms of link traversal functions. In fact, let us denote by $\theta_i(t)$ the exit time function on the link i, then we have:

$$D_r(t) = \sum_{i=1}^{n} \delta_{ir}[\theta_i(t) - \theta_{i-1}(t)],$$

where δ is the link-route incidence matrix whose entry δ_{ir} is 1 if link i is contained in route r and 0 otherwise. Therefore, we can directly use route-flow variables, which can be derived from the link-flow variables.

5. AN EXAMPLE

In this section, we present an example of a retarded model. Let us consider a network where $N = \{P_1, P_2, P_3, P_4\}$ is the set of nodes and $L = \{(P_1, P_2), (P_1, P_3), (P_2, P_4), (P_3, P_4), (P_2, P_3)\}$ is the set of links.



We assume that the origin-destination pair is represented by (P_1, P_4), so that the routes are the following:

$$R_1 = P_1P_2 \cup P_2P_4$$
$$R_2 = P_1P_3 \cup P_3P_4$$
$$R_3 = P_1P_2 \cup P_2P_3 \cup P_3P_4.$$

Let us assume that the route costs are the following:

$$C_1(F(t)) = \alpha F_1(t) + \beta$$
$$C_2(F(t)) = \alpha F_2(t) + \gamma$$
$$C_3(F(t)) = \alpha F_3(t) + \delta$$

where $\alpha, \beta, \gamma, \delta \geq 0$.

Now let us introduce the delay vector $d = (d_1(t, l_1), d_2(t, l_2), d_3(t, l_3))$ where $l_1 = l_2 = 5, l_3 = 5(2 + \sqrt{2})$ are the lengths of the routes, $d_1(t, l_1) = (\eta + 5)t + \lambda$, $d_2(t, l_2) = (\theta + 5)t + \lambda$, $d_3(t, l_3) = (\iota + 5(2 + \sqrt{2}))t + \lambda$, with $\eta, \theta, \iota, \lambda \geq 0$.

For the sake of simplicity, we suppose that no capacity restrictions are active on flows. Let us introduce the following set:

$E = \{F(t+d(t,l)) \in L^2(0,T;R_+^3) : F_r(t+d_r(t,l_r)) \geq 0$ a.e. in$[0,T], r = 1,2,3;$
$F_r(t_1 + d_r(t_1,l_r)) \leq F_r(t_2 + d_r(t_2,l_r))\ \forall t_1,t_2$ a.e. in $[0,T], r = 1,2,3,$
$\lim_{|h|\to 0} \int_0^T |\ F(t+h+d(t+h,l)) - F(t+d(t,l))\ |^2\ dt = 0$
uniformly in F, $F(t+d(t,l)) = 0$ if $t + d(t,l) \notin [0,T]\}$;

the set of feasible flows is then given by:

$$K_d(H) = \{F(t+d(t,l)) \in E : F_1(t+d_1(t,l_1)) + F_2(t+d_2(t,l_2)) +$$
$$+F_3(t+d_3(t,l_3)) = \frac{1}{T}\int_0^T (\varepsilon t + \zeta H_1(\tau))d\tau \text{ a.e. in } [0,T]\},$$

where $\varepsilon \geq 0$ and $\zeta \in [0,3[$.

The equilibrium flow is the solution of the quasi-variational inequality:

$$H(t+d(t,l)) \in K_d(H)$$
$$\int_0^T \sum_{r=1}^{3} C_r(H(t+d(t,l)))(F_r(t+d_r(t,l_r)) - H_r(t+d_r(t,l_r)))dt \geq 0, \quad (4)$$
$$\forall F(t+d(t,l)) \in K_d(H).$$

Following the procedure shown in (De Luca, 1995; De Luca and Maugeri, 1989; Maugeri, 1987) we set:

$$F_3(t+d_3(t,l_3)) = F_3((\iota + 5(2+\sqrt{2}))t + \lambda) = \frac{1}{T}\int_0^T (\varepsilon t + \zeta H_1(\tau))d\tau +$$
$$- F_1(t+d_1(t,l_1)) - F_2(t+d_2(t,l_2));$$

$\widetilde{E} = \{\widetilde{F}(t+d(t,l)) \in L^2(0,T;\mathbb{R}_+^2) : F_r(t+d_r(t,l_r)) \geq 0$ a.e. in $[0,T]$,
$i = 1,2; F_r(t_1 + d_r(t_1,l_r)) \leq F_r(t_2 + d_r(t_2,l_r))\ \forall t_1,t_2$ a.e. in $[0,T]$,
$r = 1,2, \lim_{|h|\to 0} \int_0^T |\ F(t+h+d(t+h,l)) - F(t+d(t,l))\ |^2\ dt = 0$
uniformly in F, $F(t+d(t,l)) = 0$ if $t + d(t,l) \notin [0,T]\}$;

$\widetilde{K}_d(\widetilde{H}) = \{\widetilde{F}(t+d(t,l)) \in \widetilde{E} : F_1(t+d_1(t,l_1)) - F_2(t+d_2(t,l_2)) \leq$
$\leq \frac{1}{T}\int_0^T (\varepsilon t + \zeta H_1(\tau))d\tau$ a.e in $[0,T]\}$.

Let us consider:

$$\begin{aligned}\Gamma_1(\tilde{F}(t+d(t,l)),\tilde{H}(t+d(t,l))) &= C_1(\tilde{F}(t+d(t,l)),\tilde{H}(t+d(t,l)))+\\ &\quad -C_3(\tilde{F}(t+d(t,l)),\tilde{H}(t+d(t,l)))=\\ &\quad =2\alpha F_1(t+d_1(t,l_1))+\alpha F_2(t+d_2(t,l_2))+\\ &\quad -\frac{\alpha}{T}\int_0^T(\varepsilon t+\zeta H_1(\tau))d\tau+\beta-\delta.\\ \Gamma_2(\tilde{F}(t+d(t,l)),\tilde{H}(t+d(t,l))) &= C_2(\tilde{F}(t+d(t,l)),\tilde{H}(t+d(t,l)))+\\ &\quad -C_3(\tilde{F}(t+d(t,l)),\tilde{H}(t+d(t,l)))=\\ &\quad =\alpha F_1(t+d_1(t,l_1))+2\alpha F_2(t+d_2(t,l_2))+\\ &\quad -\frac{\alpha}{T}\int_0^T(\varepsilon t+\delta H_1(\tau))d\tau+\gamma-\delta.\end{aligned}$$

Thus, the problem can be written as:

$$\begin{aligned}&\tilde{H}(t+d(t,l))\in\tilde{K}(\tilde{H})\\ &\int_0^T\sum_{r=1}^{2}\Gamma_r(\tilde{H}(t+d(t,l)))(\tilde{F}_r(t+d_r(t,l_r))-\tilde{H}_r(t+d_r(t,l_r))dt\geq 0 \quad (5)\\ &\forall\tilde{F}(t+d(t,l))\in\tilde{K}_d(\tilde{H}).\end{aligned}$$

It is immediate to show that if $\tilde{H}$ satisfies the system:

$$\begin{cases}\Gamma_1(\tilde{H},\tilde{H})=0\\ \Gamma_2(\tilde{H},\tilde{H})=0\\ \tilde{H}\in\tilde{K}(\tilde{H})\end{cases}$$

it solves the retarded variational inequality (5). We find that:

$$H_1(t+d_1(t,l_1)) = H_1((\eta+6)t+\lambda) = \frac{\varepsilon t}{3} + \frac{1}{6}\frac{\varepsilon\,\zeta(T-2\lambda)}{(\eta+6)(3-\zeta)} - \frac{2\beta-\gamma-\delta}{\alpha(3-\zeta)}$$

$$H_2(t+d_2(t,l_2)) = H_2((\theta+6)t+\lambda) = \frac{\varepsilon t}{3} + \frac{1}{6}\frac{\varepsilon\,\zeta(T-2\lambda)}{(\eta+6)(3-\zeta)} - \frac{2\beta-\gamma-\delta}{\alpha(3-\zeta)} + \frac{\beta-\gamma}{\alpha}$$

$$H_3(t+d_3(t,l_3)) = H_3((\iota+11+5\sqrt{2}))t+\lambda) = \frac{\varepsilon t}{3} + \frac{1}{6}\frac{\varepsilon\,\zeta(T-2\lambda)}{(\eta+6)(3-\zeta)} + \frac{(2\beta-\gamma-\delta)(2-\zeta)}{\alpha(3-\zeta)} - \frac{\beta-\gamma}{\alpha}$$

under the condition:

$$H_1(t+d_1(t,l_1)) + H_2(t+d_2(t,l_2)) \le \frac{1}{T}\int_0^T (\varepsilon t + \zeta H_1(\tau))d\tau.$$

A numerical example can be obtained by choosing: $\alpha = 100$, $\beta = \frac{3}{10}$, $\gamma = \frac{1}{2}$, $\delta = \frac{3}{2}$, $\varepsilon = 1$, $\zeta = \frac{4}{3}$, $\eta = \frac{1}{6}$, $\lambda = \frac{2}{5}$ and $T = 1$.

Remark 2. *We observe that we have obtained the equilibrium solution in such a way that the retarded flows on the routes R_2 and R_3 depend, implicitly, on θ and ι respectively and, explicitly, on the delay parameter of H_1. We are then led to draw the conclusion that, due to the constraints, the delay on the flow H_1 affects the delays and the distributions of flows on the other two routes.*

We also want to highlight how the flow rate is finite, it results indeed that:

$$\frac{d}{dt}H_r(t+d_r(t,l_r)) = v_r(t) = \frac{\varepsilon}{3} \quad r = 1,2,3.$$

REFERENCES

[1] Brezis, H. *Analyse fonctionelle. Théorie et applications.* Masson Editeur, Paris, 1983.

[2] Daniele, P., Maugeri, A. and Oettli, W. Time-Dependent Traffic Equilibria. *Journal of Optimization Theory and Applications*, 103(3):543–554, 1999.

[3] De Luca, M. Generalized Quasi-Variational Inequality and Traffic Equilibrium Problems. In F. Giannessi and A. Maugeri, editors, *Variational Inequalities and Network Equilibrium Problems*, pages 45–55. Plenum Publishing, New York, 1995.

[4] De Luca, M. Existence of Solutions for a Time-dependent Quasi-Variational Inequality. *Supplemento Rendiconti del Circolo Matematico di Palermo*, Serie 2 48: 101–106, 1997.

[5] De Luca, M and Maugeri, A. Quasi-Variational Inequalities and Applications to the Traffic Equilibrium Problem; Discussion of a Paradox. *Journal of Computational and Applied Mathematics*, 28:163–171, 1989.

[6] De Luca, M and Maugeri, A. Discontinuous Quasi-Variational Inequalities and Applications to Equilibrium Problems. *Nonsmooth Optimization Methods and Applications*, Gordon & Breach Sc. Pubbl.,:70–74, 1992.

[7] Friesz, T.L., Bernstein, D., Smith, T.E., Tobin, R.L. and Wie B.W. A variational inequality formulation of the dynamic network user equilibrium problem. *Operation Research*, 41:179–191, 1993.

[8] Gwinner, J. Time Dependent Variational Inequalities. Some Recent Trends. In P. Daniele, F. Giannessi and A. Maugeri, editors, *Equilibrium Problems and Variational Models*, pages 225–264, Kluwer Academic Publishers, 2003.

[9] Maugeri, A. Convex Programming, Variational Inequalities and Applications to the Traffic Equilibrium Problems. *Applied Mathematics and Optimization*, 16:169–185, 1987.

[10] Maugeri, A. Dynamic Models and Generalized Equilibrium Problems. In F. Giannessi, editor, *New Trends in Mathematical Programiing*, pages 191–202, Kluwer, Dordrecht, Holland, 1998.

[11] Raciti, F. Equilibrium in Time-dependent Traffic Networks with Delay. In F. Giannessi, A. Maugeri, P. Pardalos, editors, *Equilibrium Problems: Nonsmooth Optimization and Variational Inequality Models*, pages 247–253, Kluwer Academic Publishers, 2001.

[12] Raciti, F. and Scrimali, L. Time-dependent Variational Inequalities and Applications to Equilibrium Problems. *Journal of Global Optimization*, to appear

[13] Tan, N.X. Quasi-Variational Inequality in Topological Linear Locally Convex Hausdorff Spaces. *Mathematische Nachrichten*, 122:231–245, 1985.

HIGHER ORDER APPROXIMATION EQUATIONS FOR THE PRIMITIVE EQUATIONS OF THE OCEAN

E. Simonnet,[1] T.Tachim-Medjo[2] and R. Temam[3]
Institut Non Linéaire de Nice, CNRS, Valbonne, France ;[1] Dept. of Mathematics, Florida International University, DM413B, University Park, Miami, Florida, USA;[2] The Institute for Scientific Computing and Applied Mathematics, Indiana University, Bloomington, IN 47405, USA[3]

Abstract: In this article, we present a family of models which approximate the full primitive equations (PEs) of the ocean, with temperature and salinity, as introduced in [9]. We consider asymptotic expansions of the PEs to all orders with respect to the aspect ratio δ. At first order, we recover the well-known barotropic quasi-geostrophic (QG) equations of the ocean. At higher orders, we obtain simple linear models that share the same mathematical structure but different right-hand sides. From the computational point of view, there are two advantages. Firstly, all the higher-order expansions are linear so that they are easy to implement. Secondly, the same numerical code can be used to compute all of them. From the physical viewpoint, we expect that higher-order corrections to the first-order barotropic QG equations will capture the vertical dynamics and the thermodynamics correctly. We will address these delicate physical issues as well as the convergence of the asymptotics in a forthcoming work.

Key words and phrases. Primitive equations, geostrophic, asymptotic, barotropic flow, baroclinic flow

1991 Mathematics Subject Classification. 65N12, 49K35, 76D55.

1. INTRODUCTION

Large-scale geophysical oceanic flows can be considered in a first approximation as 2-D incompressible geostrophic flows, i.e. the balance is predominantly between the horizontal pressure gradient forces and the Coriolis force. However, over long timescales, typically of the order of several years to several decades and more, one must take into account thermodynamics processes as well as the vertical dynamics of the oceans, which is still far from being understood. Indeed, this vertical dynamics is of primary importance for explaining the pole-to-pole ultra-low frequency thermohaline circulation. This circulation plays a crucial role on the climate and is related to a subtil balance between salinity and temperature surface fluxes between the poles and the equator. In order to understand the low-frequency dynamics of the oceans, their spatial-temporal structures, as well as the impact of the oceans on the climate, a hierarchy of models exists that help to predict and understand the oceans behavior. We will consider here an intermediate model between the quasi-geostrophic (QG) equations and ocean general circulation models (OGCMs) which is often refered to as the Primitives Equations (PEs). These equations have been introduced and analysed in [5]. They are obtained from the Boussinesq equations (BEs) by replacing the vertical momentum equation by the hydrostatic equation, thanks to the fact that the ratio H/L between the vertical and the horizontal scales is very small, [5, 9, 6, 7, 8, 15]. In our study, we consider seawater as a slightly compressible fluid whose equation of state involve the salinity and the temperature. Our main objective is to determine high-order corrections to the balanced QG equations, when the vertical aspect ratio δ is small, by considering systematic asymptotic expansions of the PEs with respect to δ.

There are two essential characteristics of the ocean which are used in simplifying the PEs or the BEs. The first one is that, for large-scale geostrophic flows, the ratio δ between the vertical and the horizontal scale is very small, typically of the order of 10^{-2}. Another small parameter of primary importance is the Rossby number ϵ, which corresponds to the ratio of the typical (horizontal) oceanic current velocity to the speed of rotation of the earth around the poles axis. This ratio is known to be smaller in the ocean, than in the atmosphere, typically 10^{-2}. It is precisely the small value of the Rossby number which is responsible for the constrained QG dynamics in the oceans.

In [9], the authors considered the asymptotic expansion of the PEs of the ocean with respect to the Rossby number and obtained at first order the QG equations. In this article, we consider the PEs of the ocean presented in [9] and use asymptotic expansions with respect to $\delta = H^2/L^2$. We are able to derive the QG equations at first order as well as higher-order linear

corrections of the vertical dynamics. More precisely, we write the unknown functions $(\mathbf{v},T,S)$, namely the velocity, temperature and salinity, in the form

$$\begin{aligned} \mathbf{v} &= \mathbf{v}^0 + \delta\mathbf{v}^1 + \delta^2\mathbf{v}^2 + \delta^3\mathbf{v}^3 + \cdots, \\ T &= T^0 + \delta T^1 + \delta^2 T^2 + \delta^3 T^3 + \cdots, \\ S &= S^0 + \delta S^1 + \delta^2 S^2 + \delta^3 S^3 + \cdots, \end{aligned} \tag{1.1}$$

and we derive a simple equation for the k^{th} terms.

As in [11], the main idea is to *properly* decompose the flow $(\mathbf{v}^k, T^k, S^k)$ of the k^{th} order approximation; we write:

$$\mathbf{v}^k = \overline{\mathbf{v}}^k + \mathbf{v}^{k,\flat}, T^k = \overline{T}^k + T^{k,\flat}, S^k = \overline{S}^k + S^{k,\flat}, \tag{1.2}$$

where $\overline{\varphi}$ corresponds to the vertical average of φ and $\varphi^\flat$ to the deviation $\varphi - \overline{\varphi}$. It immediately follows that

$$\overline{\mathbf{v}}^{k,\flat} = 0, \overline{T}^{k,\flat} = 0, \overline{S}^{k,\flat} = 0. \tag{1.3}$$

In the decomposition (1.2), $(\mathbf{v}^{k,\flat}, T^{k,\flat}, S^{k,\flat})$ correspond to the baroclinic flow and $(\overline{\mathbf{v}}^k, \overline{T}^k, \overline{S}^k)$ to the barotropic flow. As we will see later on, the key point here is that the baroclinic flow $(\mathbf{v}^{k,\flat}, T^{k,\flat}, S^{k,\flat})$ is given by lower-order approximations solution of a *simple linear ordinary equation,* while the barotropic flow $(\overline{\mathbf{v}}^k, \overline{T}^k, \overline{S}^k)$ satisfies a simple linear system for $k \geq 1$ and reduces to the barotropic QG equations for $k = 0$.

Although the derivation procedure is somehow involved, the final equations for determining $(\overline{\mathbf{v}}^k, \overline{T}^k, \overline{S}^k)$ and calculating $(\mathbf{v}^{k,\flat}, T^{k,\flat}, S^{k,\flat})$ are surprisingly very natural and simple. The equations obtained are linear and are all of the same form. They are therefore very easy to implement and the same numerical code can be applied to all level $k \geq 1$.

The article is organized as follows. In section 2, we present and properly reformulate the PEs of the ocean. The third section introduces the asymptotic expansions with respect to the aspect ratio and presents the approximate models. We start with the zeroth and the first order approximations and we generalize the procedure to obtain the higher-order approximations.

2. THE PES OF THE OCEAN AND THE GEOSTROPHIC SCALING

In this section, we study the geostrophic asymptotics of the PEs with double diffusions, i.e., with the diffusion equations for both the temperature and the salinity functions.

2.1 The spatial domain

Before we introduce the governing equations, let us first describe the spatial domain $\tilde{\mathcal{M}}$ occupied by the ocean. As in [9], we assume that the space domain $\tilde{\mathcal{M}}$ is given by

$$\tilde{\mathcal{M}} = \left\{(\theta, \varphi, z); (\theta, \varphi) \in \tilde{\mathcal{M}}_s, -\tilde{h} < z < 0\right\}, \tag{2.1}$$

where the surface region is given by a simply connected open set $\tilde{\mathcal{M}}_S \subset S_a^2$, S_a^2 being the 2-D sphere of radius a.

For simplicity, we assume like in [9] that

$$\tilde{\mathcal{M}}_s \subset \tilde{\mathcal{O}} \equiv \left\{(\theta, \varphi) \in S_a^2; |\, \theta - \theta_0 \,| < \frac{\overline{\theta}}{2}, |\, \varphi - \varphi_0 \,| < \frac{\overline{\varphi}}{2}\right\}, \tag{2.2}$$

such that $\tilde{\mathcal{M}}$ and $\tilde{\mathcal{O}}$ have the same horizontal length scale L in both the θ and φ directions.

Let us recall that (θ, φ, r) stands for the spherical coordinates, where θ is the colatitude $(0 \le \theta \le \pi)$, φ is the longitude $(0 \le \varphi \le 2\pi)$, r is the radial distance and $z = r - a$ is the vertical coordinate with respect to the sea level. Here (θ_0, φ_0) and $(\overline{\theta}, \overline{\varphi})$ are given and fixed, such that $\tilde{\mathcal{M}}$ corresponds to a midlatitude domain.

The boundary of $\tilde{\mathcal{M}}$ consists of the following three parts.

$$\begin{aligned} &\tilde{\Gamma}_t(z = 0) = \text{upper boundary of the ocean (interface with air)}, \\ &\tilde{\Gamma}_l = \text{lateral boundary}, \\ &\tilde{\Gamma}_b(z = -\tilde{h}) = \text{bottom of the ocean}. \end{aligned} \tag{2.3}$$

For simplicity, we assume that $\tilde{h} > 0$ is constant. Let L be the horizontal scale of the motion. We are interested here in the mesoscale or synoptic scale motions of the ocean for which

$$L = O(100 \text{ km}). \tag{2.4}$$

Thanks to (2.2), we obtain

$$\overline{\theta} = \frac{L}{a} << 1. \tag{2.5}$$

We make the following coordinate transformation that yields the so-called $\beta -$ plane approximation, namely

$$x = a \sin \theta_0(\varphi - \varphi_0), y = (\theta - \theta_0)a. \tag{2.6}$$

The variables (x, y) correspond to the Cartesian coordinates on the β - plane and the space domain (2.2) is replaced by the following domain in the (x, y, z) Cartesian coordinates

$$\hat{\mathcal{M}} = \left\{(x, y, z); (x, y) \in \tilde{\mathcal{M}}_s, -\tilde{h} < z < 0\right\},$$
$$\tilde{\mathcal{M}}_s \subset \left\{(x, y); | x | < \frac{L}{2}, | y | < \frac{L}{2}\right\}. \tag{2.7}$$

2.2 Mean temperature and salinity distribution and the PEs with double diffusions

In order to obtain a proper geostrophic scaling of the PEs, we need to consider the standard temperature and salinity profiles, i.e., the mean temperature and salinity distributions. For the mesoscale ocean, it is legitimate to consider the vertical profile of these functions. Let $T_s(z)$ and $S_s(z)$ be the vertical stratification profiles of the temperature T_{tot} and the salinity S_{tot} respectively, which can be considered as the mean values of T_{tot} and S_{tot} at the level z. We refer the reader to [1, 12] for some typical profiles of the temperature and salinity functions.

As in [9], we write the temperature and salinity functions as follows:

$$T_{tot} = T_s + \tilde{T}, S_{tot} = S_s + \tilde{S}, \tag{2.8}$$

$\tilde{T}$ and $\tilde{S}$ being the deviations of T_{tot} and S_{tot} from T_s and S_s respectively. We also assume that the following equation of state is satisfied

$$\rho = \rho_0(1 - \beta_T(T - T_{ref}) + \beta_S(S - S_{ref})), \tag{2.9}$$

where β_T and β_S are expansion coefficients, T_{ref} and S_{ref} are the reference values of T and S respectively (see [9]); in particular (2.9) is satisfied by ρ_s, T_s, S_s and by ρ_{tot}, T_{tot}, S_{tot}, where ρ_{tot} is the total density.

We assume that the ocean is approximately hydrostatic and the vertical mean pressure $p_s(z)$ then satisfies

$$\frac{\partial p_s(z)}{\partial z} = -\rho_s(z)g, \tag{2.10}$$

where g is the gravitational constant. We also write the total density ρ_{tot} and total pressure p_{tot} in the form

$$\tilde{\rho} = \rho_{tot} - \rho_s, \tilde{p} = p_{tot} - p_s. \tag{2.11}$$

The dimensional form of the PEs using the β-plane approximation reads

$$\begin{cases} \dfrac{\partial \mathbf{v}}{\partial t} - \mu\Delta\mathbf{v} - \nu\dfrac{\partial^2 \mathbf{v}}{\partial z^2} + \tilde{f}\ \mathbf{k}_0 \times \mathbf{v} + \dfrac{1}{\rho_0}\operatorname{grad}\tilde{p} + (\mathbf{v}\cdot\nabla)\mathbf{v} + w\dfrac{\partial \mathbf{v}}{\partial z} = 0, \\ \dfrac{\partial \tilde{p}}{\partial z} = -g\tilde{\rho}, \\ \operatorname{div}\mathbf{v} + \dfrac{\partial w}{\partial z} = 0, \\ \dfrac{\partial \tilde{T}}{\partial t} - \mu_T\Delta\tilde{T} - \nu_T\dfrac{\partial^2 \tilde{T}}{\partial z^2} + (\mathbf{v}\cdot\nabla)\tilde{T} + w\dfrac{\partial \tilde{T}}{\partial z} + w\dfrac{\partial T_s}{\partial z} = \nu_T\dfrac{\partial^2 T_s}{\partial z^2}, \\ \dfrac{\partial \tilde{S}}{\partial t} - \mu_S\Delta\tilde{S} - \nu_S\dfrac{\partial^2 \tilde{S}}{\partial z^2} + (\mathbf{v}\cdot\nabla)\tilde{S} + w\dfrac{\partial \tilde{S}}{\partial z} + w\dfrac{\partial S_s}{\partial z} = \nu_S\dfrac{\partial^2 S_s}{\partial z^2}, \\ \tilde{\rho} = \rho_0(-\beta_T\tilde{T} + \beta_S\tilde{S}). \end{cases} \tag{2.12}$$

The boundary conditions are given by

$$\begin{cases} \rho_0 \nu \dfrac{\partial \mathbf{v}}{\partial z} = \tilde{\tau}_v,\ w = 0,\ C_p \rho_0 \nu_T \dfrac{\partial \tilde{T}}{\partial z} = \tilde{\alpha}_T(\tilde{T}^* - \tilde{T}), \text{and} \\ C_p \rho_0 \nu_S \dfrac{\partial \tilde{S}}{\partial z} = \tilde{\alpha}_S(\tilde{S}^* - \tilde{S}) \text{ on } \tilde{\Gamma}_i\ , \\ \dfrac{\partial \mathbf{v}}{\partial z} = 0,\ w = 0,\ \dfrac{\partial \tilde{T}}{\partial z} = 0,\ \dfrac{\partial \tilde{S}}{\partial z} = 0 \text{ on } \tilde{\Gamma}_b\ , \\ \mathbf{v} = 0,\ \dfrac{\partial \tilde{T}}{\partial n} = 0,\ \dfrac{\partial \tilde{S}}{\partial n} = 0,\ w = 0 \text{ on } \tilde{\Gamma}_l. \end{cases} \tag{2.13}$$

Here $\tilde{T}^*$ and $\tilde{S}^*$ are given functions representing the apparent temperature and salinity distribution on the upper surface of the ocean, while $\tilde{\tau}_v$ is the (given) wind stress, which accounts for the mechanical motion of the ocean; n is the outward normal on $\tilde{\Gamma}_l$ and $\mathbf{k}_0$ is the vertical unit vector.

The initial conditions are given by

$$(\mathbf{v}, \tilde{T}, \tilde{S}) = (\mathbf{v}_0, \tilde{T}_0, \tilde{S}_0) \text{ at } t = 0. \tag{2.14}$$

In (2.12)-(2.14), the unknown functions are the horizontal velocity $\mathbf{v} = (u, v)$, the vertical velocity w, the temperature deviation $\tilde{T}$, the salinity deviation $\tilde{S}$ and the pressure deviation $\tilde{p}$. The positive constant C_p corresponds to the heat capacity of the ocean, ρ_0 is the reference value of the density, g is the gravitational constant and $\tilde{f}$ is the Coriolis parameter.

Here the differential operators ∇, Δ and div are 2-D horizontal operators acting on the variables x and y. The dimensional domain is $\tilde{\mathcal{M}} = \tilde{\mathcal{M}}_s \times (-H, 0)$, where $\tilde{\mathcal{M}}_s \subset \mathbb{R}^2$ is bounded and $\tilde{H} > 0$. The positive constants ν and μ are the viscosity coefficients, $\nu_T > 0$ and $\mu_T > 0$ are the thermal diffusivity, and $\nu_S > 0$ and $\mu_S > 0$ are the diffusivity coefficients of the salinity.

The PEs (2.12)-(2.14) are derived from the Boussinesq equation using the fact that the aspect ratio δ (square of the ratio between the horizontal and the vertical length scales) is small, [5, 9, 12]. For more details on the PEs of the ocean, the reader is referred to [3, 12, 13, 16] for the physical aspect, and to [5], in which the existence and uniqueness results of the system (2.12)-(2.14) are studied; see also [14].

2.3 Geostrophic scaling

Here we first recall from [9] a standard scaling for the PEs of the ocean. We set

$$(x, y, z, t) = (Lx', Ly', Hz', \frac{L}{U}t'), \tag{2.15}$$

where U is the reference value of the horizontal velocity.

We also set

$$\mathbf{v} = U\mathbf{v}', w = \frac{H}{L}Uw', \tilde{h} = Hh, \tag{2.16}$$

$$p_{tot} = p_s(z) + L\rho_0 f_0 Up', \rho_{tot} = \rho_s(z) + \epsilon F\rho_0\rho', \tilde{f}_0 = 2\Omega\cos\theta_0,$$

$$\tilde{f} = \tilde{f}_0(1 + \epsilon_1), f_1 = \frac{1}{\epsilon}\frac{\cos\theta - \cos\theta_0}{\cos\theta_0} = O(1), f = 1 + \epsilon f_1, \tag{2.17}$$

where $F = \tilde{f}_0^2 L^2/gH$ is the Froude number, $\delta = H^2/L^2$, $\epsilon = U/\tilde{f}_0 L$ is the Rossby number and Ω is the angular velocity of the earth.

We scale T_s and S_s by

$$(T_s, S_s) = (T_{ref}T_s', S_{ref}S_s'). \tag{2.18}$$

For the temperature and salinity deviations, we set

$$\tilde{T} = \epsilon F T_{ref}T', \tilde{T} = \epsilon F S_{ref}S'. \tag{2.19}$$

Other nondimensional parameters are given by

$$\begin{cases} \dfrac{1}{R_{e_1}} = \dfrac{\mu}{LU}, \dfrac{1}{R_{e_2}} = \dfrac{\nu}{LU}, \\ \dfrac{1}{R_{t_1}} = \dfrac{\mu_T}{LU}, \dfrac{1}{R_{t_2}} = \dfrac{\nu_T}{LU}, \\ \dfrac{1}{R_{s_1}} = \dfrac{\mu_S}{LU}, \dfrac{1}{R_{s_2}} = \dfrac{\nu_S}{LU}, \\ \beta_T = \tilde{\beta}_T T_{ref}, \beta_S = \tilde{\beta}_S S_{ref}, \\ T^* = \dfrac{\tilde{T}^*}{\varepsilon F T_{ref}}, S^* = \dfrac{\tilde{S}^*}{\varepsilon F S_{ref}}, \\ \alpha_T = \dfrac{L\tilde{\alpha}_T}{C_p H U \rho_0}, \alpha_S = \dfrac{L\tilde{\alpha}_S}{C_p H U \rho_0}, \\ \tau_v = \dfrac{L}{\rho_0 U^2 H} \tilde{\tau}_v. \end{cases} \tag{2.20}$$

We introduce the nondimensional functions

$$q_1 = \frac{\nu_T g H}{\tilde{f}_0 U^2 T_{ref}} \frac{\partial^2 T_s}{\partial z^2}, q_2 = \frac{\nu_S g H}{\tilde{f}_0 U^2 S_{ref}} \frac{\partial^2 S_s}{\partial z^2}. \tag{2.21}$$

Hereafter, we assume that

$$q_1 = O(1), q_2 = O(1). \tag{2.22}$$

The nondimensional space domain $\mathcal{M}$ becomes

$$\begin{aligned} &\mathcal{M} = \{(x, y, z); (x, y) \in \mathcal{M}_s, -h < z < 0\}, \\ &\mathcal{M}_s \subset \left\{(x, y); |x| < \frac{1}{2}, |y| < \frac{1}{2}\right\}. \end{aligned} \tag{2.23}$$

We denote by Γ_i, Γ_b and Γ_l the corresponding boundaries of $\mathcal{M}$ and we assume for simplicity that $h = 1$; Γ_i is the same as $\mathcal{M}_s$.

Substituting these expressions, variables and parameters in the PEs (2.12), we obtain the following nondimensional form of the PEs of the ocean (see [9] for more details):

$$\begin{cases} \epsilon\left[\dfrac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v} + w\dfrac{\partial \mathbf{v}}{\partial z}\right] - \dfrac{\epsilon}{R_{e_1}}\Delta\mathbf{v} - \dfrac{\epsilon}{\delta R_{e_2}}\dfrac{\partial^2 \mathbf{v}}{\partial z^2} + f\mathbf{k}_0 \times \mathbf{v} + \text{grad } p = 0, \\ \dfrac{\partial p}{\partial z} = -\rho, \\ \text{div } \mathbf{v} + \dfrac{\partial w}{\partial z} = 0, \\ \epsilon\left[\dfrac{\partial T}{\partial t} + (\mathbf{v}\cdot\nabla)T + w\dfrac{\partial T}{\partial z}\right] - \dfrac{\epsilon}{R_{t_1}}\Delta T - \dfrac{\epsilon}{\delta R_{t_2}}\dfrac{\partial^2 T}{\partial z^2} + F^{-1}\dfrac{\partial T_s}{\partial z}w = \epsilon q_1, \\ \epsilon\left[\dfrac{\partial S}{\partial t} + (\mathbf{v}\cdot\nabla)S + w\dfrac{\partial S}{\partial z}\right] - \dfrac{\epsilon}{R_{s_1}}\Delta S - \dfrac{\epsilon}{\delta R_{t_2}}\dfrac{\partial^2 S}{\partial z^2} + F^{-1}\dfrac{\partial S_s}{\partial z}w = \epsilon q_2, \\ \rho = -\beta_T T + \beta_S S. \end{cases} \tag{2.24}$$

The boundary conditions (2.13) become

$$\begin{cases} \dfrac{1}{\delta R_{e_2}}\dfrac{\partial \mathbf{v}}{\partial z} = \tau_v, w = 0, \dfrac{1}{\delta R_{t_2}}\dfrac{\partial T}{\partial z} = \alpha_T(T^* - T), \dfrac{1}{\delta R_{s_2}}\dfrac{\partial S}{\partial z} = \alpha_S(S^* - S) \text{ on } \Gamma_i, \\ \dfrac{\partial \mathbf{v}}{\partial z} = 0, w = 0, \dfrac{\partial T}{\partial z} = 0, \dfrac{\partial S}{\partial z} = 0 \text{ on } \Gamma_b, \\ \mathbf{v} = 0, w = 0, \dfrac{\partial T}{\partial n} = 0, \dfrac{\partial T}{\partial n} = 0 \text{ on } \Gamma_l. \end{cases} \tag{2.25}$$

The initial conditions are

$$(\mathbf{v}, T, S) = (\mathbf{v}_0, T_0, S_0) \text{ at } t = 0. \tag{2.26}$$

Formal asymptotic expansions of the PEs with respect to the Rossby number yield the QG equations that correspond to an a priori $O(\epsilon)$ approximation to the PEs model (see [9]).

In the next subsection, we derive an $O(\delta^k)$ approximation to the PEs of the ocean (2.12)-(2.14), where k is any positive integer. Let us recall from [12] that for oceanic synoptic scales, $O(\delta) \simeq O(4.10^{-4})$.

3. QUASI-GEOSTROPHIC AND HIGHER ORDER APPROXIMATE MODELS FOR THE PES

3.1 Quasi-geostrophic approximations

We expand the unknowns $\mathbf{v}, w, \rho, T$ and S with respect to δ as

$$\begin{aligned}
&\mathbf{v} = \mathbf{v}^0 + \delta\mathbf{v}^1 + \delta^2\mathbf{v}^2 + \delta^3\mathbf{v}^3 + \cdots,\\
&T = T^0 + \delta T^1 + \delta^2 T^2 + \delta^3 T^3 + \cdots,\\
&S = S^0 + \delta S^1 + \delta^2 S^2 + \delta^3 S^3 + \cdots,\\
&w = w^0 + \delta w^1 + \delta^2 w^2 + \delta^3 w^3 + \cdots,\\
&p = p^0 + \delta p^1 + \delta^2 p^2 + \delta^3 p^3 + \cdots,\\
&\rho = \rho^0 + \delta\rho^1 + \delta^2\rho^2 + \delta^3\rho^3 + \cdots.
\end{aligned} \tag{3.1}$$

Hereafter, we use the notations

$$\overline{u} = \int_{-1}^{0} u \, dz, \quad u^b = u - \overline{u}, \tag{3.2}$$

for a given function u. In oceanography, the vertical average $\overline{u}$ is usually referred to as the barotropic flow and u^b is called the baroclinic flow [3, 16].

The zeroth order approximation.
Replacing (3.1) in (2.24)-(2.26), at the level $O(1)$ we find

$$\begin{cases}
\dfrac{\partial^2\mathbf{v}^0}{\partial z^2} = 0, \dfrac{\partial^2 T^0}{\partial z^2} = 0, \dfrac{\partial^2 S^0}{\partial z^2} = 0, \dfrac{\partial p^0}{\partial z} = -\rho^0,\\
\text{div } \mathbf{v}^0 + \dfrac{\partial w^0}{\partial z} = 0, \rho^0 = -\beta_T T^0 + \beta_S S^0,
\end{cases} \tag{3.3}$$

with the boundary conditions

$$\begin{cases} \dfrac{1}{R_{e_2}}\dfrac{\partial \mathbf{v}^0}{\partial z} = 0, w^0 = 0, \dfrac{1}{R_{t_2}}\dfrac{\partial T^0}{\partial z} = 0, \dfrac{1}{R_{s_2}}\dfrac{\partial S^0}{\partial z} = 0 \text{ on } \Gamma_i, \\ \dfrac{\partial \mathbf{v}^0}{\partial z} = 0, w^0 = 0, \dfrac{\partial T^0}{\partial z} = 0, \dfrac{\partial S^0}{\partial z} = 0 \text{ on } \Gamma_b, \\ \mathbf{v}^0 = 0, w^0 = 0, \dfrac{\partial T^0}{\partial n} = 0, \dfrac{\partial T^0}{\partial n} = 0 \text{ on } \Gamma_l. \end{cases} \tag{3.4}$$

This amounts to say that

$$\frac{\partial \mathbf{v}^0}{\partial z} = 0, \frac{\partial T^0}{\partial z} = 0, \frac{\partial S^0}{\partial z} = 0, w^0 = 0, \operatorname{div} \mathbf{v}^0 = 0. \tag{3.5}$$

The first order approximation.

At the level $O(\delta)$, we find using (3.5) and the decomposition $v^1 = \overline{v}^1 + v^{1,\flat}$, $T^1 = \overline{T}^1 + T^{1,\flat}$ and $S^1 = \overline{S}^1 + S^{1,\flat}$

$$\begin{cases} \epsilon\left|\dfrac{\partial \mathbf{v}^0}{\partial t} + (\mathbf{v}^0 \cdot \nabla)\mathbf{v}^0\right| - \dfrac{\epsilon}{R_{e_1}}\Delta \mathbf{v}^0 - \dfrac{\epsilon}{R_{e_2}}\dfrac{\partial^2 \mathbf{v}^{1,\flat}}{\partial z^2} + \not{f}\mathbf{k}_0 \times \mathbf{v}^0 + \operatorname{grad} p^0 = 0, \\ \dfrac{\partial p^1}{\partial z} = -\rho^1, \\ \operatorname{div} \mathbf{v}^1 + \dfrac{\partial w^1}{\partial z} = 0, \\ \epsilon\left|\dfrac{\partial T^0}{\partial t} + (\mathbf{v}^0 \cdot \nabla)T^0\right| - \dfrac{\epsilon}{R_{t_1}}\Delta T^0 - \dfrac{\epsilon}{R_{t_2}}\dfrac{\partial^2 T^{1,\flat}}{\partial z^2} = \epsilon q_1, \\ \epsilon\left|\dfrac{\partial S^0}{\partial t} + (\mathbf{v}^0 \cdot \nabla)S^0\right| - \dfrac{\epsilon}{R_{s_1}}\Delta S^0 - \dfrac{\epsilon}{R_{s_2}}\dfrac{\partial^2 S^{1,\flat}}{\partial z^2} = \epsilon q_2, \\ \rho^1 = -\beta_T T^1 + \beta_S S^1, \end{cases} \tag{3.6}$$

with the boundary conditions

$$\begin{cases} \dfrac{1}{R_{e_2}}\dfrac{\partial \mathbf{v}^{1,\flat}}{\partial z} = \tau_v, w^1 = 0, \dfrac{1}{R_{t_2}}\dfrac{\partial T^{1,\flat}}{\partial z} = \alpha_T(T^* - T^0), \text{and} \\ \dfrac{1}{R_{s_2}}\dfrac{\partial S^{1,\flat}}{\partial z} = \alpha_S(S^* - S^0) \text{ on } \Gamma_s, \\ \dfrac{\partial \mathbf{v}^{1,\flat}}{\partial z} = 0, w^1 = 0, \dfrac{\partial T^{1,\flat}}{\partial z} = 0, \dfrac{\partial S^{1,\flat}}{\partial z} = 0 \text{ on } \Gamma_b, \\ \mathbf{v}^1 = 0, w^1 = 0, \dfrac{\partial T^1}{\partial n} = 0, \dfrac{\partial T^1}{\partial n} = 0 \text{ on } \Gamma_l. \end{cases} \tag{3.7}$$

Taking the vertical average of $(3.6)_1$, $(3.6)_3$ and $(3.6)_4$ and using the boundary conditions (3.7), we derive the following quasi-geostrophic equations of the ocean

$$\begin{cases} \epsilon\left|\dfrac{\partial \mathbf{v}^0}{\partial t} + (\mathbf{v}^0 \cdot \nabla)\mathbf{v}^0\right| - \dfrac{\epsilon}{R_{e_1}}\Delta \mathbf{v}^0 + f\,\mathbf{k}_0 \times \mathbf{v}^0 + \text{grad}\,\overline{p}^0 = \epsilon\tau_v, \\ \text{div}\,\mathbf{v}^0 = 0, \\ \epsilon\left|\dfrac{\partial T^0}{\partial t} + (\mathbf{v}^0 \cdot \nabla)T^0\right| - \dfrac{\epsilon}{R_{t_1}}\Delta T^0 + \epsilon\alpha_T T^0 = \epsilon\overline{q}_1, + \epsilon\alpha_T T^*, \\ \epsilon\left|\dfrac{\partial S^0}{\partial t} + (\mathbf{v}^0 \cdot \nabla)S^0\right| - \dfrac{\epsilon}{R_{s_1}}\Delta S^0 + \epsilon\alpha_S S^0 = \epsilon\overline{q}_2, + \epsilon\alpha_S S^*, \end{cases} \tag{3.8}$$

with the boundary and initial conditions

$$\begin{cases} \mathbf{v}^0 = 0, \ \dfrac{\partial T^0}{\partial n} = 0, \ \dfrac{\partial T^0}{\partial n} = 0 \ \text{ on } \Gamma_l \\ (\mathbf{v}^0, T^0, S^0) = (\overline{\mathbf{v}}_0^0, \overline{T}_0^0, \overline{S}_0^0) \text{ at } t = 0. \end{cases} \tag{3.9}$$

We can easily determine $(\mathbf{v}^0, T^0, S^0)$ from (3.8)-(3.9); note that, by (3.3) and (3.5) $p^0 = -z\rho^0$, so that $\overline{p}^0 = -\dfrac{1}{2}\overline{\rho}^0 = -\dfrac{1}{2}\left(\beta_T\overline{T}^{\theta} + \beta_S\overline{S}^{\theta}\right)$. Now the baroclinic components $v^{1,\flat}, T^{1,\flat}, S^{1,\flat}$ can be solved using the following proposition.

Proposition 3.1. *The ordinary differential equation*

$$\begin{cases} \dfrac{d^2u}{dz^2} = a \text{ in } (-1,0), \\ \displaystyle\int_{-1}^{0} u dz = 0, \\ \dfrac{du}{dz}(0) = \Theta, \dfrac{du}{dz}(-1) = 0, \end{cases} \tag{3.10}$$

possesses a unique solution u *given by*

$$u(z) = \int_{-1}^{z} \left(\int_{-1}^{s} a(t)dt \right) ds - \int_{-1}^{0} \left[\int_{-1}^{z} \left(\int_{-1}^{s} a(t)dt \right) ds \right] dz, \tag{3.11}$$

provided that Θ *and the function* a *satisfy the compatibility condition*

$$\int_{-1}^{0} a(z)dz = \Theta. \tag{3.12}$$

Proof. Taking the vertical average of $(3.10)_1$ and using the boundary conditions $(3.10)_2$, we obtain that a and Θ must satisfy the condition (3.12). Under the condition (3.12), the unique solution to (3.10) is therefore given by (3.11). □

Remark 3.1. If the function a in (3.10) is constant, then (3.11) reduces to

$$u(z) = \left(\frac{z^2}{2} + z + \frac{1}{3} \right) a. \tag{3.13}$$

Remark 3.2. The compatibility condition (3.12) is automatically verified using (3.8) and (3.9) so that the baroclinic components are well defined. For the computation of the pressure p^0 that appears in (3.6), see the explanations given in Remark 3.6 for the general case with $k \geq 1$.

The second order approximation.

Before we generalize the process to the k^{th} order approximation, we consider first the second-order approximation (whose equations have a distinct form). The second-order expansion is indeed needed to recover the barotropic components $\overline{v}^1, \overline{T}^1$ and $\overline{S}^1$. At the $O(\delta^2)$ level, we find using (3.5)

$$\begin{cases} \epsilon\left|\dfrac{\partial \mathbf{v}^1}{\partial t} + (\mathbf{v}^0 \cdot \nabla)\mathbf{v}^1 + (\mathbf{v}^1 \cdot \nabla)\mathbf{v}^0\right| \\ -\dfrac{\epsilon}{R_{e_1}}\Delta \mathbf{v}^1 - \dfrac{\epsilon}{R_{e_2}}\dfrac{\partial^2 \mathbf{v}^2}{\partial z^2} + f\,\mathbf{k}_0 \times \mathbf{v}^1 + \operatorname{grad} p^1 = 0, \\ \operatorname{div} \mathbf{v}^2 + \dfrac{\partial w^2}{\partial z} = 0, \\ \epsilon\left|\dfrac{\partial T^1}{\partial t} + (\mathbf{v}^0 \cdot \nabla)T^1 + (\mathbf{v}^1 \cdot \nabla)T^0\right| - \dfrac{\epsilon}{R_{t_1}}\Delta T^1 - \dfrac{\epsilon}{R_{t_2}}\dfrac{\partial^2 T^2}{\partial z^2} + F^{-1}\dfrac{\partial T_s}{\partial z}w^1 = 0, \\ \epsilon\left|\dfrac{\partial S^1}{\partial t} + (\mathbf{v}^0 \cdot \nabla)S^1 + (\mathbf{v}^1 \cdot \nabla)S^0\right| - \dfrac{\epsilon}{R_{s_1}}\Delta S^1 - \dfrac{\epsilon}{R_{t_2}}\dfrac{\partial^2 S^2}{\partial z^2} + F^{-1}\dfrac{\partial S_s}{\partial z}w^1 = 0, \\ \rho^1 = -\beta_T T^1 + \beta_S S^1, \end{cases} \tag{3.14}$$

with the boundary conditions

$$\begin{cases} \dfrac{1}{R_{e_2}}\dfrac{\partial \mathbf{v}^2}{\partial z} = 0, w^2 = 0, \dfrac{1}{R_{t_2}}\dfrac{\partial T^2}{\partial z} = -\alpha_T T^1, \dfrac{1}{R_{s_2}}\dfrac{\partial S^2}{\partial z} = -\alpha_S S^1 \text{ on } \Gamma_s, \\ \dfrac{\partial \mathbf{v}^2}{\partial z} = 0, w^2 = 0, \dfrac{\partial T^2}{\partial z} = 0, \dfrac{\partial S^2}{\partial z} = 0 \text{ on } \Gamma_b, \\ \mathbf{v}^2 = 0, w^2 = 0, \dfrac{\partial T^2}{\partial n} = 0, \dfrac{\partial S^2}{\partial n} = 0 \text{ on } \Gamma_l. \end{cases} \tag{3.15}$$

First notice that $\operatorname{div} \overline{\mathbf{v}}^1 = 0$ by taking the vertical average of

$$\operatorname{div} \mathbf{v}^1 + \frac{\partial w^1}{\partial z} = 0, \tag{3.16}$$

and using the boundary conditions (3.7).

Finally, we obtain from (3.14) that the barotropic flow $(\overline{\mathbf{v}}^1, \overline{T}^1, \overline{S}^1)$ satisfies

$$
\begin{cases}
\epsilon\left|\dfrac{\partial \overline{\mathbf{v}}^1}{\partial t}+(\mathbf{v}^0\cdot\nabla)\overline{\mathbf{v}}^1+(\overline{\mathbf{v}}^1\cdot\nabla)\mathbf{v}^0\right| \\
-\dfrac{\epsilon}{R_{e_1}}\Delta\overline{\mathbf{v}}^1-\dfrac{\epsilon}{R_{e_2}}\dfrac{\partial^2\mathbf{v}^2}{\partial z^2}+f\,\mathbf{k}_0\times\overline{\mathbf{v}}^1+\operatorname{grad}\,p^1+K_1=0, \\
\dfrac{\partial p^1}{\partial z}=-\rho^1, \\
\operatorname{div}\,\overline{\mathbf{v}}^1=0, \\
\epsilon\left|\dfrac{\partial \overline{T}^1}{\partial t}+(\mathbf{v}^0\cdot\nabla)\overline{T}^1+(\overline{\mathbf{v}}^1\cdot\nabla)T^0\right|-\dfrac{\epsilon}{R_{t_1}}\Delta\overline{T}^1-\dfrac{\epsilon}{R_{t_2}}\dfrac{\partial^2T^2}{\partial z^2}+K_2=0, \\
\epsilon\left|\dfrac{\partial \overline{S}^1}{\partial t}+(\mathbf{v}^0\cdot\nabla)\overline{S}^1+(\overline{\mathbf{v}}^1\cdot\nabla)S^0\right|-\dfrac{\epsilon}{R_{s_1}}\Delta\overline{S}^1-\dfrac{\epsilon}{R_{s_2}}\dfrac{\partial^2S^2}{\partial z^2}+K_3=0, \\
\rho^1=-\beta_T T^1+\beta_S S^1,
\end{cases}
\tag{3.17}
$$

where K_1, K_2 and K_3 are known at this stage and are given by

$$
\begin{cases}
K_1=\epsilon\left|\dfrac{\partial \mathbf{v}^{1,\flat}}{\partial t}+(\mathbf{v}^0\cdot\nabla)\mathbf{v}^{1,\flat}+(\mathbf{v}^{1,\flat}\cdot\nabla)\mathbf{v}^0\right|-\dfrac{\epsilon}{R_{e_1}}\Delta\mathbf{v}^{1,\flat}+f\,\mathbf{k}_0\times\mathbf{v}^{1,\flat}, \\
K_2=\epsilon\left|\dfrac{\partial T^{1,\flat}}{\partial t}+(\mathbf{v}^0\cdot\nabla)T^{1,\flat}+(\mathbf{v}^{1,\flat}\cdot\nabla)T^0\right|-\dfrac{\epsilon}{R_{t_1}}\Delta T^{1,\flat}+F^{-1}\dfrac{\partial T_s}{\partial z}w^1, \\
K_3=\epsilon\left|\dfrac{\partial S^{1,\flat}}{\partial t}+(\mathbf{v}^0\cdot\nabla)S^{1,\flat}+(\mathbf{v}^{1,\flat}\cdot\nabla)S^0\right|-\dfrac{\epsilon}{R_{s_1}}\Delta S^{1,\flat}+F^{-1}\dfrac{\partial S_s}{\partial z}w^1, \\
w^1=\displaystyle\int_z^0 \operatorname{div}\,\mathbf{v}^{1,\flat}d\zeta.
\end{cases}
\tag{3.18}
$$

Taking the vertical average of (3.17) and using the boundary conditions in (3.15), we obtain the following system for $(\overline{\mathbf{v}}^1, \overline{T}^1, \overline{S}^1)$:

$$\begin{cases} \epsilon\left|\dfrac{\partial \overline{\mathbf{v}}^1}{\partial t} + (\mathbf{v}^0 \cdot \nabla)\overline{\mathbf{v}}^1 + (\overline{\mathbf{v}}^1 \cdot \nabla)\mathbf{v}^0\right| - \dfrac{\epsilon}{R_{e_1}}\Delta \overline{\mathbf{v}}^1 + f\,\mathbf{k}_0 \times \overline{\mathbf{v}}^1 + \text{grad}\ \overline{p}^1 + \overline{K}_1 = 0, \\ \text{div}\ \overline{\mathbf{v}}^1 = 0, \\ \epsilon\left|\dfrac{\partial \overline{T}^1}{\partial t} + (\mathbf{v}^0 \cdot \nabla)\overline{T}^1 + (\overline{\mathbf{v}}^1 \cdot \nabla)T^0\right| - \dfrac{\epsilon}{R_{t_1}}\Delta \overline{T}^1 + \epsilon\alpha_T \overline{T}^1 + \overline{K}_2 = -\epsilon\alpha_T T^{1,\flat}(0), \\ \epsilon\left|\dfrac{\partial \overline{S}^1}{\partial t} + (\mathbf{v}^0 \cdot \nabla)\overline{S}^1 + (\overline{\mathbf{v}}^1 \cdot \nabla)S^0\right| - \epsilon\Delta \overline{S}^1 + \epsilon\alpha_S \overline{S}^1 + \overline{K}_3 = -\epsilon\alpha_S S^{1,\flat}(0), \end{cases} \tag{3.19}$$

with the boundary and initial conditions

$$\begin{cases} \overline{\mathbf{v}}^1 = 0, \dfrac{\partial \overline{T}^1}{\partial n} = 0, \dfrac{\partial \overline{S}^1}{\partial n} = 0 \text{ on } \Gamma_l, \\ (\overline{\mathbf{v}}^1, \overline{T}^1, \overline{S}^1) = (\overline{\mathbf{v}}_0^1, \overline{T}_0^1, \overline{S}_0^1) \text{ at } t = 0. \end{cases} \tag{3.20}$$

Remark 3.3. Since $\mathbf{v}^0$ is independent of z, it is easy to check that

$$\overline{K}_1 = 0. \tag{3.21}$$

Moreover, if $\partial T_s/\partial z = \partial S_s/\partial z = 0$ then

$$\overline{K}_2 = \overline{K}_3 = 0. \tag{3.22}$$

Therefore $(\overline{\mathbf{v}}^1, \overline{T}^1, \overline{S}^1) \equiv (0,0,0)$ if $(\overline{\mathbf{v}}_0^1, \overline{T}_0^1, \overline{S}_0^1) = (0,0,0)$ and $\partial T_s/\partial z = \partial S_s/\partial z = 0$. Furthermore, since $(\mathbf{v}^0, T^0, S^0)$ is independent of the vertical variable z, Remark 3.1 and equations (3.6)-(3.7) show that the baroclinic flow $(T^{1,\flat}, S^{1,\flat})$ takes the form

$$(T^{1,\flat}, S^{1,\flat}) = (\frac{z^2}{2} + z + \frac{1}{3})(a^1, a^2), \tag{3.23}$$

where $a = (a^1, a^2)$ is an *explicit* function of $(\mathbf{v}^0, T^0, S^0)$ easily derived from (3.6)-(3.7). □

The k^{th} order approximation.

We now generalize the process described above to the k^{th} order approximation. Although the steps are very similar to the derivations of the

second order approximation, we repeat them for the sake of clarity. We first compute the baroclinic components $\mathbf{v}^{k,b}, T^{k,b}, S^{k,b}$ by considering Proposition 3.1 and the equations for $(\mathbf{v}^{k-1}, T^{k-1}, S^{k-1})$,

$$
\begin{cases}
\epsilon\left[\dfrac{\partial \mathbf{v}^{k-1}}{\partial t} + (\mathbf{v}^{k-2}\cdot\nabla)\mathbf{v}^{k-1} + (\mathbf{v}^{k-1}\cdot\nabla)\mathbf{v}^{k-2} + w^{k-1}\dfrac{\partial \mathbf{v}^{k-2}}{\partial z} + w^{k-2}\dfrac{\partial \mathbf{v}^{k-1}}{\partial z}\right] \\
-\dfrac{\epsilon}{R_{e_1}}\Delta \mathbf{v}^{k-1} - \dfrac{\epsilon}{R_{e_2}}\dfrac{\partial^2 \mathbf{v}^{k,b}}{\partial z^2} + f\,\mathbf{k}_0 \times \mathbf{v}^{k-1} + \operatorname{grad} p^{k-1} = 0, \\
\dfrac{\partial p^{k-1}}{\partial z} = -\rho^{k-1}, \\
\operatorname{div} \mathbf{v}^{k-1} + \dfrac{\partial w^{k-1}}{\partial z} = 0, \\
\epsilon\left[\dfrac{\partial T^{k-1}}{\partial t} + (\mathbf{v}^{k-2}\cdot\nabla)T^{k-1} + (\mathbf{v}^{k-1}\cdot\nabla)T^{k-2} + w^{k-1}\dfrac{\partial T^{k-2}}{\partial z} + w^{k-2}\dfrac{\partial T^{k-1}}{\partial z}\right] \\
-\dfrac{\epsilon}{R_{t_1}}\Delta T^{k-1} - \dfrac{\epsilon}{R_{t_2}}\dfrac{\partial^2 T^{k,b}}{\partial z^2} + F^{-1}\dfrac{\partial T_s}{\partial z}w^{k-1} = 0, \\
\epsilon\left[\dfrac{\partial S^{k-1}}{\partial t} + (\mathbf{v}^{k-2}\cdot\nabla)S^{k-1} + (\mathbf{v}^{k-1}\cdot\nabla)S^{k-2} + w^{k-1}\dfrac{\partial S^{k-2}}{\partial z} + w^{k-2}\dfrac{\partial S^{k-1}}{\partial z}\right] \\
-\dfrac{\epsilon}{R_{s_1}}\Delta S^{k-1} - \dfrac{\epsilon}{R_{t_2}}\dfrac{\partial^2 S^{k,b}}{\partial z^2} + F^{-1}\dfrac{\partial S_s}{\partial z}w^{k-1} = 0, \\
\rho^{k-1} = -\beta_T T^{k-1} + \beta_S S^{k-1},
\end{cases}
\tag{3.24}
$$

and the boundary conditions:

$$
\begin{cases}
\dfrac{1}{R_{e_2}}\dfrac{\partial \mathbf{v}^{k,b}}{\partial z} = 0, w^k = 0, \dfrac{1}{R_{t_2}}\dfrac{\partial T^{k,b}}{\partial z} = -\alpha_T T^{k-1}, \dfrac{1}{R_{s_2}}\dfrac{\partial S^{k,b}}{\partial z} = -\alpha_S S^{k-1} \text{ on } \Gamma_i, \\
\dfrac{\partial \mathbf{v}^{k,b}}{\partial z} = 0, w^k = 0, \dfrac{\partial T^{k,b}}{\partial z} = 0, \dfrac{\partial S^{k,b}}{\partial z} = 0 \text{ on } \Gamma_b, \\
\mathbf{v}^k = 0, w^k = 0, \dfrac{\partial T^k}{\partial n} = 0, \dfrac{\partial S^k}{\partial n} = 0 \text{ on } \Gamma_l.
\end{cases}
\tag{3.25}
$$

In order to recover the barotropic components $(\overline{\mathbf{v}}^k, \overline{T}^k, \overline{S}^k)$, we then need to consider the $(k+1)^{th}$ order. We derive the following equations for $(\mathbf{v}^k, T^k, S^k)$

$$\begin{cases} \epsilon\left[\dfrac{\partial \mathbf{v}^k}{\partial t} + (\mathbf{v}^{k-1}\cdot\nabla)\mathbf{v}^k + (\mathbf{v}^k\cdot\nabla)\mathbf{v}^{k-1} + w^k\dfrac{\partial \mathbf{v}^{k-1}}{\partial z} + w^{k-1}\dfrac{\partial \mathbf{v}^k}{\partial z}\right] \\ -\dfrac{\epsilon}{R_{e_1}}\Delta\mathbf{v}^k - \dfrac{\epsilon}{R_{e_2}}\dfrac{\partial^2\mathbf{v}^{k+1}}{\partial z^2} + f\,\mathbf{k}_0\times\mathbf{v}^k + \text{grad}\ p^k = 0, \\ \dfrac{\partial p^k}{\partial z} = -\rho^k, \\ \text{div}\ \mathbf{v}^k + \dfrac{\partial w^k}{\partial z} = 0, \\ \epsilon\left[\dfrac{\partial T^k}{\partial t} + (\mathbf{v}^{k-1}\cdot\nabla)T^k + (\mathbf{v}^k\cdot\nabla)T^{k-1} + w^k\dfrac{\partial T^{k-1}}{\partial z} + w^{k-1}\dfrac{\partial T^k}{\partial z}\right] \\ -\dfrac{\epsilon}{R_{t_1}}\Delta T^k - \dfrac{\epsilon}{R_{t_2}}\dfrac{\partial^2 T^{k+1}}{\partial z^2} + F^{-1}\dfrac{\partial T_s}{\partial z}w^k = 0, \\ \epsilon\left[\dfrac{\partial S^k}{\partial t} + (\mathbf{v}^{k-1}\cdot\nabla)S^k + (\mathbf{v}^k\cdot\nabla)S^{k-1} + w^k\dfrac{\partial S^{k-1}}{\partial z} + w^{k-1}\dfrac{\partial S^k}{\partial z}\right] \\ -\dfrac{\epsilon}{R_{s_1}}\Delta S^k - \dfrac{\epsilon}{R_{t_2}}\dfrac{\partial^2 S^{k+1}}{\partial z^2} + F^{-1}\dfrac{\partial S_s}{\partial z}w^k = 0, \\ \rho^k = -\beta_T T^k + \beta_S S^k, \end{cases} \tag{3.26}$$

with the boundary conditions

$$\begin{cases} \dfrac{1}{R_{e_2}}\dfrac{\partial \mathbf{v}^{k+1}}{\partial z} = 0, w^{k+1} = 0, \dfrac{1}{R_{t_2}}\dfrac{\partial T^{k+1}}{\partial z} = -\alpha_T T^k, \dfrac{1}{R_{s_2}}\dfrac{\partial S^{k+1}}{\partial z} = -\alpha_S S^k \text{ on } \Gamma_i, \\ \dfrac{\partial \mathbf{v}^{k+1}}{\partial z} = 0, w^{k+1} = 0, \dfrac{\partial T^{k+1}}{\partial z} = 0, \dfrac{\partial S^{k+1}}{\partial z} = 0 \text{ on } \Gamma_b, \\ \mathbf{v}^{k+1} = 0, w^{k+1} = 0, \dfrac{\partial T^{k+1}}{\partial n} = 0, \dfrac{\partial S^{k+1}}{\partial n} = 0 \text{ on } \Gamma_l. \end{cases} \tag{3.27}$$

We rewrite (3.26) like

$$
\begin{cases}
\epsilon\left|\dfrac{\partial \overline{\mathbf{v}}^k}{\partial t}+(\mathbf{v}^{k-1}\cdot\nabla)\overline{\mathbf{v}}^k+(\overline{\mathbf{v}}^k\cdot\nabla)\mathbf{v}^{k-1}\right|-\dfrac{\epsilon}{R_{e_1}}\Delta\overline{\mathbf{v}}^k \\
-\dfrac{\epsilon}{R_{e_2}}\dfrac{\partial^2\mathbf{v}^{k+1}}{\partial z^2}+f\,\mathbf{k}_0\times\mathbf{v}^k+\operatorname{grad}\,p^k+K_1=0, \\
\operatorname{div}\,\overline{\mathbf{v}}^k=0, \\
\epsilon\left|\dfrac{\partial \overline{T}^k}{\partial t}+(\mathbf{v}^{k-1}\cdot\nabla)\overline{T}^k+(\overline{\mathbf{v}}^k\cdot\nabla)T^{k-1}\right|-\dfrac{\epsilon}{R_{t_1}}\Delta\overline{T}^k-\dfrac{\epsilon}{R_{t_2}}\dfrac{\partial^2T^{k+1}}{\partial z^2}+K_2=0, \\
\epsilon\left|\dfrac{\partial \overline{S}^k}{\partial t}+(\mathbf{v}^{k-1}\cdot\nabla)\overline{S}^k+(\overline{\mathbf{v}}^k\cdot\nabla)S^{k-1}\right|-\dfrac{\epsilon}{R_{s_1}}\Delta\overline{S}^k-\dfrac{\epsilon}{R_{s_2}}\dfrac{\partial^2S^{k+1}}{\partial z^2}+K_3=0,
\end{cases}
\tag{3.28}
$$

where K_1, K_2 and K_3 are now given by

$$
\begin{cases}
K_1=\epsilon\left[\dfrac{\partial\mathbf{v}^{k,\flat}}{\partial t}+(\mathbf{v}^{k-1}\cdot\nabla)\mathbf{v}^{k,\flat}+(\mathbf{v}^{k,\flat}\cdot\nabla)\mathbf{v}^{k-1}+w^k\dfrac{\partial\mathbf{v}^{k-1}}{\partial z}+w^{k-1}\dfrac{\partial\mathbf{v}^{k,\flat}}{\partial z}\right] \\
-\dfrac{\epsilon}{R_{e_1}}\Delta\mathbf{v}^{k,\flat}+f\,\mathbf{k}_0\times\mathbf{v}^{k,\flat}, \\
K_2=\epsilon\left[\dfrac{\partial T^{k,\flat}}{\partial t}+(\mathbf{v}^{k-1}\cdot\nabla)T^{k,\flat}+(\mathbf{v}^{k,\flat}\cdot\nabla)T^{k-1}+w^k\dfrac{\partial T^{k-1}}{\partial z}+w^{k-1}\dfrac{\partial T^{k,\flat}}{\partial z}\right] \\
-\dfrac{\epsilon}{R_{t_1}}\Delta T^{k,\flat}+F^{-1}\dfrac{\partial T_s}{\partial z}w^k, \\
K_3=\epsilon\left[\dfrac{\partial S^{k,\flat}}{\partial t}+(\mathbf{v}^{k-1}\cdot\nabla)S^{k,\flat}+(\mathbf{v}^{k,\flat}\cdot\nabla)S^{k-1}+w^k\dfrac{\partial S^{k-1}}{\partial z}+w^{k-1}\dfrac{\partial S^{k,\flat}}{\partial z}\right] \\
-\dfrac{\epsilon}{R_{s_1}}\Delta S^{k,\flat}+F^{-1}\dfrac{\partial S_s}{\partial z}w^k, \\
-\dfrac{\epsilon}{R_{s_1}}\Delta S^{k,\flat}+F^{-1}\dfrac{\partial S_s}{\partial z}w^k, \\
w^k=\displaystyle\int_z^0\operatorname{div}\,\mathbf{v}^{k,\flat}d\zeta.
\end{cases}
\tag{3.29}
$$

Taking the vertical average of (3.28) and using the boundary conditions (3.27), we derive the following system for $(\overline{\mathbf{v}}^k, \overline{T}^k, \overline{S}^k)$

$$\begin{cases} \epsilon\left[\dfrac{\partial \overline{\mathbf{v}}^k}{\partial t} + (\overline{\mathbf{v}}^{k-1}\cdot\nabla)\overline{\mathbf{v}}^k + (\overline{\mathbf{v}}^k\cdot\nabla)\overline{\mathbf{v}}^{k-1}\right] - \dfrac{\epsilon}{R_{e_1}}\Delta\overline{\mathbf{v}}^k + f\,\mathbf{k}_0\times\overline{\mathbf{v}}^k + \text{grad}\,\overline{p}^k + \overline{K}_1 = 0, \\ \text{div}\,\overline{\mathbf{v}}^k = 0, \\ \epsilon\left[\dfrac{\partial \overline{T}^k}{\partial t} + (\overline{\mathbf{v}}^{k-1}\cdot\nabla)\overline{T}^k\right] - \dfrac{\epsilon}{R_{t_1}}\Delta\overline{T}^k + \epsilon\alpha_T\overline{T}^k + \overline{K}_2 = -\epsilon\alpha_T T^{k,b}(0), \\ \epsilon\left[\dfrac{\partial \overline{S}^k}{\partial t} + (\overline{\mathbf{v}}^{k-1}\cdot\nabla)\overline{S}^k\right] - \dfrac{\epsilon}{R_{s_1}}\Delta\overline{S}^k + \epsilon\alpha_S\overline{S}^k + \overline{K}_3 = -\epsilon\alpha_S S^{k,b}(0), \end{cases} \tag{3.30}$$

with the boundary and initial conditions

$$\begin{cases} \overline{\mathbf{v}}^k = 0, \dfrac{\partial \overline{T}^k}{\partial n} = 0, \dfrac{\partial \overline{S}^k}{\partial n} = 0 \text{ on } \Gamma_l, \\ (\overline{\mathbf{v}}^k, \overline{T}^k, \overline{S}^k) = (\overline{\mathbf{v}}_0^k, \overline{T}_0^k, \overline{S}_0^k) \text{ at } t = 0. \end{cases} \tag{3.31}$$

Remark 3.5. For $k \geq 2$ and K_1, K_2 and K_3 given by (3.29), equality (3.21) is not necessary satisfied since $\mathbf{v}^{k-1}$, T^{k-1} and S^{k-1} are now functions of the vertical variable z as well.

Remark 3.6. Note that in the equations (3.19) or (3.30) for the barotropic flow $\overline{\mathbf{v}}^k$, the vertical average $\overline{p}^k$ of the pressure p^k appears as a Lagrange multiplier for the constraint $\text{div}\,\overline{\mathbf{v}}^k = 0$. Hence the first two equations in (3.30) and the conditions on $\overline{\mathbf{v}}^k$ in (3.31) determine at once $\overline{\mathbf{v}}^k$ and $\overline{p}^k$. However, at each level $k \geq 1$ we need to recover the whole pressure p^{k-1} in order to compute the baroclinic component $(\mathbf{v}^{k,b}, T^{k,b}, S^{k,b})$ (see equations (3.24)). This is achieved as follows. Using the following relations (which are satisfied at each level)

$$\frac{\partial p^{k-1}}{\partial z} = -\rho^{k-1}, \rho^{k-1} = -\beta_T T^{k-1} + \beta_S S^{k-1}, \tag{3.32}$$

we obtain

$$p^{k-1}(x,y,z) = -\int_{-1}^{z} \rho^{k-1}(x,y,s)ds + p^{k-1}(x,y,-1). \tag{3.33}$$

Let us recall that

$$\overline{p}^{k-1} = \int_{-1}^{0} p^{k-1}(x,y,z)dz. \tag{3.34}$$

From (3.33)-(3.34), we obtain by integration in z from -1 to 0:

$$p^{k-1}(x,y,-1) = \overline{p}^{k-1} + \int_{-1}^{0} \left(\int_{-1}^{z} \rho^{k-1}(x,y,s)ds \right) dz. \tag{3.35}$$

Hence we have

$$p^{k-1}(x,y,z) = -\int_{-1}^{z} \rho^{k-1}(x,y,s)ds + \overline{p}^{k-1} + \int_{-1}^{0} \left(\int_{-1}^{z} \rho^{k-1}(x,y,s)ds \right) dz, \tag{3.36}$$

where ρ^{k-1} known at this stage is given by (3.32). Thus p^{k-1} is now fully known and we can proceed with the resolution. □

Remark 3.7. The existence and uniqueness of solutions $\mathbf{v}^0$ to the QG equations (3.8)-(3.9) is easily proved using the same idea as for the Navier-Stokes equations in two- dimensional space, [4, 10]. Now, assuming enough regularity on the velocity $\mathbf{v}^0$, one can then prove the existence and uniqueness of solutions T^0 and S^0 to (3.8)-(3.9).

For $k \geq 1$, the existence and uniqueness of the solutions $\overline{\mathbf{v}}^k$ to the system (3.19)-(3.20) or (3.30)-(3.31) is easily proved (provided that K_1 is regular enough) using the same method as for the linearized Navier-Stokes equations, [4, 10]. Assuming enough regularity on the velocity $\overline{\mathbf{v}}^k$ and the data K_1, K_2 and K_3, one can then prove the existence and uniqueness of solutions $\overline{T}^k$ and $\overline{S}^k$ to (3.19)-(3.20) or (3.30)-(3.31). The regularity of the solution $\overline{\mathbf{v}}^k$ to (3.19)-(3.20) or (3.30)-(3.31) and of the data K_1, K_2, and K_3 given in (3.18) or (3.29) depends on the regularity of the solutions at the lower level. The mathematical analysis of the models presented in this article as well as the convergence of the asymptotics will be addressed elsewhere.

Remark 3.8. The coefficients α_T and α_S appearing in (2.25) should also be expanded with respect to δ, namely

$$\begin{aligned} \alpha_T &= \alpha_T^0 + \delta\alpha_T^1 + \delta^2\alpha_T^2 + \cdots, \\ \alpha_S &= \alpha_S^0 + \delta\alpha_S^1 + \delta^2\alpha_S^2 + \cdots. \end{aligned} \tag{3.37}$$

In this work, we have assumed that $\alpha_T^0 \neq 0$ and $\alpha_S^0 \neq 0$, but different situations where the first components of α_T and α_S vanish may occur depending on the corresponding physical problem. Other surface flux boundary conditions for S and T may also be used, see [2]. These issues will be addressed elsewhere.

ACKNOWLEDGMENTS

This work was supported in part by NSF Grant DMS0305110, DOE Grant DE-FG02-01ER63251:A000, and by the Research Fund of Indiana University.

REFERENCES

[1] P.F. Emblid and A. Majda. Averaging over fast gravity waves for geophysical flows with arbitrary potential voticity. *Comm. Partial Differential Equations*, 21(3-4):619-658, 1996.

[2] L. Fleury and O. Thual. Stationary fronts of the thermohaline circulation in the low-aspect-ratio limit. *J. Fluid Mech.*, 349:117-147, 1997.

[3] G.J. Haltiner and R.T. Williams. *Numerical prediction and dynamic metereology*. John Wiley and Sons, New York, 1980.

[4] J.-L. Lions. *Oeuvres choisies de Jacques-Louis Lions*, volume I, Equations aux dérivées partielle-Interpolation. EDP Sciences, Paris, 2003.

[5] J.-L. Lions, R. Temam, and S. Wang. On the equations of large-scale ocean. *Nonlinearity*, 5:1007-1053, 1992.

[6] J-L. Lions, R. Temam, and S. Wang. Models of the coupled atmosphere and ocean (CAO I). *Computational Mechanics Advance*, 1:3-54, 1993.

[7] J-L. Lions, R. Temam, and S. Wang. Numerical analysis of the coupled atmosphere and ocean (CAO II). *Computational Mechanics Advance*, 1:55-120, 1993.

[8] J-L. Lions, R. Temam, and S. Wang. Mathematical study of the coupled models of atmosphere and ocean (CAO III). *Math. Pures et Appl.*, 73:105-163, 1995.

[9] J-L. Lions, R. Temam, and S. Wang. On Mathematical problems for the primitive equations of the ocean: the mesoscale midlatitude case. Lakshmikantham's legacy: a tribute on his 75th birthday. *Nonlinear Anal.*, 40(1-8, Ser. A. Theory Methods): 439-482, 2000.

[10] J-L. Lions. *Quelques méthodes de résolution des problèmes aux limites non linèaires*. Dunod; Gauthier-Villars, Paris, 1969.

[11] T.Tachim-Medjo, R. Temam, and S. Wang. Hight order approximation equations for the primitive equations of the atmosphere. *J.Engrg. Math., Special issue on Large-Scale Numerical Modeling of Problems Involving the Navier-Strokes equations.*, 32:237-256 1997.

[12] Pedlosky. *Geophisycal Fluid Dynamics*. Springer-Verlag, New-York, second edition, 1987.

[13] J.P. Peixoto and A.H. Oort. *Physics of Climate*. American Institute of Physics, New-York, 1992.

[14] R.T. Temam and M. Ziane. Some mathematical problem in geophysical fluid dynamics. In S. Friedlander and D. Serre, editors, *Handbook of Mathematical Fluid Dinamics, Vol. III* Elsevier, 2003, to appear.

[15] S. Wang. *On Solvability for the Equations of the Large-Scale atmospheric Motion.* PhD thesis, Lanzhou University, China, 1988.

[16] W.M. Washington and C. L. Parkinson. *An Introduction to Three-Dimentional Climate Modeling.* Oxford University Press, Oxford, 1986.

HAHN–BANACH THEOREMS AND MAXIMAL MONOTONICITY

S. Simons
Dept. of Mathematics, University of California, Santa Barbara, USA

1. INTRODUCTION

In this paper, we discuss new versions of the Hahn–Banach theorem that have a number of applications in different fields of analysis. We shall give applications to linear and nonlinear functional analysis, convex analysis, and the theory of monotone multifunctions. All vector spaces in this paper will be *real*.

The main result appears in Theorem 2.8, which is bootstrapped from the special case contained in Lemma 2.4.

In Section 3, we sketch how Theorem 2.8 can be used to give the main existence theorems for linear functionals in functional analysis, and also how it gives a result that leads to a minimax theorem. We also discuss three applications of Theorem 2.8 to convex analysis, pointing the reader to [26] for further details in two of these cases. One noteworthy property of proofs using Theorem 2.8 is that they allow us to avoid the problem of the "vertical hyperplane".

In Section 4, we show how Theorem 2.8 can be used to obtain considerable insight on the existence of Lagrange multipliers for constrained convex minimization problems. The usual *sufficient* condition for the existence of such multipliers is normally found using the Eidelheit separation theorem. In Theorem 4.5, we use Theorem 2.8 to derive this sufficient condition, with the added bonus that we obtain a bound on the

norm of the multiplier. Here again, the proof using Theorem 2.8 allows us to avoid the problem of the "vertical hyperplane". More to the point, the results leading up to Theorem 4.5, namely Lemma 4.1 and Theorem 4.2, use Theorem 2.8 to obtain a *necessary and sufficient* condition for the existence of Lagrange multipliers, with a *sharp lower bound* on the norm of the multiplier.

Section 5 is motivated by the theory of monotone multifunctions. Theorem 5.1 is an existence theorem without any *a priori* scalar bounds in normed spaces that has proved very useful in the investigation of these multifunctions, and will be used in Theorem 6.5. A new feature of the result as presented here is a sharp lower bound on the norm of the linear functional obtained. Theorem 5.3 is a two–stage result obtained by combining Theorems 5.1 and 2.8, and will be used in the proof of Theorem 7.4.

In Section 6, we discuss the *free convexification* technique, which has many applications to the theory of monotone multifunctions. We list several of these without proof. We also use Theorem 5.1 to derive Rockafellar's surjectivity theorem for general (i.e., not renormed) reflexive spaces, with a sharp lower bound for solutions of the problem. Apart from its intrinsic interest, we have given this result here to introduce the techniques that are used in the more difficult problem treated in Section 7.

Maximal monotone multifunctions of "type (D)" were introduced by Gossez in order to generalize to nonreflexive spaces some of the results previously known for reflexive spaces (see Gossez, [10, Lemme 2.1, p. 375] and Phelps, [15, Section 3] for an exposition). Maximal monotone multifunctions of "type (FP)" were introduced by Fitzpatrick–Phelps in [8, Section 3] under the name of "locally maximal monotone" multifunctions. The motivation for their introduction was as follows. If E is reflexive then every maximal monotone multifunction on E can be approximated by "nicer" maximal monotone multifunctions using the Moreau-Yosida approximation. If E is nonreflexive then every subdifferential can also be approximated by "nicer" subdifferentials by using the operation of inf–convolution. So the question arises whether a general maximal monotone multifunction on a nonreflexive space can also be approximated by "nicer" maximal monotone multifunctions in some appropriate sense. Fitzpatrick–Phelps defined an appropriate sense of approximation in [8], and showed that the multifunctions of type (FP) can be approximated by "nicer" maximal monotone multifunctions in their sense. There has been considerable speculation for the past several years about the relationship between multifunctions of type (D) and multifunctions of type (FP). The main result of Section 7 (in Theorem 7.4) is then that *every maximal monotone multifunction of type (D) is necessarily of type (FP).*

In the final section, we return to our consideration of abstract Hahn–Banach theorems. Noting a certain formal similarity between the statements of Theorem 5.1 and Theorem 2.8, we ask the question whether these two results can be unified. Indeed, they have a common generalization, which is given in Theorem 8.1.

2. THE MAIN RESULT

Theorem 2.8 contains the new version of the Hahn-Banach theorem that forms the main topic of this paper. Theorem 2.8 is proved by bootstrapping from the special case contained in Lemma 2.4 – most of the work is actually done in Lemma 2.3.

We start by recalling in Lemma 2.2 the classical Hahn–Banach theorem for sublinear functionals.

Definition 2.1. Let E be a nontrivial vector space. We say that $S : E \mapsto \mathbb{R}$ is *sublinear* if

$$x, y \in E \;\Rightarrow\; S(x+y) \leq S(x) + S(y)$$

and

$$x \in E \text{ and } \lambda > 0 \;\Rightarrow\; S(\lambda x) = \lambda S(x).$$

Lemma 2.2. *Let E be a nontrivial vector space and $S : E \mapsto \mathbb{R}$ be sublinear. Then there exists a linear functional L on E such that $L \leq S$ on E.*

Proof. See Kelly–Namioka, [11, 3.4, p. 21] for a proof using cones, Rudin, [20, Theorem 3.2, p. 56–57] for a proof using an extension by subspaces argument, and König, [12] and Simons, [21] for a proof using an ordering on sublinear functionals. □

Lemma 2.3. *Let E be a nontrivial vector space and $S : E \mapsto \mathbb{R}$ be sublinear. Let D be a nonempty convex subset of a vector space, $a : D \mapsto E$ be affine and $\beta := \inf_D S \circ a \in \mathbb{R}$. For all $x \in E$, let*

$$T(x) := \inf_{d \in D,\ \lambda > 0} [S(x + \lambda a(d)) - \lambda\beta]. \tag{2.3.1}$$

Then $T : E \mapsto \mathbb{R}$, T *is sublinear,* $T \leq S$ *on* E *and*
$-T(-a(d)) \geq \beta, \quad \forall d \in D,.$

Proof. If $x \in E$, $d \in D$ and $\lambda > 0$ then

$$S(x + \lambda a(d)) - \lambda\beta \geq -S(-x) + \lambda S(a(d)) - \lambda\beta \geq -S(-x) > -\infty.$$

Taking the infimum over $d \in D$ and $\lambda > 0$, $T(x) \geq -S(-x) > -\infty$. Thus $T : E \mapsto \mathbb{R}$. It is now easy to check that T is positively homogeneous, so to prove that T is sublinear it remains to show that T is subadditive. To this end, let $x_1, x_2 \in E$. Let $d_1, d_2 \in D$ and $\lambda_1, \lambda_2 > 0$ be arbitrary. Write $x := x_1 + x_2, \lambda := \lambda_1 + \lambda_2, \mu_i := \lambda_i/\lambda$ and $d := \mu_1 d_1 + \mu_2 d_2$. Then, using the fact that $\mu_1 a(d_1) + \mu_2 a(d_2) = a(d)$,

$$\begin{aligned}
[S(x_1 + \lambda_1 a(d_1)) - \lambda_1\beta] + [S(x_2 + \lambda_2 a(d_2)) - \lambda_2\beta] & \\
&\geq S(x + \lambda_1 a(d_1) + \lambda_2 a(d_2)) - \lambda\beta \\
&= \lambda S(x/\lambda + \mu_1 a(d_1) + \mu_2 a(d_2)) - \lambda\beta, \\
&= \lambda S(x/\lambda + a(d)) - \lambda\beta \\
&= S(x + \lambda a(d)) - \lambda\beta \\
&\geq T(x) = T(x_1 + x_2).
\end{aligned}$$

Taking the infimum over d_1, d_2, λ_1 and λ_2 gives

$$T(x_1) + T(x_2) \geq T(x_1 + x_2).$$

Thus T is subadditive, and consequently, sublinear. Fix $d \in D$. Let x be an arbitrary element of E. Then, for all $\lambda > 0$,

$$T(x) \leq S(x) + \lambda[S(a(d)) - \beta].$$

Letting $\lambda \to 0$, $T(x) \leq S(x)$. Thus $T \leq S$ on E. Finally, let d be an arbitrary element of D. Then, taking $\lambda = 1$ in (2.3.1),

$$T(-a(d)) \leq S(-a(d) + a(d)) - \beta = -\beta,$$

hence $-T(-a(d)) \geq \beta$, which completes the proof of Lemma 2.3. □

Lemma 2.4. *Let* E *be a nontrivial vector space and* $S : E \mapsto \mathbb{R}$ *be sublinear. Let* D *be a nonempty convex subset of a vector space and*

$a : D \mapsto E$ be affine. Then there exists a linear functional L on E such that $L \leq S$ on E and

$$\inf_D L \circ a = \inf_D S \circ a.$$

Proof. Let $\beta := \inf_D S \circ a$. If $\beta = -\infty$, the result is immediate from Lemma 2.2 (take any linear functional L on E such that $L \leq S$ on E). So we can suppose that $\beta \in \mathbb{R}$. Define T as in Lemma 2.3. From Lemma 2.2, there exists a linear functional L on E such that $L \leq T$ on E. Since $T \leq S$ on E, $L \leq S$ on E, as required. Let $d \in D$. Then

$$L(a(d)) = -L(-a(d)) \geq -T(-a(d)) \geq \beta.$$

Taking the infimum over $d \in D$,

$$\inf_D L \circ a \geq \beta = \inf_C S \circ a.$$

On the other hand, since $L \leq S$ on E, $\inf_D L \circ a \leq \inf_D S \circ a$. □

Definition 2.5. Let C be a nonempty convex subset of a vector space and $\mathcal{PC}(C)$ stand for the set of all convex functions $k : C \mapsto (-\infty, \infty]$ such that $\operatorname{dom} k \neq \emptyset$, where $\operatorname{dom} k$, the *effective domain* of k, is defined by

$$\operatorname{dom} k := \{x \in C : k(x) \in \mathbb{R}\}.$$

(The "$\mathcal{P}$" stands for "proper", which is the adjective frequently used to denote the fact that a function is finite at least at one point.)

Definition 2.6. Let E be a nontrivial vector space and $S : E \mapsto \mathbb{R}$ be sublinear. Let C be a nonempty convex subset of a vector space and $j : C \mapsto E$. We say that j is *S–convex* if

$$\left.\begin{array}{l} x_1, x_2 \in C, \mu_1, \mu_2 > 0 \\ \text{and} \qquad \mu_1 + \mu_2 = 1 \end{array}\right\} \Rightarrow \; S(j(\mu_1 x_1 + \mu_2 x_2) - \mu_1 j(x_1) - \mu_2 j(x_2)) \leq 0.$$

Note that if we define an ordering "$\leq_S$" on E by declaring that $y \leq_S z$ if $S(y - z) \leq 0$ then j is S–convex if, and only if,

$$\left.\begin{array}{l} x_1, x_2 \in C, \mu_1, \mu_2 > 0 \\ \text{and} \quad \mu_1 + \mu_2 = 1 \end{array}\right\} \Rightarrow j(\mu_1 x_1 + \mu_2 x_2) \leq_S \mu_1 j(x_1) + \mu_2 j(x_2).$$

An affine function is clearly S–convex.

Remark 2.7. Suppose that C_S is the level set $\{y \in E\colon S(y) \leq 0\}$. It is clear that the ordering $\leq_S$ on E is determined solely by C_S (though the proof of Theorem 2.8 depends on the other values of S). Now let us consider the special case when $E = \mathbb{R}$. Since C_S is a convex cone with vertex at the origin, there are exactly four possibilities for C_S, namely $\{0\}$, $(-\infty, 0], [0, \infty)$ and $\mathbb{R}$. These can be realized by $S(y) := |\, y \,|$, $S(y) := y$, $S(y) := -y$ and $S(y) := 0$, respectively. In these four cases, "S–convex" means "affine", "convex", "concave" and "arbitrary", repectively. In general, when $E \neq \mathbb{R}$, there is no analog of convex or concave function from C into E, and it makes sense to ask the question when a function $j\colon C \mapsto E$ is S–convex with respect to some nontrivial sublinear functional S on E. A solution to this problem has been provided by Giandomenico Mastroeni (personal communication).

Theorem 2.8. *Let E a nontrivial vector space and $S\colon E \mapsto \mathbb{R}$ be sublinear. Let C be a nonempty convex subset of a vector space, $k \in \mathcal{PC}(C)$ and $j\colon C \mapsto E$ be S–convex. Then there exists a linear functional L on E such that $L \leq S$ on E and*

$$\inf_C [L \circ j + k] = \inf_C [S \circ j + k]. \tag{2.8.1}$$

Proof. Let $\tilde{E} := E \times \mathbb{R}$, and define $\tilde{S}\colon \tilde{E} \mapsto \mathbb{R}$ by

$$\tilde{S}(y, \lambda) := S(y) + \lambda \quad ((y, \lambda) \in \tilde{E}).$$

Then, as the reader can easily verify, $\tilde{S}$ is sublinear. Let

$$D := \{(x, y, \lambda) \in C \times E \times \mathbb{R}\colon S(j(x) - y) \leq 0,\ k(x) \leq \lambda\},$$

and $a\colon D \mapsto \tilde{E}$ be defined by

$$a(x, y, \lambda) := (y, \lambda) \quad ((x, y, \lambda) \in D).$$

Then D is a nonempty convex set and a is an affine function. Lemma 2.4 with E replaced by $\tilde{E}$, S by $\tilde{S}$, and C by D now gives a linear functional $\tilde{L}$ on $\tilde{E}$ such that

$$\tilde{L} \leq \tilde{S} \text{ on } \tilde{E} \quad \text{and} \quad \inf_D \tilde{L} \circ a = \inf_D \tilde{S} \circ a.$$

Since $\tilde{L} \leq \tilde{S}$ on $\tilde{E}$, there exists a linear functional L on E such that

$$L \leq S \text{ on } E \quad \text{and} \quad (y, \lambda) \in \tilde{E} \Rightarrow \tilde{L}(y, \lambda) = L(y) + \lambda.$$

The result follow since, by direct computation,

$$\inf_D \tilde{L} \circ a = \inf_C [L \circ j + k] \quad \text{and} \quad \inf_D \tilde{S} \circ a = \inf_C [S \circ j + k]. \qquad \square$$

3. APPLICATIONS TO FUNCTIONAL ANALYSIS AND MINIMAX THEOREMS

In this section, we mention without proof a number of applications of Theorem 2.8 that were discussed in [26].We then state and prove in Theorem 3.5 a (necessary and sufficient) criterion for the Fenchel duality condition to hold.

Theorem 3.1 is the *sandwich theorem* (see [12, Theorem 1.7, p. 112]). It follows immediately from Theorem 2.8 with $C := E$ and $j(x) := x$.

Theorem 3.1. *Let E be a nontrivial vector space, S: $E \mapsto \mathbb{R}$ be sublinear $k \in \mathcal{PC}(E)$ and $-k \leq S$ on E. Then there exists a linear functional L on E such that $-k \leq L \leq S$ on E.*

Theorem 3.1 implies in turn two other well known existence results: the *extension form of the Hahn–Banach theorem,* Corollary 3.2, (see [12, Corollary 1.8, p. 112]) and the *Mazur–Orlicz theorem,* Corollary 3.3, (see [12, Theorem 1.9, p. 112]).

Corollary 3.2. *Let E be a nontrivial vector space, F be a linear subspace of E, S: $E \mapsto \mathbb{R}$ be sublinear, M: $F \mapsto \mathbb{R}$ be linear and $M \leq S$ on F. Then there exists a linear functional L on E such that $L \leq S$ on E and $L|_F = M$.*

Corollary 3.3. *Let E be a nontrivial vector space, $S\colon E \mapsto \mathbb{R}$ be sublinear and C be a nonempty convex subset of E. Then there exists a linear functional L on E such that $L \leq S$ on E and* $\inf_C L = \inf_C S$.

Theorem 3.4 below was essentially proved by Fan–Glicksberg–Hoffman (see [6,Theorem 1, p. 618]), and leads to a short proof of the minimax theorem proved by Fan in [5] (see [23, Theorem 3.1, p. 17] for details of this). Theorem 3.4 follows easily from Theorem 2.8 with $E := \mathbb{R}^m$, $S(\mu_1,\ldots,\mu_m) := \mu_1 \vee \cdots \vee \mu_m$, $j(c) := (f_1(c),\ldots,f_m(c))$ and $k(c) := 0$.

Theorem 3.4. *Let C be a nonempty convex subset of a vector space and $f_1,\ldots,f_m$ be convex real functions on C. Then there exist $\lambda_1,\ldots,\lambda_m \geq 0$ such that $\lambda_1 + \cdots + \lambda_m = 1$ and*

$$\inf_C[f_1 \vee \ldots \vee f_m] = \inf_C[\lambda_1 f_1 + \cdots + \lambda_m f_m].$$

Let E be a nontrivial Hausdorff locally convex space with dual E^*. If $f \in \mathcal{PC}(E)$, the *Fenchel conjugate*, f^*, of f is the function from E^* into $(-\infty,\infty]$ defined by

$$f^*(x^*) := \sup_E(x^* - f).$$

It follows easily from the definitions above that, for all $y \in E$,

$$f(y) \geq \sup_{E^*}(y - f^*). \tag{3.4.1}$$

It was proved by Moreau in [14, Section 5–6, p. 26–39] that if f is lower semicontinuous on E then, for all $y \in E$, we have equality in (3.4.1). If f is lower semicontinuous at $y \in E$ but not on E then it does *not* follow that equality holds in (3.4.1) (see [26, Remark 3.1]). On the other hand, Theorem 2.8 can be used to find a necessary and sufficient condition for equality to hold in (3.4.1) for a given $y \in E$ (see [26, Theorem 3.2]). This provides a proof of Moreau's original result with the advantage that we do not have to deal with the elimination of the "vertical hyperplane".

We now show how Theorem 2.8 leads to a version of the Fenchel duality theorem.

Theorem 3.5. *Let E be a nontrivial Hausdorff locally convex space with dual E^*, and $f, g \in \mathcal{PC}(E)$. Then*

$$\text{there exists } z^* \in E^* \text{ such that } f^*(-z^*) + g^*(z^*) \leq 0 \tag{3.5.1}$$

if, and only if, writing $\mathcal{S}(E)$ *for the family of continuous seminorms on* E,

$$\text{there exists } S \in \mathcal{S}(E) \text{ such that } x, y \in E \Rightarrow f(x) + g(y) + S(x - y) \geq 0. \tag{3.5.2}$$

Proof. Suppose first that (3.5.1) is satisfied. Then, for all $x, y \in E$,

$$\langle x, -z^* \rangle - f(x) + \langle y, z^* \rangle - g(y) \leq f^*(-z^*) + g^*(z^*) \leq 0,$$

consequently,

$$f(x) + g(y) + \langle x - y, z^* \rangle \geq 0,$$

and (3.5.2) follows with $S := \mid z^* \mid$. Suppose, conversely, that (3.5.2) is satisfied. Then we apply Theorem 2.8 with $C := E \times E$, $j(x, y) := x - y$ and $k(x, y) := f(x) + g(y)$, and obtain a linear functional L on E such that $L \leq S$ and

$$x, y \in E \Rightarrow f(x) + g(y) + L(x - y) \geq 0,$$

or equivalently,

$$x, y \in E \Rightarrow (-L)(x) - f(x) + L(y) - g(y) \leq 0.$$

(3.5.1) now follows (with $z^* = L$) by taking the supremum over x and y. □

In the normed case, Theorem 3.5 takes the following form:

Corollary 3.6. *Let* E *be a nontrivial normed space with dual* E^*, *and* $f, g \in \mathcal{PC}(E)$. *Then*

$$\text{there exists } z^* \in E^* \text{ such that } f^*(-z^*) + g^*(z^*) \leq 0 \tag{3.5.1}$$

if, and only if, there exists $M \geq 0$ *such that*

$$x, y \in E \Rightarrow f(x) + g(y) + M \parallel x - y \parallel \; \geq \; 0.$$

Corollary 3.6 leads easily to proofs of the versions of the Fenchel duality theorem and the formula for the subdifferential of a sum due to Moreau–Rockafellar (see [17, Theorem 3, p. 85]) and Attouch–Brezis (see [1, Theorem 1.1, p. 126–127] and [1, Corollary 2.1, p. 130–131]). Yet again, we

do not have to deal with the elimination of the "vertical hyperplane". We emphasize that Theorem 3.5 and Corollary 3.6 give a *necessary and sufficient* condition for the existence of the linear functional, and not merely *sufficient* conditions.

In [19], Rockafellar develops a theory of dual problems and Lagrangians that gives a very large number of results in convex analysis. It was shown in [26, Theorem 3.6] how Theorem 2.8 can be used to give an efficient proof of [19, Theorem 17(a), p. 41], one of the main existence results in [19].

4. A SHARP RESULT ON THE EXISTENCE OF LAGRANGE MULTIPLIERS

This section is about Lagrange multipliers for the constrained convex optimization problem outlined below. The main result is Theorem 4.2 which, combined with Lemma 4.1, gives a necessary and sufficient condition for the existence of a Lagrange multiplier, with a sharp lower bound on its norm. We also show in Theorem 4.5 how Theorem 4.2 implies the classical result, with an upper bound on the norm as a bonus. The analysis in this section depends only on Theorem 2.8 — it does not depend on Section 3 in any way.

Let $(E, \|\cdot\|)$ be a nontrivial normed space, C be a nonempty convex subset of a vector space, k: $C \mapsto \mathbb{R}$ be convex, j: $C \mapsto E$, and $\preceq$ be a partial ordering on E compatible with its vector space structure. Let N be the negative cone $\{y \in E\colon\ y \preceq 0\}$. Suppose that

$$\left.\begin{array}{l} x_1, x_2 \in C, \mu_1, \mu_2 > 0 \\ \text{and} \qquad \mu_1 + \mu_2 = 1 \end{array}\right\} \Rightarrow j(\mu_1 x_1 + \mu_2 x_2) \preceq \mu_1 j(x_1) + \mu_2 j(x_2) \tag{4.0.1}$$

(i.e., j is *convex* with respect to $\preceq$), and

$$\inf_{j^{-1}N} k = \inf\{k(x)\colon\ x \in C,\ j(x) \preceq 0\} = \mu_0 \in \mathbb{R}. \tag{4.0.2}$$

A *Lagrange multiplier* for the problem is an element z_0^* of E such that

$$\sup_N z_0^* \leq 0 \tag{4.0.3}$$

(i.e., z_0^* is positive with respect to $\preceq$), and

$$\inf_{x \in C}\left[\left\langle j(x), z_0^*\right\rangle + k(x)\right] = \mu_0. \tag{4.0.4}$$

Clearly 0 is a Lagrange multiplier $\Leftrightarrow \inf_C k \geq \mu_0$. In order to exclude this trivial case, we shall suppose that $\inf_C k < \mu_0$. Let

$$A := \{x \in C\colon k(x) < \mu_0\} \quad \text{and} \quad B := \{v \in C\colon j(v) \prec 0\}, \tag{4.0.5}$$

where we write $j(v) \prec 0$ to mean that $j(v) \in \operatorname{int} N$. The above conditions imply that $A \neq \emptyset$. We start off with a simple consequence of the existence of a Lagrange multiplier.

Lemma 4.1. *Let z_0^* be a Lagrange multiplier, and A be as in (4.0.5). Then*

$$0 < \sup_{x \in A} \frac{\mu_0 - k(x)}{\operatorname{dist}(j(x), N)} \leq \| z_0^* \| < \infty.$$

Proof. Let $x \in A$, and u be an arbitrary element of N. Then, from (4.0.3) and (4.0.4),

$$\| j(x) - u \| \| z_0^* \| \geq \langle j(x), z_0^* \rangle - \langle u, z_0^* \rangle \geq \langle j(x), z_0^* \rangle \geq \mu_0 - k(x) > 0.$$

Taking the infimum over $u \in N$,

$$\operatorname{dist}(j(x), N) \| z_0^* \| \geq \mu_0 - k(x) > 0.$$

The result follows on division by $\operatorname{dist}(j(x), N)$ and then taking the supremum over $x \in A$.

□

The main result of this section is the following partial converse to Lemma 4.1.

Theorem 4.2. *Suppose that* $0 < \mathrm{M} := \sup_{x \in A} \dfrac{\mu_0 - k(x)}{\operatorname{dist}(j(x), N)} < \infty$. *Then there exists a Lagrange multiplier z_0^* such that $\| z_0^* \| \leq M$. It then follows from Lemma 4.1 that* $M = \min \{\| z_0^* \| : z_0^* \text{ is a Lagrange multiplier}\}$.

Proof. Let $S\colon E \mapsto [0, \infty)$ be defined by

$$S(y) := \operatorname{dist}(y, N) = \inf_{u \in N} \| y - u \| \quad (y \in E).$$

It is easily checked from this definition that

S is sublinear, (4.2.1)

$$S \leq \|\cdot\| \quad \text{on } E, \tag{4.2.2}$$

and

$$y \in N \Rightarrow S(y) = 0. \tag{4.2.3}$$

The definition of M gives

$$x \in A \Rightarrow MS \circ j(x) + k(x) \geq \mu_0.$$

Since $k \geq \mu_0$ on $C \setminus A$ and $S \geq 0$ on E, in fact

$$x \in C \Rightarrow MS \circ j(x) + k(x) \geq \mu_0,$$

that is to say

$$\inf_C \ [MS \circ j + k] \ \geq \ \mu_0.$$

Let $x_1, x_2, \in C$, $\mu_1, \mu_2 > 0$ and $\mu_1 + \mu_2 = 1$. Then it follows from (4.0.1) that

$$j(\mu_1 x_1 + \mu_2 x_2) - \mu_1 j(x_1) - \mu_2 j(x_2) \in N,$$

and so (4.2.3) implies that j is MS-convex. Thus (4.2.1) and Theorem 2.8 give a linear functional L on E such that $L \leq MS$ on E and

$$\inf_C [L \circ j + k] = \inf_C [MS \circ j + k] \geq \mu_0. \tag{4.2.4}$$

We now derive from (4.2.2) and (4.2.3) that $L \in E^*$, $\| L \| \leq M$ and $\sup_N L \leq 0$. Since $x \in j^{-1}(N) \Rightarrow j(x) \in N \Rightarrow L \circ j(x) \leq 0$, (4.0.2) now gives

$$\mu_0 = \inf_{j^{-1}N} k \geq \inf_{j^{-1}N} [L \circ j + k] \geq \inf_C [L \circ j + k].$$

Thus we have equality in (4.2.4), which gives the required result (with $z_0^* = L$). □

Remark 4.3. At this point, we make some comments about the formulation of the preceding analysis in terms of Lagrangians. Let

$$\mathcal{P} := \{z^* \in E^* : \sup_N z^* \leq 0\},$$

and define $L : C \times \mathcal{P} \mapsto \mathbb{R}$ by $L(x, z^*) := \langle j(x), z^* \rangle + k(x)$. Then z_0^* is a Lagrange multiplier exactly when $\inf_{x \in C} L(x, z_0^*) = \mu_0$. Arguing as in the final few lines of Theorem 4.2, if $z^* \in \mathcal{P}$ then $\inf_{x \in C} L(x, z^*) \leq \mu_0$, so in fact

$$\sup_{z^* \in \mathcal{P}} \inf_{x \in C} L(x, z^*) = \inf_{x \in C} L(x, z_0^*) = \mu_0.$$

In the event that there exists $x_0 \in j^{-1}N$ such that $k(x_0) = \mu_0$ then (x_0, z_0^*) is a saddle point of L. See [13, Corollary 8.3.1, p. 219] for details of the argument.

We recall from (4.0.5) that $B := \{v \in C : j(v) \prec 0\}$. The classical sufficient condition for the existence of Lagrange multipliers is that $B \neq \emptyset$. (See [13, Theorem 8.3.1, p. 217–218].) This will be improved in Theorem 4.5. We first give a preliminary lemma.

Lemma 4.4.
(a) Let $x \in A, u \in N, v \in B, 0 < \eta < \operatorname{dist}(j(v), E \setminus N)$ and $\alpha := \|j(x) - u\|$. Then

$$j\left(\frac{\eta x + \alpha v}{\eta + \alpha}\right) \preceq 0.$$

(b) Let $x \in A$ and $v \in B$. Then

$$\operatorname{dist}(j(x), N)(k(v) - \mu_0) \geq \operatorname{dist}(j(v), E \setminus N)(\mu_0 - k(x)) > 0.$$

Proof. (a) If $\alpha = 0$ then $j(x) = u$ and so

$$j\left(\frac{\eta x + \alpha v}{\eta + \alpha}\right) = j(x) = u \preceq 0,$$

which gives the required result. If $\alpha > 0$ then

$$\left\|\frac{\eta}{\alpha}(j(x) - u)\right\| = \eta < \operatorname{dist}(j(v), E \setminus N)$$

and so

$$\frac{\eta}{\alpha}(j(x)-u)+j(v)\in N,$$

from which

$$\eta j(x)+\alpha j(v)\preceq \eta u\preceq 0.$$

(4.0.1) now gives

$$j\left(\frac{\eta x+\alpha v}{\eta+\alpha}\right)\preceq \frac{\eta j(x)+\alpha j(v)}{\eta+\alpha}\preceq 0,$$

which completes the proof of (a).

(b) Let $u\in N$ and α and η be as in (a). Using (a), the convexity of k and (4.0.2), we obtain

$$\frac{\eta k(x)+\alpha k(v)}{\eta+\alpha}\geq k\left(\frac{\eta x+\alpha v}{\eta+\alpha}\right)\geq \mu_0,$$

from which

$$\alpha(k(v)-\mu_0)\geq \eta(\mu_0-k(x)).$$

If we now let $\eta\to \operatorname{dist}(j(v),E\setminus N)$ and then take the infimum over $u\in N$, we obtain that

$$\operatorname{dist}(j(x),N)(k(v)-\mu_0)\geq \operatorname{dist}(j(v),E\setminus N)(\mu_0-k(x)),$$

and (b) follows from (4.0.5). □

Theorem 4.5. *Suppose that $B\neq\emptyset$. Then there exists a Lagrange multiplier z_0^* such that*

$$\|z_0^*\| \leq \inf_{v\in B}\frac{k(v)-\mu_0}{\operatorname{dist}(j(v),E\setminus N)}.$$

Proof. Let $x\in A$ and $v\in B$. From Lemma 4.4(b), $\operatorname{dist}(j(x),N)>0$ and

$$\frac{\mu_0-k(x)}{\operatorname{dist}(j(x),N)}\leq \frac{k(v)-\mu_0}{\operatorname{dist}(j(v),E\setminus N)}.$$

Taking the supremum over $x \in A$ and the infimum over $v \in B$,

$$\sup_{x \in A} \frac{\mu_0 - k(x)}{\operatorname{dist}(j(x), N)} \leq \inf_{v \in B} \frac{k(v) - \mu_0}{\operatorname{dist}(j(v), E \setminus N)}.$$

The result now follows from Theorem 4.2. □

5. EXISTENCE THEOREMS WITHOUT *A PRIORI* SCALAR BOUNDS FOR NORMED SPACES

The main result in this section is Theorem 5.1. The equivalence of (5.1.1) and (5.1.2) actually first appeared in [23, Theorem 7.2, p. 27-28], and was used in [23] to obtain a number of criteria for a monotone multifunction on a reflexive Banach space to be maximal monotone (including Rockafellar's "surjectivity theorem", which we revisit in Theorem 6.5), to obtain conditions for the sum of maximal monotone multifunctions on a reflexive Banach space to be maximal monotone, and to obtain some results on maximal monotone multifunctions of Gossez's type (D) on an arbitrary Banach space. For more information, see the introductions to Sections 5 and 6 of [26]. This equivalence was also used in [25] to prove other results on maximal monotonicity. We will revisit the least technical of these in Theorem 7.4, but this time using Theorem 5.3, obtained by combining Theorem 5.1 and 2.8.

The proof of the equivalence of (5.1.1) and (5.1.2) given in [23, Theorem 7.2] was quite nonconstructive, and a more contructive proof was given in [26, Theorem 5.1], together with the bound

$$\inf_{c \in C} \left[\| j(c) \| + \sqrt{k(c) + \| j(c) \|^2} \right]$$

on the norm of $\| y^* \|$ (see [26, Remark 5.6]). We now give a new proof of this equivalence, which relies on the direct Dedekind section argument (5.1.6)–(5.1.7) and is much simpler than the proofs given in [23] and [26]. Furthermore, as is clear from (5.1.4), the bound

$$\sup_{c \in C} \left[\| j(c) \| - \sqrt{k(c) + \| j(c) \|^2} \right] \vee 0$$

on the norm of $\| y^* \|$ found in Theorem 5.1 is sharp. The analysis in this section depends only on Theorem 2.8 — it does not depend on Sections 3–4 in any way.

Theorem 5.1. *Let C be a nonempty convex subset of a vector space, F be a nontrivial normed space, j: $C \mapsto F$ be affine and $k \in \mathcal{PC}(C)$. Then*

$$c \in C \Rightarrow k(c) + \| j(c) \|^2 \geq 0 \tag{5.1.1}$$

if, and only if,

$$\exists y^* \in F^* \text{ such that } c \in C \Rightarrow k(c) - 2\left\langle j(c), y^* \right\rangle \geq \| y^* \|^2 . \tag{5.1.2}$$

Furthermore, if

$$M := \sup_{c \in C} \left[\| j(c) \| - \sqrt{k(c) + \| j(c) \|^2} \right] \vee 0 \tag{5.1.3}$$

then

$$\min \left\{ \left\| y^* \right\| : y^* \text{ is as in } (5.1.2) \right\} = M. \tag{5.1.4}$$

Proof. Since the values of c in $C \setminus \operatorname{dom} k$ have no impact on (5.1.1), (5.1.2) or the definition of M, we can and will suppose that k: $C \mapsto \mathbb{R}$. We first prove the implication (5.1.2) $\Rightarrow$ (5.1.1). Suppose that y^* is as in (5.1.2). Then

$$\begin{aligned}
c \in C \Rightarrow \quad & k(c) \geq 2\left\langle j(c), y^* \right\rangle + \left\| y^* \right\|^2 \\
\Rightarrow \quad & k(c) + \left\| j(c) \right\|^2 \geq \left\| j(c) \right\|^2 + 2\left\langle j(c), y^* \right\rangle + \left\| y^* \right\|^2 \\
\Rightarrow \quad & k(c) + \left\| j(c) \right\|^2 \geq \left\| j(c) \right\|^2 - 2\left\| j(c) \right\| \left\| y^* \right\| + \left\| y^* \right\|^2 \\
\Rightarrow \quad & k(c) + \left\| j(c) \right\|^2 \geq \left(\left\| j(c) \right\| - \left\| y^* \right\| \right)^2 \geq 0 \\
\Rightarrow \quad & \sqrt{k(c) + \left\| j(c) \right\|^2} \geq \left\| j(c) \right\| - \left\| y^* \right\| \\
\Rightarrow \quad & \left\| y^* \right\| \geq \left\| j(c) \right\| - \sqrt{k(c) + \left\| j(c) \right\|^2} .
\end{aligned} \tag{5.1.5}$$

(5.1.5) gives (5.1.1) and, since $\left\| y^* \right\| \geq 0$, this also establishes that $\left\| y^* \right\| \geq M$. We now prove the implication (5.1.1) $\Rightarrow$ (5.1.2). So suppose that (5.1.1) is satisfied. We first show that

$$a,b \in C \;\Rightarrow\; \|j(b)\| - \sqrt{k(b) + \|j(b)\|^2} \le \|j(a)\| + \sqrt{k(a) + \|j(a)\|^2}. \quad (5.1.6)$$

To this end, let

$$a,b \in C,\ \lambda > \sqrt{k(a) + \|j(a)\|^2} \ge 0 \text{ and } \mu > \sqrt{k(b) + \|j(b)\|^2} \ge 0.$$

Write $\alpha := \|j(a)\| + \lambda$ and $\beta := \|j(b)\| - \mu$. Then, since j is affine,

$$0 \le \left\| j\left(\frac{\mu a + \lambda b}{\mu + \lambda}\right)\right\| = \left\|\frac{\mu j(a) + \lambda j(b)}{\mu + \lambda}\right\| \le \frac{\mu\|j(a)\| + \lambda\|j(b)\|}{\mu + \lambda} = \frac{\mu\alpha + \lambda\beta}{\mu + \lambda}.$$

Thus, from (5.1.1) applied to $c = \dfrac{\mu a + \lambda b}{\mu + \lambda} \in C$, and the convexity of k and $(\cdot)^2$,

$$0 \le k\left(\frac{\mu a + \lambda b}{\mu + \lambda}\right) + \left(\frac{\mu\alpha + \lambda\beta}{\mu + \lambda}\right)^2 \le \frac{\mu k(a) + \lambda k(b) + \mu\alpha^2 + \lambda\beta^2}{\mu + \lambda}.$$

Multiplying by $\mu + \lambda$ gives

$$\begin{aligned} 0 &\le \mu k(a) + \lambda k(b) + \mu\alpha^2 + \lambda\beta^2 \\ &= \mu(k(a) + \alpha^2) + \lambda(k(b) + \beta^2) \\ &= \mu(k(a) + \|j(a)\|^2 + 2\lambda\|j(a)\| + \lambda^2) + \lambda(k(b) + \|j(b)\|^2 - 2\mu\|j(b)\| + \mu^2) \\ &< \mu(2\lambda^2 + 2\lambda\|j(a)\|) + \lambda(2\mu^2 - 2\mu\|j(b)\|) = 2\mu\lambda(\lambda + \|j(a)\| + \mu - \|j(b)\|). \end{aligned}$$

On dividing by $2\mu\lambda$, we obtain $\|j(b)\| - \mu < \|j(a)\| + \lambda$, and (5.1.6) follows by letting $\mu \to \sqrt{k(b) + \|j(b)\|^2}$ and $\lambda \to \sqrt{k(a) + \|j(a)\|^2}$. Now (5.1.3) and (5.1.6) imply that, for all $c \in C$,

$$\|j(c)\| - \sqrt{k(c) + \|j(c)\|^2} \le M \text{ and } M \le \|j(c)\| + \sqrt{k(c) + \|j(c)\|^2}, \quad (5.1.7)$$

from which

$$\begin{aligned} c \in C &\Rightarrow \left|\|j(c)\| - M\right| \le \sqrt{k(c) + \|j(c)\|^2} \\ &\Rightarrow \left(\|j(c)\| - M\right)^2 \le k(c) + \|j(c)\|^2 \\ &\Rightarrow k(c) + 2M\|j(c)\| \ge M^2. \end{aligned}$$

It now follows from Theorem 2.8 that there exists $L \in F^*$ such that $\|L\| \le 2M$ and

$$k + L \circ j \ge M^2 \text{ on } C.$$

Thus (5.1.2) is satisfied with $y^* := -L/2$. This completes the proof of (5.1.2), and also shows that we can find y^* satisfying (5.1.2) with $\|y^*\| \le M$, establishing (5.1.4). □

Remark 5.2. We note that $y^* = 0$ satisfies (5.1.2) exactly when $k \ge 0$ on C and, in this case, $M = 0$. In all other cases, M is given by the simpler formula

$$\sup_{c \in C}\left[\|j(c)\| - \sqrt{k(c) + \|j(c)\|^2}\right].$$

Theorem 5.3. *Let C be a nonempty convex subset of a vector space, F be a nontrivial normed space, Q: $F \mapsto \mathbb{R}$ be sublinear, h: $C \mapsto F$ be Q-convex, j: $C \mapsto F$ be affine, $k \in \mathcal{PC}(C)$ and*

$$c \in C \Rightarrow k(c) + Q \circ h(c) + \|j(c)\|^2 \ge 0. \tag{5.3.1}$$

Then there exist a linear functional Λ on F such that $\Lambda \le Q$ on F, and $y^* \in F^*$ such that

$$c \in C \Rightarrow k(c) - 2\left\langle j(c), y^* \right\rangle + \Lambda \circ h(c) \ge \|y^*\|^2. \tag{5.3.2}$$

Proof. Since $k + Q \circ h$ is convex, we first apply Theorem 5.1, with k replaced by $k + Q \circ h$, and obtain $y^* \in F^*$ such that

$$\begin{aligned} c \in C &\Rightarrow k(c) + Q \circ h(c) - 2\left\langle j(c), y^* \right\rangle \ge \|y^*\|^2 \\ &\Rightarrow k(c) - 2\left\langle j(c), y^* \right\rangle + Q \circ h(c) \ge \|y^*\|^2. \end{aligned}$$

The result now follows from Theorem 2.8, with E, S, k, and j replaced by $F, Q, k - 2y^* \circ j$, and h, respectively. □

6. THE FREE CONVEXIFICATION TECHNIQUE

The main idea introduced in this section is a technique, the *free convexification* technique, which we will discuss in Definitions 6.1 and 6.2, Lemma 6.3 and Corollary 6.4. If we combine this technique with Theorem 2.8, Theorem 5.1 or Theorem 5.3, we can obtain a large number of results on (or related to) monotone multifunctions on a Banach space. (Specifically, Lemma 11.1, p. 41, Lemma 18.1, p. 65–66, Lemma 20.1, p. 77-78, Corollary 29.2, p. 114, Lemma 36.1, p.141–142, Theorem 38.2, p. 146–147 and Theorem 38.3, p. 147–149 of [23] fall into this category, as well as some of the results of [24] and [25].) These results had been obtained previously using the minimax theorem of Fan referred to before Theorem 3.4. Since Theorem 2.8, Theorem 5.1 and Theorem 5.3 use the sublinear functional (nearly always, a scalar multiple of the norm) directly, this alternative method of proof is not only shorter, but it also avoids the need for the Banach–Alaoglu theorem, required to establish the compactness needed for the minimax theorem. As an illustration, we gave in [26, Theorem 4.1] a proof using Theorem 2.8 that a maximal monotone multifunction on a normed space with bounded range necessarily has full domain. This result can also be established using the Debrunner–Flor extension theorem (which depends on Brouwer's fixed-point theorem, see Phelps, [15, Lemma 1.7, p. 4] and the comments preceding), or the Farkas Lemma (see Fitzpatrick–Phelps, [8, Lemma 2.4, p. 580–581]). In Theorem 6.5 of this section, we will show how Theorem 5.1 leads to a proof of Rockafellar's surjectivity theorem for reflexive Banach spaces, with a sharp lower bound on the norm of solutions. For more details, see the discussion preceding Theorem 6.5. In Theorem 7.4 of the next section, we will show how Theorem 5.3 leads to a proof of a more recent result on maximal monotone multifunctions on nonreflexive Banach spaces. The analysis in this section does not depend on Sections 3–4 in any way.

Definition 6.1. Let $X \neq \emptyset$ and $\mathbb{R}^{(X)}$ be the direct sum of X copies of $\mathbb{R}$, the vector space of functions $\mu\colon X \mapsto \mathbb{R}$ such that

$$\{x \in X\colon \mu(x) \neq 0\} \text{ is finite.}$$

Define the injection $\delta_X\colon X \mapsto \mathbb{R}^{(X)}$ by

$$\delta_X(x)(y) := \begin{cases} 1, & (y = x); \\ 0, & (y \neq x). \end{cases}$$

$\left(\mathbb{R}^{(X)}, \delta_X\right)$ is the *free vector space over* X. Since $\delta_X(X)$ is a Hamel basis of $\mathbb{R}^{(X)}$, if V is any vector space and $f\colon X \mapsto V$ is any function whatsoever then there exists a linear map $g\colon \mathbb{R}^{(X)} \mapsto V$ such that $g \circ \delta_X = f$. We define $\mathcal{CO}(X) := \operatorname{co} \delta_X(X)$, the convex hull of $\delta_X(X)$ in $\mathbb{R}^{(X)}$. If $h = g\,|_{\mathcal{CO}(X)}$ then h is affine and $h \circ \delta_X = f$. We call $(\mathcal{CO}(X), \delta_X)$ the *free convexification of* X. We can give the following explicit description of h: if $c \in \mathcal{CO}(X)$ then there exist uniquely determined $\alpha_1, \ldots, \alpha_m \geq 0$ and $x_1, \ldots, x_m \in X$ such that $\sum_i \alpha_i = 1$ and $c = \sum_i \alpha_i \delta_X(x_i)$. In this case, $h(c) = \sum_i \alpha_i f(x_i)$.

Definition 6.2. Let E be a nontrivial Banach space with dual E^*, and $T\colon E \rightrightarrows E^*$ be a multifunction with

$$G(T) := \{(t, t^*)\colon t \in E, t^* \in Tt\} \neq \emptyset.$$

We say that (C, δ, p, q, r) is an $(E, E^*, \mathbb{R})$-*convexification of* T if C is a convex subset of a vector space, $p : C \mapsto E$, $q : C \mapsto E^*$ and $r\colon C \mapsto \mathbb{R}$ are affine, $\delta\colon G(T) \mapsto C$ with

$$C = \operatorname{co} \delta\big(G(T)\big) \tag{6.2.1}$$

and

$$(t, t^*) \in G(T) \quad \Rightarrow \quad \begin{cases} p \circ \delta(t, t^*) = t, \\ q \circ \delta(t, t^*) = t^*, \\ \text{and} \quad r \circ \delta(t, t^*) = \langle t, t^* \rangle. \end{cases} \tag{6.2.2}$$

It is clear from Definition 6.1, applied with $V = E$, $V = E^*$ and $V = \mathbb{R}$ in turn, that there always exist $(E, E^*, \mathbb{R})$-convexifications of T.

Lemma 6.3. *Let E be a nontrivial Banach space, $T: E \rightrightarrows E^*$ be a multifunction with $G(T) \neq \emptyset$, and (C, δ, p, q, r) be an $(E, E^*, \mathbb{R})$–convexification of T. Then T is monotone if, and only if*

$$c \in C \quad \Rightarrow \quad r(c) \geq \langle p(c), q(c) \rangle. \tag{6.3.1}$$

Proof. ($\Rightarrow$) Let $c \in C$. From (6.2.1), $c = \sum_i \alpha_i \delta(t_i, t_i^*)$, where $\alpha_1, \ldots, \alpha_m \geq 0$, $\sum_i \alpha_i = 1$, and $(t_1, t_1^*), \ldots, (t_m, t_m^*) \in G(T)$. Then

$$\begin{aligned} r(c) - \langle p(c), q(c) \rangle &= \sum_i \alpha_i \langle t_i, t_i^* \rangle - \left\langle \sum_i \alpha_i t_i, \sum_i \alpha_i t_i^* \right\rangle \\ &= \sum_{i,j} \alpha_i \alpha_j \langle t_i, t_i^* \rangle - \sum_{i,j} \alpha_i \alpha_j \langle t_i, t_j^* \rangle \\ &= \sum_{i,j} \alpha_i \alpha_j \langle t_i, t_i^* - t_j^* \rangle \\ &= \sum_{i<j} \alpha_i \alpha_j \langle t_i, t_i^* - t_j^* \rangle + \sum_{j<i} \alpha_i \alpha_j \langle t_i, t_i^* - t_j^* \rangle \\ &= \sum_{i<j} \alpha_i \alpha_j \langle t_i, t_i^* - t_j^* \rangle + \sum_{i<j} \alpha_i \alpha_j \langle t_j, t_j^* - t_i^* \rangle \\ &= \sum_{i<j} \alpha_i \alpha_j \langle t_i - t_j, t_i^* - t_j^* \rangle \geq 0, \end{aligned}$$

where the final inequality follows from the monotonicity of T.
($\Leftarrow$) Let $(x, x^*), (y, y^*) \in G(T)$. Then $\frac{1}{2}\delta(x, x^*) + \frac{1}{2}\delta(y, y^*) \in C$ and so, from (6.2.2) and (6.3.1),

$$\begin{aligned} 2\langle x, x^* \rangle + 2\langle y, y^* \rangle &= 2r \circ \delta(x, x^*) + 2r \circ \delta(y, y^*) \\ &= 4r(\tfrac{1}{2}\delta(x, x^*) + \tfrac{1}{2}\delta(y, y^*)) \\ &\geq 4\langle p(\tfrac{1}{2}\delta(x, x^*) + \tfrac{1}{2}\delta(y, y^*)), q(\tfrac{1}{2}\delta(x, x^*) + \tfrac{1}{2}\delta(y, y^*)) \rangle \\ &= 4\langle \tfrac{1}{2} p \circ \delta(x, x^*) + \tfrac{1}{2} p \circ \delta(y, y^*), \tfrac{1}{2} q \circ \delta(x, x^*) + \tfrac{1}{2} q \circ \delta(y, y^*) \rangle \\ &= 4\langle \tfrac{1}{2} x + \tfrac{1}{2} y, \tfrac{1}{2} x^* + \tfrac{1}{2} y^* \rangle = \langle x + y, x^* + y^* \rangle. \end{aligned}$$

It follows from this that T is monotone. □

Corollary 6.4. *Let E be a nontrivial Banach space, $T: E \rightrightarrows E^*$ be a monotone multifunction with $G(T) \neq \emptyset$, and (C, δ, p, q, r) be an $(E, E^*, \mathbb{R})$–convexification of T. Then:*
(a) $c \in C \Rightarrow 4r(c) + (\| p(c) \| + \| q(c) \|)^2 \geq 0$.
(b) $(\tau, \tau^*) \in G(T)$ *and* $c \in C \Rightarrow r(c) \geq \langle p(c), \tau^* \rangle + \langle \tau, q(c) \rangle - \langle \tau, \tau^* \rangle$.

Proof. (a) Let $c \in C$. From Lemma 6.3,

$$\begin{aligned} 4r(c) + (\| p(c) \| + \| q(c) \|)^2 &\geq (\| p(c) \| + \| q(c) \|)^2 + 4\langle p(c), q(c)\rangle \\ &\geq (\| p(c) \| + \| q(c) \|)^2 - 4 \| p(c) \| \| q(c) \| \\ &= (\| p(c) \| - \| q(c) \|)^2 \geq 0 \end{aligned}$$

(b) Let $(\tau, \tau^*) \in G(T)$ and $c \in C$. From (6.2.2) and the monotonicity of T, for all $(t, t^*) \in G(T)$,

$$\begin{aligned} &r(\delta(t,t^*)) - \langle p(\delta(t,t^*)), \tau^*\rangle - \langle \tau, q(\delta(t,t^*))\rangle + \langle \tau, \tau^*\rangle = \\ &\langle t, t^*\rangle - \langle t, \tau^*\rangle - \langle \tau, t^*\rangle + \langle \tau, \tau^*\rangle = \langle t - \tau, t^* - \tau^*\rangle \geq 0. \end{aligned}$$

Thus, from (6.2.1) and the affineness of p, q, and r,
$r(c) - \langle p(c), \tau^*\rangle - \langle \tau, q(c)\rangle + \langle \tau, \tau^*\rangle \geq 0.$ □

Rockafellar proved in [18, Proposition 1, p. 77-78] that if E is a non–trivial reflexive Banach space with dual E^* and duality map J: $E \rightrightarrows E^*$, J and J^{-1} are single–valued and T: $E \rightrightarrows E^*$ is a monotone multifunction then T is maximal monotone $\Longleftrightarrow T + J$ is surjective. Now ($\Leftarrow$) of the above statement fails if J or J^{-1} is not single–valued (see [23, Remark 10.8, p. 39] for a discussion of this), while ($\Rightarrow$) remains true (see [23, Theorem 10.7, p. 38]). It follows from a simple translation argument that, in order to prove that $T + J$ is surjective, it suffices to prove that there exists $y \in E$ such that $Ty + Jy \ni 0$. In Theorem 6.5 below, we give a proof of this result with a sharp lower bound on $\| y \|$ obtained from Theorem 5.1. (See also [23, Theorem 10.3, Corollary 10.4 and Theorem 10.6 p. 36–37] for characterizations of maximal monotonicity that are valid in general reflexive spaces with no restriction on J.) We mention parenthetically that the result of Rockafellar mentioned above depends on results of Browder, [3], which depend, in turn, on Brouwer's fixed–point theorem. We note for future reference that

$$G(J) = \{(x, x^*) \in E \times E^* : \| x \|^2 \vee \| x^* \|^2 = \langle x, x^*\rangle\}.$$

Theorem 6.5. *Let E be a nontrivial reflexive Banach space, T: $E \rightrightarrows E^*$ be a maximal monotone multifunction, (C, δ, p, q, r) be an $(E, E^*, \mathbb{R})$-convexification of T and*

$$M := \tfrac{1}{2}\sup_{c\in C}\left[\| p(c)\| + \| q(c)\| - \sqrt{4r(c) + (\| p(c)\| + \| q(c)\|)^2}\right] \vee 0. \quad (6.5.1)$$

Then there exists $x \in E$ *such that* $Tx + Jx \ni 0$, *and*

$$M = \min\{\| x\| : x \in E,\ Tx + Jx \ni 0\}. \quad (6.5.2)$$

Proof. Write $F := E \times E^*$ with $\| (x, x^*)\| := \| x\| + \| x^*\|$ and, for all $c \in C, k(c) := 4r(c)$ and $j(c) := (p(c), q(c))$. It follows from Corollary 6.4(a) and Theorem 5.1 that there exists $y^* \in F^*$ such that

$$\left.\begin{aligned}\|y^*\| &= \sup_{c\in C}\left[\| p(c)\| + \| q(c)\| - \sqrt{4r(c) + (\| p(c)\| + \| q(c)\|)^2}\right] \vee 0\\ &= 2M\end{aligned}\right\} \quad (6.5.3)$$

and

$$c \in C \Rightarrow 4r(c) - 2\langle j(c), y^*\rangle \geq \| y^*\|^2. \quad (6.5.4)$$

Now we can write $y^* = (2x^*, 2x)$ for some $(x, x^*) \in E \times E^*$, and $\| y^*\| = 2\| x\| \vee 2\| x^*\|$. (This is where we use the reflexivity of E.) Dividing (6.5.4) by 4, we obtain

$$c \in C \Rightarrow r(c) - \langle p(c), x^*\rangle - \langle x, q(c)\rangle \geq \| x\|^2 \vee \| x^*\|^2.$$

If now $(t, t^*) \in G(T)$ and we substitute $c = \delta(t, t^*)$, we obtain from (6.2.2) that

$$\begin{aligned}(t, t^*) \in G(T) \quad &\Rightarrow \quad \langle t, t^*\rangle - \langle t, x^*\rangle - \langle x, t^*\rangle \geq \| x\|^2 \vee \| x^*\|^2,\\ &\Rightarrow \quad \langle t - x, t^* - x^*\rangle \geq \| x\|^2 \vee \| x^*\|^2 + \langle x, x^*\rangle.\end{aligned}$$

Now $\| x\|^2 \vee \| x^*\|^2 + \langle x, x^*\rangle \geq \| x\|^2 \vee \| x^*\|^2 - \| x\|\| x^*\| \geq 0$, and so the maximal monotonicity of T implies that $(x, x^*) \in G(T)$. Substituting $(t, t^*) = (x, x^*)$ yields $\| x\|^2 \vee \| x^*\|^2 + \langle x, x^*\rangle \leq 0$, from which $-x^* \in Jx$. Since $0 = x^* + (-x^*)$, it is now immediate that $Tx + Jx \ni 0$, and it follows from (6.5.3) that $\| x\| = M$.

Suppose, conversely, that $x \in E$ and $Tx + Jx \ni 0$. Then there exists $x^* \in Tx$ such that $\| x\|^2 \vee \| x^*\|^2 + \langle x, x^*\rangle = 0$. Since T is monotone, using (6.2.2),

$$\begin{aligned}(t,t^*)\in G(T) \quad &\Rightarrow \quad \langle t-x,t^*-x^*\rangle \geq \ \| x \|^2 \vee \| x^* \|^2 + \langle x,x^*\rangle \\ &\Rightarrow \quad \langle t,t^*\rangle - \langle t,x^*\rangle - \langle x,t^*\rangle \geq \ \| x \|^2 \vee \| x^* \|^2, \\ &\Rightarrow \quad r(\delta(t,t^*)) - \langle p(\delta(t,t^*)),x^*\rangle - \langle x,q(\delta(t,t^*)),x^*\rangle \\ &\qquad\qquad \geq \ \| x \|^2 \vee \| x^* \|^2 .\end{aligned}$$

It follows from (6.2.1), the affineness of p, q and r on C and the fact that $\| x^* \| = \| x \|$ that

$$\begin{aligned}c\in C \quad &\Rightarrow \quad r(c) - \langle p(c),x^*\rangle - \langle x,q(c)\rangle \geq \ \| x \|^2 \vee \| x^* \|^2 \\ &\Rightarrow \quad r(c) + \| p(c) \| \| x^* \| + \| x \| \| q(c) \| \ \geq \ \| x \|^2 \vee \| x^* \|^2 \\ &\Rightarrow \quad r(c) + (+ \| q(c) \|) \| x \| \ \geq \ \| x \|^2 .\end{aligned}$$

Or completing the square, we obtain that

$$c\in C \ \Rightarrow \ \| x \| \geq \frac{1}{2}\left[\| p(c) \| + \| q(c) \| - \sqrt{4r(c) + (\| p(c) \| + \| q(c) \|)^2}\ \right].$$

Since $\| x \| \geq 0$, it is immediate from this that $\| x \| \ \geq M$, completing the proof of Theorem 6.5. □

Remark 6.6 It was shown by Zălinescu that the existence of $x \in E$ such that $Tx + Jx \ni 0$ in Theorem 6.5 can also be established by an argument using the Fitzpatrick function on $E \times E^*$ (see [7]), a technique due to Burachik and Svaiter (see [4]), and the Moreau–Rockafellar formula for the subdifferential of a sum, though it is not clear that this argument leads easily to a sharp lower bound on $\| x \|$. See [27] for details.

7. TYPE (D) IMPLIES TYPE (FP)

In Theorem 7.4 of this section, we show how Theorem 5.3 and the free convexification technique introduced in Section 6 lead to a proof that *every maximal monotone multifunction of type (D) on a (possibly nonreflexive) Banach space is of type (FP) (i.e. locally maximal monotone)*, thus settling a question that has been open for some time. Yet again, the analysis in this section does not depend on Sections 3–4 in any way.

We now proceed to the definitions of the terms introduced above. In order to define maximal monotone multifunctions of type (D), we must

introduce a concept due to Gossez: if $W\colon E \rightrightarrows E^*$, we define the multifunction $\overline{W}\colon E^{**} \rightrightarrows E^*$ by:

$$x^* \in \overline{W}x^{**} \quad \Longleftrightarrow \quad \inf_{(w,w^*)\in G(W)} \langle w^* - x^*, \widehat{w} - x^{**} \rangle \geq 0,$$

where $\widehat{w}$ is the canonical image of w in E^{**}. In what follows, $R(W) := \bigcup_{x\in E} Wx$.

Definition 7.1. Let $W\colon E \rightrightarrows E^*$ be maximal monotone. W is said to be *of type (D)* if, for all $(x^{**}, x^*) \in G(\overline{W})$, there exists a bounded net $\{(w_\gamma, w_\gamma^*)\}$ of elements of $G(W)$ such that

$$(\widehat{w_\gamma}, w_\gamma^*) \to (x^{**}, x^*) \text{ in } w(E^{**}, E^*) \times \mathcal{T}_{\|\ \|}(E^*),$$

where $\mathcal{T}_{\|\ \|}(E^*)$ is the norm topology of E^*. Clearly

- *if E is reflexive then every maximal monotone multifunction $W\colon E \rightrightarrows E^*$ is of type (D).*

It was essentially proved by Gossez in [10] (see Phelps, [15, Theorem 3.8, p. 221] for an exposition) that

- *if W is maximal monotone of type (D) then $\overline{R(W)}$ is convex.*

It was proved by Gossez in [10, Théorème 3.1, p. 376–378] that

- *if $f\colon E \mapsto (-\infty, \infty]$ is proper, convex and lower semicontinuous then $\partial f\colon E \rightrightarrows E^*$ is maximal monotone of type (D).*

Definition 7.2. A monotone multifunction $W\colon E \rightrightarrows E^*$ is said to be *of type (FP)* or *locally maximal monotone* provided the following holds: for any open convex subset U of E^* such that $U \cap R(W) \neq \emptyset$, if $(v, v^*) \in E \times U$ is such that

$$(w, w^*) \in G(W) \text{ and } w^* \in U \ \Rightarrow \ \langle w - v, w^* - v^* \rangle \geq 0$$

then $(v, v^*) \in G(W)$. (If we take $U = E^*$, we see that every multifunction of type (FP) is maximal monotone.) It was proved by Fitzpatrick and Phelps in [8, Proposition 3.3, p. 585] that

- *if E is reflexive then every maximal monotone multifunction $W\colon E \rightrightarrows E^*$ is of type (FP).*

It was proved by Fitzpatrick and Phelps in [8, Theorem 3.5, p. 585] that

- *if W is maximal monotone of type (FP) then $\overline{R(W)}$ is convex.*

It was proved in [22, Main theorem, p. 470] and [23, Theorem 30.3, p. 120] that

- *if $f\colon E \mapsto (-\infty, \infty]$ is proper, convex and lower semicontinuous then $\partial f\colon E \rightrightarrows E^*$ is maximal monotone of type (FP).*

Finally, it was proved by Fitzpatrick and Phelps in [9, Theorem 3.7, p. 67] that

- *if W is maximal monotone and $R(W) = E^*$ then W is of type (FP).*

Most of the work for Theorem 7.4 will be done in the rather technical Lemma 7.3, below.

Lemma 7.3. *Let E be a nontrivial Banach space, $S\colon E \mapsto \mathbb{R}$ be sublinear, $T\colon E \rightrightarrows E^*$ be monotone and such that, for some $\tau^* \in R(T)$ and $\varepsilon > 0$,*

$$x \in E \Rightarrow S(x) \geq \langle x, \tau^* \rangle + \varepsilon \| x \|. \tag{7.3.1}$$

Write

$$B := \{x^* \in E^* : x^* \leq S \text{ on } E\}.$$

Then there exists $(z^*, z^{**}, y^{**}) \in B \times E^{**} \times E^{**}$ such that

$$\left.\begin{aligned} \inf_{(t,t^*)\in G(T)} \langle t^* - z^*, \hat{t} - z^{**} + y^{**} \rangle \geq \; & \| z^* \|^2 \vee \| z^{**} \|^2 + \langle z^*, z^{**} \rangle + \\ & \sup \langle B - z^*, y^{**} \rangle \geq 0. \end{aligned}\right\} \tag{7.3.2}$$

Proof. We fix $\tau \in T^{-1}\tau^*$, and then write $M := \| \tau \| \vee \| \tau^* \|$ and $N := 3M^2/\varepsilon$. Now let (C, δ, p, q, r) be an $(E, E^*, \mathbb{R})$-convexification of T and $D := C \times E \times B$. We first prove that, for all $(c, x, x^*) \in D$,

$$r(c) + S(x) + N \| q(c) - x^* \| + \tfrac{1}{4}(\| p(c) + x \| + \| q(c) \|)^2 \geq 0. \tag{7.3.3}$$

So let us suppose that $(c, x, x^*) \in D$. If $\| x \| \leq N$ then, from Lemma 6.3 and the definition of D,

$$\begin{aligned} r(c) + S(x) + N \| q(c) - x^* \| + \tfrac{1}{4}(\| p(c) + x \| + \| q(c) \|)^2 & \\ \geq r(c) + \langle x, x^* \rangle + N \| q(c) - x^* \| + \| p(c) + x \| \| q(c) \| & \\ \geq r(c) + \langle x, x^* \rangle + \langle x, q(c) - x^* \rangle - \langle p(c) + x, q(c) \rangle & \\ = r(c) - \langle p(c), q(c) \rangle \geq 0, & \end{aligned}$$

which gives (7.3.3). Suppose, on the other hand, that $\| x \| > N$. Then, from (7.3.1),

$$S(x) \geq \langle x, \tau^* \rangle + \varepsilon \| x \| \quad \geq \langle x, \tau^* \rangle + 3M^2, \tag{7.3.4}$$

and, from Corollary 6.4 (b),

$$r(c) \geq \langle p(c), \tau^* \rangle + \langle \tau, q(c) \rangle - \langle \tau, \tau^* \rangle \geq \langle p(c), \tau^* \rangle - M \| q(c) \| - M^2. \tag{7.3.5}$$

Using (7.3.4) and (7.3.5), we have

$$\begin{aligned}
&r(c) + S(x) + \tfrac{1}{4}(\| p(c) + x \| + \| q(c) \|)^2 \\
&\geq \left[\langle p(c), \tau^* \rangle - M \| q(c) \| - M^2\right] + \left[\langle x, \tau^* \rangle + 3M^2\right] + \\
&\qquad\qquad \tfrac{1}{4}\left[\| p(c) + x \|^2 + \| q(c) \|^2\right] \\
&= 2M^2 + \langle p(c) + x, \tau^* \rangle - M \| q(c) \| + \tfrac{1}{4} \| p(c) + x \|^2 + \tfrac{1}{4} \| q(c) \|^2 \\
&\geq 2M^2 - M \| p(c) + x \| - M \| q(c) \| + \tfrac{1}{4} \| p(c) + x \|^2 + \tfrac{1}{4} \| q(c) \|^2 \\
&= \left(\tfrac{1}{2} \| p(c) + x \| - M\right)^2 + \left(\tfrac{1}{2} \| q(c) \| - M\right)^2 \geq 0,
\end{aligned}$$

and (7.3.3) follows, since $N \| q(c) - x^* \| \geq 0$. Write $F := E \times E^*$, normed by

$$\| (x, x^*) \| := \| x \| + \| x^* \|,$$

and define, Q: $F \mapsto \mathbb{R}$, h: $D \mapsto F$, j: $D \mapsto F$ and k: $D \mapsto \mathbb{R}$ by

$$\begin{aligned}
Q(x, x^*) &:= S(x) + N \| x^* \|, && ((x, x^*) \in F) \\
h(c, x, x^*) &:= (x, q(c) - x^*), && ((c, x, x^*) \in D) \\
j(c, x, x^*) &:= \tfrac{1}{2}(p(c) + x, q(c)), && ((c, x, x^*) \in D)
\end{aligned}$$

and

$$k(c, x, x^*) := r(c). \qquad ((c, x, x^*) \in D)$$

We note then that (7.3.3) can be written in the form

$$(c, x, x^*) \in D \Rightarrow k(c, x, x^*) + Q \circ h(c, x, x^*) + \| j(c, x, x^*) \|^2 \geq 0.$$

We now apply Theorem 5.3 with C replaced by D, and obtain a linear functional Λ on F such that $\Lambda \leq Q$ on F, and $y^* \in F^*$, such that

$$(c, x, x^*) \in D \Rightarrow r(c) - \langle (p(c) + x, q(c)), y^* \rangle + \Lambda(x, q(c) - x^*) \geq \| y^* \|^2 .$$

Now there exists $(z^*, z^{**}) \in E^* \times E^{**}$ such that $y^* = (z^*, z^{**})$. Furthermore, the form of Q implies that there exist a linear functional L on E such that $L \leq S$ on E, and $y^{**} \in E^{**}$ with $\| y^{**} \| \leq N$ such that $\Lambda = (L, y^{**})$. Consequently,

$$\begin{aligned} & (c, x, x^*) \in D \\ \Rightarrow \quad & r(c) - \langle p(c) + x, z^* \rangle - \langle q(c), z^{**} \rangle + L(x) + \langle q(c) - x^*, y^{**} \rangle \geq \| y^* \|^2 \\ \Leftrightarrow \quad & (L - z^*)(x) + r(c) - \langle p(c), z^* \rangle - \langle q(c), z^{**} - y^{**} \rangle - (x^*, y^{**}) \geq \| y^* \|^2 \end{aligned}$$

thus, adding $\langle x^*, y^{**} \rangle$ to both sides of the above, and then taking the supremum over $x^* \in B$,

$$(c, x) \in C \times E \Rightarrow (L - z^*)(x) + r(c) - \langle p(c), z^* \rangle - \langle q(c), z^{**} - y^{**} \rangle \geq \| y^* \|^2 + \sup \langle B, y^{**} \rangle .$$

For the moment fix c. It follows by taking the infimum over $x \in E$ that $L = z^*$, thus $z^* \leq S$ on E, and so $z^* \in B$, as required. Substituting $L = z^*$ in the above, we have

$$c \in C \quad \Rightarrow \quad r(c) - \langle p(c), z^* \rangle - \langle q(c), z^{**} - y^{**} \rangle \geq \| y^* \|^2 + \sup \langle B, y^{**} \rangle .$$

It follows by taking $c = \delta(t, t^*)$ and using (5.2.2) that

$$\begin{aligned} (t, t^*) \in G(T) \Rightarrow \quad & \langle t, t^* \rangle - \langle t, z^* \rangle - \langle t^*, z^{**} - y^{**} \rangle \geq \| y^* \|^2 + \sup \langle B, y^{**} \rangle . \\ \Leftrightarrow \quad & \langle t^* - z^*, \hat{t} + y^{**} \rangle - \langle t^*, z^{**} \rangle \geq \| y^* \|^2 + \sup \langle B - z^*, y^{**} \rangle . \end{aligned}$$

We obtain the first inequality in (7.3.2) by adding (z^*, z^{**}) to both sides of the above and observing that $\| y^* \|^2 = \| z^* \|^2 \vee \| z^{**} \|^2$, and the second inequality follows since

$$\| z^* \|^2 \vee \| z^{**} \|^2 + \langle z^*, z^{**} \rangle \geq \| z^* \|^2 \vee \| z^{**} \|^2 - \| z^* \| \| z^{**} \| \geq 0 \quad (7.3.6)$$

and

$$\sup\langle B - z^*, y^{**}\rangle \geq \langle z^* - z^*, y^{**}\rangle \geq 0. \tag{7.3.7}$$

This completes the proof of Lemma 7.3. □

Theorem 7.4. *Let* $W\colon E \rightrightarrows E^*$ *be maximal monotone of type (D). Then* W *is of type (FP).*

Proof. Let U be an open convex subset of E^* such that $U \cap R(W) \neq \emptyset$ and $(v, v^*) \in E \times U$ be such that

$$(w, w^*) \in G(W) \text{ and } w^* \in U \;\Rightarrow\; \langle w - v, w^* - v^*\rangle \geq 0.$$

We want to prove that $(v, v^*) \in G(W)$. Now define $T\colon E \rightrightarrows E^*$ by $G(T) := G(W) - (v, v^*)$. Further, writing $V := U - v^*$, we have that V is an open convex subset of E^* such that $V \ni 0, V \cap R(T) \neq \emptyset$ and

$$(t, t^*) \in G(T) \text{ and } t^* \in V \;\Rightarrow\; \langle t, t^*\rangle \geq 0 \tag{7.4.1}$$

and now what we must prove is that

$$(0, 0) \in G(T). \tag{7.4.2}$$

We first find $\tau^* \in V \cap R(T)$ and choose $\varepsilon > 0$ so that

$$[0, \tau^*] + \{x^* \in E^* : \| x^* \| \leq \varepsilon\} \subset V.$$

We define the sublinear functional $S\colon E \mapsto \mathbb{R}$ by

$$S(x) := \langle x, \tau^*\rangle \vee 0 + \varepsilon \| x \|,$$

and B as in Lemma 7.3. It is then easy to see that $B = [0, \tau^*] + \{x^* \in E^* : \| x^* \| \leq \varepsilon\} \subset V$. Lemma 7.3 then gives us $(z^*, z^{**}, y^{**}) \in B \times E^{**} \times E^{**}$, such that

$$\inf_{(t,t^*)\in G(T)} \left\langle t^* - z^*, \hat{t} - z^{**} + y^{**}\right\rangle \geq \left.\begin{array}{r} \| z^* \|^2 \vee \| z^{**} \|^2 + \langle z^*, z^{**}\rangle + \\ \sup\langle B - z^*, y^{**}\rangle \geq 0. \end{array}\right\} \tag{7.3.2}$$

It follows from this that $(z^{**} - y^{**}, z^*) \in G(\overline{T})$. Since W is of type (D), the same is true of T and so there exists a bounded net $\{(t_\gamma, t_\gamma^*)\}$ of elements of $G(T)$ such that $(\widehat{t_\gamma}, t_\gamma^*) \to (z^{**} - y^{**}, z^*)$ in $w(E^{**}, E^*) \times \mathcal{T}_{\|\ \|}(E^*)$. This implies that $\langle t_\gamma^* - z^*, \widehat{t_\gamma} - z^{**} + y^{**} \rangle \to 0$, and so putting $(t, t^*) = (t_\gamma, t_\gamma^*)$ and passing to the limit in (7.3.2),

$$0 \geq \| z^* \|^2 \vee \| z^{**} \|^2 + \langle z^*, z^{**} \rangle + \sup \langle B - z^*, y^{**} \rangle.$$

(7.3.6) now implies that

$$0 \geq \sup \langle B - z^*, y^{**} \rangle, \tag{7.4.3}$$

and (7.3.7) that

$$0 \geq \| z^* \|^2 \vee \| z^{**} \|^2 + \langle z^*, z^{**} \rangle. \tag{7.4.4}$$

Since $B \supset \{x^* \in E^* : \| x^* \| \leq \varepsilon\}$, (7.4.3) gives

$$\langle z^*, y^{**} \rangle \geq \sup \langle B, y^{**} \rangle \geq \varepsilon \| y^{**} \|. \tag{7.4.5}$$

Now $z^* \in B \subset V$ and $t_\gamma^* \to z^*$ in $\mathcal{T}_{\|\ \|}(E^*)$, so by truncating the net $\{(t_\gamma, t_\gamma^*)\}$ if necessary, we may suppose that, for all γ, $t_\gamma^* \in V$. Using (7.4.1), we now derive that, for all γ, $\langle t_\gamma^*, \hat{t}_\gamma \rangle = \langle t_\gamma, t_\gamma^* \rangle \geq 0$. Passing to the limit in this, $\langle z^*, z^{**} - y^{**} \rangle \geq 0$ and, combining with (7.4.5), we obtain $\langle z^*, z^{**} \rangle \geq \varepsilon \| y^{**} \|$. If we now substitute this into (7.4.4) we obtain $0 \geq \| z^* \|^2 \vee \| z^{**} \|^2 + \varepsilon \| y^{**} \|$, hence $z^* = 0$ and $z^{**} = y^{**} = 0$. Substituting back into (7.3.2) yields $\inf_{(t,t^*) \in G(T)} \langle t^* - 0, \hat{t} - 0 \rangle \geq 0$, that is to say,

$$\inf_{(t,t^*) \in G(T)} \langle t - 0, t^* - 0 \rangle \geq 0.$$

Since T is maximal monotone, this gives (7.4.2) and completes the proof of Theorem 7.4. □

We note that it was proved in Bauschke and Borwein [2, Theorem 4.1] (see also [16, Theorem 8.1, p. 327]) that every *continuous single–valued linear* maximal monotone multifunction of type (FP) is necessarily of type (D). However, we do not know the solution to the following problem:

Problem 7.5. Is every maximal monotone multifunction of type (FP) necessarily of type (D) ?

8. AN EXISTENCE THEOREM WITHOUT *A PRIORI* SCALAR BOUNDS FOR SUBLINEAR FUNCTIONALS

We note that (5.1.1) can be written $\inf_C[k + \psi \circ S \circ j] \geq 0$, where $\psi : \mathbb{R} \mapsto \mathbb{R}$ is defined by $\psi := (\cdot)^2$ and $S := \|\cdot\|$, and $\inf_C[S \circ j + k]$ in (2.8.1) can be written $\inf_C[k + \psi \circ S \circ j]$, where $\psi : \mathbb{R} \mapsto \mathbb{R}$ is defined by $\psi := (\cdot)$. Thus it is natural to ask whether there is a result that simultaneously generalizes Theorem 2.8 and Theorem 5.1. Theorem 8.1, which is such a result, is the topic of this section. The equivalence of (8.1.3) and (8.1.4) was first proved in [26, Theorem 5.4] using a rather technical product space argument and giving a weaker bound on N than that given here. We give here a new proof of this equivalence, which relies on the much simpler Dedekind section argument (8.1.7)–(8.1.11). Furthermore, as is clear from (8.1.6), the bound on N found in Theorem 8.1 is sharp. We refer the reader to [26, Remarks 5.5 and 5.6] for the details of how Theorem 8.1 implies Theorem 2.8 and Theorem 5.1.

We first discuss the conditions (8.1.1) and (8.1.2) on the function ψ. (8.1.1) is to ensure that the quantity M defined in (8.1.5) is finite, while (8.1.2) is needed in (8.1.8). Of course, (8.1.1) is automatically true if ψ is real–valued, as is the case with the two examples mentioned above. As for (8.1.2), if $\psi := (\cdot)$, ψ is increasing on $\mathbb{R}$ and so (8.1.2) is automatic while, if $\psi := (\cdot)^2$ and $S := \|\cdot\|$, (8.1.2) is true since

$$S \circ j(c) \leq \gamma \Rightarrow S \circ j(c), \quad \gamma \in [0, \infty)$$

and ψ is increasing on $[0, \infty)$. (We note that (8.1.1) was described in [26] by saying that ψ is "S, j-compatible".)

Theorem 8.1. *Let C be a nonempty convex subset of a vector space, E be a nontrivial vector space, $S : E \mapsto \mathbb{R}$ be sublinear, $j : C \mapsto E$ be S-convex and $k \in \mathcal{PC}(C)$. Let $\psi \in \mathcal{PC}(\mathbb{R})$ satisfy*

$$(S \circ j(\operatorname{dom} k) + (0, \infty)) \cap \operatorname{dom} \psi \neq \emptyset \tag{8.1.1}$$

and

$$c \in C \text{ and } S \circ j(c) \leq \gamma \Rightarrow \psi \circ S \circ j(c) \leq \psi(\gamma). \tag{8.1.2}$$

Then

$$k + \psi \circ S \circ j \geq 0 \quad \text{on } C \tag{8.1.3}$$

if, and only if,

$$\left.\begin{array}{c} \text{there exist } N \geq 0 \text{ and a linear functional } L \text{ on } E \text{ such that} \\ L \leq NS \text{ on } E \text{ and } k + L \circ j \geq \psi^*(\mathrm{N}) \text{ on } C. \end{array}\right\} \tag{8.1.4}$$

Furthermore, if

$$M := \sup_{c \in C, \mu < 0} \frac{k(c) + \psi(S \circ j(c) + \mu)}{\mu} \vee 0 \tag{8.1.5}$$

then

$$\min\{N\colon N \text{ is as in } (8.1.4)\} = M. \tag{8.1.6}$$

Proof. Suppose first that (8.1.4) is satisfied, from which $\psi^*(N) \in \mathbb{R}$. Then, for all $c \in C$ and $\nu \in \mathbb{R}$,

$$\begin{aligned} k(c) + \psi(S \circ j(c) + \nu) &\geq k(c) + N(S \circ j(c) + \nu) - \psi^*(N) \\ &= k(c) + NS \circ j(c) - \psi^*(N) + N\nu \\ &\geq k(c) + L \circ j(c) - \psi^*(N) + N\nu \geq N\nu. \end{aligned}$$

If we put $\nu = 0$ in this, we obtain (8.1.3). On the other hand, we also derive that

$$c \in C \text{ and } \mu < 0 \quad \Rightarrow \quad \frac{k(c) + \psi(S \circ j(c) + \mu)}{\mu} \leq N$$

and, since $N \geq 0$, this also shows that $N \geq M$. Suppose, conversely, that (8.1.3) is satisfied. We first show that

$$a,b \in C \text{ and } \mu < 0 < \lambda \Rightarrow \frac{k(b) + \psi(S \circ j(b) + \mu)}{\mu} \leq \frac{k(a) + \psi(S \circ j(a) + \lambda)}{\lambda}. \tag{8.1.7}$$

To this end, let $a,b \in C$ and $\mu < 0 < \lambda$. Write $\alpha := S \circ j(a) + \lambda$ and $\beta := S \circ j(b) + \mu$. Then, from the S–convexity of j and the sublinearity of S,

$$S \circ j\left(\frac{\lambda b - \mu a}{\lambda - \mu}\right) \leq S\left(\frac{\lambda j(b) - \mu j(a)}{\lambda - \mu}\right) \leq \frac{\lambda S \circ j(b) - \mu S \circ j(a)}{\lambda - \mu} = \frac{\lambda\beta - \mu\alpha}{\lambda - \mu}.$$

Thus, using (8.1.2) with

$$c := \frac{\lambda b - \mu a}{\lambda - \mu} \quad \text{and} \quad \gamma := \frac{\lambda\beta - \mu\alpha}{\lambda - \mu},$$

(8.1.3) and the convexity of k and ψ,

$$0 \leq k\left(\frac{\lambda b - \mu a}{\lambda - \mu}\right) + \psi\left(\frac{\lambda\beta - \mu\alpha}{\lambda - \mu}\right) \leq \frac{\lambda k(b) - \mu k(a) + \lambda\psi(\beta) - \mu\psi(\alpha)}{\lambda - \mu}, \tag{8.1.8}$$

and (8.1.7) follows on multiplication by $\lambda - \mu > 0$ and substituting in the values of α and β. From (8.1.2) and (8.1.3), for all $c \in C$ and $\lambda > 0$,

$$\frac{k(c) + \psi(S \circ j(c) + \lambda)}{\lambda} \geq \frac{k(c) + \psi \circ S \circ j(c)}{\lambda} \geq 0, \tag{8.1.9}$$

and (8.1.1) provides $a \in \operatorname{dom} k$ and $\lambda > 0$ such that

$$S \circ j(a) + \lambda \in \operatorname{dom} \psi,$$

from which

$$\frac{k(a) + \psi(S \circ j(a) + \lambda)}{\lambda} < \infty. \tag{8.1.10}$$

(8.1.7) and (8.1.10) imply that $M \in [0,\infty)$, and (8.1.7) and (8.1.9) that, for all $c \in C$ and $\mu < 0 < \lambda$,

$$\frac{k(c)+\psi(S\circ j(c)+\mu)}{\mu}\leq M\leq\frac{k(c)+\psi(S\circ j(c)+\lambda)}{\lambda}. \qquad (8.1.11)$$

Combining this with (8.1.3), we obtain

$$\begin{aligned} c\in C \text{ and } \nu\in\mathbb{R} \quad &\Rightarrow \quad k(c)+\psi(S\circ j(c)+\nu)\geq M\nu \\ &\Leftrightarrow \quad k(c)+MS\circ j(c)\geq M(S\circ j(c)+\nu)-\psi(S\circ j(c)+\nu). \end{aligned}$$

Taking the supremum of the right-hand side over $\nu\in\mathbb{R}$ shows that

$$c\in C \quad \Rightarrow \quad k(c)+MS\circ j(c)\geq\psi^*(M)$$

and (8.1.4) (with N replaced by M) now follows from Theorem 2.8. This completes the proof of Theorem 8.1 □

REFERENCES

[1] H. Attouch and H. Brezis, *Duality for the sum of convex functions in general Banach spaces,* Aspects of Mathematics and its Applications, J.A. Barroso, ed., Elsevier Science Publishers(1986), 125-133.

[2] H.H. Bauschke and J.M. Borwein, *Maximal monotonicity of dense type, local maximal monotonicity, and monotonicity of the conjugate are all the same for continuous linear operator,* Pacific J. Math. **189**(1999), 1-20.

[3] F.E. Browder, *Nonlinear maximal monotone operators in Banach spaces,* Math. Annalen **175** (1968), 89-113.

[4] R.S. Burachick and B.F. Svaiter, *Maximal monotonicity, conjugation and the duality product,* IMPA Preprint 129/2002, February 28, 2002.

[5] K. Fan, *Minimax theorems,* Proc. Nat. Acad. Sci. U.S.A. **39** (1953), 42-47.

[6] K. Fan, I. Glicksberg and A. J. Hoffman, *System of inequalities involving convex functions,* Proc. Amer. Math. Soc. **8** (1957), 617-622.

[7] S. Fitzpatrick, *Representing monotone operators by convex functions,* Workshop/ Miniconference on Functionl Analysis and Optimization (Canberra, 1988), 59-65, Proc. Centre Math. Anal. Austral. Nat. Univ., **20**, Austral. Nat. Univ., Canberra, 1988.

[8] S.P. Fitzpatrick and R.R. Phelps, *Bounded approximants to monotone operators on Banach spaces,* Ann. Inst. Henri Poincaré, Analyse non Linéaire **9** (1992), 573-595.

[9] —, *Some properties of maximal monotone operators on nonreflexive Banach spaces,* Set-Valued Analysis **3**(1995), 51-69.

[10] J.-P. Gossez, *Opérateurs monotones non linéaires dans les espaces de Banach non réflexifs,* J. Math. Anal. Appl. **34** (1971), 371-395.

[11] J.L. Kelley, I. Namioka, and co-authors, *Linear Topological Spaces,* D. Van Nostrand Co., Inc., Princeton-Toronto-London-Melbourne (1963).

[12] H. König, *Some Basic Theorems in Convex Analysis*, in "Optimization and operations research", edited by B. Korte, North-Holland (1982).

[13] D.L. Luenberger, *Optimization by Vector Space Methods,* John Wiley & Sons, Inc, New York-Chichester-Brisbane-Toronto-Singapore (1969).

[14] J.-J. Moreau, *Fonctionelles convexes,* Séminaire sur les équations aux derivées partielles, Lecture notes, Collège de France, Paris 1966.

[15] R.R. Phelps, *Lectures on Maximal Monotone Operators, Extracta Mathematicae* **12** (1997), 193-230.

[16] R.R. Phelps, and S. Simons, *Unbounded linear monotone operators on nonreflexive Banach spaces,* J. Convex Analysis, **5** (1998), 303-328.

[17] R.T. Rockafellar, *Extension of Fenchel's Duality theorem for convex functions,* Duke Math. J. **33** (1966), 81-89.

[18] —, *On the Maximality of Sums of Nonlinear Monotone Operators,* Trans. Amer. Math. Soc. **149** (1970), 75-88.

[19] —, *Conjugate duality and optimization,* Conference Board of the Mathematical Sciences **16** (1974), SIAM publications.

[20] W. Rudin, *Functional anlysis,* McGraw-Hill, New York (1973).

[21] S. Simons, *Minimal sublinear functionals,* Studia Math. **37** (1970), 37-56.

[22] —, *Subdifferentials are locally maximal monotone,* Bull. Australian Math. Soc. **47** (1993), 465-471.

[23] —, *Minimax and Monotonicity,* Lecture Notes in Mathematics **1693** (1998), Springer-Verlag.

[24] —, *Maximal monotone multifunctions of Brøndsted-Rockafellar type,* Set-Valued Anal. **7** (1999), 255-294.

[25] —, *Five Kinds of maximal monotonicity,* Set –Valued Anal. **9** (2001), 391-409.

[26] —, *A new version of the Hahn-Banach theorem,* Archiv der Mathematik, (2003,630-646)

[27] S. Simons and C. Zălinescu, *A New Proof for Rockafellar's Characterization of Maximal Monotone Operators,* Proc. Amer. Math. Soc., In press.

CONCRETE PROBLEMS AND THE GENERAL THEORY OF EXTREMUM

V.M. Tikhomirov
Moscow State University, Moscow, Russia

0. INTRODUCTION

Mathematics consists of general theories and concrete facts. In the books [1]–[5] (written by myself in cooperation with my colleagues and students) general concepts and principles on which the general theory of extrema is based are accompanied by solutions of many concrete problems. In this paper intercommunication between general principles and concrete problems will be illustrated by the example of the so-called Landau–Kolmogorov-type inequalities on the real line and the half-line. These problems are discussed in my papers [6]–[8] (written jointly with A. Buslaev, G. Magaril-Il'yaev, and A. Kochurov).

Extremal problems are formulated initially in terms of the science or the field of applications which gives rise to them, i.e., in terms of engineering, physics, geometry, etc. In order to provide for their mathematical treatment, one has to translate them into analytic terms. Such translation is called *formalization*.

To formalize an extremal problem, one has to specify a *function* f (along with its *domain of definition* X, $f: X \to \overline{\mathbb{R}}, = \overline{\mathbb{R}} \cup \pm\infty$) to be minimized or maximized, as well as a *constraint* $C \subset X$. As a rule, constraints are specified by equalities and inequalities.

The problem: "minimize (maximize) f under the constraint C" is written as

$$f(x) \to \min \ (\max), \quad x \in C. \tag{P}$$

When writing $f(x) \to \text{extr}$ we mean that both the problems for maximum and minimum may be considered.

The theory of extremum consists of the following four parts: necessary conditions for extremum, perturbations of the problem and sufficient conditions, existence, and algorithms. The principal parts of the mathematical basis for the whole theory are calculus and convex analysis, as well as non-smooth calculus, actively developed nowadays.

1. LANDAU–KOLMOGOROV INEQUALITIES

- Here we set up a family of concrete problems which will be considered as a testing ground for the general theory of extrema.

Landau–Kolmogorov-type inequalities on the line and the half-line have the following form:

$$\left\|x^{(k)}(\cdot)\right\|_{L_q(T)} \le K \left\|x(\cdot)\right\|_{L_p(T)}^{\alpha} \left\|x^{(n)}(\cdot)\right\|_{L_r(T)}^{1-\alpha},$$

(1.1)

$$T = \mathbb{R} \text{ or } \mathbb{R}_+, \quad \alpha = \frac{n-k-r^{-1}+q^{-1}}{n-r^{-1}+p^{-1}} > 0,$$

where $n \in \mathbf{N}, k \in \mathbf{Z}_+ = \mathbf{N} \cup \{0\}$, $0 \le k \le n-1$, $1 \le p,q,r \le \infty$. Inequalities (1.1) are considered in the space $\mathcal{W}_{pr}^n(T)$ of functions $x(\cdot) \in L_p(T)$ with $(n-1)$ th derivative locally absolutely continuous on T and $x^{(n)}(\cdot) \in L_r(T)$.

For fixed T inequalities (1.1) depend on five parameters: n,k,p,q,r. We denote the best possible constant in this inequality by $K_T(n,k,p,q,r)$.

The first work dealing with such a problem is due to E. Landau (1913) who proved that $K_{\mathbb{R}_+}(2,1,\infty,\infty,\infty) = 2$. Kolmogorov (1938) determined the constant $K_{\mathbb{R}}(n,k,\infty,\infty,\infty)$ for $n \ge 2, 0 < k < n$. This result remains the most remarkable one for this type of problems, and this is the reason why the exact inequalities of type (1.1) are often called Kolmogorov or Landau–Kolmogorov inequalities and the constant $K_T(n,k,p,q,r)$ is called the *Kolmogorov constant.*

The determination of the Kolmogorov constant is equivalent to the following problem:

$$\left\|x^{(k)}(\cdot)\right\|_{L_q(T)} \to \max, \quad \left\|x(\cdot)\right\|_{L_p(T)} \le \gamma_1, \quad \left\|x^{(n)}(\cdot)\right\|_{L_r(T)} \le \gamma_2, \tag{1.2}$$

where γ_1 and γ_2 are arbitrary positive numbers. Such problems are called isoperimetrical ones.

2. THE LAGRANGE PRINCIPLE FOR NECESSARY CONDITIONS.

- In solving the problems (1.1), (1.2) and others we will use a unified approach which we call the *Lagrange principle.* It may be formulated as follows: *to solve an extremal problem with constrains, construct the Lagrange function of the problem, then write down the necessary condition in the similar problem on the extremum of the Lagrange function "as if the variables were independent" (in Lagrange's own words), and finally investigate the relations thus obtained.* This idea is the main principle of the first part of the theory of extremal problems. In this section we demonstrate its application to some important classes of extremal problems.

2.1 Problems without constraints.

The simplest extremal problem is a *problem without constraints*

$$f(x) \to \text{extr.} \qquad (P_1)$$

The first general method of solving (smooth) problems (P_1) in case of one variable was described by P. Fermat (even before calculus was developed). In the modern language it reads: *the main linear part of the increment of f at an extremum point equals to zero.* In this form it remains valid also in the infinite-dimensional case: *if $\hat{x}$ is a local minimum point of a function f differentiable at the point $\hat{x}$ then the following equality holds*:

$$f'(\hat{x}) = 0. \qquad (2.1)$$

2.2 The simplest problem of the calculus of variations.

After Fermat the theory of extremum made a sudden transition from one variable to infinitely many variables. This happened in 1696 when J. Bernoulli stated the *problem on brachistochrone*, where the argument was an infinite-dimensional object, viz., *a smooth curve* joining two given points of the plane. Later J. Bernoulli proposed to his student L. Euler to find a

general approach to problems of brachistochrone type. Euler summarized his results in the memoir "Methodus inveniendi..." published in 1744.

Euler considered the problem

$$\int_{t_0}^{t_1} L(t,x(t),\dot{x}(t))dt \to \text{extr}, \quad x(t_0)=x_0, \quad x(t_1)=x_1, \qquad (P_2)$$

which is referred to as the *simplest problem of the calculus of variations*. Here $L=L(t,x,y)$ is a function of three variables called the *integrand* of the problem. A necessary condition for extremum at $\hat{x}(\cdot)$ in the problem (P_2) is the following *Euler's equation*:

$$-\frac{d}{dt}L_{\dot{x}}+L_x=0 \Leftrightarrow -\frac{d}{dt}L_{\dot{x}}(t,\hat{x}(t),\dot{\hat{x}}(t))+L_x(t,\hat{x}(t),\dot{\hat{x}}(t))=0. \qquad (2.2)$$

2.3 The Lagrange multipliers rule.

A general principle for investigation of problems with constraints was first stated by J. L. Lagrange. In his book "Théorie des fonctions analytique", Paris, 1813, he wrote:

" On peut les réduire à ce principe générale. Lors qu'une fonction de plusieurs variables doit être un maximum ou minimum, et qu'il y a entre ces variables une ou plusieurs équation, il suffira d'ajouter à la fonction proposée les fonctions qui doivent être nulles, multipliées chacune par une quantité indéterminee, et là chercher ensuite le maximum ou minimum comme si les variables étaient indépendanbes; les équation qu'on trouvées serviront à déterminer toutes les inconnues."

Lagrange considers here a finite-dimensional problem

$$f_0(x)\to\text{extr}, \quad f_i(x)=0, \quad 1\le i\le m, \quad \lambda=(\lambda_0,\ldots,\lambda_m). \qquad (P_3)$$

His idea is as follows: compose the function $\mathcal{L}(x,\lambda)=\sum_{i=0}^{m}\lambda_i f_i(x)$ (we somewhat change Lagrange's formulation multiplying the functional itself by an indefinite factor too) and write down the necessary condition in the problem without constraints $\mathcal{L}(x,\lambda)\to\text{extr}$, i.e., apply equation (2.1) to obtain the relation

$$\mathcal{L}_x(\hat{x},\lambda)=0 \Leftrightarrow \sum_{i=0}^{m}\lambda_i f_i'(\hat{x})=0. \qquad (2.3)$$

(The function $\mathcal{L}(x,\lambda)$ is called the *Lagrange function*, while the numbers $\{\lambda_i\}_{i=0}^m$ are the *Lagrange multipliers*.) The result is as follows: *if the problem* (P_3) *satisfies some smoothness conditions then equality* (2.3) *holds at a local extremum point* $\hat{x}$. This result is referred to as the *Lagrange multipliers rule.*

Lagrange himself applied the idea of *elimination of constraints* by means of the Lagrange function (not only in finite-dimensional problems, but also in problems of calculus of variations) at least since 1750-th.

In the books [1]–[5] me and my co-authors tried to demonstrate the universal applicability of (somewhat extended) Lagrange's approach, according to which a meaningful necessary condition in a problem with constraints can be obtained by *writing down the Lagrange function and deriving then the necessary condition for its extremum "as if the variables were independent"*. (In [1]–[5] this approach is called the *Lagrange principle.)*

We illustrate the application of the Lagrange principle by two examples.

2.4 Lagrange's problem in calculus of variations.

Consider the problem:

$$\int_{t_0}^{t_1} f(t,x(t),u(t))\,dt \to \text{extr}, \quad \dot{x} = \varphi(t,x,u), \quad x(t_0) = x_0, x(t_1) = x_1, \tag{P_4}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^r$, f is a function of $n+r+1$ variables and φ is an n-dimensional vector function of the same variables. The variables x are called the *phase variables*, and u are the control. The problems of the form (P_4) are called the *Lagrange problems of calculus of variations*. Let us apply to them the Lagrange principle.

The Lagrange function here has the form

$$\mathcal{L} = \int_{t_0}^{t_1} L(t,x,\dot{x},u)\,dt, \quad L = \lambda_0 f(t,x,u) + p(t)\cdot(\dot{x} - \varphi(t,x,u))$$

(and it has been written in this form since Lagrange's time).

Obtaining the necessary condition in the problem for extremum of Lagrange's function without constraints reduces to writing down the Euler equation in x and u. As a result, we arrive at the equations

$$-\frac{d}{dt}L_{\dot{x}} + L_x = 0, \qquad L_u = 0, \tag{2.4}$$

which were derived in different forms for long time from brachistochrone in 1696 till forties of the last century (a book summarizing this topic was published in 1939 by the American mathematician G. Bliss).

Assuming that f and φ are sufficiently smooth, one can prove that *if a pair* $(\hat{x}(\cdot),\hat{u}(\cdot))$ *affords a local minimum in the problem* (P_4) *considered in* $C^1([t_0,t_1],\mathbb{R}^n)\times C([t_0,t_1],\mathbb{R}^r)$ *(when both nearness of phase coordinates and controls is taken into account; such extremum is said to be weak) then equations* (2.4) *hold.*

2.5 The optimal control problem

Now let a function f and a mapping φ be defined on the product $[t_0,t_1]\times\mathbb{R}^n\times U$, where U is a subset of $\mathbb{R}^r$. Assume that f and φ are continuous in all variables and smooth in x. Consider the problem of the same form as (P_4) assigning its own number to it:

$$\int_{t_0}^{t_1} f(t,x(t),u(t))\,dt \to \text{extr}, \quad \dot{x}=\varphi(t,x,u),$$
$$x(t_0)=x_0,\ x(t_1)=x_1,\ u\in U. \qquad (P_5)$$

The problems of type (P_5) are called *optimal control problems*. Their study was begun at L. S. Pontryagin's seminar during 50-s of the last century. In such problems a *strong* extremum is of interest when only nearness of phase coordinates is considered, while nearness of controls is neglected.

The set U in (P_5) may be arbitrary, for example, it may consist of finitely many points. The methods of smooth analysis used to derive equations (2.2) are inapplicable here. But the Lagrange principle remains valid.

The Lagrange function of the problem (P_5),

$$\mathcal{L}=\int_{t_0}^{t_1} L(t,x,\dot{x},u)\,dt, \quad L=\lambda_0 f(t,x,u)+p(t)\cdot(\dot{x}-\varphi(t,x,u)),$$

is the same as in the previous example. Consider two problems: first, we fix $\hat{u}(\cdot)$ and minimize over $x(\cdot)$, then we fix $\hat{x}(\cdot)$ and minimize over $u(\cdot)$.

In the first case the minimization problem is of simplest problems of calculus of variations type, and a necessary condition in this problem is Euler's equation in x:

$$-\frac{d}{dt}L_{\dot{x}}+L_x=0. \qquad (2.5)$$

In the second case we get the problem

$$\int_{t_0}^{t_1} L(t, \hat{x}(t), \dot{\hat{x}}(t), u(t))\, dt \to \min \quad \text{in} \quad u(\cdot), \tag{P_5'}$$

with $u(t) \in U$. It is easily seen that (under very mild assumptions) a criterion for minimality of $\hat{u}(\cdot)$ has the form of the following "minimum principle":

$$\min_{u \in U} L(t, \hat{x}(t), \dot{\hat{x}}(t), u) = L(t, \hat{x}(t), \dot{\hat{x}}(t), \hat{u}(t)). \tag{2.5'}$$

Changing signs we arrive at the form in which a necessary condition for the problem (P_5) was stated by Pontryagin's school:

$$\max_{u \in U} (p(t) \cdot \varphi(t, \hat{x}(t), \dot{\hat{x}}(t), u) - \lambda_0 f(t, \hat{x}(t), \dot{\hat{x}}(t), u)) =$$

$$\max_{u \in U} (p(t) \cdot \varphi(t, \hat{x}(t), \dot{\hat{x}}(t), \hat{u}(t)) - \lambda_0 f(t, \hat{x}(t), \dot{\hat{x}}(t), \hat{u}(t)). \tag{2.5''}$$

The combination of relations (2.5) and (2.5″) is known as *Pontryagin's maximum principle.*

An application of the maximum principle to the simplest variational problem leads to Legendre's and Weierstrass' necessary conditions, while its application to the second variation of the functional yields Jacobi's necessary condition.

The Lagrange principle will be the main device in the subsequent treatment of concrete problems.

3. THE LAGRANGE PRINCIPLE AND LANDAU – KOLMOGOROV INEQUALITIES.

- Now we begin to treat concrete problems on inequalities for derivatives of Landau–Kolmogorov type (see (1.1)) applying the Lagrange principle.

In all cases we proceed according to the following scheme:

1. Formalization of the problem. 2. Application of the Lagrange principle to the problem. 3. Investigation of the equations and/or inequalities resulting from the Lagrange principle. 4. Statement of the final result.

As a rule the investigation of necessary conditions leads to an identity (usually we call it the *general identity*). We obtain it heuristically, but after it has been obtained the identity can be verified directly.

3.1 Example.

Let us begin with the simplest situation: $T=\mathbb{R}_+$, $p=r=2$, $q=\infty$, $k=0$, $n=1$. This example is connected with the names of two mathematicians: B. Sz.-Nagy who found the Kolmogorov constant $K_{\mathbb{R}_+}(1,0,p,q,r)$ for arbitrary p,q and r, and V. N. Gabushin who investigated the case $p=r=2$, $q=\infty$, $n\in\mathbb{N}$, $0\le k<n$.

1. Formalization.

$$x(0)\to\max,\quad \int_{\mathbb{R}_+}x^2(t)\,dt\le\delta^2,\quad \int_{\mathbb{R}_+}\dot{x}^2(t)\,dt\le 1. \qquad (i)$$

This is an isoperimetric problem of the calculus of variations.

2. Lagrange principle. The Lagrange function of the problem (i) is

$$\mathcal{L}(x(\cdot),\lambda)=-\lambda_0x(0)+\lambda_1\int_{\mathbb{R}_+}x^2(t)\,dt+\lambda_2\int_{\mathbb{R}_+}\dot{x}^2(t)\,dt,$$

where $\lambda=(\lambda_0,\lambda_1,\lambda_2)$, $\lambda_i\ge 0$, $i=0,1,2$.

According to the main idea of Lagrange we have to use the Fermat theorem for the extremum of the Lagrange function "as if the variables were independent", i.e., for the problem

$$\mathcal{L}(x(\cdot),\lambda)=-\lambda_0x(0)+\lambda_1\int_{\mathbb{R}_+}x^2(t)\,dt+\lambda_2\int_{\mathbb{R}_+}\dot{x}^2(t)\,dt\to\min.$$

Hence (denoting by $\hat{x}(\cdot)$ a solution of the problem (i))

$$\mathcal{L}_{x(\cdot)}(\hat{x}(\cdot),\lambda)=0\Leftrightarrow$$
$$-\lambda_0x(0)+2\lambda_1\int_{\mathrm{IR}_+}\hat{x}(t)x(t)dt+2\lambda_2\int_{\mathrm{IR}_+}\dot{\hat{x}}(t)\dot{x}(t)dt=0,\quad\forall x(\cdot). \qquad (ii)$$

Integrating by parts we obtain the Euler equation and the transversality condition (the Euler equation could be obtained directly, see (2.3):

$$-2\lambda_2\ddot{\hat{x}}(t)+2\lambda_1\hat{x}(t)=0,\quad 2\lambda_2\dot{\hat{x}}(0)=-\lambda_0. \qquad (iii)$$

3. Investigation. All Lagrange multipliers are positive (for example, if $\lambda_2=0$ then $\lambda_1=\lambda_0=0$, which is impossible). So we can put $\lambda_0=1$. The general solution of the Euler equation vanishing at infinity has the form: $x(t,C,\lambda_1,\lambda_2)=C\exp(-\sqrt{\lambda_1/\lambda_2}\,t)$. We determine C,λ_1,λ_2 from the isoperimetrical conditions:

$$C^2\int_{\mathbb{R}_+}\exp\left(-2\sqrt{\frac{\lambda_1}{\lambda_2}}t\right)dt=\delta^2,\quad \frac{\lambda_1}{\lambda_2}C^2\int_{\mathbb{R}_+}\exp\left(-2\sqrt{\frac{\lambda_1}{\lambda_2}}t\right)dt=1$$

and the transversality conditions: $2\lambda_2\dot{\hat{x}}(0)=-1$. Therefore we obtain: $\hat{x}(t)=\sqrt{2\delta}\exp(-t/\delta)$, $\lambda_1=(2\delta)^{-3/2}$, $\lambda_2=2^{-1}(\delta/2)^{1/2}$. Substituting $\hat{x}(\cdot)$, λ_1, and λ_2 in (*ii*) me obtain the general identity:

$$x(0)=\frac{1}{\delta}\int_{\mathbb{R}+}e^{-t/\delta}x(t)\,dt-\int_{\mathbb{R}_+}e^{-t/\delta}\dot{x}(t)\,dt.$$

It is easy to verify (integrating by parts) that this identity holds true for all $x(\cdot)\in W^1_{22}(\mathbb{R}_+)$. From the Cauchy–Bunyakovskii inequality we have $|x(0)|\le\sqrt{2\delta}$ for all admissible $x(\cdot)$ in the problem (*i*), and for $\hat{x}(\cdot)$ we have the equality: $|\hat{x}(0)|=\sqrt{2\delta}$. Thus

4. The function $\hat{x}(\cdot)=\hat{x}(\cdot,\delta)$ is the solution of the problem (*i*) and $K_{\mathbf{R}_+}(1,0,2,\infty,2)=\hat{x}(0,1)=\sqrt{2}$.

Our reasoning here was heuristic, so, without justifying that λ_0 is non-zero, we will put λ_0 to be some fixed number (usually 1 or 1/2). Moreover, we had no need deduce the Euler equations and transversality conditions as in the example. We could use directly the relations of Section 2 and then write down the general identity. In all cases this identity may be derived as it was done in the example. This appointment gives the possibility to solve almost all problems of type (1.1) (see [7] – [8]). We will comment this in a few words.

3.2 Problems of small smoothness.

The largest amount of exact solutions were obtained for the cases when $n=1$ or 2. It was found about twenty such solutions, and the most part of them can be derived in a routine manner based on an application of the Lagrange principle (in a similar way to Section 3.1). We will describe briefly some arguments helpful in obtaining the solution. (The case of small smoothness is exposed in more detail in the book [4] and the paper [8].)

For $n=1$ the integrands of the isoperimetric problem do not depend on the independent variable, hence they admit the energy integral, which yields integrable equations. Moreover, if $q=\infty$ then the problem is convex, and the necessary conditions in this case coincide with sufficient ones, which implies immediately that the extremals thus obtained are solutions of the problem indeed. The case $q\neq\infty$ will be discussed later on.

Some remarkable solutions were obtained for $n=2$, $q=\infty$. In this case the solutions depend (for each of the domains $\mathbb{R}$ and $\mathbb{R}_+$) on two

parameters, p and r, and we can associate with each solution the point $(1/p, 1/r)$, $0 \le 1/p, 1/r \le 1$, of the unit square. So far the solutions have been obtained for the center of the square and for the points lying on its sides.

The center of the square will be discussed separately because in this case the "general solution" can be found for $n \ge 2$ for any $0 \le k \le n-1$. Here we briefly describe the solutions corresponding to the sides of the square. In all cases an application of the Lagrange principle yields the "general identity", which in principle solves the problem. For the left-hand side of the square, the solution $K_{\mathbb{R}_+}(2,1,\infty,\infty,r) = 2^{\frac{1}{r'+1}}(\frac{r'+1}{r'})^{\frac{r'}{r'+1}}$ (found for $r=\infty$ by E. Landau and for $r<\infty$ by V. V. Arestov; of course, by another methods) may be obtained because the application of the Lagrange principle leads to the general identity written down explicitly. The upper side of the square is dual to the Landau–Arestov case, which again enables one to write down the general identity and to obtain the solution: $K_{\mathbb{R}_+}(2,0,p,\infty,1) = (p+1)^{\frac{1}{p+1}}$. For the lower side of the square, the solution (found without using Lagrange's approach by A. Fuller and V. N. Gabushin) is obtained because the equations resulting from the Lagrange principle, besides the energy integral, admit one more integral, which reduces the problem to a system of two nonlinear equations in two unknowns. The right-hand side of the square is studied using its duality to the Fuller–Gabushin case.

Now it is natural to ask the following question: *can the solution found, say, for the lower side of the square be regarded indeed as a solution*?

For $p=2$ the equations resulting from the Lagrange principle can be solved explicitly, and the Kolmogorov constant is $K_{\mathbb{R}_+}(2,0,2,\infty,\infty) = 5^{2/5}2^{-3/5}(3\sqrt{33}+3)^{1/10}$. It is quite reasonable to consider this expression as a solution. But in order to compute, say, $K_{\mathbb{R}_+}(2,0,4,\infty,\infty)$ one has to solve the equation $56x^5 + 412x^4 - 11599x^3 + 59220x^2 - 98000x - 78400 = 0$, and the computation of $K_{\mathbb{R}_+}(2,0,3/2,\infty,\infty)$ requires solving the equation $\frac{-4x}{x^4-1} + 3(\pi + 2\operatorname{arctg} x + \log\frac{x+1}{x-1}) = 0$. May we consider the reduction to such an equation as an *exact solution* of the problem? (The "terrible" equations stated above were obtained by myself and my colleagues, while Fuller and Gabushin who studied the problem with parameters $n=2$, $k=0$, $1<p<\infty$, $q=r=\infty$ stopped even at an earlier stage.)

Thus the problem of specifying what should be regarded as an exact solution requires discussion. One can put forward the following judgement: the computer age will lead to essential modifications in the prevailing style of a huge number of mathematical studies.

3.3 General solutions.

In some cases the so-called "general solutions" were obtained giving, for fixed (p,q,r), the answers for any n and k. Relying again on the Lagrange principle, we will explain the reasons for this success (for more details, see the book [4] and the paper [7]).

In the Hardy–Littlewood–Polya case $(T=\mathbb{R}, p=q=r=2)$ the Fourier transform enables us to reduce the problem of finding the Kolmogorov constant to a linear programming problem. This gives a possibility to solve the following generalization of the problem (1.2):

$$\left\|\mathcal{D}^{\alpha^0}x(\cdot)\right\|_{L_2(\mathbb{R}^d)}\to\max,\quad \left\|\mathcal{D}^{\alpha^j}x(\cdot)\right\|_{L_2(\mathbb{R}^d)}\le\gamma_j, 1\le j\le N.$$

Here $\alpha^j=(\alpha_1^j,\alpha_2^j,...,\alpha_d^j)\in\mathbb{R}_+^d$, $0\le j\le N$, $\gamma_j>0$, $1\le j\le N$, $\mathcal{D}^{\alpha^j}x(\cdot)$ for $\alpha\in\mathbb{R}_+^d$ is the α-th Weyl derivative of a function $x(\cdot)\in L_2(\mathbb{R}^d)$. It is defined as follows: $\mathcal{D}^{\alpha}x(\cdot)=(F^{-1}\circ\mathcal{E}^{\alpha}\circ F)x(\cdot)$, where F and F^{-1} are direct and inverse Fourier transforms in $L_2(\mathbb{R}^d)$, and $\mathcal{E}^{\alpha}$ – an operator of multiplication to function $\tau\to(i\tau_1)^{\alpha_1}...(i\tau_d)^{\alpha_d},(i\tau_s)^{\alpha_s}=|\tau_s|^{\alpha_s}\exp(\frac{i\pi}{2}\alpha_s\operatorname{sgn}\tau_s)$, $\operatorname{sgn}=0$, $1\le s\le d$ (see [4] or [8]).

In the Taykov case $(T=\mathbb{R}, p=r=2, q=\infty)$ the Fourier transform reduces the problem to a linear-quadratic programming problem. In the Gabushin case $(T=\mathbb{R}_+, p=r=2, q=\infty)$ ("center of the square") one obtains a linear system of equations, which may be solved effectively. In the Kolmogorov case the general identity may be written down. In all of these four cases the problem is convex, therefore the necessary conditions coincide with the sufficient ones.

The Lyubich–Kuptsov case $(T=\mathbb{R}_+, p=q=r=2)$ also reduces to a linear system of differential equations, and one can use the basic Weierstrass identity for the proof of sufficiency.

The only case where the Lagrange principle fails is Stein's $(T=\mathbb{R}, p=q=r=1)$. It reduces to the Kolmogorov case.

4. CONCLUSIONS AND OPEN PROBLEMS.

The Lagrange principle is the main device in the subsequent treatment of concrete problems. But now we will briefly describe the other fundamental principles of the theory of extremum problems.

4.1 Perturbations of extremal problems and the principle of complete elimination of constraints.

A *perturbation* of a problem is its inclusion into a parametrized family of problems. The first mathematician who realized usefulness of perturbations in the calculus of variations theory was W. R. Hamilton. He wrote: "One has to compare dynamically feasible motions by variation of extreme states of the system". His interests were in optics, he studied the light propagation in nonhomogeneous media. Considering a beam of rays going from the same point, Hamilton (1836), along with trajectories of the rays (satisfying Fermat's variational principle), started to treat the wave fronts, i.e., the level curves of the S-function, which is the time needed for light to achieve a given point.

"Varying extreme states of the system" he derived a partial differential equation for the S-function in optics. In a year C. G. Jacobi applied Hamilton's approach to general problems of calculus of variations. The equation for the S-function of the simplest problem $S(\tau,\xi)=\inf\{J(x(\cdot))\,|\,x(t_0)=x_0, x(\tau)=\xi\}$, which has the form

$$\frac{\partial S}{\partial t}(t,x)+\mathcal{H}(t,x,\frac{\partial S}{\partial x}(t,x))=0$$

(where $\mathcal{H}(t,x,y)=\sup\{yu-L(t,x,u)\,|\,u\in\mathbb{R}\}$ is the Legendre transform of the function L), is known now as the *Hamilton–Jacobi equation.* The Hamilton–Jacobi equation enables us to express the increment of the functional J on the extremal in terms of the Weierstrass function. This leads to sufficient conditions in calculus of variations, which, subject to a regularity condition on the integrand (its convexity in the last argument), amount to strengthened Legendre's and Jacobi's conditions (when the inequality in Legendre's condition is strict and there is no conjugate point on the entire interval $[t_0,t_1]$).

But we can in fact formulate some general statement not only for the simplest problem, but also for any extremal problems of type (2.1)–(2.5) and others. The method of perturbations in nondegenerate cases leads to a generalization of the Lagrange principle, which can naturally be called the *generalized Lagrange principle* or the *principle of complete elimination of constraints.* It may be stated as follows: *for nondegenerate (in a neighborhood of a local extremum point) extremal problems with constraints, one can "slightly" modify the Lagrange function so that it will attain a local extremum without constraints at the extremum point.* ("The general identities" we spoke about in the section 3 and Weierstrass formula

in the calculus of variations are an realization of the principle of complete elimination of constraints.)

4.2 The compactness principle and the existence of solutions.

"I am sure that it will be possible to prove the existence theorems by means of a general principle whose idea is suggested by the Dirichlet principle. May be this general principle will help us to find an answer to the following question: does any regular variational problem have a solution if we assign an extended meaning when necessary to the very concept of solution?"

This deep view was expressed by D. Hilbert when formulating his 20th problem on the Paris Congress in 1900. The general principle that Hilbert had in mind was, of course, the *Weierstrass–Lebesgue–Baire compactness principle* according to which *a lower semicontinuous function on a compact set attains its minimum.*

We will illustrate the application of this principle in the simplest problem repeatedly discussed above.

In case of one-dimensional argument t it is natural to consider the problem (P_2) on the maximal in a certain sense space on which the very problem may be posed. This space may be taken to be the space of absolutely continuous functions (or the space $W_1^1([t_0,t_1])$. Regularity of the integrand guarantees lower semi-continuity, while the growth condition (the integrand must grow faster than a linear function) and the condition that the integrand is bounded from below insure compactness, hence, by the compactness principle, the existence. In this way numerous existence theorems, starting with works by L. Tonelli in the 20-s of the last century, were formulated.

But in multidimensional problems there is no natural "broadest" space. This difficulty was overcome by constructing appropriate spaces to the integrands of various problems. For example, such a space for the Dirichlet problem on a domain Ω is, of course, the Sobolev space $W_2^1(\Omega)$. This led to the concept of *generalized functions (distributions).* And to provide for the possibility to solve variational problems by direct methods (which will be briefly discussed in the next section), the theory of *embedding of functional spaces* was worked out. But the very style of existence theorems remained the same: if the space is constructed according to the integrand L which guarantees semicontinuity and compactness then the existence of a solution can be proved. In many important cases semicontinuity follows from regularity, compactness follows from the grows of integrand, and the boundary conditions are determined by the space embedded into the initial

one (in the Dirichlet problem on a circle (disk) D this is the space $W_2^{1/2}(\partial D)$ on the circumference).

4.3 The principles of optimization algorithms

"Thus the problem reduces to the simple one: given two points A and C and a horizontal line passing between them, find the point B on this line such that the route ABC be the fastest one."

This is a quotation from the letter of Leibniz to J. Bernoulli of July 31, 1696, regarding the brachistochrone.

This citation from Leibniz returns us to the origin. As was pointed out, calculus of variations arose "from brachistochrone", the problem to which J. Bernoulli "invited" his contemporary mathematicians. Apart from Johann Bernoulli himself, solutions were given by his brother Jacob, his pupil de l'Hospital, as well as by the founders of the contemporary mathematics Newton and Leibniz. These solutions contained the ideas which have influenced the theory of extremum from its origin till now. Johann Bernoulli solved the problem using an optico-mechanical analogy, which inspired Hamilton and then Jacobi to develop their theory. This principle was exploited by Huygens who also made a very important contribution to the Hamilton–Jacobi theory. On the other hand, as seen from the above quotation, Leibniz laid the foundations of "direct methods" by replacing an infinite-dimensional object, a curve, with a finite-parametric one, a broken line.

Leibniz was followed by Euler who derived his equation replacing a curve with broken lines. Nowadays the reduction to finite-dimensional problems is the main approach to numerical solution of extremal problems. Having done such a reduction, a (minimization) problem is treated by methods of descent (gradient methods and its modifications, the method of conjugate directions, and so on; this class includes also Danzig's simplex-method), various penalization methods, barrier methods; in convex problems section methods are applied. It is impossible to present all this in detail.

4.4 Problems.

The Lagrange principle not only enables us to examine completely (in a unified way) the majority of problems for extremum for which exact solutions have been found, but also provides for diverse and far-reaching generalizations. (This is demonstrated to some extent in [1]–[5] and [7], [8], as well as in a paper, now in preparation for publication, entitled "*Inequalities* by Hardy, Littlewood, and Pólya in 70 years".)

But the diversity of concrete problems (in particular, exact inequalities) allows us to review thoroughly the underlying theory and poses many open problems. We will discuss them in the context of Kolmogorov–Landau inequalities. Here are some open questions (from my point of view).

All the problems about inequalities on the real line and half-line are defined on non-compact sets ($\mathbb{R}$ and $\mathbb{R}_+$). The theory of such problems has not been developed adequately. For example, me and my colleagues could not use any available general theorems for the proof of the existence theorems in problems on inequalities for derivatives, and we had to prove such a theorem ourselves (see [6]). In this proof the compactness principle, which is actually involved in the most proofs of the existence theorems, required some refinement: our proof was based on the idea that it is "disadvantageous" for the functions of the sequence to be minimized to spread out over the entire unbounded domain. Thus,

1. The existence theory in the problems with arguments running over non-compact (especially, multidimensional) sets has to be substantially completed.

   There are a number of other questions regarding multidimensional problems which still have to be examined in more detail. This concerns, for example,
2. Transversality conditions "at infinity" for variational problems with non-compact domains.

   When integrating the Euler equations or the maximum principle relations (for example, in the Fuller–Gabushin case discussed above) many additional difficulties arise, which are related to instability of the solution to the Cauchy problem decreasing at infinity. Therefore it is desirable
3. To develop effective algorithms for integrating Cauchy problems and other important problems with ordinary and partial differential equations in unbounded domains. (Using some homemade methods we calculated in [8] the Kolmogorov constants in some cases of small smoothness. For example, $K_{\mathbb{R}_+}(2,0,4,\infty,4)=1.52178$, $K_{\mathbb{R}_+}(2,0,2,\infty,4)=1.64115$, $K_{\mathbb{R}_+}(2,0,4/3,\infty,4)=1.63751,\ldots$ Is it possible to consider such calculations as solutions of the corresponding problems?)
4. Even for the simplest isoperimetric problems it is desirable to have a well elaborated theory of sufficient conditions.

   The meaning of the words "well elaborated theory" may be as follows: it must be applicable at least to problems on inequalities with small smoothness, for example, for the problem $T=\mathbb{R}_+, n=1, k=0, q\neq\infty$ (to allow for obtaining general results without using ad hoc properties like explicit integrability).

A purpose of a good theory is to solve the majority of concrete problems, isn't it? When the questions 1 – 4 are answered, would it be possible to say that *all problems* (1.1) *are solved*?

5. And finally I want to mention the huge, fantastic world of *problems in several variables*, for which the Lagrange principle itself can hardly be regarded as justified.

As examples of such concrete problems one can consider the multidimensional problems of Landau-Kolmogorov-type:

$$\left\|\mathcal{D}^{\alpha^0} x(\cdot)\right\|_{L_p(\mathbb{R}^d)} \to \max, \quad \left\|\mathcal{D}^{\alpha^j} x(\cdot)\right\|_{L_{q_j}(\mathbb{R}^d)} \le \gamma_j, 1 \le j \le N.$$

I am grateful to Prof. Giannessi for the invitation to attend the conference "Variational Analysis and Applications" in Erice. I could not take part in the conference, so I considered it inappropriate to submit a paper in the conference proceedings. I am very thankful to Prof. Giannessi for his proposal to write such a paper.

REFERENCES

[1] Ioffe A.D. and Tikhomirov V. M.: Theory of extremal problems, North-Holland, Amsterdam, 1979.

[2] Alexeev V.M, Tikhomirov V. M.: and Fomin S.V., Optimal Control, NY, Cons. Bureau, 1987.

[3] Alexeev V.M, Galeev E. M., Tikhomirov V.M.: Recueil de Problemes, Moscow, Mir, 1987.

[4] Magaril-Il'yaev G.G. and Tikhomirov V.M.: Convex Analysis: Theory and Applications, AMS, 2003.

[5] Tikhomirov V.M., Stories about Maxima and Minima, NY, AMS, Math. Ass. of USA, 1990.

[6] Buslaev A.P., Magaril-Il'yaev G.G., Tikhomirov V.M.: On existense of extremal functions in inequalities for derivatives. // Mat. Zametki, 32:6, 1982, 823 – 834.

[7] Magaril-Il'yaev G. G., Tikhomirov V.M.: Kolmogorov-type inequalities for derivatives // Sbornik: Mathematics 9 – 1832.

[8] Kochurov A.C., Magaril-Il'yaev G.G., Tikhomirov V.M.: Inequalities for derivatives on a line and a halfline and problems of recovery// East Journal of Approximations, 9:2, 2004.

NUMERICAL SOLUTION FOR PSEUDOMONOTONE VARIATIONAL INEQUALITY PROBLEMS BY EXTRAGRADIENT METHODS*

F. Tinti
Dept. of Pure and Applied Mathematics, University of Padua, Padua, Italy

Abstract: In this work we analyze from the numerical viewpoint the class of projection methods for solving pseudomonotone variational inequality problems. We focus on some specific extragradient-type methods that do not require differentiability of the operator and we address particular attention to the steplength choice. Subsequently, we analyze the hyperplane projection methods in which we construct an appropriate hyperplane which strictly separates the current iterate from the solutions of the problem. Finally, in order to illustrate the effectiveness of the proposed methods, we report the results of a numerical experimentation.

1. INTRODUCTION

We consider the classical variational inequality problem VIP(F,C), which is to find a point x^* such that

$$x^* \in C \quad < F(x^*), x - x^* > \geq 0 \quad \forall x \in C, \tag{1}$$

* Italian FIRB Project, Grant n. RBAUO1JYPN

where C is a nonempty closed convex subset of $\Re^n$, $<\cdot,\cdot>$ the usual inner product in $\Re^n$ and $F:\Re^n \to \Re^n$ is a continuous function. Let C^* be the set of the solutions.

In the special case where $C=\Re^n_+$, the problem (1) is a *nonlinear complementary problem* (NCP):

$$x^* \geq 0, \qquad F(x^*) \geq 0 \quad \text{and} \quad <x^*, F(x^*)>=0. \tag{2}$$

If F is affine, $F(x)=Mx+q$ where $M \in \Re^{n\times n}$ and $q \in \Re^n$, then the problem (1) is an affine variational inequality problem and (2) is a *linear complementary problem* (LCP).

Many methods have been proposed to solve VIP(F,C). The simplest of these is the projection method, which, starting from any $x^0 \in C$, iteratively updates x according to the formula

$$x^{k+1} = P_C(x^k - \alpha F(x^k)),$$

where $P_C(.)$ denotes the orthogonal projection map onto C and α is a judiciously chosen positive steplength. Here, $P_C(x^k - \alpha F(x^k))$ is the solution of the following quadratic programming problem

$$\min_{x\in C} \frac{1}{2} x^T x - (x^k - \alpha F(x^k))^T x.$$

The projection method is based on the observation that $x^* \in C$ is a solution of (1) if and only if

$$x^* = P_C(x^* - \alpha F(x^*)). \tag{3}$$

This method is very simple; indeed it uses only function evaluations and projections onto C, then it is easy to implement, uses little storage, and can readily exploit any sparsity or separable structure in F or in C. Furthermore, the projection is easy to be obtained where C is defined by linear and/or box constraints. However, the projection methods require restrictive assumption on F for the convergence. The convergence analysis for the projection methods is based on the contractive properties of the operator $x \to x - \alpha F(x)$:
if F is strongly monotone (with constant l), i.e.

$$\exists l > 0 \ : \ <F(x)-F(y),(x-y)>\geq l\|x-y\|^2 \qquad \forall x, y \in C \qquad x \neq y,$$

and $F(x)$ Lipschitz continuous on C (with Lipschitz constant L), i.e.

$$\exists L>0 \ : \ \|F(x)-F(y)\| \le L\|x-y\| \qquad \forall x,y \in C,$$

and if $\alpha \in (0, 2l/L^2)$, the projection method determines a succession $\{x^k\}$ convergent to a solution of (1) (see page 24 [15], [16]).

Marcotte and Wu [11] have shown that the projection algorithm converges for cocoercive variational inequalities. We recall that the mapping F is cocoercive on C if there exist a positive constant $\tilde{l}$ such that

$$< F(y)-F(x), y-x > \ge \tilde{l}\|F(y)-F(x)\|^2 \quad \forall x,y \in C.$$

Any strongly monotone (with constant l) and Lipschitz continuous mapping (with Lipschitz constant L) is cocoecive with the constant $\tilde{l} = \frac{l}{L^2}$.

Furthermore, any cocoercive mapping is monotone, that is $< F(x)-F(y), x-y > \ge 0 \quad \forall x,y \in C$, and Lipschitz continuous ($L = \frac{1}{\tilde{l}}$), but the converse in not true. If $C^* \neq \varnothing$ and $\alpha \in (0, 2\tilde{l})$, the cocoercivity of the operator F is sufficient to assure the convergence of the projection algorithm.

To relax the strong hypotheses required by the projection method enlarging the class of the problems that we can solve, the extragradient method was proposed; because of (3), $x^* \in C$ is a solution of (1) if and only if

$$x^* = P_C(x^* - \alpha F(P_C(x^* - \alpha F(x^*)))),$$

then the basic idea of this method is to update x according to the double projection formula

$$x^{k+1} = P_C(x^k - \alpha F(P_C(x^k - \alpha F(x^k)))).$$

The extragradient method was proposed in the first time by Korpelevich [9] as follows. Given $x^0 \in C$, we generate a succession $\{x^k\}$ such that

$$\bar{x}^k = P_C(x^k - \alpha F(x^k)) \qquad x^{k+1} = P_C(x^k - \alpha F(\bar{x}^k)). \tag{4}$$

where α is constant for all iterations. In [1] and [19] the convergence of the extragradient method is proved under the following hypothesis: $C^* \neq \varnothing$, F

is a monotone and Lipschitz continuous mapping and $\alpha \in (0,1/L)$ where L is the Lipschitz constant.

A drawback is the choice of α when L is unknown. Indeed, if α is too small, the convergence is slow; when α is too large, there might be no convergence at all. This remark is confirmed by the numerical results shown in Table 1 where we report the number of iteration (iter), the number of function evaluations (nf), and the number of projections (np) for different choice of α when the extragradient method is applied on some test problems. The test problems are described in Table 3 of the Section 3.

α	np/nf	iter
Kojima-Shindo		
10^{-2}	442/442	221
10^{-1}	76/76	38
1	-/-	-
User OPT		
10^{-3}	1326/1326	663
10^{-2}	184/184	92
10^{-1}	-/-	-
Braess Net		
10^{-2}	472/472	236
10^{-1}	80/80	40
1	-/-	-

Table 1. Analysis of the convergence of the extragradient method (4) for different values of α.

Then, Khobotov in [8] introduces the idea to perform an adaptive choice of α, changing its value at each iteration as described in Section 2. If $C^* \neq \varnothing$, $F(x)$ is a monotone mapping and α choice suitable (see Section 2), then, the convergence of the scheme is proved.

The hypothesis on the Lipschitz continuity of F is removed and an automatic (algorithmic) rule is devised to make easy a convenient choice of the steplength.

Furthermore, as we see in the following, we can generalize the results on the convergence of the scheme to pseudomonotone VIPs, enlarging the class of the problems that we can solve.

Consequently, the general scheme of the algorithm becomes:

$$\bar{x}^k = P_C(x^k - \alpha_k F(x^k)) \qquad x^{k+1} = P_C(x^k - \eta_k F(\bar{x}^k)), \tag{5}$$

where $x^0 \in C$ is the starting point. In addition to the scheme in [8], we have analyzed other variants of (5) (see [10], [7]), in which the values of α_k, η_k are found using backtracking schemes similar to that of the Armijo steplength rule. The aim of these variants is to accelerate the convergence.

In [8], [10], the choice of the steplength rules follows an adaptive rule but they assume that $\alpha_k = \eta_k$, while in [6] and [7], the extragradient method uses $\alpha_k \neq \eta_k$ with different backtracking procedures to determine the steplength α_k. In the first case [6], one projection is required for any tentative step of the search, while in [7] only one evaluation of function is performed for any tentative step of the search. The last method is advantageous especially when the projection is computationally expensive.

Another class of the extragradient methods is the so called *projection-contraction methods* [17], where in the second projection a more general operator is used.

The idea of these algorithms is to choose a symmetric positive definite matrix $M \in R^{n \times n}$ and a starting point $x^0 \in C$, and to iteratively update x^k, as follows:

$$x^{k+1} = x^k - \gamma M^{-1}(T_\alpha(x^k) - T_\alpha(P_C(x^k - \alpha F(x^k)))), \tag{6}$$

where $\gamma \in \Re^+$ and $T_\alpha = (I - \alpha F)$ in which I is the identity matrix, α is chosen dynamically (in according to an Armijo type rule), so T_α is strongly monotone.

The geometric interpretation of the methods in [6] and [7] has been further on developed recently by Solodov in [18], devising an effective method. It consists of two steps per iteration: in the first step, an appropriate hyperplane is found which separates the current iterate from the solution of the problem; in the second step the next iterate is determined as the projection of the current iterate onto the intersection of the feasible set with the halfspace containing the solution set.

In all the algorithms with structure as in (5), (except that in [17], that requires the monotonicity of F), the convergence is stated under the assumptions that $C^* \neq \varnothing$ and the continuous mapping F is pseudomonotone. This is shown in the theorems reported in Section 2 that generalize to pseudomonotone case the results of the convergence obtained in [8], [6], [7]. See also [15] and [3].

It is not required F to be Lipschitz continuous.

We recall that the mapping F is pseudomonotone when the following condition holds

$$< F(y), x - y > \geq 0 \rightarrow < F(x), x - y > \geq 0 \quad \forall x, y \in C. \tag{7}$$

The paper is organized as follows.

In the Section 2 we give a survey of the above methods, pointing out its numerical features and we describe the different adaptive choices of α_k.

To evaluate the effectiveness of the proposed methods, we have implemented them as M-script files of MatLab, downloadable at the URL

`http://dm.unife.it/pn2o/software.html.`

Since we assume that C is defined by linear equalities and inequalities, in order to compute the projection $P_C(x)$, the quadratic program solver *quadprog.m* is used (see the MatLab optimization toolbox [13]).

In the last section we report the numerical results obtained by running these codes on a set of test problems arising from the literature and collected at URL

`http://dm.unife.it/pn2o/software.html.`

2. NUMERICAL FEATURES OF THE CLASS OF EXTRAGRADIENT METHODS

2.1 Khobotov's method

In [8] Khobotov proves that if $F(x)$ is a continuous monotone function and α suitable choice of the steplength is performed, the extragradient method (4) is convergent to a solution of (1). The proof is interesting since it includes a discussion about the choice of α_k.

We extended the Khobotov's theorem to a function $F(x)$ *pseudomonotone.*

For completeness, we report the convergence theorem:

Theorem 2.1 *(see [8]) Let the set C^* of solutions of (1) be non-empty, let C be a closed convex set, $F(x)$ a continuous pseudomonotone operator in x. Then, from any initial point $x^0 \in C$, if α_k is such that*

$$0 < \alpha_k \leq \min\left\{\overline{\alpha}, \beta \frac{\left\|x^k - \overline{x}^k\right\|}{\left\|F(x^k) - F(\overline{x}^k)\right\|}\right\} \tag{8}$$

with $\beta \in (0,1)$ and $\overline{\alpha}$ is equal to the maximum value of the step, then the extragradient method (4) is convergent to a solution x^* of (1), i.e.,

$$\lim_{k\to\infty} \min_{x^*} \left\|x^* - x^k\right\|_2 = 0 \qquad x^* \in C^*.$$

Proof. The proof of the theorem is based on the following condition

$$\left\|x^{k+1}-x^*\right\|^2 \le \left\|x^k-x^*\right\|^2 - \left\|x^k-\overline{x}^k\right\|^2 + \alpha_k^2\left\|F(x^k)-F(\overline{x}^k)\right\|^2. \tag{9}$$

We proof that this condition (9) holds under the pseudomonotonicity of the operator $F(x)$;
we see that, $\forall u, v \in C$,

$$\begin{aligned}\left\|u-v\right\|^2 &= \left\|u-P_C(u)+P_C(u)-v\right\|^2 \\ &= \left\|u-P_C(u)\right\|^2+\left\|v-P_C(u)\right\|^2-2<u-P_C(u),v-P_C(u)>;\end{aligned}$$

by the properties of the projection onto the convex set C

$$<u-P_C(u),v-P_C(u)>\le 0 \qquad \forall v\in C;\forall u\in\Re^n, \tag{10}$$

we obtain:

$$\left\|u-v\right\|^2 \ge \left\|u-P_C(u)\right\|^2+\left\|v-P_C(u)\right\|^2.$$

Taking $v=x^*, u=x^k-\alpha_k F(\overline{x}^k)$, (with $x^{k+1}=P_C(x^k-\alpha_k F(\overline{x}^k))$), we have

$$\left\|x^k-\alpha_k F(\overline{x}^k)-x^*\right\|^2 \ge \left\|x^k-\alpha_k F(\overline{x}^k)-x^{k+1}\right\|^2+\left\|x^*-x^{k+1}\right\|^2,$$

which leads to the inequality

$$\begin{aligned}&\left\|x^{k+1}-x^*\right\|^2 \le \left\|x^k-\alpha_k F(\overline{x}^k)-x^*\right\|^2-\left\|x^k-\alpha_k F(\overline{x}^k)-x^{k+1}\right\|^2 \\ &=\left\|x^k-x^*\right\|^2+\left\|\alpha_k F(\overline{x}^k)\right\|^2-2<\alpha_k F(\overline{x}^k),x^k-x^*>-\left\|x^k-x^{k+1}\right\|^2+ \\ &-\left\|\alpha_k F(\overline{x}^k)\right\|^2+2<\alpha_k F(\overline{x}^k),x^k-x^{k+1}> \\ &=\left\|x^k-x^*\right\|^2-\left\|x^k-x^{k+1}\right\|^2+2<\alpha_k F(\overline{x}^k),x^*-x^{k+1}>\end{aligned} \tag{11}$$

Recalling that the operator $F(u)$ is pseudomonotone, since $x^*\in C^*\subset C$,

$$<F(x^*),x-x^*>\ge 0 \rightarrow <F(x),x-x^*>\ge 0 \quad x\in C$$

Con sequently, if $x=\overline{x}^k$, $<F(\overline{x}^k),x^*-\overline{x}^k>\le 0$ and we have

$$\begin{aligned} < F(\overline{x}^k), x^* - x^{k+1} > &= < F(\overline{x}^k), x^* - \overline{x}^k > + < F(\overline{x}^k), \overline{x}^k - x^{k+1} > \\ &\leq < F(\overline{x}^k), \overline{x}^k - x^{k+1} > \end{aligned}$$

Then we have from (11):

$$\begin{aligned} \left\|x^{k+1} - x^*\right\|^2 &\leq \left\|x^k - x^*\right\|^2 - \left\|x^k - x^{k+1}\right\|^2 + 2\alpha_k < F(x^k), x^* - x^{k+1} > \\ &\leq \left\|x^k - x^*\right\|^2 - \left\|x^k - x^{k+1}\right\|^2 + 2\alpha_k < F(\overline{x}^k), \overline{x}^k - x^{k+1} > \\ &\leq \left\|x^k - x^*\right\|^2 - \left\|x^k - \overline{x}^k\right\|^2 - \left\|\overline{x}^k - x^{k+1}\right\|^2 + \\ &\quad -2 < x^k - \overline{x}^k, \overline{x}^k - x^{k+1} > + \\ &\quad +2\alpha_k < F(\overline{x}^k), \overline{x}^k - x^{k+1} > \\ &= \left\|x^k - x^*\right\|^2 - \left\|x^k - \overline{x}^k\right\|^2 - \left\|\overline{x}^k - x^{k+1}\right\|^2 + \\ &\quad +2 < x^k - \alpha_k F(\overline{x}^k) - \overline{x}^k, x^{k+1} - \overline{x}^k > \\ &\leq \left\|x^k - x^*\right\|^2 - \left\|x^k - \overline{x}^k\right\|^2 - \left\|\overline{x}^k - x^{k+1}\right\|^2 + \\ &\quad +2 < x^k - \alpha_k F(x^k) - \overline{x}^k, x^{k+1} - \overline{x}^k > + \\ &\quad +2 < \alpha_k F(x^k) - \alpha_k F(\overline{x}^k), x^{k+1} - \overline{x}^k > . \end{aligned}$$

Using (10), with $v = x^{k+1}, u = x^k - \alpha_k F(x^k)$, we obtain:

$$< x^k - \alpha_k F(x^k) - \overline{x}^k, x^{k+1} - \overline{x}^k > \leq 0.$$

Then, it follows

$$\begin{aligned} \left\|x^{k+1} - x^*\right\|^2 \leq \left\|x^k - x^*\right\|^2 - \left\|x^k - \overline{x}^k\right\|^2 + \left\|\overline{x}^k - x^{k+1}\right\|^2 + \\ +2\alpha_k \left\|F(x^k) - F(\overline{x}^k)\right\| \left\|x^{k+1} - \overline{x}^k\right\|. \end{aligned} \tag{12}$$

For any $x^{k+1}, x^k, \overline{x}^k, \alpha_k$, we have:

$$\left\|x^{k+1} - \overline{x}^k\right\|^2 + \alpha_k^2 \left\|F(x^k) - F(\overline{x}^k)\right\|^2 \geq 2\alpha_k \left\|F(x^k) - F(\overline{x}^k)\right\| \left\|x^{k+1} - \overline{x}^k\right\|;$$

then we obtain from (12):

$$\left\|x^{k+1} - x^*\right\|^2 \leq \left\|x^k - x^*\right\|^2 - \left\|x^k - \overline{x}^k\right\|^2 + \alpha_k^2 \left\|F(x^k) - F(\overline{x}^k)\right\|^2 .$$

Furthermore, the proof runs as in [8]. □

In the proof of the Khobotov's theorem, at each k-th iteration it is possible to find a compact subset of C, $\hat{C}_k$, where the function F is Lipschitz continuous; we denote by L_k the locally Lipschitz constant.
Since $\hat{C}_k \supset \hat{C}_{k+1} \supset \ldots$, it follows that

$$L_0 \geq L_1 \geq \ldots \geq L_k \geq .. \tag{13}$$

and it must $\alpha_k \in (0, 1/L_k)$.

Then, if $\{L_k\}$ are known, the succession $\{\alpha_k\}$ could be nondecreasing.

In the practice, estimates $\tilde{L}_k$ for L_k must be used; then for $\tilde{L}_k$, (13) does not hold and α_k is found from the following rule

$$0 < \hat{\alpha} \leq \alpha_k \leq \min\left\{\bar{\alpha}, \beta \frac{\left\|x^k - \bar{x}^k\right\|}{\left\|F(x^k) - F(\bar{x}^k)\right\|}\right\}$$

where $\bar{\alpha}$ is the maximum value of the step, $0 < \beta < 1$ (usually $\beta \approx 0.8, 0.9$) and $\hat{\alpha} = \min(\bar{\alpha}, \beta / L_0)$.

From the proof of the theorem, we can state the following Algorithm (Algorithm choice-α) for the choice of the steplength α_k.

Algorithm choice-α

a $\alpha = \alpha_{k-1}$ * ***(initial step)***

b compute $F(x^k)$

c compute $\bar{x}^k = P_C(x^k - \alpha F(x^k))$ and $F(\bar{x}^k)$

d **if** $F(\bar{x}^k) = 0$ then $\bar{x}^k \in C^*$

else if $\alpha > \beta \dfrac{\left\|x^k - \bar{x}^k\right\|}{\left\|F(x^k) - F(\bar{x}^k)\right\|}$ (14)

a ***reduction rule*** of α is applied
and go to (**c**)
else $\alpha_k = \alpha$, and $x^{k+1} = P_C(x^k - \alpha_k F(\bar{x}^k))$

endif

endif

* At the initial iteration $\alpha = \bar{\alpha}$

We enumerate several techniques for the reduction of α_k; the following reduction rule at the step (**c**) is suggested by Marcotte, in [10]:

$$\alpha = \min\left\{\frac{\alpha}{2}, \frac{\left\|x^k - \overline{x}^k\right\|}{\sqrt{2}\left\|F(x^k) - F(\overline{x}^k)\right\|}\right\}. \tag{15}$$

We note that this rule is not always effective: this arises when, at the initial iterations, α_k assumes a small value and, because of the initialization step $\alpha = \alpha_{k-1}$, this value does not change in all the next iterations.

Figure 1 shows the behavior of the stepsize α_k, as k increases, when we use the reduction rule (15); this rule does not exploit the opportunity of an adaptive alteration of the initial value of α_k.

A variant of Marcotte's algorithm consists in to modified the initialization rule at the step (**a**) of the Algorithm choice-α as follows:

$$\alpha = \alpha_{k-1} + \left(\beta \frac{\left\|x^{k-1} - \overline{x}^{k-1}\right\|}{\left\|F(x^{k-1}) - F(\overline{x}^{k-1})\right\|} - \alpha_{k-1}\right) \cdot \gamma, \tag{16}$$

where $\gamma \in (0,1), \beta \in (0,1)$.

By this rule we enable the increase of the value of α with respect to α_{k-1}. Then we devise the following reduction rule at the step (**c**)

$$\alpha = \max\left\{\hat{\alpha}, \min\left\{\xi \cdot \alpha, \beta \frac{\left\|x^k - \overline{x}^k\right\|}{\left\|F(x^k) - F(\overline{x}^k)\right\|}\right\}\right\}, \tag{17}$$

where $\xi \in (0,1)$.

Figure 2 shows the behavior of α_k for different test problems when the formulas (16), (17) are used, with $\beta = 0.7, \xi = 0.8, \gamma = 0.9$.

We observe that in general, the number of iterations decreases, since the rules (16), (17) enable to exploit the possibility to use convenient values of α_k at any iteration.

Since α_k is an estimate of the inverse of the local Lipschitz constant we can substitute the Algorithm choice-α with the following rule

$$\alpha_k = \beta \frac{\left\|x^k - \overline{x}^{k-1}\right\|}{\left\|F(x^k) - F(\overline{x}^{k-1})\right\|}, \tag{18}$$

avoiding the loop of the algorithm.

In this case, for the same test problem in Fig. 2, the behavior of α_k defined by (18), is similar to that observed for α_k stated by (16), (17) (see Fig. 3).

Nevertheless, in this case the convergence is not assured. The sequence x^k is convergent if α_k defined by (18) is such that

$$\alpha_k \leq \overline{\beta} \frac{\left\| x^k - \overline{x}^k \right\|}{\left\| F(x^k) - F(\overline{x}^k) \right\|}$$

where $\overline{\beta} > \beta$. This is not true in general, but in all the examined test problems the convergence is obtained.

2.2 The Extragradient method with $\alpha_k \neq \eta_k$

In [6], the author proposes the iterative scheme as in (5), where $\alpha_k > 0$ is located through a bracketing search and $\eta_k = \frac{<F(\overline{x}^k), x^k - \overline{x}^k>}{\left\| F(\overline{x}^k) \right\|^2}$.

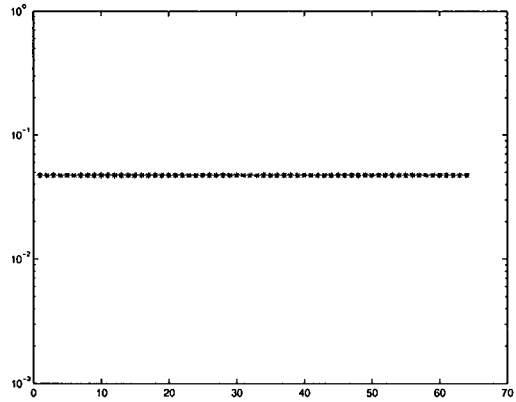
The idea behind the algorithm is the following.

Let $\partial H_k = \{x \in \Re^n \mid < F(\overline{x}^k), \overline{x}^k - x >= 0\}$ be a hyperplane normal to $F(\overline{x}^k)$ passing through $\overline{x}^k$; all solutions x^* of VIP(F,C) lie on one side of ∂H_k; indeed for the pseudomonotonicity of F, for any $x^* \in C^*$, we have $< F(x^*), \overline{x}^k - x^* >\geq 0$ and, consequently, $< F(\overline{x}^k), \overline{x}^k - x^* >\geq 0$.

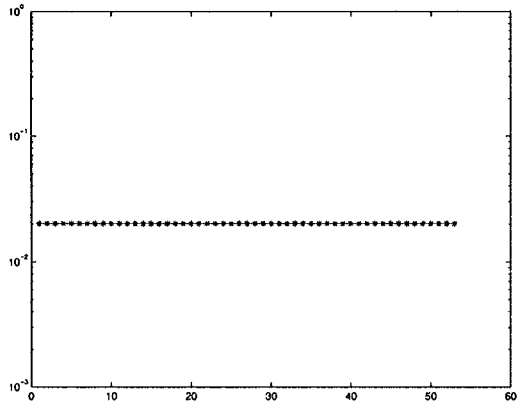
If x^k is on the other side, i.e. $< F(\overline{x}^k), \overline{x}^k - x^k >< 0$, then ∂H_k separates x^k from the solutions of VIP(F,C) (see Prop. 6, [6]).

If $\eta_k = \frac{<F(\overline{x}^k), x^k - \overline{x}^k>}{\left\| F(\overline{x}^k) \right\|^2}$, $x^k - \eta_k F(\overline{x}^k)$ is the orthogonal projection of x^k onto ∂H_k. Then x^{k+1}, obtained by the second equation of (4), is the orthogonal projection of x^k onto this hyperplane ∂H_k and onto C.

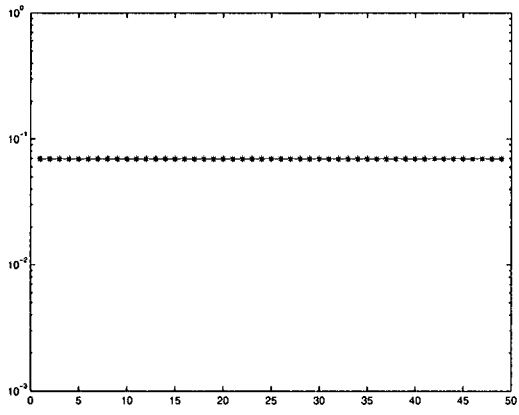
Iusem's algorithm requires three constants: $\varepsilon \in (0,1)$ and $\hat{\alpha}, \tilde{\alpha}$ such that $\tilde{\alpha} \geq \hat{\alpha} > 0$; the sequence α_k is computed so that $< F(\overline{x}^k), \overline{x}^k - x^k >\leq 0$, which is guaranteed to happen when $\alpha_k \in [\hat{\alpha}, \tilde{\alpha}]$.



Kojima Shindo α_k=0.0472, $\forall k$

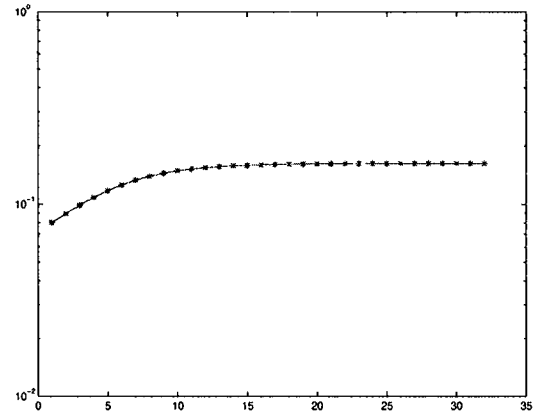


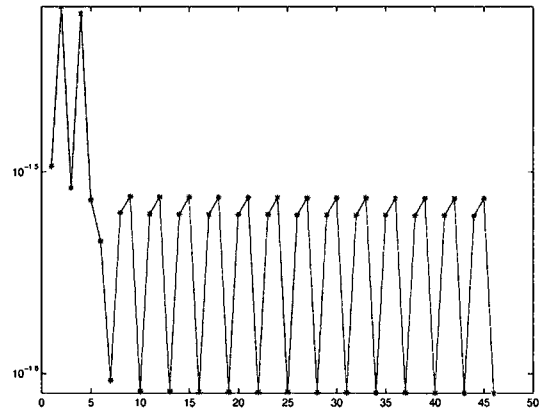
User OPT α_k=0.0201, $\forall k$



Braess Network $\alpha_k = 0.0697, \forall k$

Figure 1. Behavoir of α_k with reduction rule (15).





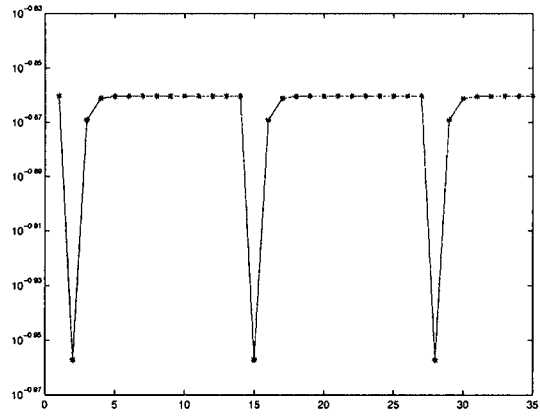
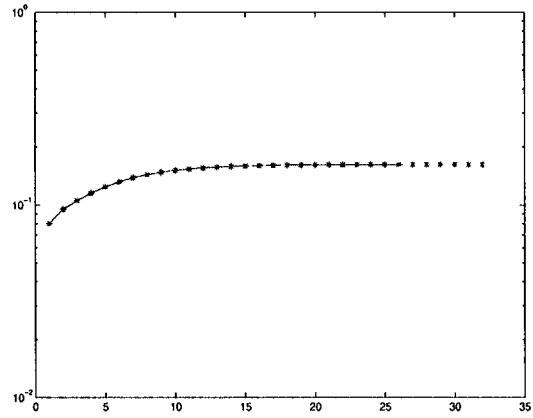
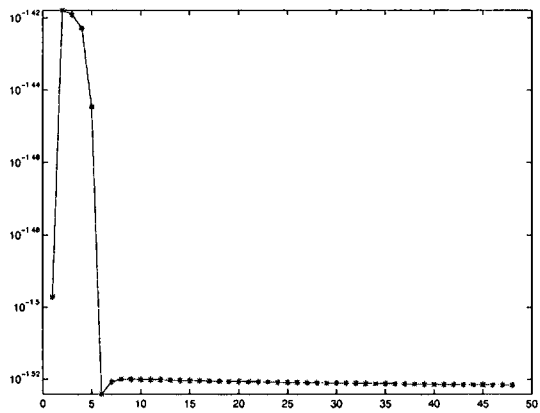


Figure 2. Behaviour of α_k with rules (16),(17); $\beta = 0.7, \xi = 0.8, \gamma = 0.9$.





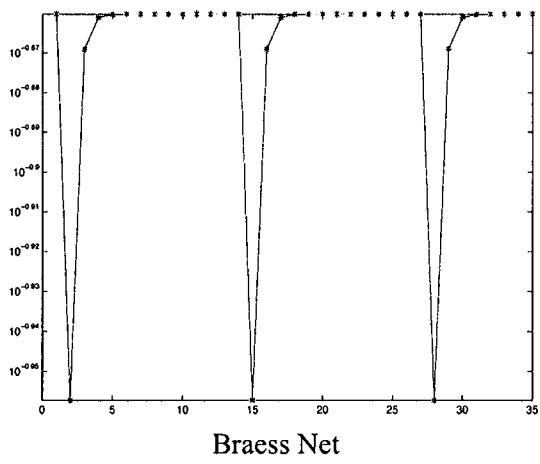


Figure 3. Behavior of α_k with rule (18); $\beta = 0.7, \tilde{\beta} = 0.9$

Then the algorithm can be stated as follows [6]:

Algorithm I

a given $x^0 \in C$, $k=0$, $rx = e$ [†]; $\varepsilon \in (0,1)$,

b **if** $\|rx\| < TOL$ [‡] **then** stop

else

chosen the initial value of the bracketing procedure $\tilde{\alpha}_k \in [\hat{\alpha}, \tilde{\alpha}]$, where $\tilde{\alpha}_k$ denote certain "candidate" of the steplength α_k.

c compute $\tilde{x}^k = P_C(x^k - \tilde{\alpha}_k F(x^k))$ and $F(\tilde{x}^k)$

d **if** $F(\tilde{x}^k) = 0$ **then** $\tilde{x}^k \in C^*$ stop

else (selection of α_k trough a finite bracketing procedure:)

if $\left\|F(\tilde{x}^k) - F(x^k)\right\| \le \dfrac{\left\|\tilde{x}^k - x^k\right\|^2}{2\tilde{\alpha}_k^2 \left\|F(x^k)\right\|}$

then $\bar{x}^k = \tilde{x}^k$

else find $\alpha_k \in (0, \tilde{\alpha}_k)$, such that

$$\varepsilon \frac{\left\|\tilde{x}^k - x^k\right\|^2}{2\tilde{\alpha}_k^2 \left\|F(x^k)\right\|} \le \left\|F(P_C(x^k - \alpha_k F(x^k)) - F(x^k)\right\| \le \frac{\left\|\tilde{x}^k - x^k\right\|^2}{2\tilde{\alpha}_k^2 \left\|F(x^k)\right\|} \tag{19}$$

$\bar{x}^k = P_C(x^k - \alpha_k F(x^k))$

endif

if $F(\bar{x}^k) = 0$ **then** $\bar{x}^k \in C^*$ stop

else compute $x^{k+1} = P_C\left(x^k - \dfrac{< F(\bar{x}^k), x^k - \bar{x}^k >}{\left\|F(\bar{x}^k)\right\|^2} F(\bar{x}^k)\right)$ (20)

$rx = x^{k+1} - x^k$;

$k = k+1$;

and go to (**b**).

endif

endif

endif

[†] e is a vector with entries equal to one

[‡] TOL is a final tolerance

In the step (**b**) of the Iusem's algorithm, one possible rule to choose the initial value $\tilde{\alpha}_k$ is

$$\tilde{\alpha}_k = median(\hat{\alpha}, \theta_k, \tilde{\alpha}),$$

where θ_k is suitably chosen.

In order to determine the stepsize α satisfying the required inequality (19), it is necessary to evaluate $P_C(x^k - \alpha F(x^k))$ at any step of the search procedure.

This means that the projections at the k-th iteration are those required for the bracketing search to determine α, plus one more in the computation of x^{k+1}.

In [6] (see Prop. 7), Ius em proves that if $C^* \neq 0$ and $F(x)$ is a continuous monotone function then this method is convergent to a solution of (1).

We extended the Prop. 7 to a function $F(x)$ continuous *pseudomonotone*, as follows:

Proposition 2.2 *(in [6]) Let the set C^* of solutions of (1) be non-empty, let C be a closed convex set, $F(x)$ a continuous pseudomonotone operator in x. Then, from any initial point $x^0 \in C$, the sequence $\{x^k\}$ generated by Algorithm I is convergent to a solution of (1).*

Proof. The proof of this proposition is based on the following condition

$$\left\|x^* - x^{k+1}\right\|^2 \leq \left\|x^* - x^k\right\|^2 - \left\|P_{H_k}(x^k) - x^k\right\|^2 - \left\|x^{k+1} - P_{H_k}(x^k)\right\|^2, \qquad (21)$$

where $x^* \in C^*, H_k = \{x \in \Re^n \mid < F(\overline{x}^k), \overline{x}^k - x > \geq 0\}$.

We proof the condition (21) under the pseudomonotonicity of the operator $F(x)$.

From (7) with $x = \overline{x}^k, y = x^*$ we obtain

$$< F(x^*), \overline{x}^k - x^* > \geq 0 \rightarrow < F(\overline{x}^k), \overline{x}^k - x^* > \geq 0;$$

then $x^* \in C \cap H_k$, so $P_C(P_{H_k}(x^*)) = P_{H_k}(x^*) = x^*$.

Let $v^k = x^k - \eta_k F(\overline{x}^k)$ the orthogonal projection of x^k onto the hyperplane ∂H_k, where ∂H_k separes x^k from the solution of VIP(F,C); by Prop. 6 in [6], we obtain $x^k \notin H_k$, then $v_k = P_{H_k}(x^k)$.

It follows from (20) that $x^{k+1} = P_C(P_{H_k}(x^k))$, then

$$\left\|x^* - x^{k+1}\right\|^2 = \left\|P_C(P_{H_k}(x^*)) - P_C(P_{H_k}(x^k))\right\|^2.$$

We apply the propriety of the projection onto the convex set C (Prop. 2(ii) in [6]):

$$\left\|P_C(x) - P_C(y)\right\|^2 \le \left\|x - y\right\|^2 - \left\|P_C(x) - x + y - P_C(y)\right\|^2 \quad \forall x, y \in \Re^n \qquad (22)$$

first with $P_C(.)$ and then with $P_{H_k}(.)$ as follows

$$\begin{aligned}
\left\|x^* - x^{k+1}\right\|^2 &\le \left\|P_{H_k}(x^*) - P_{H_k}(x^k)\right\|^2 + \\
&\quad - \left\|P_C(P_{H_k}(x^*)) - P_{H_k}(x^*) + P_{H_k}(x^k) - P_C(P_{H_k}(x^k))\right\|^2 \\
&\le \left\|x^* - x^k\right\|^2 - \left\|P_{H_k}(x^*) - x^* + x^k - P_{H_k}(x^k)\right\|^2 + \\
&\quad - \left\|P_C(P_{H_k}(x^*)) - P_{H_k}(x^*) + P_{H_k}(x^k) - P_C(P_{H_k}(x^k))\right\|^2 \\
&\le \left\|x^* - x^k\right\|^2 - \left\|P_{H_k}(x^k) - x^k\right\|^2 - \left\|x^{k+1} - P_{H_k}(x^k)\right\|^2
\end{aligned}$$

Then, the proof runs as in Prop. 7 in [6]. □

In [7], Iusem and Svaiter present a method with the scheme similar to the previous algorithm but that requires just one projection onto C for the computation of $\overline{x}^k$ and another one for x^{k+1}, i.e. only two projections per iteration, as in Korpelevich's method.

The algorithm requires the following parameters: $\varepsilon \in (0,1)$ and $\hat{\alpha}, \tilde{\alpha}$ such that $\tilde{\alpha} \ge \hat{\alpha} > 0$; the sequence α_k must be contained in $[\hat{\alpha}, \tilde{\alpha}]$; the scheme of the algorithm is:

Algorithm I-S

a given $x^0 \in C$, $k = 0$, $rx = e$;

b **if** $\|rx\| < TOL$ **then** stop

else

take an arbitrary stepsize $\alpha_k \in [\hat{\alpha}, \tilde{\alpha}]$,

c compute $z^k = x^k - \alpha_k F(x^k)$, $v^k = P_C(z^k)$

d **if** $F(v^k) = 0$ **then** $v^k \in C^*$ stop

e **else**

- compute

$$\bar{j} = \min_{j \in Z^+} \left\{ < F(2^{-j} P_C(z^k) + (1-2^{-j})x^k), x^k - P_C(z^k) > \geq \frac{\varepsilon}{\alpha_k} \left\| x^k - P_C(z^k) \right\|^2 \right\} \quad (23)$$

- compute $\beta_k = 2^{-\bar{j}}$
- compute $y^k = \beta_k v_k + (1-\beta_k)x^k$
- compute $\eta_k = \frac{<F(y^k), x^k - y^k>}{\|F(y^k)\|^2}$
- compute the orthogonal projection of x^k onto the hyperplane ∂H_k:

$$w^k = x^k - \eta_k F(y^k) \quad (24)$$

- compute

$$x^{k+1} = P_C(w^k) \quad (25)$$

$rx = x^{k+1} - x^k$;
$k = k+1$;
then go to (**b**).
endif

endif

In [7], Iusem and Svaiter observe that $\alpha_{k-1}\beta_{k-1}$ is an upper bound for the actual stepsize of the whole step from x^{k-1} to x^k, and they suggest that α_{k-1}, in the step (**b**), should be taken as

$$\alpha_{k-1} = median\{\hat{\alpha}, \theta\beta_{k-1}\alpha_{k-1}, \tilde{\alpha}\}$$

where $\theta > 1$ but not too large (for example $\theta = 2$).

Note that along the search for the appropriate β_k, the right hand side of (23) is kept constant; then we evaluate F at several points in the segment between v^k and x^k, no orthogonal projection onto C is required during the search, besides the computation of v^k and x^{k+1}.

We observe that a too small value of ε might induce a loss of precision of the algorithm; on the other hand, a value of ε close to 1, make the inequality in (23) too tight, increasing the value of j, and therefore decreasing β_k, and lengthening the bracketing search. It follows that ε should not be close to either 0 or 1.

In [7] (see Prop. 4), Iusem and Svaiter prove that if $C^* \neq 0$ and $F(x)$ is a continuous monotone function then this method is convergent to a solution of (1).

We extended the Prop. 4 to a function $F(x)$ continuous *pseudomonotone*, as follows:

Proposition 2.3 *(in [7]) Let the set C^* of solutions of (1) be non-empty, let C be a closed convex set, $F(x)$ a continuous pseudomonotone operator in x. Then from any initial point $x^0 \in C$, the sequence $\{x^k\}$ generated by Algorithm I-S is convergent to a solution of (1).*

Proof. The proof of this proposition is based on the following condition

$$\left\|x^{k+1} - x^*\right\|^2 \leq \left\|x^k - x^*\right\|^2 - \left\|w^k - x^k\right\|^2 - \left\|P_C(w^k) - w^k\right\|^2, \tag{26}$$

where $x^* \in C^*$.

Let $L_k = \left\{x \in \Re^n \mid < F(y^k), x - y^k > \leq 0\right\}$; using the pseudomonotonicity of F,

$$< F(x^*), y^k - x^* > \geq 0 \rightarrow < F(y^k), y^k - x^* > \geq 0,$$

we obtain that $x^* \in L_k$; on the other hand, $P_C(x^*) = x^*$.

By Prop. 3(iii) in [7], x^k does not belong to L_k; then using (24), it follows

$$P_{L_K}(x^k) = P_{\partial H_K}(x^k) = w^k$$

Then, from the propriety of the projection (22) and from (25) we obtain

$$\begin{aligned}
\left\|x^{k+1} - x^*\right\|^2 &= \left\|P_C(w^k) - P_C(x^*)\right\|^2 \\
&\leq \left\|w^k - x^*\right\|^2 - \left\|P_C(w^k) - w^k\right\| \\
&= \left\|P_{L_K}(x^k) - P_{L_K}(x^*)\right\|^2 - \left\|P_C(w^k) - w^k\right\|^2 \\
&\leq \left\|x^k - x^*\right\|^2 - \left\|P_{L_K}(x^k) - x^k\right\|^2 - \left\|P_C(w^k) - w^k\right\|^2.
\end{aligned}$$

Then, the proof runs as in Prop. 4 in [7]. □

From the computational point of view this method appears not effective since the convergence is very slowly, then we do not report in Section 3 the numerical results of this method, because they were rather poor. Indeed, we observe that frequently the hyperplane ∂H_k is near to the point x^k and the

next iteration $x^{k+1} = P_C(\overline{x}^k)$ is not much different from x^k and the convergence of the algorithm is very slow.

The interest forward the methods in [6] and [7] is justified by the fact that they are based on the same idea of the method of Solodov and Svaiter, discussed later in 2.4.

2.3 Solodov and Tseng (S-T) method

In [17], Solodov and Tseng propose a new class of methods for solving variational inequality problem, called *projection-contraction methods*, where the second projection is a more general operator:

$$\overline{x}^k = P_C(x^k - \alpha_k F(x^k)), \qquad x^{k+1} = x^k - \gamma M^{-1}(T_\alpha(x^k) - T_\alpha(P_C(\overline{x}^k)),$$

where $\gamma \in \Re^+$ and $T_\alpha = (I - \alpha F)$; here I is the identity matrix, α is chosen dynamically (in according to an Armijo type rule), such that T_α is strongly monotone.

Unlike the classical extragradient method (5), these methods require only one projection per iteration, rather then two, and they have an additional parameter, the scaling matrix M, that can be chosen to accelerate the convergence.

M must be a symmetric positive matrix.

The scheme of the method is the following.

Algorithm S-T

a choose $x^0 \in \Re^n, \alpha_{-1} > 0, \theta \in (0,2), \rho \in (0,1), \beta \in (0,1), M \in \Re^{n \times n}$

b $\overline{x}^0 = 0, k = 0, rx = e$

c **if** $\|rx\| < TOL$ **then** stop

else

$\alpha = \alpha_{k-1}, flag = 0$;

d **if** $F(x^k) = 0$ **then** $x^k \in C^*$ stop

else

while

$$(\alpha(x^k - \overline{x}^k)^T (F(x^k) - F(\overline{x}^k)) > (1-\rho)\left\|x^k - \overline{x}^k\right\|^2) or (flag = 0) \quad (27)$$

if $flag \neq 0$ **then** $\alpha = \alpha_{k-1}\beta$ **endif**;

update $\overline{x}^k = P_C(x^k - \alpha F(x^k))$, compute $F(\overline{x}^k)$

$flag = flag + 1;$

endwhile

f update $\alpha_k = \alpha$;

g compute $\gamma = \theta\rho\left\|x^k - \overline{x}^k\right\|^2 / \left\|M^{-1/2}(x^k - \overline{x}^k - \alpha_k F(x^k) + \alpha_k F(\overline{x}^k))\right\|^2$

h compute $x^{k+1} = x^k - \gamma M^{-1}(x^k - \overline{x}^k - \alpha_k F(x^k) + \alpha_k F(\overline{x}^k))$

$rx = x^{k+1} - x^k$;

k=k+1;

go to (**c**)

i **endif**

endif

In this algorithm the condition (27) may be viewed as a local approximation to the condition $\alpha < 1/L_k$, where the local Lipschitz constant L_k is given by

$$L_k = (x^k - \overline{x}^k)^T (F(x^k) - F(\overline{x}^k)) / \left\|x^k - \overline{x}^k\right\|^2 .$$

Then (27) reduces to $\alpha \le (1-\rho)/L_k$.

The convergence is proved under the assumption that a solution of (1) exists and that the operator F is monotone.

The rule (27) requires one projection and one function evaluation for any step of the search procedure. Another function evaluation is required to complete any iteration.

In Table 2, we shown, for $\beta = 0.3$ and $M = I$, the behavior of the method as θ and ρ assumes different values. In general, the choice of these parameters significantly affects the effectiveness of the method.

		Test Problem					
Parameters		**Kojima-Shindo**		**User-OPT Tranf. Pattern**		**Braess Network**	
$\beta = 0.3, M = I$		np/nf	iter	np/nf	iter	np/nf	iter
$\theta = 1.5$	$\rho = 0.1$	533/1064	530	71/140	68	61/121	59
	$\rho = 0.5$	84/166	81	90/117	86	89/176	86
$\theta = 1.9$	$\rho = 0.1$	416/830	413	55/108	52	48/95	46
	$\rho = 0.5$	59/116	56	21/40	18	67/132	64
$\theta = 1.0$	$\rho = 0.1$	808/1614	805	107/212	104	45/89	43
	$\rho = 0.5$	146/290	143	158/313	154	150/298	147

Table 2. Results for Projection-Contraction Methods

2.4 Solodov and Svaiter (S-S) method

Finally, we have analyzed a projection algorithm that was proposed by Solodov and Svaiter, in [18].

This algorithm allows a geometric interpretation as in [6] and [7] (see Fig. 4): let x^k be the current approximation of the solution of VIP(F,C); first, we compute the point $P_C(x^k - \mu_k F(x^k))$; next, we search the line segment between x^k and $P_C(x^k - \mu_k F(x^k))$ for a point z^i such that the hyperplane

$$\partial H_k = \left\{ x \in \Re^n \mid < F(z^k), x - z^k >= 0 \right\}$$

strictly separes x^k from the solution of the VIP(F,C) x^*.

To compute z^k, an Armijo-type procedure is used, i.e., $z^k = x^k - \eta_k r(x^k, \mu_k)$ where $\eta_k = \gamma^{\bar{i}} \mu_k$ with $\bar{i}$ being the smallest nonnegative integer i satisfying

$$< F(x^k - \gamma^i \mu_k r(x^k, \mu_k)), r(x^k, \mu_k) > \geq \frac{\sigma}{\mu_k} \left\| r(x^k, \mu_k) \right\|^2$$

and $r(x^k, \mu_k) = x^k - P_C(x^k - \mu_k F(x^k))$ is the projected residual function; after the hyperplane ∂H_k is constructed, the next iterate x^{k+1} is computing by projecting x^k onto the intersection between the feasible set C with the halfspace $H_k = \{x \in \Re^n \mid < F(z^k), x - z^k > \leq 0\}$ which contain the solution set C^*.

The scheme of the Solodov and Svaiter algorithm is reported in the following.

Algorithm S-S

a choose $x^0 \in C, \eta_{-1} > 0, \gamma \in (0,1), \sigma \in (0,1), \theta > 1, k = 0, rx = e$

b **if** $\|rx\| < TOL$ **then** stop

else

compute $\mu_k = min\{\theta \eta_{k-1}, 1\}$

c **if** $r(x^k, \mu_k) := x^k - P_C(x^k - \mu_k F(x^k)) = 0$ **then** $x^k \in C^*$ stop

d **else** compute

$$\bar{i} = \min_{i \in Z^+} \{ < F(x^k - \gamma^i \mu_k r(x^k, \mu_k)), r(x^k, \mu_k) > \geq \frac{\sigma}{\mu_k} \left\| r(x^k, \mu_k) \right\|^2 \}$$

where $\eta_k = \gamma^{\bar{i}} \mu_k$

e compute $z^k = x^k - \eta_k r(x^k, \mu_k)$

f compute the halfspace $H_k = \{x \in \Re^n \mid < F(z^k), x - z^k > \leq 0\}$

g compute $x^{k+1} = P_{C \cap H_k}(x^k)$

$rx = x^{k+1} - x^k$;

$k = k + 1$;

go to (**b**)

h **endif**

endif

Also in this method are needed only two projection per iteration.

This method should be especially effective when feasible sets are "no simpler" than general polyhedra; in this case, adding one more linear constraint to perform a projection onto $C \cap H_k$ doesn't increase the cost compared to projecting onto the feasible set C. In Figure 4, we analyze the differences between the Iusem and Svaiter method in [7] and the Solodov's method.

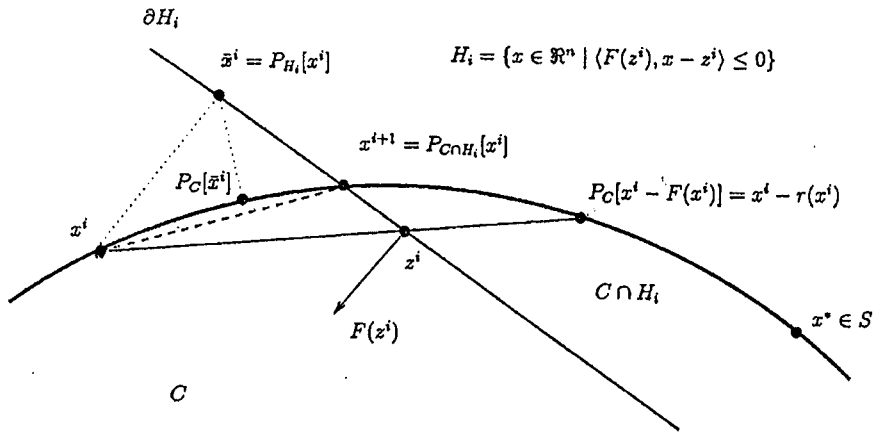


Figure 4. Comparison between Iusem Svaiter method [7] and Solodov and Svaiter method [18]

In [7], x^k is projected first onto the separating hyperplane ∂H_k and then onto C. If x^* near ∂H_k, $P_C(\bar{x}^k)$ can computationally are equal to x^k and the algorithm does not converge.

In [18], the second projection step in our method is onto the intersection $C \cap H_k$. We can observe that the iterate x^{k+1} is closer to the solution set C^* than the iterate computed by the method in [7].

In [18] it is shown that this method is convergent to a solution of the variational inequality problem under the only assumption that F is continuous and pseudomonotone.

3. COMPUTATIONAL EXPERIENCE

In order to evaluate the effectiveness of the extragradient methods discussed in the previous section, we consider a set of test problems arising from the literature (see the list in Table 3).

The M-function files implementing the considered test problems are downloadable at URL (http://dm.unife.it/pn2o/software.html).

We report in Table 4 the numerical results obtained by the MatLab M-script files implementing the considered methods. These codes can be downloadable at the URL (http://dm.unife.it/pn2o/software.html).

For the test problems with the suffix 'box' in the name of the input script files, the feasible region is given by the nonnegative orthant $x \in \Re^n_+$; they are NCPs. The other test problems are VIPs.

We choose very simple feasible regions so that the solver for inner quadratic programming problem has a low cost.

The starting point for all methods are feasible.

But, if we start from an unfeasible point, the first projection enables us to determine a feasible point that can be used as initial iterate.

All MatLab codes are run on a Notebook personal computer (ACER TravelMate 435LC, P-IV 3.06GHz) under MatLab version 6.5.0.180913a R13.

The following remarks can be drawn:

- between the three variants of the extragradient method, those related to (16)-(17) and (18) are more effective; the scheme related to (18) has near the same number of iterations with respect that related to (16)-(17) but the number of the projections and the number of the function evaluations are smaller; we remark the effectiveness of the extragradient method combined with (18) when we have to solve an NCP;
- the convergence of the S-T method is holds for monotone maps; the method has a better performance with respect the extragradient methods and it is very efficient for an **affine** VIP (see the test problem HPHard); for several test problems the number of iterations of this method appears convenient with respect to the S-S method; nevertheless the execution time of the S-S method can be smaller than that of the S-T method; but half of the projections of the S-S method has a different feasible region and then the number of projections are not comparable. Furthermore, the behavior of the S-T method strongly depends on the choices of its parameters (see Table 2).

  For monotone VIPs, we can be find convenient parameters so that the method is competitive with the others.
- For pseudomonotone VIPs, the S-S method appears in general very

effective (only for the test problem HpHard the behavior of the S-S method is poor); indeed, the numbers of iterations of the S-S method is less than those of all the other methods (except for the S-T method, however, that requires the monotonicity of F); but the complexity of each iteration can be larger of that of the other methods. Indeed the number of function evaluations can be greater than those of the extragradient method combined with the rule method (18) or (16) (17) and half of the projections has a different computational complexity since the feasible region is complicated by an additional (linear) constraint.

Then the effectiveness of the S-S method can depend on the structure of the feasible region, on the performance of the solver for the inner quadratic programming problem and on the analytical form of the mapping F.

We remark, in particular, the loss of the efficient for the NCPs, where the feasible region given by the nonnegative orthant significantly changes by the addition of a linear inequality.

4. CONCLUSION

In this paper we reported a numerical analysis of the behavior of a set of extragradient-type methods that enable us to solve pseudomonotone VIPs and NCPs. In particular, we devised a convenient variant of the Khobotov's extragradient method that appears numerically effective above all for NCPs where one projection on the nonnegative orthant is very simple.

We compared other two extragradient-type methods: the first proposed by Solodov and Tseng can be very convenient for monotone VIPs while the second proposed by Solodov and Svaiter and called hyperplane projection method can be solve also pseudomonotone VIPs. This method appears very effective when the addition of a linear inequality constraint to the original feasible region does not increase too much the computational complexity of the special projections required by the scheme.

All the numerical results are reproducible by the codes available on the web site URL(http://dm.unife.it/pn2o/software.html).

This work is in progress, since we intend to update in the site by adding new significant test problems and by collecting further numerical results on

Name Problem	dimension	reference	x^0	x^*
Mathiesen	3	[12]	$[0.1, 0.8, 0.1]^T$	$[0.50, 0.08, 0.41]^T$
		[12]	$[0.4, 0.3, 0.3]^T$	$[0.50, 0.08, 0.41]^T$
Kojima-Shindo	4	[4]	$[2, 0, 0, 2]^T$	$[1.22, 0, 0, 2.77]^T$
Kojima-Shindo box	4	[4]	$[2, 0, 0, 2]^T$	$[1.22, 0, 0, 0.50]^T$
Harker's Nash-C	5	[5]	$[1, 1, 1, 1, 1]^T$	$[0.97, 0.99, 1, 1.01, 1.01]^T$
Harker's Nash-C box	5	[5]	$[1, 1, 1, 1, 1]^T$	$[15.41, 12.50, 9.66, 7.16, 5.13]^T$
Harker's Nash-C	10	[5]	$[1, ..., 1]^T$	$[1.20, 1.12, 0.83, 0.55, 1.58, 1.12, 0.64, 1.17, 0.95, 0.79]^T$
Harker's Nash-C box	10	[5]	$[1, ..., 1]^T$	$[7.44, 4.09, 2.59, 0.93, 17.93, 4.09, 1.3, 5.59, 3.22, 1.67]^T$
Pang and Murphy's Nash-C	5	[4], [14]	$[1, 1, 1, 1, 1]^T$	$[0.95, 0.97, 0.99, 1.02, 1.04]^T$
Pang and Murphy's Nash-C box	5	[4], [14]	$[1, 1, 1, 1, 1]^T$	$[36.92, 41.73, 43.68, 42.68, 39.19]^T$
Pang and Murphy's Nash-C	10	[4], [14]	$[1, ..., 1]^T$	$[0.96, 1.1, 0.76, 0.97, 1.22, 1.10, 0.83, 1.03, 0.89, 1.10]^T$
Pang and Murphy's Nash-C box	10	[4], [14]	$[1, ..., 1]^T$	$[35.37, 46.57, 4.72, 19.91, 120.93, 46.57, 12, 42.56, 20.59, 32.98]^T$
Braess Network	5	[10]	$[6, 0, 6, 0, 6]^T$	$[4, 2, 2, 2, 4]^T$
User-Optimized Traffic Pattern	5	[2]	$[70, 70, 70, 60, 60]^T$	$[120, 90, 0, 70, 50]^T$
HPHard	20	[5]	$[1, ..., 1]^T$	$[0, 0, 1.71, 3.22, 1.95, 0, 0, 2.37, 0, 1.86,$ $1.93, 1.18, 0, 0, 0, 0.39, 1.68, 0.36, 1.44, 1.84]^T$
HPHard box	20	[5]	$[1, ..., 1]^T$	$[0.09, 1.31, 4.81, 23.31, 1.12, 0, 0, 22.35, 0, 12.60,$ $5.50, 6.36, 0, 15.69, 0, 0, 6.16, 12.23, 4.81, 11.94]^T$
HPHard	30	[5]	$[1, ..., 1]^T$	$[0, 0, 1.13, 2.61, 0, 0.51, 0, 1.31, 2.52, 0.16,$ $3.43, 1.88, , 0, 0, 0.80, 0, 0.61, 0, 3.36, 2.17,$ $0, 0, 0, 1.16, 1.09, 2.06, 2.80, 0.79, 0, 1.52]^T$
HPHard box	30	[5]	$[1, ..., 1]^T$	$[0, 0, 5.28, 9.84, 0, 2.35, 0.61, 3.83, 11.06, 0,$ $8.08, 3.71, 0, 0.19, 1.57, 0, 0.05, 7, 10.95, 6.31,$ $0.42, 0, 0, 5.42, 2.13, 5.11, 7.35, 2.90, 0, 5.08]^T$

Table 4. Numerical Results

Test Problems	Extragradient Method						Algorithm S-T			Algorithm S-S		
	(16)		(16),(17) $\beta = 0.7, \xi = 0.8, \gamma = 0.9$		(18) $\beta = 0.7, \beta = 0.9$		$\theta = 1.9, \rho = 0.5$			$\theta = 4, \gamma = 0.5, \sigma = 0.3, \eta_{-1} = 1$		
	np/nf	iter	np/nf	iter	np/nf	iter	time	np/nf	iter	time	np/nf	iter	time
Mathieson $x^0 = [0.1, 0.8, 0.1]^T$	-/-	-	80/80	31	59/60	29	0.35	96/188	91	0.45	27/62	13	0.13
$(n = 3)$ $x^0 = [0.4, 0.3, 0.3]^T$	-/-	-	65/65	24	47/48	23	0.31	34/62	27	0.25	25/60	12	0.13
Kojima-Shindo $(n = 4)$	131/132	64	67/67	32	67/68	32	0.41	59/116	56	0.37	25/60	12	0.13
Kojima-Shindo box $(n = 4)$	292/293	145	139/139	69	125/126	61	0.58	86/170	83	0.44	100/273	54	0.52
Harker's Nash-C $(n = 5)$	41/42	20	45/45	19	41/42	20	0.27	17/29	11	0.18	37/94	18	0.16
Harker's Nash-C box $(n = 5)$	1457/1458	726	175/175	69	146/147	69	0.50	941/859	941	2.09	139/349	69	0.50
Harker's Nash-C $(n = 10)$	113/114	55	82/82	39	84/85	40	0.41	46/88	41	0.30	79/200	39	0.34
Harker's Nash-C box $(n = 10)$	2383/2384	1191	743/743	296	617/618	308	1.75	669/1334	664	2	847/2119	423	3.09
P. M.'s Nash-C $(n = 5)$	44/45	21	48/48	22	45/46	22	0.26	17/30	12	0.20	43/109	21	0.17
P. M.'s Nash-C box $(n = 5)$	9486/9487	4739	143/143	58	132/133	65	0.48	2811/5618	2806	7.76	67/169	28	0.20
P.M.'s Nash-C $(n = 10)$	58/59	28	63/63	29	62/63	30	0.33	13/22	6	0.19	35/88	17	0.16
P.M.'s Nash-C box $(n = 10)$	16464/16465	8229	237/237	108	231/232	109	0.72	14198/28392	14193	40	417/634	208	1.42
Braess Network $(n = 5)$	100/101	49	74/74	36	71/72	35	0.36	67/132	64	0.34	45/111	22	0.14
User-Opt. Traf. Pat. $(n = 5)$	108/109	53	111/111	46	98/99	48	0.50	21/40	18	0.22	69/173	34	0.30
HPHard $(n = 20)$	456/457	227	392/392	157	365/366	182	4.19	119/233	113	1.58	579/1450	289	10
HPHard box $(n = 20)$	1497/1498	747	1142/1142	455	924/925	461	8.34	725/1441	717	10.70	1347/3371	673	19.04
HPHard $(n = 30)$	986/987	182	305/305	122	252/253	125	5.75	144/283	138	3.27	443/1110	221	24.50
HPHard box $(n = 30)$	617/618	307	534/534	213	437/438	217	6.92	270/535	264	7.17	645/1616	322	16.54

np is the number of projection

nf is the number of evaluation function

iter is the number of iteration; the stopping criterion is $\|r(x^k)\| = \|x^k - x^{k-1}\| \le 10^{-4}$

- denotes that the method does not converge.

the considered schemes and on the new schemes in the framework of extragradient-type methods.

REFERENCES

[1] D.P. Bertsekas, J.N. Tsitsiklis *Parallel and Distribuited Computation, Numerical Method*, Prentice-Hall, London 1989.

[2] S. Dafermos *Traffic equilibrium and variational inequalities*, Transportation Science, 14 (1980), pp. 42-54.

[3] F. Facchinei, J.-S. Pang *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer Series in Operation Research, Springer (2003).

[4] S.A. Gabriel, J.-S. Pang *NE/SQP: A robust algorithm for the nonlinear complementary problem*, Mathematical Programming, 60 (1993), pp. 295-337.

[5] P.T. Harker *Accelerating the convergence of the diagonalization and projection algorithms for finite-dimensional variational inequalities*, Mathematical Programming 41, (1988), pp. 29-59.

[6] A.N. Iusem *An iterative algorithm for the variational inequality problem*, Computational and Applied Mathematics ,Vol.13, N.2, (1994), pp. 103-114.

[7] A.N. Iusem, B.F. Svaiter *A variant of Korpelevich's method for variational inequalities with a new search strategy*, Optimization, Vol. 42, (1997), pp. 309-321.

[8] E.N. Khobotov *Modification of the extra-gradient method for solving variational inequalities certain optimization problems*, U.S.S.R. Computational Mathematics and Matematical Physics, 27, (1987), pp. 120-127.

[9] G.M. Korpelevich *The extragradient method for finding saddle points and other problems*, Matekon, 13, (1977), pp. 35-49.

[10] P. Marcotte *Application of Khobotov's algorithm to variational inequalities and network equilibrium problems*, INFORM, 29, (1991).

[11] P. Marcotte, J.H. Wu *On the convergence of projection methods application to the decomposition of affine variational inequalities*, Journal of Optimization Theory and Applications, Vol.85, N.2, (1995), pp. 347-362.

[12] L. Mathiesen *An algorithm based on a sequence of linear complementary problems applied to a Walrasian equilibrium model: an example*, Mathematical Programming, 37, (1987), pp. 1-18.

[13] The Math Works Inc. *MatLab Optimization Toolbox v.2.0*, User's Guide, (2002).

[14] F.H. Murphy, H.D. Sherali, A.L. Soyster *A mathematical programming approach for determining oligopolistic market equilibrium*, Mathematical Programming, 24, (1982), pp. 92-106.

[15] M. Patriksson *Nonlinear Programming and Variational Inequality Problems, A Unified Approach*, Applied Optimization, Vol.23, Kluwer Academic Pubblishers (1999).

[16] M. Sibony *Méthodes itératives pour les équations et inéquations aux dérivés partielles nonlinéares de type monotone* Calcolo, N.7, (1970), pp. 65-183.

[17] M.V. Solodov, P. Tseng *Modifield projection type methods for monotone variational inequalities* SIAM Journal Control Optimization, Vol.34, N.5, (1996), pp. 1814-1830.

[18] M.V. Solodov, B.F. Svaiter *A new projection method for variational inequality problems*, SIAM Journal Control Optimization, Vol.37, N.3, (1999), pp. 756-776.

[19] P. Tseng *On linear convergence of iterative methods for variational inequality problem*, Journal of Computational and Applied Mathematics 60, (1995), pp. 237-252.

REGULARITY AND EXISTENCE RESULTS FOR DEGENERATE ELLIPTIC OPERATORS

C. Vitanza[1] and P. Zamboni[2]
University of Messina, Dept. of Mathematics, Messina, Italy;[1] University of Catania, Dept. of Mathematics, Catania, Italy[2]

Abstract: In the first section of this paper we study the Hölder-continuity of solutions of the Schrödinger degenerate equation

$$-\sum_{i,j=1}^{n}\left(a_{ij}u_{x_i}\right)_{x_j}+cu=0, \qquad (*)$$

assuming the potential c belonging to appropriate *degenerate* Morrey spaces. In the second section we obtain the existence and the uniqueness of the solution of a variational inequality associated to the degenerate operator

$$Lu=-\sum_{i,j=1}^{n}\left(a_{ij}(x)u_{x_i}+d_ju\right)_{x_j}+\sum_{i=1}^{n}b_iu_{x_i}+cu \qquad (**)$$

assuming the coefficients of the lower terms and the known term belonging to a suitable *degenerate* Stummel-Kato class. In both cases the weight w, which gives the degeneration, belongs to the Muckenoupt class A_2.

1. REGULARITY RESULTS FOR DEGENERATE ELLIPTIC OPERATORS

It is known that the regularity theory of linear, as well as of quasilinear second order elliptic equations in divergence form, with lower order terms and the known term in L^p spaces, was settled in the sixties by the contributions of many Authors (see e.g. [6],[21], [11], [19]). Ladyzhenskaya and Ural'tseva claim that: *the necessity of the restrictions indicated below should be understood in the sense that if one of the restrictions is weakened, then the class of equations in question will include one with a solution not possessing the property in question* (see e.g. [11] p. 10).

To be more specific, let us consider the following equations in Ω bounded open set of $\mathbb{R}^n$ $(n \geq 3)$

$$-\Delta u + cu = 0 \tag{1.1}$$

$$-\Delta u = f. \tag{1.2}$$

It is known that the solutions of (1.1) and (1.2) are locally Hölder continuous under the following assumptions

$$c, f \in L^p(\Omega), \quad p > \frac{n}{2}.$$

The above recalled results are sharp in the classical L^p theory. Indeed both equations (1.1) and (1.2) have solutions which are unbounded if the assumptions

$$c, f \in L^{\frac{n}{2}}(\Omega)$$

hold (see e.g. [11]).

We also wish to recall the paper [15], where it is obtained the Hölder continuity of the solutions of (1.2) assuming

$$f \in L^{\frac{n}{2}}(\Omega)$$

and

$$\int_{\Omega \cap B(x,r)} |f(y)|^{\frac{n}{2}} \, dy \leq K r^{\alpha} \quad \forall x \in \Omega, \forall r > 0 \tag{1.3}$$

for some positive constants K and α. Here $B(x,r)$ denotes the ball centered in x and radius r.

At this time it may be convenient to recall the definition of the Morrey spaces.

Definition 1.1 *Let* $1 \le p < \infty$, $0 < \lambda < n$. *We say that* $f \in L^p(\Omega)$ *belongs to the Morrey space* $L^{p,\lambda}(\Omega)$ *if*

$$\sup_{\substack{x\in\Omega,\\ r>0}} \frac{1}{r^\lambda} \int_{\Omega\cap B(x,r)} |f(y)|^p \, dy \equiv \|f\|^p_{L^{p,\lambda}(\Omega)} < +\infty.$$

Here and in the sequel, $\subseteq$ and $\subset$ will denote inclusion and strict inclusion, respectively.

Remark 1.2 The condition (1.3) means that $f \in L^{\frac{n}{2},\alpha}(\Omega)$. Also it is known that $L^p(\Omega) \subset L^{\frac{n}{2},\alpha}(\Omega)$, for some opportune positive α if $p > \frac{n}{2}$ (see e.g. [18]).

Now we can consider the question if it is possible to find some subspace of L^1 to which the known term and the coefficients of the lower order terms must belong so that the previous regularity results continue to hold. The first regularity result obtained under not L^p assumptions is due to H. Lewy and G. Stampacchia in [12], where it is proved that the solutions of equation (1.2) are Hölder continuous if

$$f \in L^{1,\lambda}(\Omega)$$

for some $\lambda > n-2$.

The greatest contribution, after Lewy and Stampacchia, in the direction outlined above, seems to be given by Aizenmann and B. Simon in [1], where they were able to prove the Harnack inequality for positive weak solutions of equation (1.1) assuming c in the so called Stummel-Kato class defined as follows

Definition 1.3. *We say that* $f \in L^1(\Omega)$ *belongs to the Kato-Stummel class* $S(\Omega)$ *if*

$$\sup_{x\in\Omega} \int_{\Omega\cap B(x,r)} \frac{|f(y)|}{|x-y|^{n-2}} dy \equiv \eta(f,r) < +\infty, \forall r > 0$$

and

$$\lim_{r\to 0^+}\eta(f,r)=0.$$

Remark 1.4 It is worth to note that

$$L^{1,\lambda}\subseteq S\subseteq L^{1,n-2},\quad \lambda>n-2$$

(see e.g. [7]).

In 1986 F. Chiarenza, E. Fabes and N. Garofalo in [3] extended the previous result to the equation

$$-\sum_{i,j=1}^{n}\left(a_{ij}u_{x_i}\right)_{x_j}+cu=0 \tag{1.4}$$

where the coefficients a_{ij} are such that

$$a_{ij}=a_{ji},$$
$$\exists\lambda>0:\lambda^{-1}|\xi|^2\le a_{ij}\xi_i\xi_j\le\lambda|\xi|^2\ \forall\xi\in\mathbb{R}^n,$$

and the potential c is in $S(\Omega)$.

A further step in the study of (1.4) was done by Di Fazio in [7] (see also [20]), where it is shown that if c is taken in $L^{1,\lambda}$, with $\lambda>n-2$, then u is hölder continuous.

The results obtained in the previous works show that the L^p assumptions are not the best possible if other classes (like the Morrey spaces or the Stummel Kato class) different from L^p spaces, are taken in consideration. In fact it is worth to note that

$$L^p\subset S\quad \text{if } p>\frac{n}{2},$$
$$L^p\subset L^{1,\lambda}\quad \text{for some } \lambda>n-2 \text{ if } p>\frac{n}{2}.$$

In the degenerate case, the regularity of solutions of equation

$$-\sum_{i,j=1}^{n}\left(a_{ij}u_{x_i}\right)_{x_i}=0$$

was studied by E. Fabes, C Kenig and R. Serapioni in [8]. There the operator is assumed to be degenerate in the sense that the uniform ellipticity condition is substituted by the following

$$\exists \lambda > 0 : \lambda^{-1} w |\xi|^2 \leq a_{ij} \xi_i \xi_j \leq \lambda w |\xi|^2 \ \forall \xi \in \mathbb{R}^n.$$

The weight w is assumed to belong to the Muckenoupt class A_2 (see [16], [5] and [9]).

Subsequently C. Gutierrez gave an extension of the previous result for the degenerate Schrödinger equation

$$-\sum_{i,j=1}^{n} \left(a_{ij} u_{x_i} \right)_{x_j} + cu = 0$$

(see [10]).

In his work Gutierrez obtained the Harnack inequality, assuming $\frac{c}{w}$ in the *degenerate* space $S(\Omega, w)$ whose definition is given below.

Definition 1.4. *We say that* $f \in L^1(\Omega, w)$ *belongs to the Kato-Stummel class* $S(\Omega, w)$ *if*

$$\sup_{x \in \Omega} \int_{\Omega \cap B(x,r)} |f(y)| \int_{|x-y|}^{4R} \frac{s^2}{w(B(x,r))} \frac{ds}{s} w(y)\, dy \equiv \eta(f,r) < +\infty, \forall r > 0$$

and

$$\lim_{r \to 0^+} \eta(f,r) = 0,$$

where Ω *is a bounded open set in* $\mathbb{R}^n$, $\Omega \subseteq B(0,R)$ *and* $w(B(x,r)) = \int_{B(x,r)} w(x)\, dx$

Remark 1.5. $S(\Omega, w)$ is an appropriate modification of the Stummel-Kato class. Indeed $S(\Omega, 1) \equiv S$.

We refer to the survey [2] for more details on this matter.

Now we present some results obtained in our note [22] which extend to the degenerate case the analogous ones contained in [7] and [20].

In [22] we consider the equation (1.4) where the coefficients $a_{ij}(x)$ are measurable functions such that

$$a_{ij}(x) = a_{ji}(x) \qquad i,j = 1,2,...,n$$

and

$$\exists \nu > 0 : \nu^{-1} w(x) |\xi|^2 \le \sum_{i,j=1}^{n} a_{ij}(x)\xi_i\xi_j \le \nu w(x) |\xi|^2 \ \forall \xi \in \mathbb{R}^n,$$

with the weight w belonging to the A_2 class.

The first problem one has to face is to understand what a "degenerate" Morrey space is. We introduce two such notions of degenerate Morrey space, $M_\sigma(\Omega, w)$ and $L^{1,\varepsilon}(\Omega, w)$, whose definitions are the following

Definition 1.6 *Let* $\sigma > 0$, $C > 0$ *and* $0 < r < 2R$. *We set*

$$M_\sigma(\Omega, w) = \{f \in L^1(\Omega, w) :$$
$$\sup_{x\in\Omega} \int_{\{y\in\Omega : |x-y|<r\}} |f(y)| \int_{|x-y|}^{4R} \frac{s^2}{w(B(x,s))} \frac{ds}{s} w(y)dy \le Cr^\sigma\}.$$

Definition 1.7 *Let* $\varepsilon \in \mathbb{R}$. *We set*

$$L^{1,\varepsilon}(\Omega, w) = \{f \in L^1(\Omega, w) :$$
$$\|f\|_{1,\varepsilon} = \sup_{\substack{x\in\Omega \\ 0<r<2R}} \frac{r^{2-\varepsilon}}{w(B(x,r))} \int_{\{y\in\Omega : |x-y|<r\}} |f(y)| w(y)dy < +\infty\}.$$

Remark 1.8 We note that in the nondegenerate case, i.e. $w = 1$, $M_\sigma(\Omega, w)$ and $L^{1,\varepsilon}(\Omega, w)$ coincide with the classical Morrey space $L^{1,\lambda}$ for some opportune λ; in particular for $\sigma = \varepsilon > 0$ we obtain $M_\sigma(\Omega, 1) \equiv L^{1,\sigma}(\Omega, 1) \equiv L^{1,n-2+\sigma}$.

It is also interesting to note that if $2 < \varepsilon$ then $L^{1,\varepsilon}(\Omega, w) = \{0\}$. If $\varepsilon < 2 - 2n$ then $L^{1,\varepsilon}(\Omega, w) = L^1(\Omega, w)$.

We wish now to compare the spaces introduced above.

Proposition 1.9. (see [22] *We have*[1]

i) $M_\sigma(\Omega, w) \subseteq S(\Omega, w)$;

ii) $M_\sigma(\Omega, w) \subseteq L^{1,\sigma}(\Omega, w), \sigma > 0$;

[1] With K we denote the so called *reversed doubling* constant. We recall that a weight $w \in A_2$ satisfies the following property:

$\exists 0 < K < 1 : w(B(x,r)) \le Kw(B(x,2r))$

(see [22]).

iii) $L^{1,\varepsilon}(\Omega,w) \subseteq M_{\varepsilon}(\Omega,w)$, $\varepsilon > 0$, *if* $K < \frac{1}{4}$.

In [22] we prove the following result

Theorem 1.10. *Let* u *be a local weak solution of (1.4) in* Ω. *If* $\frac{c}{w} \in S(\Omega,w)$, *then there exist positive numbers* α, r_0 *and C, independent of* u, *such that for any ball* $B(x_0,r)$, *with* $B(x_0,16r) \subseteq \Omega$, $0 < r \leq \frac{r_0}{8}$ *and any* $x \in B(x_0,r)$ *we have*

$$|u(x)-u(x_0)| \leq C\left(\sup_{B(x_0,4r)} |u|\right)\left[|x-x_0|^{\frac{\alpha}{2}} r^{-\frac{\alpha}{2}}\eta(2r) + \right.$$

$$\left. + |x-x_0|^{\alpha} r^{-\alpha} + \eta(r^{\frac{1}{2}}|x-x_0|^{\frac{1}{2}} + |x-x_0|)\right].$$

By the inclusion $M_{\sigma}(\Omega,w) \subseteq S(\Omega,w)$ and Theorem 1.10 we obtain the following hölder-continuity result for the local solutions of equation (1.4) that extends to the degenerate case the analogous result contained in [7] and [20].

Theorem 1.11 *Let* u *be a local weak solution of (1.4) in* Ω. *If* $\frac{c}{w} \in M_{\sigma}(\Omega,w)$ *then* u *is locally hölder-continuous in* Ω.

Finally we wish to stress that the space $L^{1,\varepsilon}(w)$ in turn gives some interesting necessary conditions for hölder-continuity of solutions of (1.4). In fact it holds the following result

Theorem 1.12 (see [22]) *Let* $c \leq 0$, $c \in L^1(\Omega)$ *and let* $u \in C^{0,\alpha}(\Omega)$, $\alpha \in]0,1[$, $0 < l < u$, *be a local weak solution of equation (1.4). Then* $\frac{c}{w} \in L^{1,\alpha}(\Omega,w)$.

We wish to point out that in general the spaces $M_{\sigma}(\Omega,w)$ and $L^{1,\alpha}(\Omega,w)$ are different even if they are the same in many non trivial situations.

2. UNIQUENESS AND EXISTENCE RESULTS FOR DEGENERATE ELLIPTIC OPERATORS

In this section we provide some results concerning a variational inequality associated to a degenerate elliptic operator. These results are contained in the note [23].

Let Ω be a bounded open set in $\mathbb{R}^n$. We consider the linear differential operator

$$Lu = -\sum_{i,j=1}^{n} \left(a_{ij}(x) u_{x_i} + d_j u \right)_{x_j} + \sum_{i=1}^{n} b_i u_{x_i} + cu$$

where a_{ij}, d_j, b_i and c $(i, j = 1, ..., n)$ are measurable functions such that

$$a_{ij} = a_{ji} \tag{2.1}$$

$$\exists \nu > 0 : \nu w |\xi|^2 \le a_{ij}\xi_i\xi_j \le \frac{1}{\nu} w |\xi|^2 \ a.e. \ in \ \Omega, \forall \xi \in \mathbb{R}^n, w \in A_2 \tag{2.2}$$

$$\left(\frac{d_i}{w}\right)^2, \left(\frac{b_i}{w}\right)^2 \in S(\Omega, w), \quad \frac{c}{w} \in S(\Omega, w) \tag{2.3}$$

$$\int_\Omega \left(d_i \varphi_{x_i} + c\varphi \right) dx \ge 0 \quad \forall \varphi \in C_0^\infty(\Omega), \varphi \ge 0 \, in \, \Omega \tag{2.4}$$

Under assumptions (2.1), (2.2) and (2.3) it is possible to prove that the bilinear form

$$a(u, v) = < Lu, v >$$

is continuous in $H_0^{1,2}(\Omega, w) \times H_0^{1,2}(\Omega, w)$.

Given $\psi \in H_0^{1,2}(\Omega, w)$, $\psi \le 0$ in $\partial\Omega$ and $T \in H^{-1,2}(\Omega, w)$, in the convex

$$\mathbb{K} = \left\{ v \in H_0^{1,2}(\Omega, w) : v \ge \psi \ \ a.e. \ in \ \Omega \right\}$$

let us consider the following problem

$$u \in \mathbb{K} : a(u, v - u) \ge < T, v - u > \quad \forall v \in \mathbb{K}. \tag{2.5}$$

Under L^p assumptions on the coefficients, variational inequality (2.5) was studied in [14] and [4],

In [14] the weight w giving the degeneration was assumed under convenient hypothesis such as Murthy-Stampacchia (see [17]). In [4] instead, $w \in A_2$.

Remark 2.1 We can observe that if we consider $w(x) = |x|^\alpha$, the results obtained in [4] hold if $\alpha \in]-n, n[$, $n \geq 2$, instead, it is worth to note that in [14] the correspondent results hold if $\alpha \in [0, 2[$.

Our existence and uniqueness result needed various basic properties concerning a subspace of $S(\Omega, w)$ defined as follows

Definition 2.2 *We say that* $f \in S(\Omega, w)$ *belongs to* $S'(\Omega, w)$ *if*

$$\eta(f) \equiv \sup_{r>0} \eta(f, r) < +\infty.$$

We wish to recall some basic properties for functions belonging to S', which will play a crucial role in our proofs.

Lemma 2.3 *Let* $\frac{f}{w} \in S'(\Omega, w)$. *Then* $\forall \varepsilon > 0$, $\exists \sigma > 0$ $: E \subseteq \Omega$, $|E| < \sigma$ $\Rightarrow$ $\eta(f\chi_E) < \varepsilon$.

Lemma 2.4 *Let* $\frac{f}{w} \in S'(\Omega, w)$ *and* $\varepsilon > 0$. *Then there exist two functions* f_1 *and* f_2 *such that* $f = f_1 + f_2$, $f_2 \in L^\infty(\Omega)$, $\eta(f_1) < \varepsilon$.

Lemma 2.5. *Let* $\frac{f}{w} \in S'(\Omega, w)$, *then* $\forall \varepsilon > 0, \exists c_\varepsilon > 0$, *depending on* $\varepsilon, \Omega, w, n, c$ *such that*

$$\int_\Omega u^2 |c| dx \leq \varepsilon \int_\Omega |\nabla u|^2 w dx + c_\varepsilon \int_\Omega u^2 w dx$$
$$\forall u \in H_0^{1,2}(\Omega, w).$$

Using the previous properties we get the following

Theorem 2.6 *Let* $u \in \mathbb{K}$ *be a solution to problem* (2.5). *Then*

$$|Du|_{2,w} \leq c \left| \|T\|_{-1,w} + \left(\|\psi\|_{2,w}^2 + \|D\psi\|_{2,w}^2 \right)^{\frac{1}{2}} \right|$$

where c *is a positive constant depending on* n, ν *and* $\sum_{i=1}^{n} \frac{|b_i - d_i|^2}{w^2}$.

The uniqueness of solution of variational inequality (2.5) is proved by the following

Theorem 2.7 *Let* $u \in \mathbb{K}$ *be a solution to problem* (2.5). *Let* $u_1 \in H_0^{1,2}(\Omega, w)$, $u_1 \geq \psi$ *a.e. in* Ω, $a(u,\phi) \geq < T, \phi > \forall \phi \in H_0^{1,2}(\Omega, w), \phi \geq 0$. *Then* $u \leq u_1$ *a.e. in* Ω.

Our existence result follows from the previous estimate and the following compactness embedding theorem

Theorem 2.8 *Let* $w \in A_2$. *There exist a constant* C_Ω, *depending on* u, *the* A_2 *constant of* w, *and* $\varepsilon > \frac{1}{2}$ *such that for some* $u \in H_0^{1,2}(\Omega, w)$ *and* $1 \leq k \leq \frac{n}{n-2\varepsilon} = n'$ *we have*

$$|u|_{2k,w} \leq \|u\|_{1,w}.$$

For $1 \leq k < n'$ *the embedding of* $H_0^{1,2}(\Omega, w)$ *in* $L^{2k}(\Omega, w)$ *is compact.*

Precisely we obtain

Theorem 2.9 *Under assumptions (2.1), (2.2), (2.3) and (2.4) there exists the solution of variational inequality* (2.5).

Proof Using Lemmas 2.3,2.4 and 2.5 we have that there exists a positive constant μ, depending on the previous arguments, such that

$$a(v,v) + \mu \mid v \mid_{2,w}^2 \geq \frac{\nu}{2} \|v\|_{H_0^{1,2}(\Omega,w)}^2, \qquad \forall v \in H_0^{1,2}(\Omega, w).$$

Then (see e.g.[13]) $\forall \phi \in L^2(\Omega, w)$ the problem

$$u \in \mathbb{K} : a(u, v-u) + \mu \int_\Omega u(v-u)w\,dx \geq < T, v-u > + \mu \int_\Omega \phi(v-u)w\,dx \quad \forall u \in \mathbb{K}$$

have unique solution $u = S(\varnothing)$

In this way we have defined the operator

$$S : L^2(\Omega, w) \longrightarrow L^2(\Omega, w).$$

that results continuous and compact. Moreover, if we consider $\alpha \in [0,1]$ and ϕ such that $\phi = \alpha S(\phi)$, assuming $u = S(\phi)$, we get

$$a(u, v-u) + \mu(1-\alpha)\int_\Omega u(v-u)w\,dx \geq < T, v-u > \quad \forall v \in \mathbb{K}.$$

From the apriori estimate we can estimate $\|S(\phi)\|_{1,w}$ and $|\phi|_{2,w}$ with constants independent from ϕ and α. Then the desired conclusion follows from Leray-Schauder theorem. □

REFERENCES

[1] M. Aizenman and B. Simon *Brownian motion and Harnack inequality for Schrödinger operators* Comm. Pure Appl. Math. 35 1982 209–273

[2] F. Chiarenza *Regularity for solutions of quasilinear elliptic equations under minimal assumptions* Potential Analysis 4 1995 325–334

[3] F. Chiarenza, E. Fabes and N. Garofalo *Harnack's inequality for Schrödinger operators and continuity of solutions* Proc. A.M.S. 98 1986 415–425

[4] F. Chiarenza, M. Frasca *Una disequazione variazionale associata a un operatore ellittico con degenerazione di tipo* A_2 Le Matematiche 37 1982 239–250

[5] R. Coifman and C. Fefferman *Weighted norm inequalities for maximal functions and singular integrals* Studia Math. 51 1974 241–250

[6] E. De Giorgi *Sulla differenziabilitá e l'analicitá delle estremali degli integrali multipli regolari* Mem. Accad. Sci. Torino Cl. Sci. Fis. Mat. Nat. 3 1957 25–43

[7] G. Di Fazio *Hölder continuity of solutions for some Schrödinger equations* Rend. Sem. Mat. Univ. Padova 79 1988 173–183

[8] E. Fabes, C. Kenig and R. Serapioni *The local regularity of solutions of degenerate elliptic equations* Comm. P.D.E. 7 1982 77–116

[9] J. Garcia Cuerva and J.L. Rubio De Francia *Weighted norm inequalities and related topics* (North-Holland, Amsterdam, 1985)

[10] C. Gutierrez *Harnack's inequality for degenerate Schrödinger operators* Trans. A.M.S. 312 1989 403–419

[11] O. Ladyzhenskaya and N. Ural'tseva *Linear and quasilinear elliptic equations* (Accad. Press 1968)

[12] H. LEWY and G. Stampacchia *On the smoothness of superharmonics which solve a minimum problem* J. Analyse Math. 23 1970 227–236

[13] J.-L. Lions and G. Stampacchia *Variational inequalities* Comm. Pure Appl. Math. 20 1967 493–519

[14] M.E. Marina *Una diseguaglianza variazionale associata a un operatore ellittico che può degenerare e con condizioni al contorno di tipo misto* Rend. Sem. Mat. Padova 54 1975 107–121

[15] C.B. Morrey *Multiple integrals in the calculus of variations* (Springer Verlag 1966)

[16] B. Muckenoupt *Weigthed inequalities for the Hardy maximal functions* Trans. A.M.S. 165 1972 207–226

[17] K.V. Murthy and G. Stampacchia *Boundary value problems for some degenerate elliptic operators* Ann. Mat. Pure Appl. 80 1968 1–122

[18] L. Piccinini *Inclusioni tra spazi di Morrey* Boll. Un.Mat. It. 2 1969 95–99
[19] J. Serrin *Local behavior of solutions of quasilinear equations* Acta Math.111 1964 247–302
[20] C. Simader *An elementary proof of Harnack's inequality for Schrödinger operators and related topics* Math. Z. 203 1990 129–152
[21] G. Stampacchia *Le probleme de Dirichlet pour les equations elliptiques du second ordre a coefficients discontinus* Ann. Inst. Fourier Grenoble 15 1965 198–258
[22] C. Vitanza and P. Zamboni *Necessary and sufficient conditions for Hölder continuity of solutions of degenerate Schrödinger operators* Le Matematiche 52 1997 393–409
[23] C. Vitanza and P. Zamboni *A variational inequality for a degenerate elliptic operator under minimal assumptions on the coefficients* (preprint)

VECTOR VARIATIONAL INEQUALITIES AND DYNAMIC TRAFFIC EQUILIBRIA[1]

X.Q. Yang[1] and H. Yu[2]
Dept. of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong;[1] School of Economics and Management, Tsinghua University, Beijing, China[2]

1. INTRODUCTION

Variational inequality problems were first investigated in the study of elliptic problems and obstacle problems etc. The pioneer work was summarized in the book by Kinderlehrer and Stampacchia [10]. One important feature of variational inequalities is that many practical problems are firstly formulated as variational inequalities, and only under further conditions, they are formulated as optimization problems.

The vector variational inequality (VVI, in short) problem as a generalization of scalar variational inequalities was firstly introduced by Giannessi [6]. This problem has received extensive attentions in the last two decades. Many important results of various kinds of vector variational inequalities have been established, such as existence of a solution, relations with vector optimization, gap functions, stability, characterizations of solution sets, duality theory and applications. The research of vector variational inequalities has been advanced by the recent book Giannessi [8].

[1] This research is supported by the Research Grants Council of Hong Kong (Poly U 5141/01 E).

As a generalization of Wardrop's principle [12], the multiple criteria (or vector) Wardrop's principle is formulated as: the traffic flow along a path joining an origin node and a destination node in a road network is greater than zero only if the resulting multiple criteria cost is efficient amongst all the paths that join the pair of nodes, see [2,13]. However, the corresponding vector variational inequality problem is of finite dimension as the path vector as the variable for the problem is of finite dimension.

This paper will review some recent results on the existence of a solution of VVI and relations between a solution of VVI and that of a vector optimization problem. A new gap function for VVI will be introduced. To illustrate the application of VVI in infinite dimensional spaces, a vector dynamic traffic equilibrium principle is introduced. As a result, a new VVI is formulated. Finally the existence of a vector dynamic traffic equilibrium flow is obtained. It is worth noting that scalar dynamic traffic equilibrium problems have been investigated in [4] and [5].

2. EXISTENCE OF A SOLUTION OF VVI

Let X be a Banach space and (Y, C) an ordered Banach space with the orderings defined by the closed and convex cone C as follows:

$$y_1 \leq_C y_2 \Longleftrightarrow y_2 - y_1 \in C.$$

$$y_1 \not\leq_{C\setminus\{0\}} y_2 \Longleftrightarrow y_2 - y_1 \notin C \setminus \{0\}.$$

Assume furthermore that $intC \neq \emptyset$. The weak orderings in Y are defined by

$$y_1 \not\leq_{intC} y_2 \Longleftrightarrow y_2 - y_1 \notin intC.$$

$$y_1 \leq_{intC} y_2 \Longleftrightarrow y_2 - y_1 \in intC.$$

It is worth nothing that the partial ordering $\not\leq_{intC}$ is closed in the sense that if

$$y_n \to y \text{ as } n \to \infty, \quad y_n \not\leq_{intC} 0,$$

then $y \not\leq_{intC} 0$. But the partial ordering $\not\leq_{C\setminus\{0\}}$ is not closed.

Let $L(X,Y)$ be the set of all linear and bounded operators from X to Y. The value of $l \in L(X,Y)$ at $x \in X$ is denoted by (l,x).

Let $K \subset X$ be a closed and convex set and $T : X \to L(X,Y)$.

The Weak Vector Variational Inequality problem (WVVI, in short) is defined as: finding $x \in K$ such that

$$(T(x), y-x) \not\leq_{intC} 0, \quad \forall y \in K. \tag{1}$$

The Vector Variational Inequality problem (VVI, in short) is defined as: finding $x \in K$ such that

$$(T(x), y-x) \not\leq_{C\setminus\{0\}} 0, \quad \forall y \in K. \tag{2}$$

Definition 2.1 *Let X be a Banach space, and (Y,C) an ordered Banach space with $intC^* \neq \emptyset$. Let K be a nonempty unbounded, closed and convex subset of X and $T : X \to L(X,Y)$.*

(i) T is said to be weakly coercive on K if there exist $x_0 \in K$ and an $s \in intC^*$ such that

$$(s \circ T(x) - s \circ T(x_0), x - x_0)/\|x - x_0\| \to +\infty$$

whenever $x \in K$, and $\| x \| \to \infty$, see [3].

(ii) T is said to satisfy the v-coercive condition if there exist a weakly compact subset $B \subset X$ and $y_0 \in B \cap K$, such that $(T(x), y_0 - x) \leq_{intC} 0, \quad \forall x \in K \setminus B$.

(iii) T is said to satisfy the s-coercive condition if there exist an $s \in intC^*$, a weakly compact subset $B \subset X$ and $z_0 \in B \cap K$, such that $(s \circ T(x), z_0 - x) < 0, \forall x \in K \setminus B$.

(iv) T is said to be monotone if

$$(T(x) - T(y), x - y) \geq_C 0, \quad \forall x, y \in K.$$

Remark 2.1 *In Definition 2.1, weak coercivity of $T \Rightarrow$ v-coercivity of $T \Rightarrow$ s-coercive. If $Y = \mathbb{R}$, so $intC^* = \{r \in \mathbb{R} \mid r > 0\}$, weak coercivity of $T \Rightarrow$ s-coercivity of $T \iff$ v-coercivity of T.*

Definition 2.2 *Let X and Y be Banach spaces and $T : X \to L(X,Y)$. T is said to be v-hemicontinuous if one of the following two conditions is satisfied:*

(i) for every $x, y \in X$, *the map* $t \to (T(x+ty), y)$ *is continuous at* 0^+;

(ii) for every $x, y, z \in X$, *the map* $t \to (T(x+t(y-x)), z)$ *is continuous at* 0^+.

Remark 2.2 *In Definition 2.2, condition (ii) is stronger than (i), i.e.,(ii)* $\Rightarrow$ *(i).*

Theorem 2.1 *[3] Let* X *be a reflexive Banach space,* (Y, C) *an ordered Banach space with* $intC \neq \emptyset$. *Let* K *be a nonempty closed and convex subset of* X, *and let* $T : X \to L(X, Y)$ *be a monotone and* v*-hemicontinuous map on* X. *Assume that*

(i) K *is bounded, or*

(ii) $intC^* \neq \emptyset$ *and* T *is weakly coercive on* K.

Then WVVI(1) is solvable.

This result is proved as follows: Define the closed and convex set

$$K(y) = \{x \in K : (T(x), y - x) \not\leq_{intC} 0\}, \quad y \in K.$$

Since the partial ordering $\not\leq_{intC}$ is closed, $K(y)$ is a closed set. Every $x \in K$ satisfying

$$x \in \cap_{y \in K} K(y) \neq \emptyset$$

is a solution of the problem WVVI(1). This nonemptiness is established by applying the Knaster-Kuratowski-Mazurkiewicz Theorem and a Minty linearization lemma.

Theorem 2.2 *[3] Let* X *be a reflexive Banach space,* (Y, C) *an ordered Banach space with* $intC^* \neq \emptyset$. *Let* K *be a nonempty, bounded, closed and convex subset of* X, *and let* $T : X \to L(X, Y)$ *be a continuous map on* X. *Then WVVI(1) is solvable.*

The existence of a solution of a vector variational inequality problem with a set-valued mapping can be established via that of a vector variational inequality problem with a single-valued mapping.

Let $\mathcal{T} : K \Rightarrow L(X, Y)$ be a set-valued mapping. $T : K \to L(X, Y)$ is said to be a selection of $\mathcal{T}$ if

$$T(x) \in \mathcal{T}(x), \quad x \in K.$$

Consider the problem of finding $x \in K$ such that there is a $\overline{t} \in \mathcal{T}(x)$ satisfying

$$(\overline{t}, y - x) \not\leq_{intC} 0, \forall y \in K, \tag{3}$$

and the problem of finding $x \in K$ such that

$$(T(x), y - x) \not\leq_{intC} 0, \forall y \in K. \tag{4}$$

Lemma 2.1 *[13] Every solution of (4) is a solution of (3).*

This lemma allows us to derive existence results of a solution of (3) by that of (4) as long as certain selection of $\mathcal{T}$ exists.

3. RELATIONS WITH VECTOR OPTIMIZATION

Consider the following vector optimization problem:

$$\text{Min}_{intC} f(x), \quad \text{subject to } x \in K, \tag{5}$$

where Min_{intC} means that $x \in K$ is a solution iff

$$f(y) \not\leq_{intC} f(x), \quad \forall y \in K.$$

Theorem 3.1 *[3] Let f be continuously Gâteaux differentiable on an open set containing K and $T(x) = \nabla f(x)$. Then x is a solution of WVVI(1) if and only if x is a solution of (5).*

Consider the following vector optimization problem:

$$\text{Min}_{C\setminus\{0\}} f(x), \quad \text{subject to } x \in K, \tag{6}$$

where $\text{Min}_{C\setminus\{0\}}$ means that $x \in K$ is a solution iff

$$f(y) \not\leq_{C\setminus\{0\}} f(x), \quad \forall y \in K.$$

Theorem 3.2 *[3] Let f be continuously Gâteaux differentiable on an open set containing K and $T(x) = \nabla f(x)$. Then x is a solution of VVI(2) only if x is a solution of (6).*

The following example shows that a solution of (6) is not necessarily a solution of VVI(2).

Example 3.1 *[7] Let* $K = [-1, 0],$, $f(x) = (f_1(x), f_2(x)) = (x, x^2)$ *and* $T(x) = (1, 2x)$. *Then every* $x \in K$ *is a solution of (6), but* $x = 0$ *is not a solution of VVI as, for* $y = -1$,

$$(f_1'(x)(y-x), f_2'(x)(y-x))^{\top} = [-1\ 0]^{\top} \leq_{\mathbb{R}^2_+ \setminus \{0\}} [0\ 0]^{\top}.$$

The following result provides a necessary and sufficient condition between a solution of (6) and that of a Minty VVI. Consider the Minty VVI:

$$T(y)(y-x) \not\leq_{\mathbb{R}^{\ell}_+ \setminus \{0\}} 0, \quad \forall y \in K, \tag{7}$$

where $T : \mathbb{R}^n \to \mathbb{R}^{\ell \times n}$.

Theorem 3.3 *[7] Let* $K \subset \mathbb{R}^n$ *be a closed and convex set with nonempty interior,* $f : \mathbb{R}^n \to \mathbb{R}^{\ell}$ *be differentiable on an open set containing* K, $T(x) = \nabla f(x)$ *and* $C = \mathbb{R}^{\ell}_+$. *Then* x *is a solution of Minty VVI(7) if and only if* x *is a solution of (6).*

4. GAP FUNCTION APPROACH

The gap function was first introduced in Auslender [1]. Let $T = [T_1, \cdots, T_{\ell}]^{\top} : \mathbb{R}^n \to \mathbb{R}^{\ell \times n}$ and $K \subset \mathbb{R}^n$ be a closed and convex set. Consider the following WVVI of finding $x \in K$ such that

$$T(x)(y-x) \not\leq_{int \mathbb{R}^{\ell}_+} 0, \quad \forall y \in K. \tag{8}$$

Let $y \in K$. Note that,

$$T(x)(y-x) \not\leq_{int \mathbb{R}^{\ell}_+} 0$$

if and only if there is an index i such that

$$T_i(x)(y-x) \geq 0$$

if and only if

$$\max_{1\le i\le \ell} T_i(x)(y-x) \ge 0.$$

So

$$\min_{y\in K} \max_{1\le i\le \ell} T_i(x)(y-x) \ge 0.$$

It is clear that

$$\min_{y\in K} \max_{1\le i\le \ell} T_i(x)(y-x) \le 0.$$

A function G is called a gap function for WVVI (8) on the set K if (i) $G(x) \le 0, \forall x \in K$ and (ii) $G(x) = 0$ if and only if x is a solution of WVVI (8).

Define the following gap function

$$G(x) = \min_{y\in K} \max_{1\le i\le \ell} T_i(x)(y-x)$$

Thus $G(x) \le 0, \quad \forall x \in K$. So we have

Theorem 4.1 *$x \in K$ is a solution of the WVVI(8) if and only if $G(x) = 0$.*

Now we study the gap function properties for the following VVI of finding $x \in K$ such that

$$T(x)(y-x) \not\le_{\mathbb{R}^\ell_+\setminus\{0\}} 0, \quad \forall y \in K. \tag{9}$$

Logically, we can prove that if $G(x) > 0$, then x is a solution of VVI (9). But, by definition, $G(x) \le 0, \ \forall x \in K$. So we need to modify the definition of $G(x)$ as follows:

$$G_1(x) = \min_{y\in K, y\ne x} \max_{1\le i\le \ell} T_i(x)(y-x).$$

Then we have

Theorem 4.2 *$x \in K$ is a solution of VVI(9) if $G_1(x) > 0$.*

But this is only a sufficient condition as shown by the following example.

Example 4.1. *$x \in K$ is a solution of the VVI(9), but $G_1(x) = 0$.*

Let $K=[-1,0]$, $T(x)=(T_1(x),T_2(x))=(1,2x)$. *Then* $x=-1\in K$ *is a solution of (VVI): for any* $y\in K$,

$$(T_1(x)(y-x),T_2(x)(y-x))^{\top}=(y+1,-2(y+1))^{\top}\not\leq_{\mathbb{R}^2_+\setminus\{0\}}[00]^{\top}.$$

But

$$\begin{aligned}G_1(-1)&=\min_{y\in[-1,0],y\neq -1}\max\{(y+1),-2(y+1)\}\\&=\min_{y\in[-1,0],y\neq -1}y+1\\&=0.\end{aligned}$$

In the same way, we can define a gap function for the Minty VVI (7):

$$G_2(x)=\min_{y\in K,y\neq x}\max_{1\leq i\leq \ell}T_i(y)(y-x).$$

Theorem 4.3 $x\in K$ *is a solution of Minty VVI (7) if* $G_2(x)>0$.

5. APPLICATION: DYNAMIC TRAFFIC EQUILIBRIUM

This section formulates dynamic vector equilibrium principles as an infinite dimensional VVI.

5.1 Notation

Let $\mathcal{G}=(\mathcal{N},\mathcal{A})$ be a directed graph, I denote the set of given origin-destination (O-D) pairs in $\mathcal{G}$ and, $P_i(i\in\mathcal{I})$ denote the set of available paths joining O-D pair i.

Let $\Omega=[0,t_f]$ be the time period under consideration. For $i\in\mathcal{I}$ and a given path $k\in P_i$, let $h_k(t)$ denote the traffic flow on this path at time $t\in\Omega$ and $M=\sum_{i\in I}|P_i|$. Then, at time t,

$$h(t)=[h_k(t):k\in P_i,i\in\mathcal{I}]$$

is a M-dimensional column vector.

For technical reasons, we only take account of the functional setting for the set of flow trajectories. This set is assumed to be a reflexive Banach space $L^p(\Omega,\mathbb{R}^M)$ with $p>1$. The dual space of $L^p(\Omega,\mathbb{R}^M)$ is $L^q(\Omega,\mathbb{R}^M)$,

where $1/p + 1/q = 1$. On $L^q(\Omega, \mathbb{R}^M) \times L^p(\Omega, \mathbb{R}^M)$, we define the canonical bilinear form by

$$\langle G, h \rangle = \int_\Omega G(t)h(t)dt, \quad G \in L^q(\Omega, \mathbb{R}^M), h \in L^p(\Omega, \mathbb{R}^M)$$

For $i \in \mathcal{I}$, the demand $d_i(t) \geq 0$ on this O-D pair i depends on the time $t \in \Omega$. At time $t \in \Omega$, let

$$d(t) = [d_i(t) : i \in I].$$

Also, for technical reasons, we think of the demand trajectories in $L^p(\Omega, \mathbb{R}^{|\mathcal{I}|})$.

A flow trajectory, for convenience, a flow $h \in L^p(\Omega, \mathbb{R}^M)$ satisfying the demand, is called a feasible flow. Let $\mathcal{H}$ be the set of feasible path flows, i.e.,

$$\mathcal{H} = \{h \in L^p(\Omega, \mathbb{R}^M) \mid h(t) \geq 0 \text{ and } \sum_{k \in P_i} h_k(t) = d_i(t) \text{ a.e. on } \Omega, \ \forall i \in \mathcal{I}\}.$$

A path flow vector $h(t)$ induces an arc flow column vector $v(t) = [v_a(t)]_{a \in A}$, where, for each arc $a \in \mathcal{A}$,

$$v_a(t) = \sum_{i \in \mathcal{I}} \sum_{k \in P_i} \delta_{ak} h_k(t)$$

where

$$\Delta = [\delta_{ak}] \in \mathbb{R}^{|\mathcal{A}| \times M}$$

is the arc path incidence matrix with $\delta_{ak} = 1$ if arc a belongs to path k and 0 otherwise. Hence

$$v(t) = \Delta h(t).$$

Let $\mathcal{V}$ be the set of feasible arc flows, i.e.,

$$V = \{v \in L^p(\Omega, \mathbb{R}^{|\mathcal{A}|}) \mid v(t) \geq 0, \ v(t) = \Delta h(t) \text{ and } \sum_{k \in P_i} h_k(t) = d_i(t) \text{ a.e. on } \Omega, \ \forall i \in \mathcal{I}\}.$$

Let $(\mathbb{R}^\ell, \mathbb{R}^\ell_+)$ be an ordered space with the ordering cone $\mathbb{R}^\ell_+$ and, for each t, $c_a(v(t)) \in \mathbb{R}^\ell$ be a vector cost functional on arc a (arc weight); let $c(v(t)) = [c_a(v(t)) : a \in \mathcal{A}]$ be a $l \times |\mathcal{A}|$-matrix. The vector weight along a path $k \in P_i$ is assumed to be the sum of all the arc weights along this path; thus

$$\tau_k(h(t)) = \sum_{a \in \mathcal{A}} \delta_{ak} c_a(v(t)) \in \mathbb{R}^\ell.$$

Set

$$T(h(t)) = c(v(t))\Delta$$

which is an $l \times M$ matrix with columns given by $\tau_k(h(t))$.

So we know that, for each $h \in \mathcal{H}$, $T(h(\cdot))$ is a functional from Ω to $\mathbb{R}^{l \times M}$. We assume that, for all $h \in H$, $T(h(\cdot))$ is in $L^q(\Omega, \mathbb{R}^{l \times M})$ where $1/p + 1/q = 1$. Define a multi-cost path functional $U : L^p(\Omega, \mathbb{R}^M) \to L(L^p(\Omega, \mathbb{R}^M), \mathbb{R}^\ell)$ by

$$\left\langle U(h), \overline{h} \right\rangle = \int_\Omega T(h(t))\overline{h}(t)dt, \quad h, \overline{h} \in L^p(\Omega, \mathbb{R}^M).$$

And define a multi-cost arc functional $S : L^p(\Omega, \mathbb{R}^{|\mathcal{A}|}) \to L(L^p(\Omega, \mathbb{R}^{|\mathcal{A}|}), \mathbb{R}^\ell)$ by

$$\left\langle S(v), \overline{v} \right\rangle = \int_\Omega \sum_{a \in \mathcal{A}} c_a(v(t))\overline{v}_a(t)dt, \quad v, \overline{v} \in L^p(\Omega, \mathbb{R}^{|\mathcal{A}|}).$$

Assumption 5.1 *T is one-to-one, that is, if $h_1, h_2 \in \mathcal{H}$ and $T(h_1) = T(h_2)$, then $h_1(t) = h_2(t)$ a.e. on Ω.*

Note: It can be shown that if S is one-to-one and Δ is a square and nonsingular matrix, then the assumption 5.1 holds.

Proposition 5.1 *If the assumption 5.1 holds, then the multi-cost path functional U is one-to-one on $\mathcal{H}$.*

Proof. Proving U is one-to-one on $\mathcal{H}$ is equivalent to show that if, for $h_1, h_2 \in \mathcal{H}$ and $h_1(t) \neq h_2(t)$ *a.e. on* Ω, then $U(h_1) \neq U(h_2)$. Suppose $U(h_1) = U(h_2)$. Then, from the definition of functional U, we have

$$\int_{\Omega} (T(h_1(t)) - T(h_2(t)))\overline{h}(t)dt = 0, \quad \forall \overline{h} \in L^p(\Omega, \mathbb{R}^M).$$

From the Hahn-Banach theorem,

$$T(h_1(t)) - T(h_2(t)) = 0, \quad a.e.\ on\ \Omega.$$

From the assumption 5.1,

$$h_1(t) - h_2(t) = 0, \quad a.e.\ on\ \Omega.$$

5.2 Vector dynamic traffic equilibria

Definition 5.1 *Given an* $h \in \mathcal{H}$, *we say that a path* $k \in P_i$ *for an O-D pair* i *is efficient if there does not exist another path* $k' \in P_i$ *such that* $\tau_k(h(t)) - \tau_{k'}(h(t)) \in \mathbb{R}^{\ell}_{+} \setminus \{0\}, a.e.\ on\ \Omega$.

Given an $h \in \mathcal{H}$, let $\Gamma_i(h) = \{\tau_k(h(t)) : k \in P_i\}$ denote the (discrete) set of vector cost functionals of all paths for O-D pair i, and $\mathcal{I}_i(h) = \{k \in P_i \mid \tau_k(h(t)) - \tau_{k'}(h(t)) \notin \mathbb{R}^{\ell}_{+} \setminus \{0\},\ a.e.\ \text{on}\ \Omega\ \forall k' \in P_i\} \subseteq P_i$ denote the index set of all efficient paths for O-D pair i.

We define the *efficient frontier* for O-D pair i to be the set of efficient points in the cost-space of O-D pair i:

$$\text{Min}_{\mathbb{R}^{\ell}_{+}\setminus\{0\}} \left(\Gamma_i(h)\right) = \{\tau_p(h) \in \mathbb{R}^{\ell} \mid p \in I_i(h)\}.$$

Note *that* $\text{Min}_{\mathbb{R}^{\ell}_{+}\setminus\{0\}} \left(\Gamma_i(h)\right)$ is a discrete set because $\mathcal{I}_i(h)$ is a discrete set.

Definition 5.2 (Dynamic vector equilibrium principle) *A continuous path flow vector* $h \in \mathcal{H}$ *is said to be in dynamic vector equilibrium if,* $\forall i \in I, \forall k, k' \in P_i$,

$$h_k(t) = 0 \text{ whenever } \tau_k(h(t)) - \tau_{k'}(h(t)) \in \mathbb{R}^{\ell}_{+} \setminus \{0\},\ a.e.\ on\ \Omega. \qquad (10)$$

A flow h in dynamic vector equilibrium is often referred to as a dynamic vector equilibrium flow.

Definition 5.3 (Dynamic weak vector equilibrium principle) *A continuous path flow vector* $h(t) \in \mathcal{H}$ *is said to be in dynamic weak vector equilibrium if,* $\forall i \in I, \forall k, k' \in P_i$,

$$h_k(t) = 0 \text{ whenever } \tau_k(h(t)) - \tau_{k'}(h(t)) \in int\mathbb{R}^{\ell}_{+}, \text{ a.e. on } \Omega. \tag{11}$$

A flow h in dynamic weak vector equilibrium is often referred to as a dynamic weak vector equilibrium flow.

Remark 5.1

(i) If $\ell = 1$, *(10) is reduced to the dynamic (scalar) Wardrop's principle in Daniele et al [4].*

(ii) The dynamic vector equilibrium principle can be stated in an equivalent form as: the path flow vector h *is in dynamic vector equilibrium if,* $\forall i \in I, \forall p \in P_i$,

$$h_p(t) = 0 \text{ whenever } \tau_p(h(t)) \notin Min_{\mathbb{R}^{\ell}_{+}\setminus\{0\}}\left(\Gamma_i(h)\right), \text{ a.e. on } \Omega.$$

The following are infinite dimensional versions of the assumptions used in [9].

Assumption 5.2 Let $h \in \mathcal{H}$. *Assume that*

$$\mathrm{Min}_{\mathbb{R}^{\ell}_{+}\setminus\{0\}}(\Gamma_i(h)) \subset \mathrm{Min}_{\mathbb{R}^{\ell}_{+}\setminus\{0\}}(Co(\Gamma_i(h))), \text{ a.e. on } \Omega.$$

Remark 5.2 *Assumption 2 is equivalent to assert that there exists a null set* E_1 *such that, for any* $t \in \Omega \setminus E_1$,

$$\mathrm{Min}_{\mathbb{R}^{\ell}_{+}\setminus\{0\}}(\Gamma_i(h)) \subset \mathrm{Min}_{\mathbb{R}^{\ell}_{+}\setminus\{0\}}(Co(\Gamma_i(h))).$$

Definition 5.4 We say that the vector cost function c_a is conservative if $\partial c^k_a/\partial v_{a'} = \partial c^k_{a'}/\partial v_a$, *a.e. on* Ω, $\forall a' \in \mathcal{A}, \forall \mathrm{k} = 1, \ldots, l$.

Assumption 5.3 *The cost* c_a *is conservative for all* $a \in \mathcal{A}$.

Assumption 5.4 *Each row* $c^k(v(t))$ *of the cost matrix* $c(v(t))$ *is monotone, i.e., for all* $k = 1, 2, \ldots, l, v_1(t), v_2(t) \in \mathbb{R}^{|\mathcal{A}|}$,

$$(c^k(v_1(t)) - c^k(v_2(t)))(v_1(t) - v_2(t)) \geq 0, \text{ a.e. on } \Omega.$$

Remark 5.3 *Assumption 5.4 is to say that there exists a null set* E_2 *such that for any* $t \in \Omega \setminus E_2$, $c^k(v(t))$ *is monotone.*

5.3 Necessary and sufficient conditions of a vector dynamic flow

We need the following definition in the proof of our main result.

Definition 5.5 (Static vector equilibrium principle) *[9] Let* d_i *be the demand for the O-D pair* $i \in \mathcal{I}$ *and*

$$H = \{h \in \mathbb{R}^M \mid h \geq 0 \text{ and } \sum_{p \in P_i} h_p = d_i, \forall i \in \mathcal{I}\}.$$

A flow vector $h \in \mathcal{H}$ *is said to be in vector equilibrium if*

$$\forall i \in \mathcal{I}, \forall k, k' \in P_i, h_k = 0 \text{ whenever } \tau_k(h) - \tau_{k'}(h) \in int\mathbb{R}^{\ell}_{+}.$$

In the following, an infinite dimensional VVI problem is established as a necessary condition of a vector equilibrium flow.

Proposition 5.2 (Necessary condition) *If Assumptions 5.2, 5.3 and 5.4 hold and* h *is in dynamic vector equilibrium, then* h *is a solution of the following WVVI of finding* $h \in \mathcal{H}$ *such that:*

$$\langle U(h), g - h \rangle \not\leq_{int\mathbb{R}^{\ell}_{+}} 0, \quad \forall g \in \mathcal{H}. \tag{12}$$

Proof. Since h is in dynamic vector equilibrium, then there exists a null set E_3 such that, for any $t \in \Omega \setminus E_3$,

$$\forall i \in \mathcal{I}, \forall k, k' \in P_i, h_k(t) = 0 \text{ whenever } \tau_k(h(t)) - \tau_{k'}(h(t)) \in \mathbb{R}^{\ell}_{+} \setminus \{0\},$$

and $h(t) \geq 0$ and $h_i(t) = d_i(t), \forall i \in \mathcal{I}$. That is to say, $h(t)$ is in vector equilibrium in the sense of [2]. So, for any $t \in \Omega \cap (E_1 \cup E_2 \cup E_3)$, $h(t)$ is in vector equilibrium, and all the assumptions of Theorem 3.3 in [9] are satisfied. Hence, $T(h(t))(g(t) - h(t)) \not\leq_{int\mathbb{R}^{\ell}_{+}} 0$ holds, for every $g \in H(t)$, where

$$H(t) = \{g \in \mathbb{R}^M \mid g \geq 0 \text{ and } \sum_{p \in P_i} g_p = d_i(t), \forall i \in \mathcal{I}\}.$$

Since the union of finitely many null sets is a null set, $E_1 \cup E_2 \cup E_3$ is a null set. So,

$$T(h(t))(g(t)-h(t)) \not\leq_{int\mathbb{R}^l_+} 0$$

a.e. on Ω, for any $g \in \mathcal{H}$. That is to say,

$$\langle U(h), g-h\rangle = \int_\Omega T(h(t))(g(t)-h(t))dt \not\leq_{int\mathbb{R}^l_+} 0.$$

Proposition 5.3 (Sufficient condition) The flow $h \in \mathcal{H}$ is in dynamic vector equilibrium if h solves the following VVI of finding $h \in \mathcal{H}$ such that:

$$\langle U(h), \overline{h}-h\rangle \not\leq_{\mathbb{R}^l_+\setminus\{0\}} 0, \quad \forall \overline{h} \in \mathcal{H}. \tag{13}$$

Proof. Let $h \in \mathcal{H}$ satisfy (13), choose $\overline{h} \in L^p(\Omega, \mathbb{R}^M)$ to be such that

$$\overline{h}_j(t) = \begin{cases} h_j(t), & \text{if } j \neq k \text{ or } k' \\ 0, & \text{if } j = k \\ h_k(t) + h_{k'}(t), & \text{if } j = k' \end{cases}$$

a.e. on Ω. Clearly, $\overline{h} \in \mathcal{H}$, since $h(t) \geq 0$ and $\Delta h(t) = d(t)$ a.e. on Ω. Now,

$$\begin{aligned} \langle U(h), \overline{h}-h\rangle &= \int_\Omega T(h(t))(\overline{h}(t)-h(t))dt \\ &= \int_\Omega \sum_{i\in\mathcal{I}} \sum_{j\in P_i} (\overline{h}_j(t)-h_j(t))\tau_j(\overline{h}(t))dt \\ &= \int_\Omega (\overline{h}_k(t)-h_k(t))\tau_k(h(t)) + (\overline{h}_{k'}(t)-h_{k'}(t))\tau_{k'}(h(t))dt \\ &= \int_\Omega h_k(t)(\tau_{k'}(h(t)) - \tau_k(h(t))dt \not\leq_{\mathbb{R}^l_+\setminus\{0\}} 0. \end{aligned} \tag{14}$$

If

$$\tau_k(h(t)) - \tau_{k'}(h(t)) \geq_{\mathbb{R}^l_+ \setminus \{0\}} 0, \text{ a.e. on } \Omega,$$

then (14) implies that $h_k(t) = 0$ a.e. on Ω. Thus h is in vector dynamic equilibrium.

Proposition 5.4 (Sufficient condition) *The flow $h \in \mathcal{H}$ is in dynamic weak vector equilibrium if h solves the WVVI(12)*

Proof: The proof is similar to that Proposition 5.3 and omitted.

5.4 Existence of a vector dynamic traffic flow

In this subsection we apply the results in [3] to establish the existence of a dynamic weak vector equilibrium flow.

Proposition 5.5 *Suppose the multi-cost arc functional S is monotone and v-hemi-continuous, then there exists a path flow $h \in \mathcal{H}$, which is in dynamic weak vector equilibrium.*

Proof: Note that

$$\mathcal{H} = \{h \in L^p(\Omega, \mathbb{R}^M) \mid h(t) \geq 0 \text{ and } \sum_{k \in P_i} h_k(t) = d_i(t) \text{ a.e. on } \Omega, \forall i \in \mathcal{I}\}.$$

It is clear that $\mathcal{H}$ is bounded, convex and closed, i.e. weakly compact.

For any $h, \overline{h} \in \mathcal{H}$, set $v = \Delta h, \overline{v} = \Delta \overline{h}$. Then, from the monotonicity of the multi-cost arc functional S,

$$\begin{aligned}
\left\langle U(h) - U(\overline{h}), h - \overline{h} \right\rangle &= \int_\Omega (T(h(t)) - T(\overline{h}(t)))(h(t) - \overline{h}(t))dt \\
&= \int_\Omega (c(v(t))\Delta - c(\overline{v}(t))\Delta)(h(t) - \overline{h}(t))dt \\
&= \int_\Omega (c(v(t)) - c(\overline{v}(t)))(\Delta h(t) - \Delta \overline{h}(t))dt \\
&= \int_\Omega (c(v(t)) - c(\overline{v}(t)))(v(t) - \overline{v}(t))dt \\
&= \left\langle S(v) - S(\overline{v}), v - \overline{v} \right\rangle \geq_C 0,
\end{aligned}$$

i.e. the multi-cost path functional U is monotone on $\mathcal{H}$. Similarly,

$$\langle S(v+t\overline{v}),\overline{v}\rangle = \langle U(h+t\overline{h}),\overline{h}\rangle.$$

So, from the v-hemi-continuity of S,

$$\lim_{t\to 0^+}\langle U(h+t\overline{h}),\overline{h}\rangle = \lim_{t\to 0^+}\langle S(v+t\overline{v}),\overline{v}\rangle = \langle S(v),\overline{v}\rangle = \langle U(h),\overline{h}\rangle.$$

So from the v-hemi-continuity of S, we have the multi-cost path functional U is v-hemi-continuous. Then from theorem 2.1, the WVVI (12) has one solution $h \in \mathcal{H}$. So by proposition 5.4, h is in dynamic weak vector equilibrium.

Proposition 5.6 *Suppose the multi-cost arc functional S is continuous. Then there exists a path flow $h \in \mathcal{H}$, which is in dynamic weak vector equilibrium.*

Proof. From the continuity of S, U is continuous. By using theorem 2.2, there is $h \in \mathcal{H}$ solving the following WVVI:

$$\langle U(h),\overline{h}-h\rangle \not\leq_{int\mathbb{R}^{\ell}_{+}} 0, \quad \forall \overline{h} \in \mathcal{H}.$$

Then by proposition 5.4, h is in dynamic weak vector equilibrium.

REFERENCES

[1] Auslender, A., Optimization: Numerical Methods, Masson, Paris, France, 1976.

[2] Chen, G.Y. and Yen, N.D., On the variational inequality model for network equilibrium, Internal Report 3.196 (724), Department of Mathematics, University of Pisa, 1993.

[3] Chen, G.Y. and Yang, X.Q., The vector complementary problem and its equivalences with the weak minimal element in ordered spaces. *Journal of Mathematical Analysis and Applications*, Vol. 153, pp. 136-158, 1990.

[4] Daniele, P., Maugeri, A. and Oettli, W, Time-dependent traffic equilibria, *Journal of Optimization Theory and Applications*, Vol. 103, pp. 543-555, 1999.

[5] Daniele, P. and Maugeri, A. Variational inequalities and discrete and continuum models of network equilibrium problems, *Mathematical and Computer Modelling*, Vol. 35, pp. 689-708, 2002.

[6] Giannessi, F., Theorems of the alternative, quadratic programs, and complementarity problems, *Variational Inequalities and Complementarity Problems*, edited by R.W. Cottle, F. Giannessi, and J.L. Lions, Wiley, New York, New York, pp. 151-186, 1980.

[7] Giannessi, F., On Minty variational principle, in "New Trend in Mathematical Programming", Kluwer Academic Publishers, 1997.

[8] Giannessi, F., Vector Variational Inequalities and Vector Equilibria, Kluwer Academic Publisher 2000.

[9] Goh, C.J. and Yang, X.Q, Vector equilibrium problem and vector optimization, *European Journal of Operational Research*, Vol. 116, pp. 615-628, 1999.

[10] Kinderlehrer, D. and Stampacchia, G., An Introduction to Variational Inequalities and Their Applications Academic Press, 1980.

[11] Lee, G.M., Kim, D.S., Lee, B.S. and CHO, S.J., On vector variational inequality, *Bull. Korean Math. Soc.* Vol. 33, pp. 553-564, 1996.

[12] Wardrop, J., Some theoretical aspects of road traffic research. Proceedings of the Institute of Civil Engineers, Part II, Vol. 1, pp. 325-378, 1952.

[13] Yang X.Q. and Goh C.J., On vector variational inequality. Its applications to vector equilibria. *Journal of Optimization Theory and Applications*, Vol. 95, pp. 431-443, 1997.

[14] Yang, X.Q. and Yao, J.C., Gap functions and existence of solutions of set-valued vector variational inequalities. *Journal of Optimization Theory and Applications*, Vol. 115, pp. 407-417, 2002.

A NEW PROOF OF THE MAXIMAL MONOTONICITY OF THE SUM USING THE FITZPATRICK FUNCTION*

C. Zălinescu
Faculty of Mathematics, University "Al. I. Cuza" Iaşi, Iaşi, Rumania

1. INTRODUCTION

A classical result of Rockafellar [7] states that the sum of two maximal monotone multifunctions on a reflexive Banach space is maximal monotone when the interior of the domain of one of them intersects the domain of the other. The original proof of Rockafellar [7] uses some results of Browder [1]; put together, the proof is quite involved. Rockafellar's theorem is the companion of the result stating that the subdifferential of the sum of two lower semicontinuous convex functions on a Banach space is the sum of their subdifferentials when the interior of the domain of one of them intersects the domain of the other. In fact one could observe that the conditions under which these two important results were stated developed in parallel: having a new (more general) condition ensuring the result on the subdifferential of the sum of convex functions in short time a similar condition was used for the maximal monotonicity of the sum. (Note that the conditions imposed for the functions ensure that the conjugate of the sum is the exact convolution of the conjugates; as observed for a long time by Hiriart-Urruty [3], when such a formula holds for the conjugate of the sum,

* The results of this paper were obtained during author's (Spring 2003) stay at University of Pau, France.

the subdifferential of the sum is the sum of the subdifferentials.) However, for deriving the maximal monotonicity of the sum specific methods were used. The natural question is if one could use the results for convex functions in order to deduce those for the maximal monotone multifunctions. Simons in his book [8] uses convex functions associated to monotone multifunctions, but not the result on the subdifferential (or conjugate) of the sum (in fact he uses minimax theorems). It is our aim to give a proof of Rockafellar's sum theorem for monotone multifunctions using a result on the conjugate of the sum. This will be possible using the Fitzpatrick function associated to a monotone multifunction. We also show that several conditions met in the literature are equivalent. In a similar way we obtain a result on the maximal monotonicity of the composition of maximal monotone multifunctions with continuous linear operators. As a by-product of one of the results we furnish another proof for Simons' version of Rockafellar's characterization of maximal monotone multifunctions.

2. NOTATION AND PRELIMINARY RESULTS

We recall first some notation and results related to convex analysis. For this propose, consider a separated locally convex space E and E^* its topological dual; we get so the dual system $(E,E^*,\langle\cdot,\cdot\rangle)$, where $\langle x,x^*\rangle := x^*(x)$ for $x\in E$ and $x^*\in E^*$. We endow E^* with the weak-star topology $w^* := \sigma(E^*,E)$, and so the topological dual of E^* is identified with E. As usual, having a subset A of E, we use the notation $\operatorname{int}A$, $\operatorname{cl}A$ or $\overline{A}$, $\operatorname{co}A$, $\overline{\operatorname{co}}A$ and $\operatorname{aff}A$ for the *interior, closure, convex hull, closed convex hull,* and the *affine hull* of A, respectively; moreover, A^i and ${}^iA$ denote the *core (algebraic interior)* and the *intrinsic core* of A, while ${}^{ic}A$ is ${}^iA$ when $\operatorname{aff}A$ is closed and ${}^{ic}A$ is the empty set otherwise. The *domain*, the *epigraph* and the *conjugate* of $f: E\to\overline{\mathbb{R}}$ are introduced by

$$\operatorname{dom}f := \{x\in X \,|\, f(x)<\infty\},$$
$$\operatorname{epi}f := \{(x,t)\in X\times\mathbb{R} \mid f(x)\le t\}$$

and

$$f^*: E^*\to\overline{\mathbb{R}},\quad f^*(x^*) := \sup\{\langle x,x^*\rangle - f(x) \mid x\in E\},$$

respectively; the function f is *proper* if $\operatorname{dom}f\neq\emptyset$ and f does not take the value $-\infty$. We denote by $\Lambda(E)$ the class of proper convex functions defined on E and by $\Gamma(E)$ the class of those functions in $\Lambda(E)$ which are

lower semicontinuous (lsc for short). We also consider the *convex hull* of $f: E \to \overline{\mathbb{R}}$ as being the function

$$\operatorname{co} f: E \to \overline{\mathbb{R}}, \quad \operatorname{co} f(x) := \inf\left\{t \in \mathbb{R} \;\middle|\; (x,t) \in \operatorname{co}(\operatorname{epi} f)\right\}$$

with the convention $\inf \varnothing := \infty$. We have that

$$\operatorname{co}(\operatorname{epi} f) \subset \operatorname{epi}(\operatorname{co} f) \subset \overline{\operatorname{co}}(\operatorname{epi} f) =: \operatorname{epi}(\overline{\operatorname{co}} f),$$

and so $\overline{\operatorname{co}} f \le \operatorname{co} f \le f$; moreover $(\overline{\operatorname{co}} f)^* = (\operatorname{co} f)^* = f^*$. For the function $g: E^* \to \overline{\mathbb{R}}$ we take always its conjugate with respect to $\left(E, E^*, \langle \cdot, \cdot \rangle\right)$, and so g^* is defined on E. Having $f: E \to \overline{\mathbb{R}}$, it is well-known that $f^{**} := \left(f^*\right)^* = \overline{\operatorname{co}} f$ whenever $\overline{\operatorname{co}} f$ is proper, and $\overline{\operatorname{co}} f$ is proper if and only if $\overline{\operatorname{co}} f$ is finite somewhere. The *indicator function* of $A \subset E$ is $\iota_A: E \to \overline{\mathbb{R}}$ defined by $\iota_A(x) := 0$ for $x \in A$ and $\iota_A(x) := \infty$ for $x \in E \setminus A$. The *convolution* of the functions $f, g: E \to \overline{\mathbb{R}}$ is defined by

$$f \square g: E \to \overline{\mathbb{R}}, \quad (f \square g)(x) := \inf\left\{f(u) + g(x-u) \mid u \in X\right\};$$

the convolution is *exact* when the infimum is attained for every $x \in E$.

From now on, let $\left(X, \|\cdot\|\right)$ be a reflexive Banach space and X^* its topological dual endowed with the dual norm $\|\cdot\|_*$. Recall that the *duality mapping* of X is the multifunction

$$J_X: X \rightrightarrows X^*, \quad J_X(x) := \left\{x^* \in X^* \;\middle|\; \langle x, x^* \rangle = \|x\|^2 = \|x^*\|_*^2\right\}.$$

For notational convenience, the coupling function of the dual system $\left(X, X^*, \langle \cdot, \cdot \rangle\right)$ will be denoted by c ;so

$$c(x, x^*) := \langle x, x^* \rangle := x^*(x) \quad \forall x \in X, \forall x^* \in X^*.$$

The dual space of $X \times X^*$ is identified with $X^* \times X$ by the coupling

$$\left\langle (x, x^*), (u^*, u) \right\rangle := \langle x, u^* \rangle + \langle u, x^* \rangle$$

for $(x, x^*) \in X \times X^*$ and $(u^*, u) \in X^* \times X$.

Consider the multifunction $T: X \rightrightarrows X^*$, whose graph, i.e. $\operatorname{gph} T := \left\{(x, x^*) \mid x^* \in T(x)\right\}$, is nonempty; of course, the *domain* of T is the set $\operatorname{dom} T := \left\{x \in X \;\middle|\; T(x) \neq \varnothing\right\}$. As usual, we say that $T: X \rightrightarrows X^*$ is *monotone* if $\langle x - y, x^* - y^* \rangle \ge 0$ for all $(x, x^*), (y, y^*) \in \operatorname{gph} T$; of

course, T is *maximal monotone* if $T = S$ whenever $S : X \rightrightarrows X^*$ is monotone and $\operatorname{gph} T \subset \operatorname{gph} S$.

With $T : X \rightrightarrows X^*$ we associate the function $c_T := c + \iota_{\operatorname{gph} T}$ and the *Fitzpatrick function* (see [2])

$$f_T : X \times X^* \to \overline{\mathbb{R}}, \quad f_T(x,x^*) := \sup\left\{\langle x,u^*\rangle + \langle u,x^*\rangle - \langle u,u^*\rangle \,\middle|\, (u,u^*) \in \operatorname{gph} T\right\}.$$

Therefore,

$$f_T(x,x^*) = (c_T)^*(x^*,x) = (\overline{\operatorname{co}}\, c_T)^*(x^*,x) \quad \forall (x,x^*) \in X \times X^*; \tag{1}$$

f_T is convex and lower semicontinuous. It is obvious that

$$f_T(x,x^*) \geq \langle x,x^*\rangle \quad \forall (x,x^*) \in \operatorname{gph} T. \tag{2}$$

The next characterization of the monotonicity of T is established by Penot [5, Prop. 3]; we furnish its proof for readers convenience.

Proposition 1. *Let $T : X \rightrightarrows X^*$ have nonempty graph. Then*

$$T \text{ is monotone} \Leftrightarrow f_T \leq c_T \Leftrightarrow c \leq \operatorname{co} c_T. \tag{3}$$

Proof Assume that T is monotone and fix $(x,x^*) \in \operatorname{gph} T$. Because

$$\langle x,x^*\rangle \geq \langle x,u^*\rangle + \langle u,x^*\rangle - \langle u,u^*\rangle \quad \forall (u,u^*) \in \operatorname{gph} T,$$

we obtain that $f_T(x,x^*) \leq \langle x,x^*\rangle$. Therefore $f_T \leq c_T$.

Assume now that $f_T \leq c_T$. Because f_T is convex, $f_T \leq \varphi := \operatorname{co} c_T$. Using (1) we get

$$2\varphi(x,x^*) \geq \varphi(x,x^*) + f_T(x,x^*) = \varphi(x,x^*) + \varphi^*(x^*,x) \geq \langle (x,x^*),(x^*,x)\rangle = 2\langle x,x^*\rangle$$

for every $(x,x^*) \in X \times X^*$. Hence $\operatorname{co} c_T \geq c$.

Assume now that $\varphi := \operatorname{co} c_T \geq c$. Consider $(x,x^*),(y,y^*) \in \operatorname{gph} T$. Then

$$\langle x,x^*\rangle \leq \varphi(x,x^*) = \operatorname{co} c_T(x,x^*) \leq c_T(x,x^*) = \langle x,x^*\rangle.$$

Hence $\varphi(x,x^*)=\langle x,x^*\rangle$ and $\varphi(y,y^*)=\langle y,y^*\rangle$. Because φ is convex, it follows that

$$\begin{aligned}\left\langle \tfrac{1}{2}(x+y),\tfrac{1}{2}(x^*+y^*)\right\rangle &\le \varphi\left(\tfrac{1}{2}(x+y),\tfrac{1}{2}(x^*+y^*)\right)=\varphi\left(\tfrac{1}{2}(x,x^*)+\tfrac{1}{2}(y,y^*)\right)\\ &\le \tfrac{1}{2}\varphi(x,x^*)+\tfrac{1}{2}\varphi(y,y^*)=\tfrac{1}{2}\langle x,x^*\rangle+\tfrac{1}{2}\langle y,y^*\rangle,\end{aligned}$$

whence $\langle x-y,x^*-y^*\rangle\ge 0$. Therefore T is monotone. □

Because c is continuous for the product of the norm topologies on $X\times X^*$, to the characterizations in Proposition 1 we can add the following

$$T \text{ is monotone} \Leftrightarrow c\le \overline{\text{co}}\, c_T. \tag{4}$$

Note that Proposition 1 is true for X a general normed vector space (or even a locally convex space). Taking into account (2), the first equivalence in (3) can be written as follows:

$$T \text{ is monotone} \Leftrightarrow \text{gph } T\subset\left\{(x,x^*)\mid f_T(x,x^*)=\langle x,x^*\rangle\right\}.$$

It follows (see also [2, Th. 3.8, Cor. 3.9]) that

$$T \text{ is maximal monotone} \Leftrightarrow f_T\ge c \text{ and gph } T=\left\{(x,x^*)\mid f_T(x,x^*)=\langle x,x^*\rangle\right\}. \tag{5}$$

It is useful to observe that when $T,S:X\rightrightarrows X^*$ are such that $S(x)=T(x+v)-v^*$ for every $(x,x^*)\in X\times X^*$ and for some $(v,v^*)\in X\times X^*$, one has that

$$\begin{aligned}&\text{gph } S=\text{gph } T-(v,v^*),\ \ \text{dom } S=\text{dom } T-v,\\ &c_S(x,x^*)=c_T(x+v,x^*+v^*)-\langle x,v^*\rangle-\langle v,x^*\rangle-\langle v,v^*\rangle,\\ &\overline{\text{co}}\, c_S(x,x^*)=\overline{\text{co}}\, c_T(x+v,x^*+v^*)-\langle x,v^*\rangle-\langle v,x^*\rangle-\langle v,v^*\rangle\end{aligned}$$

for all $(x,x^*)\in X\times X^*$. In particular

$$\text{dom}(\text{co } c_S)=\text{dom}(\text{co } c_T)-(v,v^*),\text{dom}(\overline{\text{co}}\, c_S)=\text{dom}(\overline{\text{co}}\, c_T)-(v,v^*). \tag{6}$$

3. THE RESULTS

The conclusion of the next result is that of [10, Th. 3.11.4] (which corresponds to some results in Simons [8]), but the hypothesis is different.

Theorem 2. *Assume that* $T_1, T_2 : X \rightrightarrows X^*$ *are monotone multifunctions. If*

$$0 \in {}^{ic}\Pr_X(\mathrm{dom}(\overline{\mathrm{co}}\, c_{T_1}) - \mathrm{dom}(\overline{\mathrm{co}}\, c_{T_2})), \tag{7}$$

then there exist $x \in X$ *and* $x_1^*, x_2^* \in X^*$ *such that*

$$f_{T_1}(x, x_1^*) + f_{T_2}(x, x_2^*) + \tfrac{1}{2}\|x\|^2 + \tfrac{1}{2}\|x_1^* + x_2^*\|_*^2 \le 0.$$

Moreover, if T_1 *and* T_2 *are maximal monotone, then there exist* $x \in X$ *and* $x_1^*, x_2^* \in X^*$ *such that* $x_i^* \in T_i(x)$ *for* $i = 1, 2$, *and*

$$\|x\|^2 + \|x_1^* + x_2^*\|_*^2 + 2\langle x, x_1^* + x_2^*\rangle = 0;$$

in particular, $\mathrm{dom}\, T_1 \cap \mathrm{dom}\, T_2 \neq \varnothing$.

Proof Let $g_1, g_2 : X \times X^* \times X^* \to \overline{\mathbb{R}}$ be defined by

$$g_1(x, x^*, y^*) = \overline{\mathrm{co}}\, c_{T_1}(x, x^*) + \tfrac{1}{2}\|x\|^2, \quad g_2(x, x^*, y^*) = \overline{\mathrm{co}}\, c_{T_2}(x, y^*) + \tfrac{1}{2}\|x^* + y^*\|_*^2.$$

It is obvious that g_1 and g_2 are proper lsc convex functions. Moreover,

$$\mathrm{dom}\, g_1 = \{(x_1, x_1^*, y_1^*) \mid (x_1, x_1^*) \in \mathrm{dom}(\overline{\mathrm{co}}\, c_{T_1}), y_1^* \in X^*\},$$
$$\mathrm{dom}\, g_2 = \{(x_2, x_2^*, y_2^*) \mid (x_2, y_2^*) \in \mathrm{dom}(\overline{\mathrm{co}}\, c_{T_2}), x_2^* \in X^*\},$$

and so

$$\mathrm{dom}\, g_1 - \mathrm{dom}\, g_2 = \Pr_X(\mathrm{dom}(\overline{\mathrm{co}}\, c_{T_1}) - \mathrm{dom}(\overline{\mathrm{co}}\, c_{T_2})) \times X^* \times X^*.$$

Therefore $(0,0,0) \in {}^{ic}(\mathrm{dom}\, g_1 - \mathrm{dom}\, g_2)$. Using [10, Th. 2.8.7], it follows that $(g_1 + g_2)^* = g_1^* \square g_2^*$ and the convolution is exact. Using (4) we get

$$g_1(x,x^*,y^*)+g_2(x,x^*,y^*)\geq\langle x,x^*\rangle+\tfrac{1}{2}\|x\|^2+\langle x,y^*\rangle+\tfrac{1}{2}\|x^*+y^*\|_*^2$$
$$\geq\tfrac{1}{2}\|x\|^2-\|x\|\cdot\|x^*+y^*\|_*+\tfrac{1}{2}\|x^*+y^*\|_*^2\geq 0$$

for all $(x,x^*,y^*)\in X\times X^*\times X^*$; this proves that $(g_1+g_2)^*(0,0,0)\leq 0$. It follows that there exists $(x^*,x,y)\in(X\times X^*\times X^*)^*=X^*\times X\times X$ such that $g_1^*(-x^*,-x,-y)+g_2^*(x^*,x,y)\leq 0$. Let us compute g_i^*. As g_i is the sum of two convex functions, one of them being continuous, we have that

$$g_1^*(x^*,x,y)=\begin{cases}\min\left\{f_{T_1}(x,x^*-u^*)+\tfrac{1}{2}\|u^*\|_*^2\mid u^*\in X^*\right\} & \text{if } y=0,\\ +\infty & \text{if } y\neq 0,\end{cases}$$

and

$$g_2^*(x^*,x,y)=f_{T_2}(y-x,x^*)+\tfrac{1}{2}\|x\|^2.$$

Hence, there exists $(x,x^*)\in X\times X^*$ such that $g_1^*(-x^*,x,0)+g_2^*(x^*,-x,0)\leq 0$, and so there exists also $u^*\in X^*$ such that $f_{T_1}(x,u^*-x^*)+\tfrac{1}{2}\|u^*\|_*^2+f_{T_2}(x,x^*)+\tfrac{1}{2}\|x\|^2\leq 0$. Taking $x_1^*:=u^*-x^*$ and $x_2^*:=x^*$ we get the conclusion.

Assume now that T_1 and T_2 are maximal monotone. Then, by (5), $f_{T_i}\geq c$, and so

$$0\leq\langle x,x_1^*+x_2^*\rangle+\tfrac{1}{2}\|x\|^2+\tfrac{1}{2}\|x_1^*+x_2^*\|_*^2$$
$$\leq f_{T_1}(x,x_1^*)+f_{T_2}(x,x_2^*)+\tfrac{1}{2}\|x\|^2+\tfrac{1}{2}\|x_1^*+x_2^*\|_*^2\leq 0.$$

It follows that $f_{T_i}(x,x_i^*)=c(x,x_i^*)$; using again (5), we get $(x,x_i^*)\in\operatorname{gph}T_i$ for $i\in\{1,2\}$. □

Taking $T_1=0$ (that is $T_1(x)=\{0\}$ for every $x\in X$) and $T_2=T$ or $\operatorname{gph}T_2:=\operatorname{gph}T-(v,v^*)$ in the preceding theorem one obtains immediately the first part or the necessity of the second part of the next result, respectively; the proof of the sufficiency of the second part follows directly (and easily), and can be found in [8, Ths. 10.3, 10.6], [10, Ths. 3.11.5, 3.11.6] or [9]).

The result in the second part of the next theorem is nothing else but Simons' version of Rockafellar's surjectivity theorem [7, Cor. 1]. For another proof of this theorem using Fitzpatrick's function see [9].

Theorem 3. *Let $T : X \rightrightarrows X^*$ be a monotone multifunction with nonempty graph. Then there exists $(x, x^*) \in X \times X^*$ such that*

$$\langle y - x, y^* - x^* \rangle \geq \tfrac{1}{2}\|x\|^2 + \tfrac{1}{2}\|x^*\|_*^2 + \langle x, x^* \rangle \quad \forall (y, y^*) \in \text{gph } T.$$

Moreover, M is maximal monotone if and only if for every $(v, v^*) \in X \times X^*$ there exists $(x, x^*) \in \text{gph } T$ such that

$$\tfrac{1}{2}\|x - v\|^2 + \tfrac{1}{2}\|x^* - v^*\| + \langle x - v, x^* - v^* \rangle = 0,$$

or equivalently, M is maximal monotone if and only if

$$\text{gph } T + \text{gph}(-J_X) = X \times X^*.$$

The preceding two theorems yield the following criterion for the maximality of the sum of two maximal monotone multifunctions; the proof is similar to that of [10, Th. 3.11.9].

Corollary 4. *Assume that $T_1, T_2 : X \rightrightarrows X^*$ are maximal monotone and condition (7) is satisfied. Then $T_1 + T_2$ is maximal monotone.*

Proof Set $T := T_1 + T_2$; from Theorem 2 we have that $\text{dom } T = \text{dom } T_1 \cap \text{dom } T_2 \neq \varnothing$. Take $(v, v^*) \in X \times X^*$ and $S_1, S_2 : X \rightrightarrows X^*$ defined by $S_i(x) := T_i(x + v) - \tfrac{1}{2}v^*$; S_i is maximal monotone for $i = 1, 2$. Moreover, by (6),

$$\text{dom}(\overline{\text{co}}\, c_{S_1}) - \text{dom}(\overline{\text{co}}\, c_{S_2}) = \text{dom}(\overline{\text{co}}\, c_{T_1}) - \text{dom}(\overline{\text{co}}\, c_{T_2}).$$

Hence S_1, S_2 satisfy the conditions of Theorem 2. Therefore, there exists $z \in X$ and $z_1^*, z_2^* \in X^*$ such that $(z, z_i^*) \in \text{gph } S_i$ and

$$\|z\|^2 + \|z_1^* + z_2^*\| + 2\langle z, z_1^* + z_2^* \rangle = 0.$$

Taking $x := z + v$ and $x_i^* := z_i^* + \tfrac{1}{2}v^*$, we obtain that $(x, x_i^*) \in \text{gph } T_i$ and

$$\|x - v\|^2 + \|x_1^* + x_2^* - v^*\|_*^2 + 2\langle x - v, x_1^* + x_2^* - v^* \rangle = 0.$$

Using now Theorem 3 we obtain that T is maximal monotone. □

Note that condition (7) is satisfied if $0 \in (\text{dom } T_1 - \text{dom } T_2)^i$ because, obviously,

$$\begin{aligned} \operatorname{dom} T_1 - \operatorname{dom} T_2 &\subset \operatorname{co}(\operatorname{dom} T_1) - \operatorname{co}(\operatorname{dom} T_2) \\ &\subset \operatorname{Pr}_X(\operatorname{dom}(\overline{\operatorname{co}}\, c_{T_1}) - \operatorname{dom}(\overline{\operatorname{co}}\, c_{T_2})). \end{aligned} \tag{8}$$

Hence Corollary 4 covers the classical Rockafellar's result [7, Th. 1] which is obtained under the hypothesis that $\operatorname{dom} T_1 \cap \operatorname{int}(\operatorname{dom} T_2) \neq \varnothing$. In the next result, similarly to [8, Th. 23.2] and [10, Th.3.11.11], we give several equivalent conditions which ensure the maximal monotonicity of $T_1 + T_2$.

Theorem 5. *Let $T_1, T_2 : X \rightrightarrows X^*$ be maximal monotone multifunctions. Then*

$$^{ic}(\operatorname{dom} T_1 - \operatorname{dom} T_2) = {}^{ic}(\operatorname{co}(\operatorname{dom} T_1) - \operatorname{co}(\operatorname{dom} T_2)) \tag{9}$$

$$= {}^{ic}\operatorname{Pr}_X\left(\operatorname{dom}\left(\overline{\operatorname{co}}\, c_{T_1}\right) - \operatorname{dom}\left(\overline{\operatorname{co}}\, c_{T_2}\right)\right). \tag{10}$$

Therefore $^{ic}\left(\operatorname{dom} T_1 - \operatorname{dom} T_2\right)$ *is a convex set and the following statements are equivalent:*

$$0 \in {}^{ic}\operatorname{Pr}_X(\operatorname{dom}(\overline{\operatorname{co}}\, c_{T_1}) - \operatorname{dom}(\overline{\operatorname{co}}\, c_{T_2})), \tag{11}$$

$$0 \in {}^{ic}(\operatorname{co}(\operatorname{dom} T_1) - \operatorname{co}(\operatorname{dom} T_2)), \tag{12}$$

$$0 \in {}^{ic}\left(\operatorname{dom} T_1 - \operatorname{dom} T_2\right), \tag{13}$$

$\operatorname{dom} T_1 - \operatorname{dom} T_2$ *is neighborhood of the origin in* $\overline{\operatorname{lin}\left(\operatorname{dom} T_1 - \operatorname{dom} T_2\right)}$, (14)

$\bigcup_{\lambda \geq 0} \lambda\left(\operatorname{dom} T_1 - \operatorname{dom} T_2\right)$ *is a closed linear subspace,* (15)

each of these conditions ensuring that $T_1 + T_2$ is maximal monotone.

Furthermore, if $^{ic}(\operatorname{dom} T_1 - \operatorname{dom} T_2) \neq \varnothing$ *then*

$$\overline{^{ic}\left(\operatorname{dom} T_1 - \operatorname{dom} T_2\right)} = \overline{\operatorname{dom} T_1 - \operatorname{dom} T_2} = \overline{\operatorname{Pr}_X(\operatorname{dom}(\overline{\operatorname{co}}\, c_{T_1}) - \operatorname{dom}(\overline{\operatorname{co}}\, c_{T_2}))},$$

and $\overline{\operatorname{dom} T_1 - \operatorname{dom} T_2}$ *is a convex set.*

Proof Taking into account that the inclusions in (8) hold, let us prove that

$$^{ic}\Pr_X(\operatorname{dom}(\overline{\operatorname{co}}\, c_{T_1}) - \operatorname{dom}(\overline{\operatorname{co}}\, c_{T_2})) \subset \operatorname{dom} T_1 - \operatorname{dom} T_2. \tag{16}$$

Consider $v \in {}^{ic}\Pr_X(\operatorname{dom}(\overline{\operatorname{co}}\, c_{T_1}) - \operatorname{dom}(\overline{\operatorname{co}}\, c_{T_2}))$ and take the multifunction T_1' defined by $T_1'(x) := T_1(x+v)$. By (6) we have that

$$\operatorname{dom}(\overline{\operatorname{co}}\, c_{T_1'}) = \operatorname{dom}(\overline{\operatorname{co}}\, c_{T_1}) - (v,0),$$

which proves that

$$0 \in {}^{ic}\Pr_X(\operatorname{dom}(\overline{\operatorname{co}}\, c_{T_1'}) - \operatorname{dom}(\overline{\operatorname{co}}\, c_{T_2})).$$

As T_1' and T_2 are maximal monotone, using Theorem 2, we obtain that $\operatorname{dom} T_1' \cap \operatorname{dom} T_2 \neq \varnothing$, and so $v \in \operatorname{dom} T_1 - \operatorname{dom} T_2$. We obtain that (16) holds. From (8) and (16) we obtain that (9) and (10) are satisfied when ${}^{ic}\Pr_X(\operatorname{dom}(\overline{\operatorname{co}}\, c_{T_1}) - \operatorname{dom}(\overline{\operatorname{co}}\, c_{T_2}))$ is nonempty.

Observe now that

$$\begin{aligned}\operatorname{aff}(\operatorname{dom} T_1 - \operatorname{dom} T_2) &\subset \operatorname{aff}(\Pr_X(\operatorname{dom}(\overline{\operatorname{co}}\, c_{T_1}) - \operatorname{dom}(\overline{\operatorname{co}}\, c_{T_2}))) \\ &\subset \overline{\operatorname{aff}(\operatorname{dom} T_1 - \operatorname{dom} T_2)}.\end{aligned} \tag{17}$$

The first inclusion is obvious. For the second one set $V := \operatorname{aff}(\operatorname{dom} T_1 - \operatorname{dom} T_2)$. We have that

$$\begin{aligned}&\{(x_1,x_1^*,y_1^*) \mid (x_1,x_1^*) \in \operatorname{gph} T_1,\ y_1^* \in X^*\} - \{(x_2,x_2^*,y_2^*) \mid (x_2,y_2^*) \in \operatorname{gph} T_2,\ x_2^* \in X^*\} \\ &\qquad \subset (\operatorname{dom} T_1 - \operatorname{dom} T_2) \times X^* \times X^* \subset V \times X^* \times X^*.\end{aligned}$$

As for subsets $D,E,F \subset X$ with $D - E \subset F$ one has that $\overline{\operatorname{co}}\, D - \overline{\operatorname{co}}\, E \subset \overline{\operatorname{co}}\, F$, and taking into account that V is convex, from the preceding inclusion we obtain that

$$\begin{aligned}&\Pr_X(\operatorname{dom}(\overline{\operatorname{co}}\, c_{T_1}) - \operatorname{dom}(\overline{\operatorname{co}}\, c_{T_2})) \times X^* \times X^* \\ &\quad \subset \Pr_X(\overline{\operatorname{co}}\,(\operatorname{gph} T_1) - \overline{\operatorname{co}}\,(\operatorname{gph} T_2)) \times X^* \times X^* \\ &\quad = \{(x_1,x_1^*,y_1^*) \mid \big((x_1,x_1^*) \in \overline{\operatorname{co}}\,(\operatorname{gph} T_1),\ y_1^* \in X^*\big) \\ &\qquad - \{(x_2,x_2^*,y_2^*) \mid (x_2,y_2^*) \in \overline{\operatorname{co}}\,(\operatorname{gph} T_2),\ x_2^* \in X^*\} \\ &\quad \subset \overline{V} \times X^* \times X^*.\end{aligned}$$

The desired inclusion follows.

From (8), (16) and (17) we obtain that (9) and (10) hold, that conditions (11), (12) and (13) are equivalent, and that $^{ic}(\operatorname{dom} T_1 - \operatorname{dom} T_2)$ is a convex set.

It is obvious that (14) $\Rightarrow$ (15) $\Rightarrow$ (12). Also, because the relative interior of $\operatorname{dom} g_1 - \operatorname{dom} g_2$ coincides with $^{ic}(\operatorname{dom} g_1 - \operatorname{dom} g_2)$ when this is nonempty, we have that the relative interior of $\operatorname{dom} T_1 - \operatorname{dom} T_2$ coincides with $^{ic}(\operatorname{dom} T_1 - \operatorname{dom} T_2)$ when this is nonempty. Hence (13) $\Rightarrow$ (14).

Using Corollary 4 and what was proved before we get that everyone of conditions (11)–(15) is sufficient for the maximal monotonicity of $T_1 + T_2$.

Because

$$^{ic}(\operatorname{co}(\operatorname{dom} T_1) - \operatorname{co}(\operatorname{dom} T_2)) = {}^{ic}(\operatorname{dom} T_1 - \operatorname{dom} T_2) \subset \operatorname{dom} T_1 - \operatorname{dom} T_2 \subset \operatorname{co}(\operatorname{dom} T_1) - \operatorname{co}(\operatorname{dom} T_2),$$

and for a convex set A with $^{ic}A \neq \varnothing$ one has that $\overline{A} = \overline{^{ic}A}$ it follows that $\overline{\operatorname{dom} T_1 - \operatorname{dom} T_2}$ is convex when $^{ic}(\operatorname{dom} T_1 - \operatorname{dom} T_2) \neq \varnothing$. □

The next result refers to composition with linear operators and corresponds to Theorem 2. The maximal monotonicity of $A^* \circ T \circ A$ was obtained by Pennanen [4, Cor. 4.4(c)] under the same condition (using the result for the sum) and by Penot [6] under the condition $0 \in \operatorname{core}(\operatorname{dom} T - \operatorname{Im} A)$ with a different proof.

Theorem 6. *Let Y be another reflexive Banach space, $T : X \rightrightarrows X^*$ a monotone multifunction and $A : Y \to X$ a continuous linear operator. If*

$$0 \in {}^{ic}(\operatorname{Pr}_X(\operatorname{dom} \overline{\operatorname{co}}\, c_T) - \operatorname{Im} A), \tag{18}$$

then for every $(w, w^*) \in Y \times Y^*$, there exist $v \in Y$ and $x^* \in X^*$ such that

$$f_T(Av, x^*) - \langle Av, x^* \rangle + \tfrac{1}{2}\|v - w\|^2 + \tfrac{1}{2}\|A^*x^* - w\|_*^{*2} + \langle v - w, A^*x^* - w^* \rangle \leq 0. \tag{19}$$

Moreover, if T is maximal monotone then $A^* \circ T \circ A$ is maximal monotone.

Proof Fix $(w, w^*) \in Y \times Y^*$; consider $g : X \times X^* \times Y \times Y^* \to \overline{\mathbb{R}}$ defined by

$$g(u,u^*,y,y^*) := \overline{\text{co}}\, c_T(u,u^*) + \tfrac{1}{2}\|y+w\|^2 + \tfrac{1}{2}\|y^*+w^*\|_*^2 + \langle y, w^*\rangle + \langle w, y^*+w^*\rangle,$$

and $B: X^* \times Y \to X \times X^* \times Y \times Y^*$ defined by $B(x^*,v) := (Av, -x^*, -v, A^*x^*)$. Then

$$g^*(x^*,x,v^*,v) = f_T(x,x^*) + \tfrac{1}{2}\|v-w\|^2 + \tfrac{1}{2}\|v^*-w^*\|_*^2 - \langle v, w^*\rangle - \langle w, v^*-w^*\rangle$$

and $B^*(x^*,x,v^*,v) = (Av-x, A^*x^*-v^*)$.

Because T is monotone, using (3), we have that

$$\begin{aligned}(g\circ B)(x^*,v) &= \overline{\text{co}}\, c_T(Av,-x^*) + \tfrac{1}{2}\|v-w\|^2 + \\ &\quad + \tfrac{1}{2}\|A^*x^*+w^*\|_*^2 - \langle v, w^*\rangle + \langle w, A^*x^*+w^*\rangle \\ &\geq \tfrac{1}{2}\|v-w\|^2 + \tfrac{1}{2}\|A^*x^*+w^*\|_*^2 - \langle v-w, A^*x^*+w^*\rangle \geq 0,\end{aligned}$$

for all $x^* \in X^*$ and $v \in Y$. It follows that $(g\circ B)^*(0,0) \leq 0$. In order to use [10, Th. 2.8.3] (with $f=0$), we need to calculate $D := \text{dom}\, g - \text{Im}\, B$. So,

$$\begin{aligned}D &= \text{dom}(\overline{\text{co}}\, c_T) \times Y \times Y^* - \{(Av,-x^*,-v,A^*x^*) \mid v\in Y, x^*\in X^*\} \\ &= (\text{dom}(\overline{\text{co}}\, c_T) - \text{Im}\, A \times X^*) \times Y \times Y^* \\ &= (\text{Pr}_X(\text{dom}(\overline{\text{co}}\, c_T)) - \text{Im}\, A) \times X^* \times Y \times Y^*.\end{aligned}$$

Hence $0 \in {}^{ic}(\text{dom}\, g - \text{Im}\, B)$. Applying [10, Th. 2.8.3] mentioned above, we get some (x^*,x,v^*,v) such that $B^*(x^*,x,v^*,v) = (Av-x, A^*x^*-v^*) = (0,0)$ and $g^*(x^*,x,v^*,v) \leq 0$, that is, we find $v\in Y$ and $x^*\in X^*$ such that (19) holds.

Assume now that T is maximal monotone. Taking $(w,w^*) \in Y \times Y^*$, we find $v\in Y$ and $x^*\in X^*$ such that (19) holds. Because T is maximal monotone we have that $f_T \geq c$, and so we obtain that $f_T(Av,x^*) = \langle Av, x^*\rangle$ and

$$\tfrac{1}{2}\|v-w\|^2 + \tfrac{1}{2}\|A^*x^*-w^*\|_*^2 + \langle v-w, A^*x^*-w^*\rangle = 0.$$

The first relation implies that $x^* \in T(Av)$ and the second relation implies that $w^* - A^*x^* \in -J_Y(w-v)$. It follows that $(w,w^*) \in \text{gph}(A^*\circ T\circ A) - \text{gph}(-J_Y)$. Using now Theorem 3 we obtain that $A^*\circ T\circ A$ is maximal monotone. □

The proof of the next theorem is similar to that of Theorem 5, so we omit it.

Theorem 7. *Let X, Y be reflexive spaces, $A: Y \to X$ a continuous linear operator and $T: X \rightrightarrows X^*$ a maximal monotone multifunction. Then*

$$^{ic}(\operatorname{dom} T - \operatorname{Im} A) = {}^{ic}(\operatorname{co}(\operatorname{dom} T) - \operatorname{Im} A) = {}^{ic}(\operatorname{Pr}_X(\operatorname{dom} \overline{\operatorname{co}}\, c_T) - \operatorname{Im} A).$$

Therefore $^{ic}(\operatorname{dom} T - \operatorname{Im} A)$ *is a convex set and the following statements are equivalent:*

$0 \in {}^{ic}(\operatorname{Pr}_X(\operatorname{dom} \overline{\operatorname{co}}\, c_T) - \operatorname{Im} A)$,

$0 \in {}^{ic}(\operatorname{co}(\operatorname{dom} T) - \operatorname{Im} A)$,

$0 \in {}^{ic}(\operatorname{dom} T - \operatorname{Im} A)$,

$\operatorname{dom} T - \operatorname{Im} A$ *is a neighborhood of the origin in* $\overline{\operatorname{lin}(\operatorname{dom} T - \operatorname{Im} A)}$,

$\bigcup_{\lambda \geq 0} \lambda(\operatorname{dom} T - \operatorname{Im} A)$ *is a closed linear subspace,*

each of these conditions ensuring that $A^* \circ T \circ A$ is maximal monotone.

Furthermore, if $^{ic}(\operatorname{dom} T - \operatorname{Im} A) \neq \varnothing$ *then*

$$\overline{^{ic}(\operatorname{dom} T - \operatorname{Im} A)} = \overline{\operatorname{dom} T - \operatorname{Im} A} = \overline{\operatorname{Pr}_X(\operatorname{dom} \overline{\operatorname{co}}\, c_T - \operatorname{Im} A}, \tag{20}$$

and $\overline{\operatorname{dom} T - \operatorname{Im} A}$ *is a convex set.*

REFERENCES

[1] F.E. Browder, Nonlinear maximal monotone operators in Banach spaces, Math. Ann. **175** (1968), 89–113.

[2] S. Fitzpatrick, Representing monotone operators by convex functions, Workshop/Miniconference on Functional Analysis and Optimization (Canberra, 1988), Austral. Nat. Univ., Canberra, 1988, pp. 59–65.

[3] Hiriart-Urruty, J.-B., ε -subdifferential calculus, In: Convex Analysis and Optimization (London, 1980), Pitman, Boston, Mass., 1982, pp. 43–92.

[4] T. Pennanen, Dualization of generalized equations of maximal monotone type, SIAM J. Optim. **10** (2000), 809–835.

[5] J.-P. Penot, The relevance of convex analysis for the study of monotonicity, Tech. report, University of Pau, Pau, 2002.

[6] J.-P. Penot, Private communication (2003).

[7] R. T. Rockafellar, On the maximality of sums of nonlinear monotone operators, Trans. Amer. Math. Soc. **149** (1970), 75–88.

[8] S. Simons, Minimax and monotonicity, Springer-Verlag, Berlin, 1998.
[9] S. Simons, C. Zălinescu, A new proof for Rockafellar's characterization of maximal monotone operators, Proc. Amer. Math. Soc. (accepted).
[10] C. Zălinescu, Convex analysis in general vector spaces, World Scientific, Singapore, 2002.

CONTRIBUTORS

Alber Ya. I.	Department of Mathematics, Technion - Israel Institute of Technology, 32000 Haifa, Israel e-mail: alberya@tx.technion.ac.il
Allevi E.	Dipartimento di Matematica, Statitistica, Informatica e Applicazioni, Università di Bergamo, Via dei Caniana 2, Bergamo 24127, Italy e-mail: allevi@unibg.it
Barbu V.	University of Iasi, 6600 Iasi, Rumania
Beirão da Veiga H.	Dipartimento di Matematica Applicata "U. Dini", Università di Pisa, via Bonanno Pisano 25/B, 56126 Pisa, Italy e-mail: bveiga@dma.unipi.it
Bensoussan A.	Université Paris Dauphine, Paris, France e-mail : alain.bensoussan@cnes.fr
Bigi G.	Dipartimento di Informatica, Università di Pisa, via F. Buonarroti, 2, 56127 Pisa, Italia; e-mail: bigig@di.unipi.it
Boltyanski V.	CIMAT, Guanajuato, Mexico e-mail: boltian@cimat.mx
Bussotti P.	Humboldt Foundation, Institute of Science History, Ludwig-Maximilians University of Munich, Munich, Germany e-mail: crbpm@tin.it
Butnariu D.	Department of Mathematics, University of Haifa, 31905 Haifa, Israel e-mail:dbutnaru@math.haifa.ac.il
Cammaroto F.	Dipartimento di Matematica, Università di Messina, 98166 Sant'Agata-Messina, Italy

e-mail: filippo@dipmat.unime.it

Caruso A.O. — Dipartimento di Matematica e Informatica, Universita' di Catania, Viale A.Doria 6-I, 95125, Catania, Italy
e-mail: Aocaruso@Dmi.Unict

Cottle R. W. — Department of Management and Engineering (MS&E) Terman Engineering Center, Stanford University, Stanford, CA 94305-4026, USA
e-mail: rwc@stanford.edu

Chen G.Y. — Institute of Systems Science, Chinese Academy of Sciences, Beijing 100080, P.R. China
e-mail: chengy@mail.amss.ac.cn

Chinnì A. — Dipartimento di Matematica, Università di Messina, Contrada Papardo, Salita Sperone 31,98166 Sant'Agata (Messina), Italy
e-mail: chinni@dipmat.unime.it

Crespi G.P. — Università della Valle d'Aosta, Facoltà di Economia, Aosta, Italy
e-mail: giovanni.crespi@uni-bocconi.it

Daniele P. — Dipartimento di Matematica, Università di Catania, Città Universitaria, Viale Doria 6, 95125 Catania, Italy
e-mail: daniele@dmi.unict.it

Da Prato G. — Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy
e-mail: g.daprato@sns.it

Demyanov V.F. — Applied Mathematics Dept., St. Petersburg State University, Staryi Peterhof, St. Petersburg, Russia
e-mail:
vladimir.demyanov@pobox.spbu.ru

Di Bella B. — Dipartimento di Matematica, Università di Messina, Contrada Papardo, Salita Sperone 31, 98166 Sant'Agata-Messina, Italy
e-mail: beatrice@dipmat.unime.it

Ernst E. — Laboratoire de Modelisation en Méchanique et Termodynamique, Faculté de Science et Techniques de Saint Jérôme, Case 332, Avenue Escadrille Normandie-Niemen 13397 Cedex 20
e-mail: Emil.Ernst@Univ.U-3mrs.Fr

Fanciullo M.S. — Dipartimento di Matematica e Informatica, Università di Catania, Viale A.Doria 6-I - 95125, Catania, Italy
e-mail: fanciullo@dmi.unict.it

Faraci F. — Dipartimento di Matematica, Università di Catania, Città Universitaria, Viale Doria 6, 95125 Catania, Italy
e-mail:ffaraci@dmi.unict.it

Fattorusso L. — Facoltà di Ingegneria, Università degli Studi di Reggio Calabria, Corso V.Emanuele 109,89127 Reggio Calabria, Italy
e-mail: fattorusso@ing.unirc.it

Ferrari P. — Dipartimento di Vie e Trasporti, Facoltà di Ingegneria, Università di Pisa, via Diotisalvi 1, 56100 Pisa, Italy
e-mail: p.ferrari@ing.unipi.it

Garroni M. G. — Dipartimento di Matematica "Guido Castelnuovo", Università degli Studi di Roma "La Sapienza", Piazzale A. Moro 2, 00185 Roma, Italy
e-mail: ggarroni@mat.uniroma1.it

Giannessi F. — Dipartimento di Matematica, Università di Pisa, Largo B. Pontecorvo 2, 56127 Pisa, Italy
e-mail: gianness@dm.unipi.it

Ginchev I.	Technical University of Varna, Department of Mathematics, 9010 Varna, Bulgaria e-mail: ginchev@ms3.tu-varna.acad.bg
Giuffrè S.	D.I.M.E.T. Facoltà di Ingegneria, Università di Reggio Calabria, Via Graziella 1, Località Feo di Vito - 89060 Reggio Calabria, Italy e-mail:giuffre@ing.unirc.it
Gnudi A.	Dipartimento di Matematica, Statistica, Informatica e Applicazioni, Università, di Bergamo, Via dei Caniana 2, Bergamo 24127, Italy e-mail: adri@unibg..it
Guerraggio A.	Dipartimento di Economia, Università dell'Insubria, via Ravasi 2, 21100 Varese, Italy e-mail: aguerraggio@eco.uninsubria.it
Hastings S.	Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260, USA e-mail: sph+@pitt.edu
Huang X. X.	Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong, P.R. of China e-mail: mahuangx@polyu.edu.hk
Idone G.	D.I.M.E.T. Facoltà di Ingegneria, Università di Reggio Calabria, Via Graziella 1, Località Feo di Vita, 89060 Reggio Calabria, Italy e-mail: idone@ing.unirc.it
Ioffe A.D.	Department of Mathematics, Technion, Haifa 32000, Israel e-mail: ioffe@math.technion.ac.il

Jofre A. — Center for Mathematical Modelling and Dept. of Mathematical Engineering, University of Chile, Casilla 170/3, Correo 3, Santiago, Chile
e-mail: ajofre@dim.uchile.cl

Kassay G. — Faculty of Mathematics, University Babes-Bolyai, Cluj-Napoca, Rumania
e-mail: kassay@math.ubbcluj.ro

Khan A. — Department of Mathematical Sciences, Fisher Hall, Michigan Technological University, 1400 Towsend Drive,Houghton MI 49931-1295, USA
e-mail: aakhan@mtu.edu

Kinderlehrer D. — Center for Nonlinear Analysis and Department of Mathematical Sciences Carnegie Mellon University, Pittsburgh, PA 15213, USA,
e-mail: davidk@andrew.cmu.edu

Konnov I.V. — Department of Applied Mathematics, Kazan University, ul. Kremlevskaya,18, Kazan 420008, Russia
e-mail: igor.konnov@ksu.ru

Kurzhanski A.B. — Moscow State (Lomonosov) University, Faculty of Computational Mathematics and Cybernetics, 119992 Moscow, Russia
e-mail: kurzhans@mail.ru

Li S. J. — Department of Information and Computer Sciences, College of Sciences, Chongqing University, Chongqing, 400044,China

Lieberman G.M. — Department of Mathematics, Iowa State University, Ames, Iowa 50011, USA
e-mail: lieb@iastate.edu

Magenes E. — Dipartimento di Matematica "F. Casorati", Università, via Ferrata 1, 27100 Pavia, Italy
e-mail: magenes@imati.cnr.it

Mancino O.G. — Dipartimento di Matematica Applicata "U. Dini", Università, via Bonanno Pisano 25b, 56126 Pisa, Italy
e-mail: mancino@dma.unipi.it

Marino A. — Dipartimento Di Matematica "L.Tonelli" , Università di Pisa, Largo B. Pontecorvo 2, I56127 Pisa, Italia
e-mail:marino@dm.unipi.it

Maugeri A. — Dipartimento di Matematica e Informatica, Università di Catania, Viale A. Doria, 6 - 95125 Catania, Italy
e-mail: maugeri@dmi.unict.it

Mazurkevich E.O. — Department of Applied Mathematics, Kazan University, ul. Kremlevskaya,18, Kazan 420008, Russia

Mazzone S. — Dipartimento di Matematica "G. Castelnuovo" Università di Roma "La Sapienza", Piazzale A. Moro 2, 00185 Roma, Italy
e-mail: silvia.mazzone@uniroma1.it

Medjo Tachim T. — Department of Mathematics, Florida International University, DM413B, University Park, Miami, Florida 33199, USA

Meneses C. N. — Dept. of Industrial and Systems Engineering, University of Florida, Gainesville, FL, USA
e-mail: claudio@ufl.edu

Milasi M. — Università di Messina, Dipartimento di Matematica, Contrada Papardo, Salita Sperone 31, 98166 Sant'Agata (Messina), Italy
E-mail: monica@dipmat.unime.it.

Morandi Cecchi M. — Università di Padova, Dipartimento di Matematica Pura ed Applicata, Via Belzoni 7, 35131 Padova, Italy
e- mail: mcecchi@math.unipd.it

Mordukhovich B. S. — Department of Mathematics, Wayne State University, Detroit, Michigan 48202, USA
e-mail: boris@math.wayne.edu

Murthy M.K. V. — Dipartimento di Matematica, Università di Pisa, Largo B. Pontecorvo 2, 56127 Pisa, Italy
e-mail: murthy@dm.unipi.it

Nirenberg L. — Courant Institute, 251 Mercer Street, New York, NY 10012, USA
e-mail: nirenl@cims.nyu.edu

Oliveira C. A.S. — Dept. of Industrial and Systems Engineering, University of Florida, Gainesville, FL, USA
e-mail: oliveira@ufl.edu

Palagachev D. — Dipartimento Interuniversitario di Matematica, Politecnico di Bari, via Orabona 4, 70125 Bari, Italy
e-mail: dian@dm.uniba.it

Pallaschke D. — Institut für Statistik und Mathematische Wirtschaftstheorie, Universität von Karlsruhe, Kaiserstrasse 12, D-76128 Karlsruhe, Germany
e-mail: lh09@rz.uni-karlsruhe.de

Panicucci B. — Dipartimento di Matematica, Università di Pisa, Largo B. Pontecorvo 2, 56127 Pisa, Italy
e-mail: panicucc@mail.dm.unipi.it

M. Pappalardo — Dipartimento di Matematica Applicata, Università di Pisa, via Bonanno Pisano, 25/b, 56126 Pisa, Italy

e-mail: pappalardo@dma.unipi.it

Pardalos P. M. — Dept. of Industrial and Systems Engineering, University of Florida, 303 Weil Hall, Gainesville, FL 32611-9083, USA
e-mail: pardalos@ufl.edu

Penot J.-P. — Laboratoire de Mathématiques appliquéès, CNRS ERS 2055, Facultè des Sciences, av. De l'Universitè, 6400 PAU, France
e-mail: jean-paul.penot@univ-pau.fr

Pia S. — D.I.M.E.T. Facoltà di Ingegneria, Università di Reggio Calabria, Via Graziella 1, Località Feo di Vito - 89060 Reggio Calabria, Italy
e-mail: stephane.pia@tiscali.it

Pulvirenti G — Dipartimento di Matematica e Informatica, Università di Catania, Città Universitaria Viale A. Doria 6, 96125 Catania, Italy
e-mail: pulvirenti@dmi.unict.it

Pusillo L. — Dipartimento di Matematica, Università di Genova, via Dodecaneso 35, 16146 Genova, Italy
e-mail: pusillo@dima.unige.it

Qi L. — Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
e-mail: maqilq@inet.polyu.edu.hk

Raciti F. — Dipartimento di Matematica e Informatica, Università di Catania, and Facoltà di Ingegneria dell'Università di Catania, sede di Enna, V.le A.Doria, 6, 95125 Catania,Italy
e-mail: fraciti@dmi.unict.it

Ragusa M.A.	Dipartimento di Matematica e Informatica, Università di Catania, Viale Andrea Doria 6, 95125 Catania, Italy e-Mail: Maragusa@dmi.unict.it
Rapcsák T.	Computer and Automation Research Institute, Hungarian Academy of Sciences, Kende ut.13-17, 111 Budapest, Hungary e-mail: rapcsak@oplab.sztaki.hu
Ricceri B.	Dipartimento di Matematica, Università di Catania, Viale A. Doria 6, 95125 Catania, Italy e-mail: ricceri@dmi.unict.it
Robinson S.M.	Department of Industrial Engineering, University of Wisconsin–Madison, 1513 University Avenue, Madison, WI 53706-1572, USA e-mail: smrobins@wisc.edu.
Rocca M.	Dipartimento di Economia, Università dell'Insubria, via Ravasi 2, 21100 Varese, Italy e-mail: mrocca@eco.uninsubria.it
Rockafellar R.T.	Department of Mathematics, University of Washington, Seattle, WA 98195-4350, USA e-mail: rtr@math.washington.edu
Russo R.	Università di Padova, Dipartimento di Matematica Pura ed Applicata, Via Belzoni 7, 35131 Padova, Italy e-mail: mrrusso@math.unipd.it
Saccon C.	Dipartimento di Matematica Applicata, Viale Bonanno Pisano, I56126 Pisa, Italia e-mail: saccon@mail.dm.unipi.it
Sachs G.	Institute of Flight Mechanics and Flight Control, Technical University of Monaco, Garching, Germany e-mail: kindermann@lfm.mw.tum.de

Santagati G.	Dipartimento di Matematica e Informatica, Università di Catania, Viale A. Doria 6, Catania, Italy
Sbordone C.	Dipartimento di Matematica ed Applicazioni "R. Caccioppoli", via Cintia, I-80126 Napoli, Italy e-mail: sbordone@unina.it
Scrimali L.	Dipartimento di Matematica e Informatica, Università di Catania, Viale A. Doria 6, 95125 Catania, Italy e-mail: scrimali@dmi.unict.it
Simonnet E.	Institut Non Linéaire de Nice, CNRS, 1361, route des Lucioles 06560 Valbonne, France
Simons S.	Department of Mathematics, University of California, Santa Barbara - CA 93106-3080, USA e-mail: simons@math.ucsb.edu
Stampacchia Giulia	Azienda Ospedaliera Pisana, Dipartimento di Neuroscienze, via Paradisa (Cisanello), Pisa, Italy e-mail: g.stampacchia@ao-pisa.toscana.it
Tachim-Medjo T.	Department of Mathematics, Florida International University, DM 413B, University Park, Miami, Florida, USA
Tamasyan G.Sh.	Applied Mathematics Department, St. Petersburg State University, Bibliotechnaya pl.,2 Staryi Peterhof, 198904 St. Petersburg, Russia
Temam R.	The Institute for Scientific Computing and Applied Mathematics, Indiana University, Bloomington, IN 47405, USA e-mail: roger.temam@math.u-psud.fr

Teo L.	Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong e-mail: teokl@polyu.edu.hk
Théra M.	Laco, Université de Limoges, 123 Avenue A. Thomas, 87060 Limoges Cedex, France e-Mail: Michel.Thera@Unilim.Fr
Tikhomirov V.M.	Department of Mechanics and Mathematics, Moscow State University, Vorob'evy Gori, 11899 Moscow, Russia e-mail: tikh@tikhomir.mccme.ru
Tinti F.	Dipartimento di Matematica, Università di Padova, via Belzoni Padova, Italy e-mail: ftinti@math.unipd.it
Urbanski R.	Wydza l Matematyki i Informatyki, Uniwersytet im Adama Mickiewicza, ul. Umultowska 87, PL-61-614 Pozna´n, Poland
Varaiya P.	University of California at Berkeley, EECS, ERL, Berkeley, USA
Villani A.	Dipartimento di Matematica e Informatica, Università di Catania, Viale A. Doria 6, Catania, Italy e-mail: villani@dmi.unict.it
Vitanza C.	Università di Messina, Dipartimento di Matematica, Contrada Papardo, Salita Sperone 31 - 98166 Sant'Agata, Messina, Italy e-mail: vitanzac@dipmat.unime.it.
Wets R.J.-B.	Department of Mathematics, University of California, Davis, CA 95616, USA e-mail: rjbwets@ucdavis.edu
Yang X.Q.	Dept. of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong e-mail. mayangxq@polyu.edu.hk

Yu H. — School of Economics and Management, Tsinghua University, Beijing, P.R. China

Zălinescu C. — Faculty of Mathematics, University "Al. I.Cuza" Iasi, 700506 Iasi, Rumania
e-mail: zalinesc@uaic.ro

Zamboni P. — Università di Catania, Dipartimento di Matematica, viale Andrea Doria 6, 95125 Catania, Italy
e-mail: zamboni@dmi.unict.it