Lecture Notes in Economics and Mathematical Systems

Managing Editors: M. Beckmann and W. Krelle

255

Nondifferentiable Optimization: Motivations and Applications

Proceedings, Sopron, Hungary, 1984

Edited by V. F. Demyanov and D. Pallaschke

Springer-Verlag
Berlin Heidelberg New York Tokyo

Lecture Notes in Economics and Mathematical Systems

255

Nondifferentiable Optimization: Motivations and Applications

Proceedings of an IIASA (International Institute for Applied Systems Analysis) Workshop on Nondifferentiable Optimization
Held at Sopron, Hungary, September 17–22, 1984

Edited by V. F. Demyanov and D. Pallaschke

Springer-Verlag
Berlin Heidelberg New York Tokyo

Editors

Prof. Dr. Vladimir F. Demyanov
International Institute for Applied Systems Analysis
Schlossplatz 1, A-2361 Laxenburg, Austria and

Applied Mathematical Department
Leningrad University, Leningrad, USSR

Prof. Dr. Diethard Pallaschke
Institut für Statistik und Mathematische Wirtschaftstheorie
Universität Karlsruhe, Postfach 6380, D-7500 Karlsruhe 1, FRG

ISBN 3-540-15979-7 Springer-Verlag Berlin Heidelberg New York Tokyo
ISBN 0-387-15979-7 Springer-Verlag New York Heidelberg Berlin Tokyo

Printing and binding: Beltz Offsetdruck, Hemsbach/Bergstr.
2142/3140-543210

PREFACE

The International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria, has been involved in research on nondifferentiable optimization since 1976. IIASA-based East–West cooperation in this field has been very productive, leading to many important theoretical, algorithmic and applied results. Nondifferentiable optimization has now become a recognized and rapidly developing branch of mathematical programming.

To continue this tradition, and to review recent developments in this field, IIASA held a Workshop on Nondifferentiable Optimization in Sopron (Hungary) in September 1984.

The aims of the Workshop were:

1. To discuss the state-of-the-art of nondifferentiable optimization (NDO), its origins and motivation;
2. To compare various algorithms;
3. To evaluate existing mathematical approaches, their applications and potential;
4. To extend and deepen industrial and other applications of NDO.

The following topics were considered in separate sessions:

- General motivation for research in NDO: nondifferentiability in applied problems, nondifferentiable mathematical models.
- Numerical methods for solving nondifferentiable optimization problems, numerical experiments, comparisons and software.
- Nondifferentiable analysis: various generalizations of the concept of subdifferentials.
- Industrial and other applications.

This volume contains selected papers presented at the Workshop. It is divided into four sections, based on the above topics:

I. Concepts in Nonsmooth Analysis

II. Multicriteria Optimization and Control Theory

III. Algorithms and Optimization Methods

IV. Stochastic Programming and Applications

We would like to thank the International Institute for Applied Systems Analysis, particularly Prof. V. Kaftanov and Prof. A.B. Kurzhanski, for their support in organizing this meeting.

We would also like to thank Helen Gasking for her help in preparing this volume.

V. Demyanov
D. Pallaschke

CONTENTS

I. CONCEPTS IN NONSMOOTH ANALYSIS

II. MULTICRITERIA OPTIMIZATION AND CONTROL THEORY

III. ALGORITHMS AND OPTIMIZATION METHODS

IV. STOCHASTIC PROGRAMMING AND APPLICATIONS

I. CONCEPTS IN NONSMOOTH ANALYSIS

ATTEMPTS TO APPROXIMATE A SET-VALUED MAPPING

V.F. Demyanov[1], C. Lemaréchal[2] and J. Zowe[3]
[1] *International Institute for Applied Systems Analysis, Laxenburg, Austria and Leningrad State University, Leningrad, USSR*
[2] *INRIA, P.O. Box 105, 78153 Le Chesnay, France*
[3] *University of Bayreuth, P.O. Box 3008, 8580 Bayreuth, FRG*

Abstract. Given a multi-valued mapping F, we address the problem of finding another multi-valued mapping H that agrees locally with F in some sense. We show that, contrary to the scalar case, introducing a derivative of F is hardly convenient. For the case when F is convex-compact-valued, we give some possible approximations, and at the same time we show their limitations. The present paper is limited to informal demonstration of concepts and mechanisms. Formal statements and their proofs will be published elsewhere.

1. INTRODUCTION

Consider first the problem of solving a nonlinear system:

$$f(x) = 0 \tag{1}$$

where f is a vector-valued function. If we find a *first order approximation* of f near x, i. e. a vector-valued bi-function h such that

$$h(x;d) = f(x+d) + o(d) \tag{2}$$

(where $o(d)/\|d\| \to 0$ when $d \to 0$) then we can apply the *Newton principle:* given a current iterate x, solve for d

$$h(x;d) = 0 \tag{3}$$

(supposedly simpler than (1)) and move to x + d.

Everybody knows that if f is differentiable and if, in addition to satisfying (2), h is required to be affine in d, then it is *unambiguously* defined by

$$h(x;d) := f(x) + f'(x)d \tag{4}$$

Merging (2) and (4) and subtracting f(x) gives also a nonambiguous definition of f' (the jacobian operator of f) by:

$$f'(x)d := f(x+d) - f(x) + o(d).$$

Part of this research was performed at the Mathematics Research Center of the University of Wisconsin under Contract # DAAG 29-80-C-0041

Suppose now that we have to solve

$$0 \in F(x) \tag{5}$$

where F is a multi-valued mapping, i. e. $F(x) \subset R^n$. A possible application of (5) is in nonsmooth optimization, when F is the (approximate) subdifferential of an objective function to be minimized. To apply the same principle as in the single valued case, $F(x+d)$ must be approximated by some *set* $H(x;d) \subset R^n$. Continuing the parallel and requiring H to be affine in d (whatever it means), we must express it as a *sum of two sets:* $H(x,d) = F(x) + G$. In summary, we want to find a set G such that, for all $\varepsilon > 0$ and $\|d\|$ small enough:

$$F(x+d) \subset F(x) + G + \varepsilon\|d\|\, U \tag{6.a}$$

and

$$F(x) + G \subset F(x+d) + \varepsilon\|d\|\, U \tag{6.b}$$

where U is the unit ball of R^n. Unfortunately, such a writing is already worthless. First, it does not help defining the "linearization" G: just because the set of subsets is not a group, F(x) cannot be substracted in (6). Furthermore, (6) is extremely restrictive: for $n = 1$, consider the innocent mapping $F(x) := [0,3x]$ (defined for $x \geqslant 0$). Take $x = 1$, $\varepsilon = 1$ and $d < 0$. It is impossible to find a set G satisfying (6.b), i. e. $[0,3] + G \subset [d,3+2d]$. For example, $G = \{d\}$ is already too "thick".

A conclusion of this section is that a first order approximation to a multivalued mapping cannot be readily constructed by a standard linearization; the definition of such an approximation is at present ambiguous. For a deep insight into differentiability of sets, we refer to [6] and its large bibliography. Here, for want of a complete theory, we will give in the next sections two possible proposals. None of them is fully satisfactory, but they are rather complementary, in the sense that each one has a chance to be convenient when the other is not. We will restrict ourselves to the convex compact case. Furthermore, as is usual in nondifferentiable optimization, we will consider only directional derivatives. Therefore we adopt simpler notations: x and the direction d being fixed, we call F(t) the image by F of $x + td$, $t \geqslant 0$. We say that H *approximates* F *to* 1st *order near* $t = 0^+$ if for every $\varepsilon > 0$, there is $\delta > 0$ such that $t \in [0,\delta]$ implies

$$F(t) \subset H(t) + \varepsilon t\, U \text{ and } H(t) \subset F(t) + \varepsilon t\, U \tag{7}$$

Note that, among others, F approximates itself!

2. MAPPINGS DEFINED BY A SET OF CONSTRAINTS

As a first illustration, suppose F is defined by:

$$F(t) := \{z \in R^n \mid c_j(t,z) \leqslant 0 \textit{ for } j = 1,\ldots,m\}$$

where the "constraints" c_j are convex in z. Assume the existence of $c_j'(0,z)$, the right derivative of $c_j(\cdot,z)$ at $t = 0$ ($c_j'(0^+,z)$ would be more suggestive). Then it is natural to consider approximating F(t) by

$$H(t) := \{z \mid c_j(0,z) + t\, c_j'(0,z) \leqslant 0 \textit{ for } j = 1,\ldots,m\}. \tag{8}$$

An algorithm based on this set would then be quite in the spirit of [7].

It is possible to prove that the H of (8) does satisfy (7), provided some hypotheses hold, for example

(i) $[c_j(t,z) - c_j(0,z)]/t \rightarrow c'_j(0,z)$ *uniformly in* z, *when* $t \downarrow 0$,

(ii) *there exists* z_o *such that* $c_j(0,z_o) < 0$ *for* $j = 1,\dots,m$.

A weak point of (8) is that it is highly non-canonical. For example, perturbing the constraints to $(1+a_jt)c_j(t,z)$ gives the same F but does change H.

3. A DIRECT SET-THEORETIC CONSTRUCTION

If we examine (6) again, we see that there would be no difficulty if F(x) were a singleton: then (6) would always be consistent because F(x+d) would never be less thick than F(x), and F(x) could be subtracted. This leads to differentiating F at an *arbitrary but fixed* $y \in F(0)$. Define

$$F'_y(0) := \left\{ z \;\middle|\; \begin{array}{l} \textit{there exist } t_n \textit{ and } y_n \in F(t_n) \textit{ for } n \in \mathbb{N} \\ \textit{with } t_n \downarrow 0 \textit{ and } (y_n - y)/t_n \rightarrow z \end{array} \right\}$$

or, in a set-theoretic notation (see [2], Chapter VI):

$$F'_y(0) := \limsup_{t\downarrow 0} \, [F(t)-y]/t$$

This set is called the *contingent derivative* in [1], the (radial) *upper Dini derivative* in [6] and the *feasible set of first order* in [3]. We refer to [1] for an extensive study of F', but some remarks will be useful:

a) $F'_y(0)$ depends on the behaviour of F near y only. If we take an arbitrary $\alpha > 0$ and set $G(t) := F(t) \cap \{y + \alpha U\}$, then $G'_y(0) = F'_y(0)$.

b) If $F(t) \equiv F(0)$ does not depend on t, $F'_y(0)$ is just the tangent cone to F(O) at y.

c) Let A be a convex set in R^n, and $f : [0,1] \longmapsto R^n$ a differentiable mapping (with $f(0) = 0$ for notational simplicity). Consider $F(t) := \{f(t)\} + A$. Given $y \in F(0) = A$, call T_y the tangent cone to $F(0) = A$ at y. Then it can be shown that $F'_y(0) = \{f'(0)\} + T_y$. This is the situation when F is the approximate subdifferential of a convex quadratic function (see [4]).

d) Let n = 2. Given $r \in R$, consider $F(t) := P(t) \cap U$ with the halfspace $P(t) := \{y = (y_1,y_2) \mid y_2 \geqslant r t y_1\}$. It can be shown that, for $y = 0 \in F(0)$, $F'_o(0) = \{z = (z_1,z_2) \mid z_2 \geqslant 0\}$; $F'_o(0)$ is the same as it would be if r were 0 (in which case F(t) would be fixed), and *does not predict the rotation* of F(t) around y = 0.

Because a convex set is the intersection of the cones tangent to it, our remark b) above suggests to approximate F(t) by

$$H(t) := \cap \{y + tF'_o(0) \mid y \in F(0)\} \qquad (9)$$

Of course, this will be possible only under additional assumptions (not only due to the multi-valuedness of F; for example $F(t) := \{t \sin \log t\}$ has

$F(0) = \{0\}$, $F'_o(0) = [-1,+1]$ and $H(t) = [-t,+t]$).

Before mentioning the assumptions in question, we introduce another candidate to approximate F: for $p \in R^n$, denote by $s_p(t) := \sup \{\langle p,y\rangle \mid y \in F(t)\}$ the support function of $F(t)$. It is known that F can be described in terms of s, namely $F(t) = \{y \mid \langle p,y\rangle \leqslant s_p(t)\ \forall\, p \in R^n\}$. Then, if s_p has a (directional) derivative $s'_p(0)$, the following set is natural (see [5]):

$$G(t) := \{y \mid \langle p,y\rangle \leqslant s_p(0) + t\, s'_p(0) \quad \forall\, p \in R^n\}. \tag{10}$$

To assess these candidates (9) and (10), the following assumptions can be considered:

(i) $[s_p(t) - s_p(0)]/t \to s'_p(0)$ uniformly for $p \in U$, when $t \downarrow 0$;

(ii) $F(0)$ has a nonempty interior.

They allow to prove:

If (i) holds, then $H(t) = G(t)$; if (ii) also holds, then (7) holds.

We remark that (i) alone suffices to prove the second half of (7), which is the important one for (5) (solving $0 \in H(t)$ gives some among the possible Newton iterates); however $H(t)$ may be void if (ii) does not hold. It is also interesting to remark that, if $s'_p(0)$ is assumed to be convex in p (in which case (ii) is not needed), then it is the support function of a convex set *that we are entitled to call* $F'(0)d$ because there holds $H(t) = F(0) + t\, F'(0)d$ (due to additivity of support functions). In other words, convexity of $s'_p(0)$ gives the "easy" situation in which (6) holds.

The role of assumption (i) is more profound. It is natural to require that $F'_y(0)$ does predict the behaviour of $F(t)$ near y; this behaviour is trivial when $y \in \operatorname{int} F(0)$ (then $F(t)$ *must* contain y for all t small enough); if y is on the boundary of $F(0)$ then there is a normal cone $N_y(0)$ to $F(0)$ at y, and $s_p(0) = \langle p,y\rangle$ for $p \in N_y(0)$; hence the behaviour of $F(t)$ near y is naturally related to the behaviour of $s_p(t)$ for these normal p's (incidentally, a key result is that $F'_y(0) = \{z \mid \langle p,z\rangle \leqslant s'_p(0)\ \forall\, p \in N_y(0)\}$; (i) is essential for this). However, it is not only some technicalities in the proof that require the *uniformity* stated in (i), but rather the deficiency of F' suggested by d) above: consider the innocent mapping

$$F(t) := \{y = (y_1,y_2) \mid 0 \leqslant y_1 \leqslant 1,\ ty_1 \leqslant y_2 \leqslant 1\}.$$

Given $\alpha \in R$ and $p = (\alpha,-1)$, $s_p(t) = \max\{(\alpha-t)y_1 \mid 0 \leqslant y_1 \leqslant 1\}$ and thus, (i) is violated: when $\alpha \downarrow 0$, $s'_p(0)$ jumps from -1 to 0. For this example, $H(t) = G(t) = [0,1] \times [t,1]$, which is a poor approximation of $F(t)$. This is rather disappointing, but observe that Section 2 is well-suited for the present F.

REFERENCES

[1] J. P. Aubin, "Contingent derivatives of set-valued maps and existence of solutions to nonlinear inclusions and differential inclusions". *MRC Technical Summary Report* 2044, University of Wisconsin, Madison (1980). See also: same title, in: L. Nachbin (ed.) *Mathematical Analysis and Applications*, Academic Press (1981) 159 - 229.

[2] C. Berge, *Topological Spaces*, Mac Millan, London, 1963.

[3] V. F. Demyanov and I. M. Lupikov, "Extremal functions over the ε-subdifferential mapping", *Vestnik Leningrad University* 1 (1983) 27 - 32.

[4] J. B. Hiriart-Urruty, "ε-subdifferential calculus", in: J. P. Aubin, R. B. Vinter (eds.) *Convex Analysis and Optimization*, Pitman (1982) 43 - 92.

[5] C. Lemaréchal and E. A. Nurminskii, "Sur la différentiabilité de la fonction d'appui du sous-différentiel approché", *C. R. Acad. Sci. Paris* 290, 18 (1980) 855 - 858.

[6] J. P. Penot, "Differentiability of relations and differential stability of perturbed optimization problems", *SIAM Control* 22, 4 (1984) 529 - 551.

[7] S. M. Robinson, "Extension of Newton's method to nonlinear functions with values in a cone, *Numerische Mathematik* 19 (1972) 341 - 347.

MISCELLANIES ON NONSMOOTH ANALYSIS AND OPTIMIZATION

J.-B. Hiriart-Urruty
Paul Sabatier University, 118 route de Narbonne, 31062 Toulouse, France

People who work in the area of research concerned with the *analysis and optimization of nonsmooth functions* know they now have a panoply of "generalized subdifferentials" or "generalized gradients" at their disposal to treat optimization problems with nonsmooth data. In this short paper, which we wanted largely introductory, we develop some basic ideas about *how nonsmoothness is handled* by the various concepts introduced in the past decade.

For the sake of simplicity, we assume that the functions f considered throughout are defined and *locally Lipschitz* on some finite-dimensional space X (take $X = \mathbb{R}^n$ for example). To avoid technicalities, we suppose moreover that the (*usual*) *directional derivative*

$$d \to f'(x;d) = \lim_{\lambda \to 0^+} \frac{f(x+\lambda d) - f(x)}{\lambda} \tag{0.1}$$

exists for f at all x and for all d. As the reader easily imagines, all these assumptions have been removed in the different generalizations proposed by the mathematicians, but this is not our point here.

Clearly, f'(x;d) can also be expressed as :

$$\lim_{\substack{\lambda \to 0^+ \\ v \to d}} \frac{f(x+\lambda v) - f(x)}{\lambda}. \tag{0.2}$$

f'(x;d) is a genuine approximation of f around x. The graph of the function

$d \to f'(x;d)$ is, roughly speaking, the tangent cone to the graph of f at $(x,f(x))$. So, we have our "primal" mathematical object for approximating f,

$$\begin{aligned} f' : X \times X &\to \mathbb{R} \\ (x,d) &\to f'(x;d), \end{aligned} \tag{0.3}$$

which plays the role of a substitute for the linear mapping $d \to \langle\nabla f(x),d\rangle$. The "dual" corresponding concept is some multifunction, denoted generically by ∂f,

$$\begin{aligned} \partial f : X &\rightrightarrows X^* \\ x &\rightrightarrows \partial f(x), \end{aligned} \tag{0.4}$$

which, hopefully, will act as the gradient mapping does for differentiable functions.

1. NEEDS

Any primal object, denoted generically by $f^V(x;d)$ (i.e., $f'(x;d)$ or some generalization of it), and the corresponding dual object $\partial f(x)$ should satisfy the following properties :

. To *pass* easily from the *primal* object to the *dual* one ; the support function of $\partial f(x)$ has to be built up, in some manner, from $f'(x;d)$.

. To allow *first-order developments* and *mean-value theorems*. For the directional derivative f', we do have :

$$f\ (x+\lambda d) = f(x) + \lambda f'(x;d) + o(\lambda). \tag{1.1}$$

What is expected for ∂f to verify is :

$$f(y) - f(x) \in \langle\partial f(z),y-x\rangle \text{ for some } z \in]x,y[. \tag{1.2}$$

. In view of the properties of $(x,d) \to f'(x;d)$ or $x \rightrightarrows \partial f(x)$, one should be able to *recognize* the function f, and to *recover* it through some *integral representation* of $f(y) - f(x)$. We have that

$$f(y) = f(x) + \int_0^1 f'(x+t(y-x)\ ;\ y-x)\ dt, \tag{1.3}$$

and we expect

$$f(y) \underset{(\text{or } \in)}{=} f(x) + \int_0^1 \langle \partial f(x+t(y-x)), y-x \rangle \, dt. \qquad (1.4)$$

. *Semicontinuity* properties of the function $(x,d) \to f^{V}(x;d)$ and of the multifunction $x \rightrightarrows \partial f(x)$. These requirements are of a particular importance for algorithmic purposes.

. $f^{V}(x;d)$ and $\partial f(x)$ should be *tractable* from the computational viewpoint ; in effect, elements of $\partial f(x_n)$ are used to devise x_{n+1} in all first-order methods.

Consider for example the case of *convex* functions f. $f'(x;d)$ is itself a *convex* function of d so that the concept $\partial f(x)$, dual of $f'(x;d)$, is the so-called *subdifferential* of f at x,

$$\partial f(x) = \{x^* \mid \langle x^*, d \rangle \leq f'(x;d) \text{ for all } d \in X\}. \qquad (1.5)$$

∂f enjoys all the properties listed above. One is able to recognize a convex function when f' is at our disposal since : *f is convex if and only if* $f'(x;y-x) + f'(y;x-y) \leq 0$ *for all* x *and* y. If, instead, the generalized gradient ∂f of f is considered (cf. section 2), *f is convex if and only if* ∂f *is monotone,* that is

$$\langle \partial f(x) - \partial f(y), x-y \rangle \geq 0 \text{ for all } x,y. \qquad (1.6)$$

Mean-value theorems, integral representations, semicontinuity properties of f' and ∂f are basic facts in Convex Analysis.

Another class of functions which has played an important role in the development of nonsmooth analysis and optimization is that of maximums of C^1 functions :

$$f = \max_{i=1,\dots,k} f_i \ , \ f_i \in C^1(X).$$

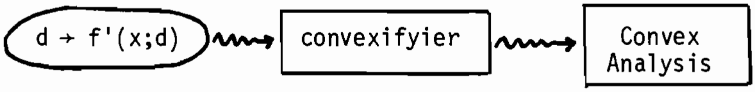
$f'(x;d)$ is a convex function of d ; it is the support function of

$$\partial f(x) = \text{co}\{\nabla f_i(x) \mid f_i(x) = f(x)\}.$$

Actually, f behaves locally like a convex function, so that handling such functions brings us back to Convex Analysis.

2. SOME ASPECTS OF THE EVOLUTION OF IDEAS (1974-1984)

Our 1977 survey paper on the various "diconvexifying" processes ([12]) remains of the present day. We will schematize here the enlightenments which have been brought up since.
Typically, dealing with nonconvex nonsmooth functions leads to the following :



With the linear mapping $d \to \ell(d) = \langle x^*,d\rangle$ is associated the dual element x^*. In a similar way, with the positively homogeneous convex function $d \to h(d)$ is associated the dual set of x^* for which $\langle x^*,d\rangle \leq h(d)$ for all d. But, since $d \to f'(x;d)$ is not convex for general nonsmooth functions f, some convexifying process has firstly to be devised for building up a positively homogeneous convex function $f^V(x;d)$. Once this step is carried out, defining $\partial f(x)$ and deriving calculus rules for it belong to the realm of Convex Analysis. So, treating of nonconvex functions relies heavily, in fine, on techniques from Convex Analysis ; that explains why researches in nonsmooth analysis and optimization are prominent in countries where there is a long standing tradition in Convex Analysis.

2.1 - Generalized subdifferentials (J.-P. PENOT, 1974)

Roughly speaking, the approach of PENOT consisted in skipping over the "convexifying operation" on $f'(x;d)$ so that the primal object $f^V(x;d)$ is $f'(x;d)$ itself. That led to the *generalized subdifferential* of f at x,

$$\partial^{\leq} f(x) = \{x^* \mid \langle x^*,d\rangle \leq f'(x;d) \text{ for all } d\}, \tag{2.1}$$

and to the generalized *superdifferential* of f at x,

$$\partial^{\geq} f(x) = \{x^* \mid \langle x^*,d\rangle \geq f'(x;d) \text{ for all } d\}. \tag{2.2}$$

Evidently $\partial^{\geq} f(x) = -\partial^{\leq}(-f)(x)$. The support function of $\partial^{\leq} f(x)$ is the biconjugate function of $d \to f'(x;d)$ and, therefore, may "slip" to $-\infty$ for all d. If $f(x) \leq g(x)$ in a neighborhood of x_0 and $f(x_0) = g(x_0)$, we then have that $\partial^{\leq} f(x_0) \subset \partial^{\leq} g(x_0)$. The vocable "generalized subdifferential" is appropriate for $\partial^{\leq} f(x_0)$ here since one is looking for the x^* such that the linear mapping $\langle x^*,d\rangle$ is a minorant of $f'(x;d)$.

f is said to be *tangentially convex* at x if $d \to f'(x;d)$ is convex, that is to say the *tangent problem* at x is convex [Following B.N. PSHENICHNYI's terminology [21], f is quasidifferentiable at x]. Tangential convexity is a property which allows to develop calculus rules on $\partial^{\leq} f$.
As we will do it for each concept, we list some advantages and drawbacks of $\partial^{\leq} f$.

Advantages

. sharp necessary conditions for optimality, keeping apart conditions for minimality ($0 \in \partial^{\leq} f(x)$) and conditions for maximality ($0 \in \partial^{\geq} f(x)$).

. nice relationship with the classical conical approximations of a set ; for example, the contingent cone to epi f (resp. hyp f) at $(x,f(x))$ is the epigraph (resp. hypograph) of $f'(x;.)$.

. mean-value theorems ; integral representations of $f(y) - f(x)$ (under some additional assumptions on f).

....

Drawbacks

. $\partial^{\leq} f(x)$ is empty too often, due to the lack of convexity of $f'(x;.)$.

. necessity of imposing assumptions like tangential convexity for the calculus to be robust.

. lack of semicontinuity of $f'(x;d)$ as a function of x.

....

2.2 - Generalized gradients (F.H. CLARKE, 1973, 1975)

The "convexifyier" of CLARKE can be described shortly as :

$$f^{o}(x;d) = \limsup_{x' \to x} f'(x';d). \tag{2.3}$$

$f^{o}(x;d)$ is therefore a regularized version of $f'(x;d)$. $f^{o}(x;.)$ is *convex* so that the *generalized gradient* of f at x, $\partial f(x)$, is the dual object associated, in a natural way, to $f^{o}(x;d)$:

$$\begin{aligned} \partial f(x) &= \{x^* \mid \langle x^*,d\rangle \leq f^{o}(x;d) \text{ for all } d\}, \\ f^{o}(x;d) &= \max_{x^* \in \partial f(x)} \langle x^*,d\rangle. \end{aligned} \tag{2.4}$$

By setting $f_0(x;d) = \liminf_{x' \to x} f'(x';d)$, we get nothing else than $-(-f)^{0}(x;d)$. Thus, the set of x^* for which $\langle x^*,d\rangle \geq f_0(x;d)$ boilds down to $\partial f(x)$ [a fact apparently missed by some authors].

Various appellations have been proposed for ∂f : epidifferential or peridifferential of f, multigradient of f, etc. "Peridifferential of f at x" is not so bad since it reminds us of the information on f we are looking for *around* x. "Generalized subdifferential" should be proscribed [∂f is the superdifferential for a concave function f]. Anyway, we stand by the original appellation "generalized gradient of f".

$\partial f(x)$ is conceptually close to the notion of derivative of f ; $\partial f(x)$ reduces to $\{Df(x)\}$ whenever f is *strictly differentiable* at x. A function f for which $f^{0}(x;d) = f'(x;d)$ for all d is called *strictly tangentially* convex at x [there is between "strict tangential convexity" and "tangential convexity" the same kind of gap there exists between "strict differentiability" and "differentiability"]. If one could rewrite mathematical history, one would say "f is tangentially linear at x" for "f is differentiable at x" [i.e., the tangent problem at x is linear] and "f is strictly tangentially linear at x" for "f is strictly differentiable at x".

Note that if $f(x) \leq g(x)$ in a neighborhood of x_0 and $f(x_0) = g(x_0)$, we only have that $\partial f(x_0) \cap \partial g(x_0) \neq \phi$.

Advantages

. $\partial f(x)$ is nonempty at all x for a very large class of functions.

. the calculus is robust ; virtually all the results holding for Df have their counterparts in terms of ∂f.

. the function $(x,d) \to f^{0}(x;d)$ as well as the multifunction $x \rightrightarrows \partial f(x)$ are upper-semicontinuous

....

Drawbacks

. $\partial f(x)$ is sometimes too large a set.

. the associated geometrical concepts (like the tangent cone) are not well adapted for nonsmooth manifolds.

. calculating effectively elements of $\partial f(x_n)$ at the $n^{\underline{th}}$ step of an algorithm might be difficult.

....

Note incidentally there is an integral estimate of f(y) - f(x) via ∂f since

$$f(y) - f(x) \in \int_0^1 <\partial f(x+t(y-x)), y-x> dt. \tag{2.5}$$

This representation is however loose, since the right-hand side may be too large and the resulting estimate not much informative.

A final remark to mention is there is a generalization of the concept of generalized gradient to vector-valued functions $F : (f_1,\ldots, f_m)^T : \mathbb{R}^n \to \mathbb{R}^m$.

The so-called *generalized Jacobian matrix* of F at x is a nonempty convex set of (n,m) matrices which take into account the possible relationships between the component functions f_i. All the other concepts extended to vector-valued $F = (f_1,\ldots, f_m)^T$ amount to considering $\overset{m}{\underset{i=1}{X}} \partial f_i(x_o)$, that is the generalized derivatives of the components f_i taken separately. This possibility of handling globally all the f_i is definitely an advantage for CLARKE's generalized derivatives. Its consequences are conspicuous in what can be called "multidifferential calculus".

2.3 - The $\star$-generalized derivatives (E. GINER, 1981)

Given f'(x;d), we are looking for a *convex*, positively homogeneous function h such that

$$h(d) \geq f'(x;d) \text{ for all } d, \tag{2.6}$$

what B.N. PSHENICHNYI calls *"an upper convex approximation of f at x"* ([23]). CLARKE's generalized directional derivative $f^o(x;.)$ is an example of such h. There is another automatic way of selecting an upper convex approximation of f at x, initiated by GINER (1981). When I moved to TOULOUSE in october 1981, GINER showed me the following way of "convexifying" a positively homogeneous function p :

$$h(d) = \sup_{u \in X} \{p(d+u) - p(u)\}. \tag{2.7}$$

h is a positively homogeneous *convex* function which majorizes p. h is moreover Lipschitz whenever p is Lipschitz over X. The functional operation $p \rightsquigarrow h$ has a geometrical interpretation by means of the so-called $\star$-difference of sets (of cones, in the present case). Given two subsets A and B, the

$\star$-difference of A and B, denoted by $A \overset{\star}{-} B$ is defined as the set of x for which $x + B \subset A$. This operation was introduced by PONTRYAGIN (1967) when dealing with linear differential games and further exploited by PSHENICHNYI (1971) in the context of Convex Analysis. It now comes clearly that :

$$\begin{aligned} \text{epi } h &= \text{epi } p \overset{\star}{-} \text{epi } p \\ &= \{x \in X \mid x + \text{epi } p \subset \text{epi } p\} \qquad (2.8) \\ &= \{x \in \text{epi } p \mid x + \text{epi } p \subset \text{epi } p\}. \end{aligned}$$

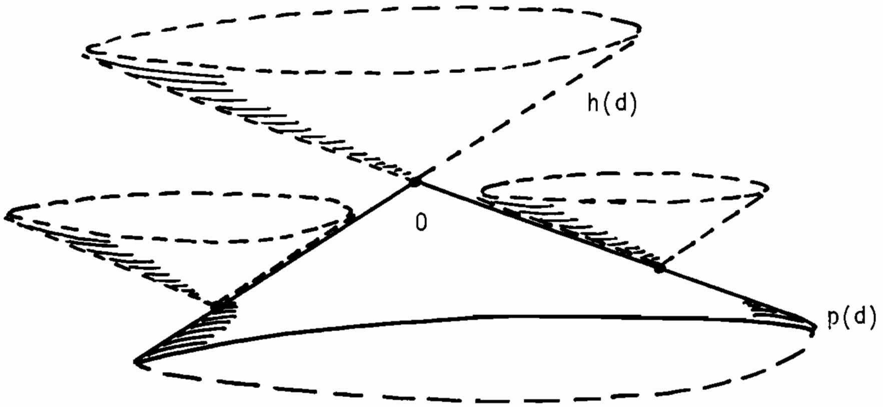
That is the reason why the convex function h built up from p in (2.7) bears the name ${}^{\star}p$. Needless to say, there is a *concave* counterpart ${}_{\star}p$ built up from p mutatis mutandis.

In a certain sense, ${}^{\star}p$ is the "minimal convex function majorizing p". To be more precise, given $d_0 \in X$,

$$p(d_0) + {}^{\star}p(d-d_0) \geq p(d) \text{ for all } d, \qquad (2.9)$$

and $h \geq {}^{\star}p$ for any positively homogeneous convex function h satisfying

$$p(d_0) + h(d-d_0) \geq p(d) \text{ for all } d_0 \text{ and } d.$$



We denote by ${}^{*}f'(x;d)$ what should be written as $[{}^{*}(f'(x;.)](d)$. The corresponding $*$-generalized derivative of f at x is defined by :

$$\partial^{*}f(x) = \{x^{*} \mid \langle x^{*},d\rangle \leq {}^{*}f'(x;d) \text{ for all } d\}. \qquad (2.10)$$

H. FRANKOWSKA (1983) got independently at the same concepts she called asymptotic directional derivative of f (= ${}^{*}f'$) and asymptotic gradient of f (= $\partial^{*}f$) respectively. The terminology comes from the fact that the asymptotic (or recession) cone of a closed convex set C is precisely $C \overset{*}{-} C$.

A wonderful thing about $\partial^{*}f$ and the generalized gradients in CLARKE's sense is the following :

<u>THEOREM</u> : *The generalized gradient of $d \to f'(x;d)$ at 0 is exactly $\partial^{*}f(x)$.*

That means, among other things, that the generalized directional derivative (in CLARKE's sense) of a positively homogeneous function p can be calculated via the formula (2.7). Furthermore, calculus rules on generalized gradients may be used for deriving calculus rules on $*$-generalized derivatives.

The proof of the theorem above is based upon the following geometrical result : CLARKE's tangent cone to a cone K at its apex is $K \overset{*}{-} K$ (cf. [5] for example).

As expected, the advantages and drawbacks of $\partial^{*}f$ are pretty much alike those of the generalized gradient ∂f.

Advantages

. $\partial^{*}f(x)$ is nonempty at all x for a large class of functions ; $\partial^{*}f(x) \subset \partial f(x)$.

. $\partial^{*}f(x)$ reduces to $\{Df(x)\}$ whenever f is differentiable at x.

. good calculus ; mean-value theorems, integral representations (without any further assumption on f).

....

Drawbacks

. lack of upper-semicontinuity of $x \to {}^{*}f'(x;d)$ [and therefore of $x \rightrightarrows \partial^{*}f(x)$].

. difficulties of calculating ${}^{*}f'(x;d)$ when f (or $f'(x;d)$) is at our disposal.

....

If $f(x) \leq g(x)$ in a neighborhood of x_0 and $f(x_0) = g(x_0)$, we have that $\partial^* f(x_0) \cap \partial^* g(x_0) \neq \phi$ (and not $\partial^* f(x_0) \subset \partial^* g(x_0)$!).

2.4 - Bidifferentials of tangentially d.c. functions (V.F. DEMYANOV and A.M. RUBINOV, 1980)

DEMYANOV and RUBINOV consider the class functions f for which $f'(x;d)$ can be written as a *difference* of two positively homogeneous convex functions :

$$f'(x;d) = p(d) - q(d). \tag{2.11}$$

The so-called d.c. functions (differences of convex functions) belong to this class as well as functions whose directional derivatives $f'(x;d)$ can be expressed as a minimum of two positively homogeneous convex functions.
DEMYANOV and RUBINOV use the vocable quasidifferentiable for the functions for which (2.11) holds true, a term borrowed from PSHENICHNYI ([21]). In accordance with the terminology used earlier in this paper, we call these functions *tangentially d.c.* (i.e., the tangent problem is d.c.).
$f'(x;d)$ is thus the difference of two support functions p and q,

$$f'(x;d) = \max_{x^* \in A} \langle x^*, d\rangle - \max_{x^* \in B} \langle x^*, d\rangle \tag{2.12}$$

$$= \max_{x^* \in A} \langle x^*, d\rangle + \min_{x^* \in -B} \langle x^*, d\rangle. \tag{2.13}$$

The sets A and B are not uniquely determined since one could add a support function to the support function of A and cut if off from the support function of B, without altering $f'(x;d)$. However, provided a relation of equivalence is used, the sets A and B are associated with $f'(x;d)$ and the pair (A,B) [actually $(A,-B)$ in the formulation (2.13) used by DEMYANOV and RUBINOV], is the *bidifferential* of f at x. This bidifferential, denoted as $(\underline{\partial} f(x), \overline{\partial} f(x))$, includes a subdifferential $\underline{\partial} f(x)$ (taking into account the convex part of $f'(x;d)$) and a superdifferential $\overline{\partial} f(x)$ (reflecting the concave contribution of $f'(x;d)$).
Now, calculus on $(\underline{\partial} f(x), \overline{\partial} f(x))$ amounts to using Convex Analysis *twice* !

Advantages

. conceptually close to the usual directional derivative $f'(x;d)$.

. separates the "convex part" and the "concave part" of $f'(x;d)$; sharp optimality conditions.

. mean-value theorems, etc.

....

Drawbacks

. the bidifferential is actually a class of equivalence ; there is no automatic way of selecting a representative of it.

. heavy calculus rules.

. no geometrical interpretation for $(\underline{\partial}f(x), \overline{\partial}f(x))$.

. lack of upper-semicontinuity of $x \rightrightarrows (\underline{\partial}f(x), \overline{\partial}f(x))$.

....

A way of taking something which is unambiguously associated with the class of equivalence $(\underline{\partial}f(x), \overline{\partial}f(x))$ is to consider $\underline{\partial}f(x) \overset{*}{-} \overline{\partial}f(x)$ and $\overline{\partial}f(x) \overset{*}{-} \underline{\partial}f(x)$. It is an easy exercise to verify that

$$\begin{aligned} \underline{\partial}f(x) \overset{*}{-} \overline{\partial}f(x) &= \partial^{\leq}f(x) \\ \overline{\partial}f(x) \overset{*}{-} \underline{\partial}f(x) &= -\partial^{\geq}f(x). \end{aligned} \qquad \text{(see §2.1)}$$

So, for tangentially d.c. functions, necessary conditions for optimality become :

$$0 \in \partial^{\leq}f(x) \Longleftrightarrow 0 \in \underline{\partial}f(x) \overset{*}{-} \overline{\partial}f(x) \Longleftrightarrow \overline{\partial}f(x) \subset \underline{\partial}f(x)$$

(necessary condition for minimality)

$$0 \in \partial^{\geq}f(x) \Longleftrightarrow 0 \in \overline{\partial}f(x) \overset{*}{-} \underline{\partial}f(x) \Longleftrightarrow \underline{\partial}f(x) \subset \overline{\partial}f(x)$$

(necessary condition for maximality).

The problem of selecting a representative of $(\underline{\partial}f(x), \overline{\partial}f(x))$ is related to that of finding the "best" decomposition of $f'(x;d)$ as a difference of two support functions p and q ; the same problem arises in decomposing d.c. functions ([6, 14]).

When we say there is no geometrical base for $(\underline{\partial}f(x), \overline{\partial}f(x))$, we are actually posing a question : is there some tangent "bicone" around ?

3. RECOGNIZING FUNCTIONS f AND RECOVERING THEM FROM f', ∂f

Given a multifunction $\Gamma : X \to X^*$, is Γ the generalized derivative (in some sense) of a function $f : X \to \mathbb{R}$? There is no full answer to this question, whatever the kind of generalized derivative we are considering. In particular, the generalized gradient multifunction (in CLARKE's sense) may be very "bizarre". A more sensible question is : knowing that Γ is a generalized derivative multifunction of a function f, what kind of properties of Γ could serve to characterize f ?

$$\boxed{\ \Gamma = \partial f \text{ is } \dots\ } \Longleftrightarrow \boxed{\ f \text{ is } \dots\ }$$

A strongly related question is : how to recover f from ∂f ?

$$f(y) - f(x) = \int_0^1 <\partial f(x+t(y-x)), y-x> dt \ ? \tag{3.1}$$

Recovering f from the directional derivative offers no problem ; but properties of "derivatives" are better expressed in terms of ∂f, so that the question (3.1) arises.

Classifying nonsmooth functions can be splitted up into two parts :

(1) Having the definition of a class of functions, what is the characterization of such functions in terms of ∂f or f'(.,.) ?

(2) Defining a class of functions via ∂f, what is an equivalent definition in terms of the function f itself ?

Let us mention some classes of functions used in nonsmooth optimization :

Conv(X) : convex functions on X ;

QC(X) : quasi-convex functions on X ;

$LC^k(X)$: lower - C^k functions on X ;

SS(X) : semi-smooth functions on X ;

DC(X) : differences of convex functions on X.

We have that :

$$\left.\begin{matrix} \mathrm{Conv}(X) \\ C^2(X) \end{matrix}\right\} \subset LC^2(X) \subset DC(X) \subset SS(X).$$

Convex or lower-C^2 functions enjoy a characterization via f or CLARKE's generalized gradient ∂f of f :

f is convex if and only if ∂f is monotone ;

f is lower-C^2 if and only if ∂f is strictly hypomotone ([25]).

D.c. functions are, *by definition*, differences of convex functions. To characterize them in terms of ∂f is a difficult task ; see [6, Ch. II] for the first fruits in that respect. Even for d.c. functions, it may happen that $\partial^* f$ differs from ∂f ; see [14, §1] for an example of d.c. function for which $\partial^* f(x_0) = \{Df(x_0)\}$ and $\partial f(x_0)$ contains other elements than $Df(x_0)$.

Semismooth functions are, on the contrary, *defined* through a property of ∂f or $f'(.,.)$; what such properties mean equivalently on f is unclear.

Quasi-convex functions are defined analytically,

$$f(\lambda x+(1-\lambda)y) \leq \max\{f(x),f(y)\} \text{ for all } x,y \text{ and } \lambda \in [0,1],$$

or geometrically

$$\{x \in X \mid f(x) \leq \alpha\} \text{ is convex for all } \alpha \in \mathbb{R}.$$

A characterization of quasi-convex functions, similar to the one known for differentiable quasi-convex functions, is a follows :

<u>THEOREM</u> ([10, Ch. III]) : *Let f be merely locally Lipschitz on X. Then f is quasi-convex on X if and only if the following property holds true :*

$$\forall x,\ x' \in X \qquad f(x') < f(x) \Longrightarrow \langle x'-x,\ \partial f(x)\rangle \leq 0.$$

Unfortunately, this characterization uses both f and ∂f. It is desirable to find a characterization based upon ∂f only ; this has been done by HASSOUNI ([10, Ch. III]).

Following HASSOUNI, a multifunction $\Gamma : X \rightrightarrows X^*$ is said to be *quasi-monotone* in the direction $d \in X$ if, for all $x \in X$, there exists $\bar{\lambda} \in \mathbb{R}$ such that

$$\operatorname{sign}(\lambda-\bar{\lambda}) \,.\, \langle\Gamma(x+\lambda d),\ d\rangle \subset \mathbb{R}^+ \text{ for all } \lambda \in \mathbb{R},$$

where sign u = 1 if u > 0, -1 if u<0, 0 if u = 0.

Observe that $\bar{\lambda}$ may be $+\infty$ or $-\infty$ in the requirement above. Also all the x' on the line $x_0 + \mathbb{R}d$ give rise to the same condition ; only the direction d is relevant.

Γ is called quasi-monotone if it is quasi-monotone in all directions of X. As expected, a monotone Γ is quasi-monotone.

THEOREM ([10, Ch. III]) : *A locally Lipschitz f is quasi-convex if and only if the generalized gradient multifunction ∂f is quasi-monotone.*

The proof reduces to the one-dimensional case since quasi-convexity is a "radial" notion ; it has however to overcome the difficulty that the generalized gradient of $f_{x,d} : \lambda \to f(x+\lambda d)$ does not necessarily equal $<\partial f(x+\lambda d), d>$.

IV. CONCLUSION AND CURRENT TRENDS

The presentation we have made here is somewhat sketchy. Virtually all the mathematicians who have contributed substantially to the area of nonsmooth analysis and optimization have proposed their own "generalized derivative" or "generalized subdifferential". The reader interested in going more deeply in the subject will find in the bibliographies [9] and [18] most of the appropriate references.

Concerning the first-order generalized differentiation of nonsmooth functions, we think the golden age is over for researches in this area, even if several problems remain unsolved. Theories are now solidifyied at least for *real-valued* functions. The researches which are pursued can be described in the following manner :

. *classification* of nonsmooth functions and optimization problems, this classification using in most of the cases the various concepts of generalized derivatives we discussed about.

. *applications* of the new tools and methods to problems which are nonsmooth "by nature" : problems from Mathematical Economy, Optimal Control and Calculus of Variations, as also Mechanics. In spite of continuous efforts, the studies in view of dealing with *vector-valued* functions (i.e., functions taking values in an infinite-dimensional space) are neither quite satisfactory nor complete. There is a strong demand from Nonlinear Analysis (bifurcation theory, etc) for tools like implicit function theorems, inverse function theorems for nonsmooth data.

. *Fall-out in Nonsmooth Analysis and Geometry.* New geometrical notions of "tangency" and "normality" are associated with the generalized gradients. For "thin" sets like Lipschitz manifolds, all the *convex* normal cones derived from first-order differentiation are too small (they reduce to {0} at the corners of the manifold). Attempts by the author to define a "normal subcone" to the set $S = \{x \mid h(x) = 0\}$, h Lipschitz function, depend on the function h used for representating S as an equality constraint.
It is clear that much more work should be done to better understand the geometrical structure of Lipschitz manifolds.

A very promising area of research is now the *generalized second-order differentiation* of nonsmooth functions. Various generalized second-order directional derivatives have been studied in the literature, some of them quite recently. It remains that no satisfactory (= tractable) definition of $\partial^2 f(x)$ has come out as yet.

REFERENCES

[1] F.H. CLARKE, *Generalized gradients and applications,* Trans. Amer. Math. Soc. 205, (1975) 247-262.

[2] F.H. CLARKE, *Nonsmooth analysis and optimization,* J. Wiley Interscience, 1983.

[3] V.F. DEMYANOV and A.M. RUBINOV, *On quasidifferentiable functionals,* Soviet Math. Dokl. Vol. 21, (1980), 14-17.

[4] V.F. DEMYANOV and A.M. RUBINOV, *On quasidifferentiable mappings,* Math. Operationsforsch. u. Stat. ser. Optimization 14, (1983), 3-21.

[5] S. DOLECKI, *Hypertangent cones for a special class of sets,* in "Optimization : theory and algorithms", J.-B. Hiriart-Urruty, W. Oettli, J. Stoer, ed., Marcel Dekker, Inc., (1983) 3-11.

[6] R. ELLAIA, *Contribution à l'analyse et l'optimisation de différences de fonctions convexes,* Thèse de 3$^{\underline{e}}$ cycle de l'Université Paul Sabatier, 1984.

[7] H. FRANKOWSKA, *Inclusions adjointes associées aux trajectoires minimales d'inclusions différentielles*, Note aux C.R. Acad. Sc. Paris, t. 297, Série I, (1983) 461-464.

[8] E. GINER, *Ensembles et fonctions étoilés ; applications à l'optimisation et au calcul différentiel généralisé*, Manuscript Université Paul Sabatier (1981).

[9] J. GWINNER, *Bibliography on nondifferentiable optimization and nonsmooth analysis*, J. of Computational and Applied Mathematics 7, (1981) 277-285.

[10] A. HASSOUNI, *Sous-différentiels des fonctions quasi-convexes*, Thèse de 3º cycle de l'Université Paul Sabatier, 1983.

[11] J.-B. HIRIART-URRUTY, *Conditions nécessaires d'optimalité en programmation non différentiable*, Note aux C.R. Acad. Sc. Paris, t. 283, Série A, (1976) 843-845.

[12] J.-B. HIRIART-URRUTY, *New concepts in nondifferentiable programming*, Actes des journées d'analyse non convexe (Pau, 1977), Bull. Soc. Math. de France, Mémoire n°60, (1979) 57-85.

[13] J.-B. HIRIART-URRUTY, *Un concept récent pour l'analyse et l'optimisation de fonctions non différentiables : le gradient généralisé*. Publications de l'I.R.E.M. de Clermont-Ferrand, (1980) 28 p.

[14] J.-B. HIRIART-URRUTY, *Generalized differentiability, duality and optimization for problems dealing with differences of convex functions*, Lecture Notes in Mathematics, to appear in 1985.

[15] A. IOFFE, *Nonsmooth analysis : differential calculus of nondifferentiable mappings*, Trans. Amer. Math. Soc. 266, (1981) 1-56.

[16] A. IOFFE, *New applications of nonsmooth analysis to nonsmooth optimization*, in "Mathematical Theories of Optimization", J.P. Cecconi and T. Zolezzi, ed., Lecture Notes in Mathematics 979, (1983) 178-201.

[17] R. MIFFLIN, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control and Optimization 15, (1977) 959-972.

[18] E. NURMINSKII, *Bibliography on nondifferentiable optimization*, in "Progress in nondifferentiable optimization", E. Nurminskii, ed., Pergamon Press, (1981) 215-257.

[19] J.-P. PENOT, *Sous-différentiels de fonctions numériques non convexes*, Note aux C.R. Acad. Sc. Paris, t. 278, Série A, (1974) 1553-1555.

[20] J.-P. PENOT, *Calcul sous-différentiel et optimisation*, J. of Funct. Analysis 27, (1978) 248-276.

[21] B.N. PSHENICHNYI, *Necessary conditions for an extremum*, Marcel Dekker, N.Y., 1971.

[22] B.N. PSHENICHNYI, in "*Contrôle optimal et jeux différentiels*", Cahiers de l'I.R.I.A. n°4 (1971).

[23] B.N. PSHENICHNYI and R.A. HAČATRJAN, *Constraints of equality type in nonsmooth optimization problems*, Soviet. Math. Dokl. 26, (1982) 659-662.

[24] R.T. ROCKAFELLAR, *The theory of subgradients and its applications to problems of optimization : convex and nonconvex functions*, Helderman Verlag, W. Berlin, 1981.

[25] R.T. ROCKAFELLAR, *Favorable classes of Lipschitz continuous functions in subgradient optimization*, in "Progress in nondifferentiable optimization", E. Nurminskii, ed., Pergamon Press, (1981) 125-143.

BUNDLE METHODS, CUTTING-PLANE ALGORITHMS AND σ-NEWTON DIRECTIONS

C. Lemaréchal[1] and J.J. Strodiot[2]
[1] *INRIA, P.O. Box 105, 78153 Le Chesnay, France*
[2] *FNDP, Rempart de la Vièrge 8, 5000 Namur, Belgium*

1. INTRODUCTION

Recently Lemaréchal and Zowe [7] have introduced a theoretical second-order model for minimizing a real, not necessarily differentiable, convex function defined on $\mathbb{R}^n$. This model approximates the convex function f along any fixed direction d and is based on the variation with respect to σ of the perturbed directional derivative $f'_\sigma(x,d)$ (all definitions in convex analysis used in this paper can be found in the classical book by Rockafellar [9]). With this help, a second-order expansion of $f(x+d) - f(x)$, depending on $\sigma \geq 0$, is obtained at the current iterate x and a σ-Newton direction is naturally defined as a direction which minimizes this expansion (when f is twice continuously differentiable on a neighborhood of x and $\sigma = 0$, then this direction coincides with the classical Newton direction).

If the subdifferential $\partial f(x)$ is approximated by a singleton $\{g_k\}$ and the σ-subdifferential $\partial_\sigma f(x)$ by some convex compact set G_σ such that $0 \notin G_\sigma$, then a σ-Newton direction (relative to g_k and G_σ) is a vector d of norm 1 satisfying :

$$\max_{g^* \in G_\sigma} \langle g^*, d\rangle = \langle t_\sigma\, g_k, d\rangle < 0 \tag{1}$$

where $\langle\ ,\ \rangle$ denotes the usual scalar product and t_σ the smallest number $t > 0$ such that $t\, g_k \in G_\sigma$. Condition (1) means that the hyperplane defined by d in $\mathbb{R}^n$ supports G_σ at $t_\sigma\, g_k$ and separates G_σ strictly from the origin. As observed in [6], the model is really interesting when $t_\sigma < 1$, in the sequel it will be assumed that $0 < t_\sigma < 1$.

Our purpose in this paper is to prove that if G_σ is the usual polyhedral approximation of many bundle methods (see, e.g. [6], [4], [8], [3]) then finding a σ-Newton direction is equivalent to solving a variant of the cutting plane problem, in which one of the linear pieces is imposed to be active. We also show that a σ-Newton direction can be interpreted in terms of the perturbed second order derivative given in [5], [1].

2. PRELIMINARY RESULTS

Let $x_1, \ldots, x_k$ be the iterates generated by the algorithm and let $g_1, \ldots, g_k$ be the corresponding subgradients. As usual, at each subgradient g_i, $1 \le i \le k$, is associated a weight p_i [6] defined by : $p_i = f(x_k) - f(x_i) - \langle g_i, x_k - x_i \rangle$. For $\sigma \ge 0$, $\partial_\sigma f(x_k)$ is approximated by the convex compact polyhedron

$$G_\sigma = \{ \sum_{i=1}^{k} \lambda_i g_i \mid \lambda_i \ge 0, 1 \le i \le k, \sum_{i=1}^{k} \lambda_i = 1, \sum_{i=1}^{k} \lambda_i p_i \le \sigma \}$$

Throughout, we will assume that $p_k = 0$ and $p_i > 0$, $1 \le i \le k-1$; observe that g_k belongs to G_σ. The following lemma gives the extreme points of G_σ when σ is small (throughout, we use the notation Σ_+ for $\sum_{i=1}^{k}$ and Σ_- for $\sum_{i=1}^{k-1}$).

LEMMA 1. If $\sigma \le p_i$, $1 \le i \le k-1$, then

$$G_\sigma = \{ \Sigma_+ \mu_i \gamma_i \mid \mu_i \ge 0, 1 \le i \le k, \Sigma_+ \mu_i = 1 \}$$

where $\gamma_i = g_k + \frac{\sigma}{p_i} (g_i - g_k)$ for $i < k$ and $\gamma_k = g_k$.

PROOF. Let $g = \mu_k g_k + \Sigma_- \mu_i \gamma_i$ with $\mu_i \ge 0$ and $\Sigma_+ \mu_i = 1$. Then

$$g = [1 - \Sigma_- \sigma \mu_i / p_i] g_k + \sigma \Sigma_- \mu_i g_i / p_i .$$

Set $\lambda_o = 1 - \Sigma_- \mu_i \sigma / p_i \ge 1 - \Sigma_- \mu_i p_i / p_i = \mu_k \ge 0$ and $\lambda_i = \mu_i \sigma / p_i \ge 0$,

to observe that $\Sigma_+ \lambda_i = 1$ and

$$\Sigma_+ \lambda_i p_i = 0 + \Sigma_- \lambda_i p_i = \Sigma_- \mu_i \sigma = (1 - \mu_k)\sigma \le \sigma, \text{ so } g \in G_\sigma .$$

The converse inclusion is proved through a similar calculation. ■

The next lemma relates G_σ and the function used in the cutting plane algorithm.

LEMMA 2

$$G_\sigma = \partial_\sigma \tilde{f}(x_k)$$

where

$$\tilde{f}(x) = \max_{1 \le i \le k} \{ f(x_k) - p_i + \langle g_i, x - x_k \rangle \}$$

PROOF. Set $f_i(x) = f(x_k) - p_i + \langle g_i, x - x_k \rangle$ and observe that $\tilde{f}(x) = \max f_i(x)$; for all $\alpha \ge 0$, $\partial_\alpha f_i(x) = \{g_i\}$ and $f_i(x_k) = \tilde{f}(x_k) - p_i$.

Then use a result of Hiriart- Urruty ([2], see also [10]) to obtain the desired result. ∎

In a bundle algorithm, the direction is computed by minimizing $\tilde{f}(x) + \frac{1}{2} u \, ||x - x_k||^2$ for given $u \geq 0$. Choosing $u=0$ gives the cutting plane algorithm. Here, a variant of the cutting plane algorithm is considered, in which the last linear function is imposed to be active at the optimum. More precisely, consider the problem.

$$\begin{cases} \underset{x}{\text{Minimize}} \ \tilde{f}(x) \\ \\ \text{s.t.} \ \tilde{f}(x) = f(x_k) + \langle g_k, x-x_k \rangle , \end{cases}$$

or equivalently

$$(CP) \begin{cases} \underset{v,x}{\text{Minimize}} \ v \\ \\ \text{s.t.} \quad v = f(x_k) + \langle g_k, x-x_k \rangle \\ \\ \qquad\quad v \geq f(x_k) - p_i + \langle g_i, x-x_k \rangle \qquad i = 1,\dots,k-1. \end{cases}$$

Eliminating v, this problem is nothing else than finding $d = x - x_k$ solution of the following program :

$$(P) \begin{cases} \underset{d}{\text{Minimize}} \ \langle g_k, d \rangle \\ \\ \text{s.t.} \ \langle g_i - g_k, d \rangle \leq p_i \qquad\qquad i = 1,\dots,k-1. \end{cases}$$

It is a linear programming problem whose dual is

$$(D) \begin{cases} \text{Minimize} \ \Sigma\_ \ \lambda_i \, p_i \\ \\ \text{s.t.} \ \Sigma\_ \ \lambda_i (g_i - g_k) + g_k = 0 \\ \\ \qquad \lambda_i \geq 0 \qquad\qquad\qquad i = 1,\dots,k-1. \end{cases}$$

When $0 < \sigma \leq p_i$, $1 \leq i \leq k-1$, then, using the definition of γ_i in Lemma 1 and setting $\lambda_i \, p_i = \sigma \, \mu_i$, one sees that (D) can be written :

$$(D') \begin{cases} \text{Minimize} \ \Sigma\_ \ \mu_i \\ \\ \text{s.t.} \ \Sigma\_ \ \mu_i (\gamma_i - g_k) + g_k = 0 \\ \\ \qquad \mu_i \geq 0, \ i = 1,\dots,k-1. \end{cases}$$

The following lemma characterizes the length t_σ in terms of the solution of (D) or (D').

LEMMA 3. If $0 < t_\sigma < 1$ and $0 < \sigma \le p_i$, $1 \le i \le k-1$, then (D) and (D") are feasible and there exists at least one solution to problems (P), (D) and (D'). Moreover if d^* denotes a soltution to (P), $\lambda^* = (\lambda_1^*,\ldots,\lambda_{k-1}^*)$ a solution to (D) and $\mu^* = (\mu_1^*,\ldots,\mu_{k-1}^*)$ a solution to (D') then $d^* \neq 0$,

$$t_\sigma = 1 - \sigma/\Sigma_- \lambda_i^* p_i = 1 - 1/\Sigma_- \mu_i^*$$

and

$$<g_k,d^*> = - \Sigma_- \lambda_i^* p_i = -\sigma \Sigma_- \mu_i^*.$$

PROOF. Take $t < 1$ such that $tg_k \in G_\sigma$. By Lemma 1, there exist $\nu_i \ge 0$, $1 \le i \le k$ such that $\Sigma_+ \nu_i = 1$ and

$$tg_k = \Sigma_- \nu_i \gamma_i + \nu_k g_k = \Sigma_- \nu_i(\gamma_i - g_k) + g_k.$$

Hence $\{\nu_i/(1-t)\}$ is feasible in (D'), which has an optimal solution $\{\mu_i^*\}$ satisfying :

$$0 \le \Sigma_- \mu_i^* \le \Sigma_- \nu_i/(1-t)$$

so that

$$t \Sigma_- \mu_i^* \ge \Sigma_- \mu_i^* - \Sigma_- \nu_i \ge \Sigma_- \mu_i^* - 1. \qquad (2)$$

Now let $\{\mu_i\}$ be feasible in (D'). Then

$$(1 - \Sigma_- \mu_i) g_k + \Sigma_- \mu_i \gamma_i = 0.$$

Because we have assumed $0 \notin G_\sigma$, this implies that $\Sigma_- \mu_i > 1$ and, dividing by $\Sigma_- \mu_i$, we obtain

$$(1 - 1/\Sigma_- \mu_i) g_k = \Sigma_- \mu_i \gamma_i/\Sigma_- \mu_i \in G_\sigma.$$

Hence $t_\sigma \le 1 - 1/\Sigma_- \mu_i$; equality follows from (2), and the rest of the Lemma is a consequence of duality theory. ∎

3. CHARACTERIZATION OF σ-NEWTON DIRECTIONS

The next theorem makes precise the relationship between σ-Newton directions and solutions of problem (P).

THEOREM 1 . If $0 < t_\sigma < 1$ and $0 < \sigma \le p_i$, $1 \le i \le k-1$, then

(i) for each σ-Newton direction d, αd is a solution of (P) where

$$\alpha = - \frac{\text{optimal value (D)}}{<g_k, d>} > 0$$

(ii) for each solution d of (P), the direction $d/||d||$ is a σ-Newton direction.

PROOF.

(i) By the strong duality theorem in linear programming and Lemma 3, it is sufficient to prove that αd is feasible for (P) and $\langle g_k, \alpha d\rangle = \Sigma_- \lambda_i^* p_i$ where λ^* is a solution of (D).

The above equality just results from the definition of α, and it remains to prove that αd is feasible for (P), i.e.,

$$\langle g_i - g_k, \alpha d\rangle \le p_i \qquad i = 1,\dots,k-1,$$

or equivalently that

$$\sigma\langle g_i - g_k, \alpha d\rangle/p_i \le \sigma \qquad i = 1,\dots,k-1 \tag{3}$$

As $\gamma_i = g_k + \sigma(g_i - g_k)/p_i \in G_\sigma$ (see Lemma 1) and as d is a σ-Newton direction we deduce successively for each $i = 1,\dots,k-1$ that

$$\begin{aligned}\sigma\langle g_i-g_k, \alpha d\rangle/p_i &= \langle \gamma_i, \alpha d\rangle - \langle g_k, \alpha d\rangle,\\ &\le t_\sigma\langle g_k, \alpha d\rangle - \langle g_k, \alpha d\rangle,\\ &= (t_{\sigma-1})\ \langle g_k, \alpha d\rangle,\end{aligned}$$

which is precisely inequality (3) if we replace α and t_σ by their value.

(ii) Let d be a solution of (P). As $\langle g_k, d\rangle < 0$ and $t_\sigma\, g_k \in G_\sigma$ it is sufficient to prove that

$$\langle g, d\rangle \le t_\sigma\ \langle g_k, d\rangle \qquad \forall g \in G_\sigma.$$

Let $g \in G_\sigma$. Then $g = \Sigma_+\lambda_i\, g_i$ with $\lambda_i \ge 0$, $1 \le i \le k$, $\Sigma_+ \lambda_i = 1$ and $\Sigma_+ \lambda_i\, p_i \le \sigma$. As d is a solution of (P) we deduce successively that

$$\begin{aligned}\langle g, d\rangle = \langle \Sigma_+ \lambda_i\, g_i, d\rangle &= \Sigma_- \lambda_i\ \langle g_i-g_k, d\rangle + \langle g_k, d\rangle\\ &\le \Sigma_- \lambda_i\, p_i + \langle g_k, d\rangle \ \le \sigma + \langle g_k, d\rangle\end{aligned} \tag{4}$$

On the other hand, by using Lemma 3, we obtain that

$$\begin{aligned}t_\sigma\langle g_k, d\rangle = \left(1 - \frac{\sigma}{\Sigma_- \lambda_i^* p_i}\right) \langle g_k, d\rangle &= \langle g_k, d\rangle - \frac{\sigma}{\Sigma_- \lambda_i^* p_i}\ \langle g_k, d\rangle\\ &= \langle g_k, d\rangle + \sigma\end{aligned} \tag{5}$$

The result follows then from (4) and (5). ■

Because (P) may have several solutions there may exist several σ-Newton directions. In that case,Lemarechal and Zowe [7] suggest to select the best hyperplane which supports G_σ at $t_\sigma\, g_k$ and separates G_σ strictly from the origin. They solve

$$(N)\quad \begin{cases} \text{Maximize } \frac{1}{2}\, t_\sigma^2 <g_k, d>^2 \\ \text{s.t. } d \in \mathcal{D} \end{cases}$$

and show that (N) has a unique solution ; here $\mathcal{D}$ denotes the set of σ-Newton directions. The next result relates (N) and (P).

COROLLARY 1. Let d^*, be the unique solution of

$$\begin{cases} \text{Minimize } ||d|| \\ \text{s.t. } d \text{ is a solution of (P).} \end{cases}$$

Then $d^*/||d^*||$ solves (N).

PROOF. By theorem 1, $d^*/||d^*||$ is a σ-Newton direction and it remains to prove that $<g_k, d>^2 \le <g_k, \frac{d^*}{||d^*||}>^2$ for each $d \in \mathcal{D}$. Let $d \in \mathcal{D}$. Then $||d|| = 1$ and by Theorem 1, αd is a solution of (P) for α satisfying the relation

$$0 < \alpha = \frac{-(\text{optimal value of (D)})}{<g_k, d>} = \frac{<g_k, d^*>}{<g_k, d>}$$

By definition of d^*, we have $||d^*|| \le ||\alpha d|| = \alpha$ and consequently $||d^*|| \le \frac{|<g_k, d^*>|}{|<g_k, d>|}$, which is just the announced result. ∎

In terms of problem (CP), selecting the best hyperplane means choosing, among all the solutions x of (CP), the one which is nearest to x_k.

We conclude this paper with a further interpretation of σ-Newton directions.

A way to introduce the classical Newton method is to consider the second derivative $(f''(x)\, d,d)$ as the square of a norm to compute the steepest descent direction by solving

$$\begin{cases} \text{Minimize}(f'(x), d) \\ \text{s.t. } (f''(x)\, d, d) \le 1 \end{cases}$$

Here we can do the same. Taking $\tilde{f}$ instead of f (in order to obtain something implementable) and considering the perturbed second order directional derivation $f_\sigma''(x, d, d)$ (given in [5], [1]).we are led to compute the direction by solving

$$(P')\ \begin{cases}\text{Minimize } \tilde{f}'(x_k,d)\\ \text{s.t. } \tilde{f}''_\sigma(x_k, d, d) \le M^2\end{cases}$$

Because of positive homogeneity, the direction thus obtained is independent of $M > 0$. We claim that (P') is equivalent to (P). For this, we need to characterize $\tilde{f}''_\sigma(x_k, d, d)$.

LEMMA 4. Assume $0 < t_\sigma < 1$ and $0 < \sigma \le p_i$, $i \le k-1$. Let d be such that $<g_k,d> < 0$. Then

(i) there exists $i \le k-1$ such that $<g_i-g_k,d> > 0$

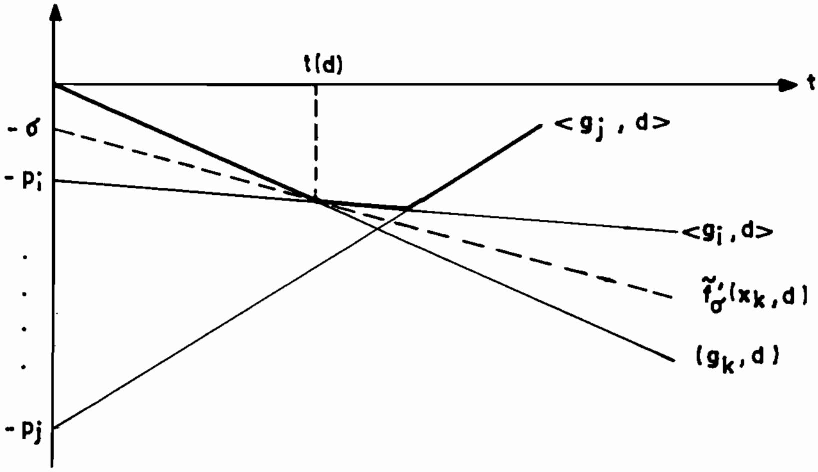
(ii) $\tilde{f}'_\sigma(x_k, d) = \tilde{f}'(x_k, d) + \sigma/t(d)$

(iii) $\tilde{f}''_\sigma(x_k, d, d) = [\tilde{f}'_\sigma(x_k, d) - \tilde{f}'(x_k,d)] / t(d)$

where $t(d) = \min \{\frac{p_i}{<g_i-g_k, d>} / <g_i-g_k, d> > 0\}$

PROOF. If (i) were false, then (P) would have no optimal solution, contradicting Lemma 3.

Then, drawing the graph of the functions $-p_i + t <g_i,d>$, $i \le k$ and of the function $-\sigma + t\, \tilde{f}'_\sigma(x_k, d)$,



it can be seen that t(d) is the smallest solution of

$$\inf_{t>0} [\tilde{f}(x_k + td) - \tilde{f}(x_k) + \sigma]/ t$$

ant that $\tilde{f}(x_k + td) = \tilde{f}(x_k) + t <g_k,d>$ if $0 \le t \le t(d)$.

This implies (ii) and then (iii) is just the definition of $\tilde{f}''_\sigma(x_k,d,d)$. ∎

If d is such that $\langle g_i-g_k, d\rangle \le 0$, $1 \le i \le k$, then $\tilde{f}'_\sigma(x_k,d) = \tilde{f}'(x_k,d)$ and $\tilde{f}''_\sigma(x_k,d,d) = 0$. Lemma 4 says that, in this case, $\langle g_k,d\rangle \ge 0$.

THEOREM 2. If $0 < t_\sigma < 1$, $0 < \sigma \le p_i$, $i = 1,\dots,k-1$ and $M = \sqrt{\sigma}$, then(P') is (P).

PROOF. Because d = 0 is feasible, we have to consider in (P') only those d for which $f'(x_k,d) = \langle g_k,d\rangle < 0$. Thus we can apply Lemma 4 to write (P') in the form

$$\begin{cases} \text{Minimize } \langle g_k,d\rangle \\ \text{s.t. } t(d) \text{ exists} \\ \dfrac{\sigma}{t^2(d)} \le M^2 \end{cases}$$

in which the last constraint can be expressed as

$$\sqrt{\sigma}\, \langle g_i-g_k, d\rangle / p_i \le M \text{ for } i \text{ such that } \langle g_i-g_k, d\rangle > 0.$$

Obviously, any d satisfying this condition does satisfy the same condition for all i. In other words, (P') can be written

$$\begin{cases} \text{Minimize } \langle g_k, d\rangle \\ \langle g_i-g_k, d\rangle \le p_i \quad M / \sqrt{\sigma} \qquad 1 \le i \le k \end{cases}$$

which is (P) if $M = \sqrt{\sigma}$ ∎

REFERENCES

[1] Auslender A. On the differential properties of the support function of the ε-subdifferential of a convex function. Math. Prog. 24, 3 (1982) 257-268

[2] Hiriart-Urruty J.B. ε-subdifferential calculus. In: Convex Analysis and Optimization, Research Notes in Mathematics Series 57 (Pitman Publishers, 1982).

[3] Kiwiel K.C. An aggregate subgradient method for nonsmooth convex minimization. Math. Prog. 27, 3 (1983) 320-341.

[4] Lemaréchal C. Nonsmooth optimization and descent methods. IIASA Report RR 78-4 (1978).

[5] Lemaréchal C., Nurminskii E.A. Sur la différentiabilité de la fonction d'appui du sous-différentiel approché. Comptes-rendus Acad. Sc. Paris 290 A (1980) 855-858.

[6] Lemaréchal C., Strodiot J.J., Bihain A. On a bundle algorithm for nonsmooth optimization. in : Mangasarian, Meyer, Robinson (Eds) Nonlinear Programming 4 (Academic Press, 1981) 245-282.

[7] Lemaréchal C., Zowe J. Some remarks on the construction of higher order algorithms in convex optimization. J. Appl. Math. and Opt. 10 (1983) 51-68.

[8] Mifflin R. A modification and an extension of Lemaréchal's algorithm for nonsmooth optimization. Math. Prog. Study 17 (1982) 77-90.

[9] Rockafellar R.T. Convex Analysis. Princeton University Press (1970).

[10] Strodiot J.J., Nguyen V.H., Heukemes N. ε-optimal solutions in nondifferentiable convex programming and some related questions. Math. Prog. 25 (1983) 307-328.

THE SOLUTION OF A NESTED NONSMOOTH OPTIMIZATION PROBLEM

Robert Mifflin
Washington State University, Pullman, WA 99164-2930, USA

1. INTRODUCTION

This paper reports on the successful solution of a nonsmooth version of a practical optimization problem using a recently developed algorithm for single variable constrained minimization. The problem is a single resource allocation problem with five bounded decision variables. The algorithm is used in a nested manner on a dual (minimax) formulation of the problem, i.e., a single variable dual (outer) problem is solved where each function evaluation involves solving a five variable Lagrangian (inner) problem that separates into five independent single variable problems.

A sufficiently accurate solution is obtained with a very reasonable amount of effort using the FORTRAN subroutine PQ1 (Mifflin 1984b) to solve both the outer problem and inner subproblems. PQ1 implements the algorithm in Mifflin (1984a) which solves nonsmooth single variable single constraint minimization problems. The method combines polyhedral and quadratic approximation of the problem functions, an automatic scale-free penalty technique for the constraint and a safeguard. The algorithm is rapidly convergent and reliable in theory and in numerical practice.

Research sponsored by the Air Force Office of Scientific Research, Air Force System Command, USAF, under Grant Number AFOSR-83-0210. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

The smooth version of the problem is due to Heiner, Kupferschmid and Ecker (1983) and is solved there and in Mifflin (1984b). The nonsmooth version is defined in the next section and its solution is discussed in section 3.

2. THE RESOURCE ALLOCATION PROBLEM AND ITS DUAL

The nonsmooth problem solved here is a modification of a smooth applied problem given in detail in Heiner, Kupferschmid and Ecker (1983).

The general problem is to find values for J decision variables $v_1, v_2, \ldots, v_J$ to

$$\text{maximize} \qquad \Sigma_{j=1}^{J} R_j(v_j)$$

$$\text{subject to} \qquad \Sigma_{j=1}^{J} c_j \, v_j \leqq B$$

$$\text{and} \quad 0 \leqq v_j \leqq V_j \quad \text{for} \quad j = 1,2,\ldots,J$$

where

$$R_j(v_j) = \max\{Y_j - 4S_jV_j[v_j^{-1} - (2V_j)^{-1}]^{1/2}, 0\} - c_j \, v_j. \qquad (1)$$

The specific problem of interest has $J = 5$, a budget value $B = 150{,}000$ and the data Y_j, S_j, $2\,V_j$, c_j for $j = 1,2,\ldots,5$ as given in the "Hospitals" table on page 14 of Heiner et al. (1983). Actually, the real application requires integer values for the variables, but rounded continuous solutions appear to be quite adequate for this application.

The nonsmooth problem solved in this paper is the above problem with with R_j and its derivative R_j' replaced by P_j and P_j^+, respectively, where for $v_j \geqq 0$

$$P_j(v_j) = R_j(\underline{v}_j) + P_j^+(v_j)(v_j - \underline{v}_j), \qquad (2)$$

$$P_j^+(v_j) = R_j(\underline{v}_j + 1) - R_j(\underline{v}_j),$$

and $\underline{v}_j$ is the largest whole number not exceeding v_j. Note that P_j is a piecewise affine approximation of R_j which agrees with R_j at integer values of v_j and that P_j^+ is the derivative of P_j at noninteger values of v_j and the right derivative at integer values. The above defined problem is referred to as the primal problem in the sequel.

Each R_j is not a concave function, but R_j does consist of two concave pieces, one of which is linear and the other of which is strictly concave. P_j inherits a piecewise affine version of this R_j structure. The fact that the objective function is a sum of P_j's each having the above special structure allow for attempting to solve this problem via a dual approach.

Let $x \geqq 0$ be a dual variable associated with the linear budget constraint, define the Lagrangian function L by

$$L(v_1,v_2,\ldots,v_5;x) = \Sigma_{j=1}^{5} P_j(v_j) + (B - \Sigma_{j=1}^{5} c_j\ v_j)\ x$$

$$= \Sigma_{j=1}^{5} (P_j(v_j)-c_jv_jx) + Bx \qquad (3)$$

and define the dual function f by

$$f(x) = \max[L(v_1,v_2,\ldots,v_5;x):$$

$$0 \leqq v_j \leqq V_j,\ j = 1,2,\ldots,5]. \qquad (4)$$

The associated dual or <u>outer</u> problem is to find a value for x to

$$\text{minimize}\quad f(x)\quad \text{subject to } -x \leqq 0. \qquad (5)$$

The Lagrangian or <u>inner</u> problem defined by (3) and (4) separates into 5 independent single variable single constraint problems indexed by j and equivalent to

$$\text{minimize}\quad -P_j(v_j) + c_j\ v_j\ x$$

$$\text{subject to}\quad \max\ [-v_j,v_j-V_j] \leqq 0. \qquad (6)$$

Note that these five inner problems could be solved in parallel if one has the facility for parallel processing. The nonconvexity of $-P_j$ gives the possibility of two local minimizers of the jth inner problem (6), one of which is at $v_j = 0$ where $P_j = 0$. The dual approach can be carried out on this problem, because both local minimizers can be found and the better one chosen. Since f is a pointwise maximum over a compact family of affine functions f is a convex function.

Let $V_j(x) \subset [0,V_j]$ be the set of minimizing solutions to the jth inner subproblem depending on the nonnegative parameter (outer variable) x. Then for $x \geqq 0$ and $v_j(x) \in V_j(x)$

$$f(x) = \Sigma_{j=1}^{5}[P_j(v_j(x))-c_jv_j(x)x] + Bx$$

and a subgradient of f at x, denoted g(x), is given by

$$g(x) =-\Sigma_{j=1}^{5}\ c_j\ v_j(x) + B.$$

In general, the outer problem is solved at a point of nondifferentiability of f, say x*. Hence, there exist subgradients of f at x*, say g^- and g^+, and a multiplier $\lambda^* \in [0,1]$ such that

$$(1-\lambda^*)g^- + \lambda^* g^+ = 0. \tag{7}$$

From the inner subproblems

$$g^- = -\Sigma_{j=1}^{5} c_j v_j^- + B$$

and

$$g^+ = -\Sigma_{j=1}^{5} c_j v_j^+ + B$$

where

$$v_j^-,\ v_j^+ \in V_j(x^*) \quad \text{for} \quad j = 1,2,\ldots,5.$$

From the convex combination in (7)

$$0 = -\Sigma_{j=1}^{5} c_j [(1-\lambda^*)v_j^- + \lambda^* v_j^+] + B$$

and a solution to the primal problem is given by

$$v_j = (1-\lambda^*)v_j^- + \lambda^* v_j^+ \text{ for } j = 1,2,\ldots,5$$

provided that for each j

$$(1-\lambda^*)v_j^- + \lambda^* v_j^+ \in V_j(x^*) \tag{8}$$

In general, (8) could be violated, because $V_j(x^*)$ is not a convex set when the primal objective function is not concave. Fortunately, for the particular problem considered here it turns out that (8) is satisfied, i.e., there is no duality gap.

3. THE SOLUTION VIA NESTED OPTIMIZATION

Since the outer problem and each inner subproblem defined above are single variable single constraint minimization problems they can be solved numerically using the FORTRAN subroutine PQ1 of Mifflin (1984b) which implements the algorithm in Mifflin (1984a).

PQ1 requires the user to supply a starting point and a starting stepsize. The starting vector supplied to the multivariable nonlinear programming algorithms used by Heiner et al. (1983) to solve the smooth primal problem was given by $v_j = \frac{1}{2} V_j$ for $j = 1,2,\ldots,5$ (Ecker and Kupferschmid 1984).

To determine a related starting point x and a starting step d for the outer problem v_j was set equal to $\frac{1}{2} V_j \overline{b}$ where $\overline{b}$ was chosen so that

$$\frac{1}{2} \Sigma_{j=1}^{5} c_j V_j \overline{b} = B.$$

This gave the values

$$(v_1, v_2, \ldots, v_5) = (883.1, 240.5, 570.2, 1127.1, 54.0) \qquad (9)$$

that satisfy the budget constraint with equality. Then five values for x were computed such that

$$-P_j^+(v_j) + c_j x = 0 \quad \text{for} \quad j = 1,2,\ldots,5.$$

If these five values had been the same positive number, then this common value and (9) would have been the solution to the minimax problem defined by (4) and (5). This was not the case and the starting x was set to the median value 0.57 and the starting stepsize was set to 0.57 also, so as not to go infeasible if g(0.57) were positive. However, g(0.57) was negative, so the second outer point was 0.57 + 0.57 = 1.14.

For the first set of five inner subproblems, the starting points were set as in (9). For the subsequent inner subproblems when the outer variable was changed from x to x+d, the previous inner solution $v_j(x)$ was used as the starting point in the search for the next inner solution $v_j(x+d)$. Note that the inner objective and right derivative values at the starting point $v_j(x)$ can be updated simply by addition when x is replaced by x+d without evaluating P_j and P_j^+ again. For all of the inner subproblems the starting stepsizes were set to 1.0.

The problem was solved using single precision FORTRAN on a VAX 11/750 computer. For both the outer and inner problems, the numerical parameters STHALF and PENLTY required by PQ1 were set as in Mifflin (1984b) to the values 0.2 and 5×10^{-8}, respectively. The termination criteria were set so that the outer problem was solved to the point where f appeared to be numerically stationary in single precision and the inner subproblems were solved to a corresponding degree of accuracy.

The computer run terminated with two points $x_L = 1.539$ and $x_R = 1.564$ having $f(x_L) = 3{,}975{,}041.$, $f(x_R) = 3{,}975{,}051.$, $g(x_L) = -13.3$, $g(x_R) = 833.9$,

$$(v_1(x_L),\ldots,v_5(x_L)) = \\ (196.5, 0.0,\ 409.8,\ 2015.0,\ 346.0) \qquad (10)$$

and

$$(v_1(x_R),\ldots,v_5(x_R)) = \\ (195.5,\ 0.0,\ 407.2,\ 2001.6,\ 346.0). \qquad (11)$$

To approximate the optimal multiplier λ^* in (7) λ was defined by

$$(1-\lambda)g(x_L) + \lambda g(x_R) = 0.$$

This gave $\lambda = 0.04$ and the corresponding convex combination of (10) and (11) gave the approximate primal solution $(v_1,\ldots,v_5) =(196.5,\ 0.0,\ 409.7,\ 2014.8,\ 346.0)$ with corresponding primal objective value 3,975,041.

This v-solution has v_2 at its lower bound, v_5 at its upper bound, and is very close to the feasible integer solution that is the best known integer solution to this problem (Heiner, et. al., 1983).

The run required 6 outer iterations and, hence, a total of 30 inner subproblems were solved. The total number of evaluations of the P_j's and P_j^+'s was 102. Since evaluating P_j and P_j^+ at a point requires two evaluations of R_j, the total number of evaluations of the R_j's was 204. This is a reasonable amount of work, because 440 such evaluations were used to solve the corresponding smooth primal problem by the code GRG2 (Lasdon et. al., 1978) with double precision arithmetic and function value difference approximations of the partial derivatives (Heiner et al., 1983, Ecker and Kupferschmid,1984).

The smooth version of this problem also was solved using PQ1 in a nested manner on the corresponding dual formulation with only 100 evaluations of the R_j's and R_j''s (Mifflin 1984b). This result represents less work than evaluating the R_j's 204 times, because evaluating R_j and R_j' at a point requires considerably less effort than evaluating R_j twice, due to the same square root being used to calculate R_j and its derivative at a point.

4. CONCLUDING REMARKS

One could imagine problems where the objective function is only given at a finite number of points and some approximation to the function needs to be made before the optimization problem can be solved. As observed here a problem with a smooth approximation of the objective probably could be solved with less effort in the optimization phase than a problem with a piecewise affine approximation of the objective. However, the latter problem does not require the initial phase of setting up and running some procedure to find the smooth approximation. Hence, in terms of overall effort the piecewise affine version might be preferred for some problems where the objective is described only by data points.

5. REFERENCES

Ecker, J.G. and Kupferschmid, M. (1984). Private communication.

Heiner, K.W., Kupferschmid, M., and Ecker, J.G. (1983) Maximizing restitution for erroneous medical payments when auditing samples from more than one provider. Interfaces, 13(5): 12-17.

Lasdon, L.S., Waren, A., Jain, A., and Ratner, M.W. (1978). Design and testing of a generalized reduced gradient code for nonlinear programming. ACM Transactions on Mathematical Software, 4(1): 34-50.

Mifflin, R. (1984a). Stationarity and superlinear convergence of an algorithm for univariate locally Lipschitz constrained minimization. Mathematical Programming, 28: 50-71.

Mifflin, R. (1984b). An implementation of an algorithm for univariate minimization and an application to nested optimization. Dept. of Pure and Applied Mathematics, Washington State University, Pullman, WA, to appear in Mathematical Programming Studies.

VARIATIONS ON THE THEME OF NONSMOOTH ANALYSIS: ANOTHER SUBDIFFERENTIAL

Jean-Paul Penot
Faculty of Science, Avenue de l'Université, 64000 Pau, France

Making one's way through various kinds of limits of differential quotients in order to define generalized derivativesis a rather dull task : one has to be very careful about the moving or fixed ingredients. Formulas such as the following one [11] may be thrilling for some readers :

$$f^{\overset{\uparrow}{\circ}}(a,x) = \sup_{\substack{w \in X \\ r \in \mathbb{R}}} \ \sup_{U \in \mathcal{V}(x)} \ \limsup_{\substack{(v,s,t) \to (w,r,0_+) \\ f(a)+ts \geqslant f(a+ts)}} \ \inf_{u \in U} \frac{1}{t} [f(a+tu+tv) - f(a) - ts] \cdot$$

But for most readers and for most listeners of a lecture with rapidly moving slides, the lure of such a limit may not resist when compared with the clarity and attractiveness of a simple drawing. Thus we choose to focus our attention on a more geometrical aspect of the same problem : the study of tangent cones. It appears that this point of view is also quite rewarding when one has to give the proofs of the calculus rules one may hope to dispose of : these proofs are clearer and simpler when given in geometrical terms instead of analytical calculations ; but this advantage will not appear here. For the sake of clarity in our slides and in this report we adopt rather unusual notations using capital letters instead of subscripts or superscripts (although a systematic use of superscripts as $T^{\uparrow}, T^{o}, T^{\underset{+}{\circ}}, \ldots, f^{\uparrow}, f^{o}, f^{\underset{+}{\circ}} \ldots$ would be elegant). A general agreement on notations and terminology is still ahead ; it may be difficult to realize in a period of fast growing interest and use.

In the sequel E is a subset of a normed vector space X and e is an element of the closure $\mathrm{cl}\ E$ of E . It would be useful to consider the more general situation in which E is a vector space endowed with two topologies but we refrain to do so here.

1 - WELL KNOWN TANGENT CONES

1-1 Definition

The contingent cone to E at e is the set $K(E,e) = \limsup_{t \to 0_+} t^{-1}(E - e)$.

The classical tangent cone to E at e is the set $T(E,e) = \liminf_{t \to 0_+} t^{-1}(E-e)$.

The strict tangent cone to E at e is the set $S(E,e) = \liminf_{t \to 0_+,\ e' \to e} t^{-1}(E-e')$.

This latter cone is also known as the Clarke's tangent cone and the first one is often called the Bouligand's tangent cone or tangent cone in short. The following two characterizations are useful and well known.

1-2 Proposition

(a) A vector v belongs to $K(E,e)$ iff there exist sequences (t_n), (v_n) in $\mathbb{R}^+_o =]0,+\infty[$ and X with limits 0 and v respectively such that $e + t_n v_n \in E$ for each $n \in \mathbb{N}$.

(b) A vector v belongs to $T(E,e)$ iff for each sequence (t_n) in $\mathbb{R}^+_o$ with limit 0 there exists a sequence (v_n) in X with limit v such that $e + t_n v_n \in E$ for each $n \in \mathbb{N}$.

(c) A vector v belongs to $S(E,e)$ iff for each sequence (t_n) in $\mathbb{R}^+_o$ with limit 0 and each sequence (e_n) in E with limit e there exists a sequence (v_n) with limit v in X such that $e_n + t_n v_n \in E$ for each $n \in \mathbb{N}$.

1-3 Proposition

(a) A vector v belongs to $T(E,e)$ iff there exists a curve $c : [0,1] \to X$ with $c(0) = e$, $c(t) \in E$ for $t > 0$ and $v = \dot{c}_+(0) := \lim_{t \to 0_+} t^{-1}(c(t) - c(0))$.

(b) A vector v belongs to $K(E,e)$ iff there exists a curve $c : [0,1] \to X$ with $c(0) = e$, $v = \dot{c}_+(0)$, 0 being an accumulation point of $c^{-1}(E)$.

A characterization of $S(E,e)$ in terms of curves is more delicate ([24],[25]). A characterization of each of the preceding cones can be given in terms of the generalized derivative of the distance function d_E to E (defined by $d_E(x) = \inf \{d(x,e) : e \in E\}$) through the equivalence :

$$v \in C(E,e) \iff d_E^C(e,v) \leqslant 0 \quad \text{for} \quad C = K,T,S \ .$$

Here the C-derivative f^C of a function $f : X \to \overline{\mathbb{R}}$ finite at $a \in X$ is defined through the formula

$$E(f^C(a,.)) = C(E(f),e) \quad \text{for} \quad C = K,T,S$$

where $e = (a,f(a))$, $E(f) = E_f = \{(x,r) \in X \times \mathbf{R} : r \geqslant f(x)\}$ is the epigraph of f . The introduction of generalized derivatives through concepts of tangent cones is well established ([1],[13],[21] for instance) ; see the lecture by K.E. Elster in these proceedings for a systematic treatment along this line. Let us observe that a reverse procedure is possible as long as one is able to define generalized derivatives of an arbitrary function $f : X \to \overline{\mathbf{R}}$ finite at $a \in E$: if i_E is the indicator function of $E \subset X$ (given by $i_E(x) = 0$ if $x \in E, i_E(x) = +\infty$ if $x \in X \setminus E$) and if some generalized derivative $(i_E)^D(a,.)$ of i_E is an indicator function, one can define the related tangent cone $D(E,a)$ as the set D such that

$$i_D(v) = (i_E)^D(a,v) .$$

We will not pursue this line of thoughthere since we insist on the first process we described above.

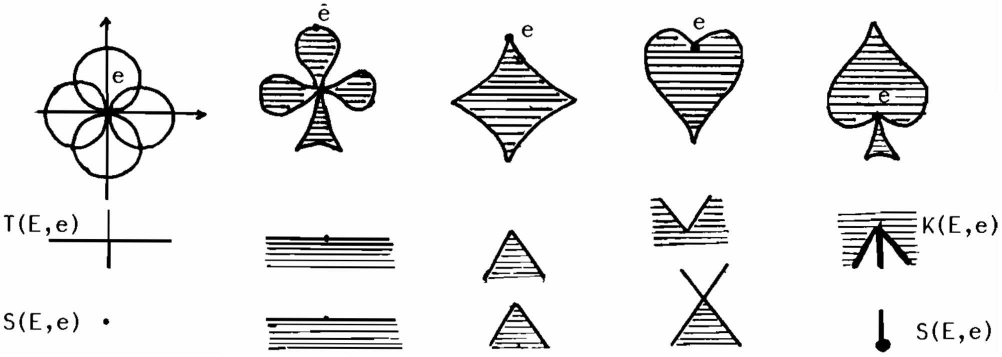
The obvious inclusions

$$K(E,e) \supset T(E,e) \supset S(E,e)$$

yield the following inequalities for an arbitrary function $f : X \to \overline{\mathbf{R}}$ finite at a :

$$f^K(a,.) \leqslant f^T(a,.) \leqslant f^S(a,.)$$

In many cases of interest the preceding inclusions and inequalities are equalities. However they are strict inclusions in general, even if $K(E,e)$ and $T(E,e)$ are seldom different. As a matter of fact $K(E,e)$ and $T(E,e)$ give a closer approximation to E at e than $S(E,e)$ as shown by the following figures and the example $X = \mathbf{R}^2$, $e = (0,0)$, $E = \{(x,y) \in \mathbf{R}^2 : (x - \alpha)^2 + (y - \beta^2) = 1, \alpha,\beta \in \{-1,0,1\}, |\alpha| + |\beta| = 1\}$, for which $K(E,e) = T(E,e) = \mathbf{R} \times \{0\} \cup \{0\} \times \mathbf{R}$, $S(E,e) = \{(0,0)\}$.



2 - THE INTERPLAY BETWEEN THESE NOTIONS

Corresponding to the accuracy of the geometric approximation of E near $\bar{e}$ by a (translated) cone is the precision of the approximation of f by a translated positively homogeneous mapping. We believe this accuracy is of fundamental importance when one is aiming at necessary conditions : as a good detective indicts a small number of suspects, a good necessary condition has to clear most of innocent points of the suspicion of bëing a minimizer. In this respect it is easy to construct a lipschitzian function $f : \mathbb{R} \to \mathbb{R}$ with a unique minimizer at 0 for which one has

$$0 \in \partial^K f(x) \quad \text{iff} \quad x = 0 \ ,$$

whereas

$$\partial^S f(x) = [-10^{100}, 10^{100}] \quad \text{for each} \quad x \in \mathbb{R} \ ,$$

where for $C = K,T,S$, $\bar{x} = (x, f(x))$, $Q^o = \{x^* \in X^*, \langle x^*, x\rangle \leqslant 0 \quad \forall x \in Q\}$

$$\partial^C f(x) = \{x^* \in X^* : x^* \leqslant f^C(x,.)\}$$
$$= \{x^* \in X^* : (x^*,-1) \in C(E_f,\bar{x})^o\}$$

is the C-subdifferential of f associated with C . One cannot claim that the relation $0 \in [-10^{100}, 10^{100}]$ is very informative, especially from a numerical point of view.

Thus we propose to add accuracy to the list of six requirements presented by R.T. Rockafellar in this conference as the goals of subdifferential analysis. These seven goals are certainly highly desirable.

Of course if there were a proposal meeting these seven requirements, this seventh marvel would withdraw nonsmooth analysis from most rights to be entitled as nonsmooth analysis. Our conclusion is that a multiplicity of viewpoints is likely to be the most fruitful approach to this topic, while the lure of a messianic, miraculous generalized derivative may lead to delusion for what concerns necessary conditions (for other aims of nonsmooth analysis as inverse function results, the situation may be quite different as the strict derivative approach seems to be strictly better than anything else).

What precedes will be more clearly understood if we add that the contingential or tangential calculus for sets or functions is relatively poor (see [13],[14] for instance) while the strict tangential calculus is more tractable : accuracy is in balance with handability. This is due to the build-in convexity carried by strict tangency. Contingential or tangential calculus cannot reach such an handability without some added assumptions. One such assumption can be tangential convexity (i.e. $K_e E$ or $T_e E$, f^K or f^T are supposed to be convex); this is not too restrictive, as this assumption encompasses the convex case and the differentiable case. Another kind of assumption which seems to be rather mild is presented in proposition 5.3 below. On the other hand more precise calculus rules can be achieved with strict tangency when one adds regularity conditions in the form : $f^S(a,.)$ coïncides with $f^K(a,.)$ or $f^T(a,.)$; then one is able to replace inclusions by equalities (see [1],[20] for instance).

Here are some more reasons why not forsaking the tangential or congential points of view (see also recent works of J.P. Aubin and the author on differentiability of multifunctions) :

1) in contrast with the strict tangent cone concept these notions are compatible with inclusion : for $E \subset F$ we have $K(E,e) \subset K(F,e)$, $T(E,e) \subset T(F,e)$ but not $S(E,e) \subset S(F,e)$;

2) tangent or contingent concepts are easier to define as the relevant point e is kept fixed ;

3) this fixity of the relevant point permits easier interpretations in marginal analysis for instance or in defining natural directions of decrease ;

4) higher order contingent or tangent cones and derivatives are easy to define and use ([16],...) whereas no strict counter-part are known to the author ;

5) tangent or contingent quotients are basic ingredients in more refined generalized subdifferential calculus as the "fuzzy" calculus of Ioffe [8],[9], Kruger and Mordhukovich ;

6) there is a close link between strict tangent cones and derivatives and contingent cones, at least if the space X is finite dimensional (or reflexive, with some adaption of the preceding concepts). Let us make clear this sixth assertion.

2-1 Proposition [22]

If $f : X \to \overline{\mathbb{R}}$ is finite at $a \in X$ and lower semi-continuous on the Banach space X then for each $v \in X$, denoting by $B(v,\varepsilon)$ the closed ball with center v and radius ε ,

$$f^S(a,v) \leqslant \lim_{\varepsilon \to 0_+} \limsup_{\substack{x \to a \\ f(x) \to f(a)}} \inf_{u \in B(v,\varepsilon)} f^K(x,u) \leqslant \limsup_{\substack{x \to a \\ f(x) \to f(a)}} f^K(x,v) \leqslant \limsup_{x \to a} f^T(x,v)$$

If X is finite dimensional the first inequality is an equality.
If f is locally lipschitzian around a the opposite inequalities holds and

$$f^S(a,v) = \limsup_{x \to a} f^K(x,v) = \limsup_{x \to a} f^T(x,v) = \limsup_{x \to a} f^S(x,v)$$

Proof

The first assertion of the preceding proposition is a consequence of the relation

$$\liminf_{e \to \bar{e}, e \in E} K(E,e) \subset S(E,\bar{e})$$

proved in [23] and [5] ; it becomes an equality if X is finite dimensional ([15] corol. 3.4 and 3.5 and [2]). Let us prove the last assertion : let $r > f^S(a,v)$ and let k be a lipschitz constant of f on some neighborhood X_o of a . By definition of f^S ([21], relation 4.6) we have

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \forall t \in]0,\delta[\quad \forall x \in B(a,\delta) \quad \exists u \in B(v,\varepsilon) \ : \ f(x+tu) - f(x) \leqslant tr$$

As δ can be taken so small that $B(a,\delta) + [0,\delta] B(v,\delta) \subset X_o$ we get

$$\forall \varepsilon > 0 \quad \exists \delta > 0 \quad \forall x \in B(a,\delta) \quad \sup_{0<t<\delta} t^{-1}(f(x+tv) - f(x)) \leqslant r + \varepsilon k$$

Thus $f^T(x,v) \leqslant r + \varepsilon k$ for each $x \in B(a,\delta)$ and $\limsup_{x \to a} f^T(x,v) \leqslant f^S(a,v)$.□

3 - NEW SPECIES OF TANGENT CONES

Let us try now to conciliate the two antagonistic aims of defining convex tangent cones and keeping these approximations related to the set as closely as possible. We incorporate our proposals in a general scheme for obtaining tangent cones ; initially they appeared as an intermediate step in the calculus of tangent and strictly tangent cones in singular cases ([17]). They were preceded by [7] and followed by [6] which contains applications to optimal control theory.

Let us suppose we are given a convergence C on $\mathbb{R}_o^+ \times E$ for each subset E of a n.v.s. X : this is a relation (multifunction) C from $(\mathbb{R}_o^+ \times E)^{\mathbb{N}}$ into $\mathbb{R} \times E$ written $((t_n,e_n)) \overset{C}{\to} (t,e)$ satisfying the usual laws of limits ([10]) (a subsequence of a converging sequence converges to the same limit and so on ...). In fact we are only interested in the case $(t_n) \to 0_+$ in the usual sense ; moreover supposing that (e_n) converges too in X would not alter our present purposes.
Moreover we suppose that if E is a subset of $F \subset X$ then the convergence

C relatively to E is the convergence induced on $\mathbb{R}_o^+ \times E$ by the convergence C on $\mathbb{R}_o^+ \times F$.

The point here is that the convergences $(t_n) \to 0_+$, $(e_n) \to e$ are tied together. We suppose that the following condition is satisfied for each $r \in \mathbb{R}_o^+$:

$$(t_n, e_n) \overset{C}{\to} (0,e) \implies (rt_n, e_n) \overset{C}{\to} (0,e) .$$

In other words the convergence C_t on E associated with a sequence $t = (t_n)$ by

$$(e_n) \overset{C_t}{\to} e \quad \text{iff} \quad ((t_n, e_n)) \overset{C}{\to} (0,e)$$

depends only on the class of (t_n) up to homotheties. The case of primary interest is the case of directional convergence i.e. the case in which $((t_n,e_n)) \overset{C}{\to} (0,e)$ iff $(t_n) \to 0_+$ and $(t_n^{-1}(e_n - e))$ converges. Now we are able to introduce our definition.

3-1 <u>Definition</u>

The C-tangent cone to E at e is the set

$$\mathbf{C(E,e)} = \bigcap_{((t_n,e_n)) \overset{C}{\to} (0,e)} \liminf t_n^{-1}(E - e_n)$$

In other words, $v \in C(E,\tilde{e})$ iff for each sequence $((t_n,e_n)) \overset{C}{\to} (0,e)$ there exists a sequence (v_n) in X with limit v such that $e_n + t_n v_n \in E$ for each $n \in \mathbb{N}$. Thanks to the condition we imposed on C above, $C(E,e)$ is seen to be a closed cone. It is convex in the three last examples below ; to each example we affect a particular letter to denote the convergence C.

<u>Example 1</u> :

$((t_n,e_n)) \overset{I}{\to} (0,e)$ iff $(t_n) \to 0_+$, $e_n = e$ for n large enough ; then $C(E,e)$ is nothing but $T(E,e)$.

<u>Example 2</u> :

$((t_n,e_n)) \overset{S}{\to} (0,e)$ iff $(t_n) \to 0_+$, $(e_n) \to e$ in the topology of X ; then $C(E,e)$ is nothing but $S(E,e)$.

<u>Example 3</u> :

$((t_n,e_n)) \overset{P}{\to} (0,e)$ iff $(t_n) \to 0_+$, $(e_n) \to e$ and $(t_n^{-1}(e_n - e))$ converges in X ; we denote $C(E,e)$ by $P(E,e)$ in this case and call it the prototangent cone or pseudo-strict tangent cone.

<u>Example 4</u> :

$((t_n,e_n)) \overset{Q}{\to} (0,e)$ iff $(t_n) \to 0_+$, $(e_n) \to e$ and $(t_n^{-1}(e_n - e))$ converges to some element of $T(E,e)$; the corresponding cone, denoted by $Q(E,e)$

is called the quasi-strict tangent cone. Comparison of the strength of the convergences occuring in the previous examples shows the following inclusions :

$$S(E,e) \subset P(E,e) \subset Q(E,e) \subset T(E,e) \subset K(E,e) .$$

4 - INTERIORLY TANGENT CONES

Up to now we have only looked at the "male" version of tangent cones. By analogy with the concept of interiorly contingent cone (or interior displacements or feasible directions) recalled below we intend to give an interior partner to each of the cones we introduced above.

4-1 Definition

The interiorly contingent cone to E at e is the set $IK(E,e) = X \backslash K(X \backslash E,e)$: $v \in IK(E,e)$ iff for any sequences $(t_n),(v_n)$ in $\mathbb{R}_o^+$ and X with limits 0 and v respectively one has $e + t_n v_n \in E$ for n large enough.

4-2 Definition

The interiorly C-tangent cone to E at e is the set $IC(E,e)$ of vectors v in X such that for each sequence $((t_n,e_n)) \overset{C}{\to} (0,e)$ and each sequence (v_n) of X with limit v one has $e_n + t_n v_n \in E$ for n in an infinite subset of $\mathbb{N}$ (or equivalently for n large enough).

For $C = T$ we get $IT(E,e) = IK(E,e)$; for $C = S$ we find a cone which is open and closely related to the cone of hypertangent vectors in the sense of Rockafellar ; in fact this cone plays a key role in the proofs of [20] and is called in [21] the hypertangent cone. The cases $C = P,Q,T$ will also be of interest. Obviously

$$IC(E,e) \subset C(E,e) .$$

4-3 Proposition

Suppose the convergence C is directionally stable in the following sense :

if $((t_n e_n)) \overset{C}{\to} (0,e)$, if $d \in C(E,e)$ and if $(d_n) \to d$ with $e_n + t_n d_n \in E$ for each $n \in \mathbb{N}$ then $((t_n,e_n + t_n d_n)) \overset{C}{\to} (0,e)$.

Then $C(E,e)$ and $I(C,e)$ are convex and

$$\mathbf{IC(E,e) + C(E,e) \subset IC(E,e) .}$$

This occurs in particular for $C = P,Q,S$ (but not T) .

Let us prove the inclusion above ; the proof of the convexity of

$C(E,e)$ and $IC(E,e)$ are similar. Let $u \in IC(E,e)$, $v \in C(E,e)$ and let $w = u + v$. Let (w_n) be a sequence with limit w in X and let $(t_n,e_n) \overset{C}{\to} (0,e)$ in $\mathbb{R}_0^+ \times E$. There exists (v_n) with limit v such that $e_n + t_n v_n \in E$ for each n . By assumption we have $(t_n, e_n + t_n v_n) \overset{C}{\to} (0,e)$. As $(u_n) := (w_n - v_n)$ converges to $u = w - v$ we have $e_n + t_n v_n + t_n u_n = e_n + t_n w_n \in E$ for n in an infinite subset of $\mathbb{N}$, hence $w \in IC(E,e)$.

4-4 <u>Corollary</u>

When C is directionally stable and $IC(E,e)$ is nonempty $C(E,e)$ is the closure of $IC(E,e)$ and one has

$$\operatorname{int} C(E,e) \subset IC(E,e) \subset C(E,e) .$$

In fact if $u \in IC(E,e)$, for each $v \in C(E,e)$ and each $t \in \mathbb{R}_0^+$ we have $v + tu \in IC(E,e)$ and $v + tu \to v$ as $t \to 0_+$. On the other hand, for each $w \in \operatorname{int} C(E,e)$ we can write $w = (w - tu) + tu$ with $w - tu \in C(E,e)$ for $t \in \mathbb{R}_0^+$ small enough, so that $w \in IC(E,e)$.

For a function $f : X \to \overline{\mathbb{R}}$ finite at a let us set, with $e = (a,f(a))$

$$f^{IC}(a,v) = \inf \{r \in \mathbb{R} : (v,r) \in IC(E_f,e)\} .$$

4-5 <u>Corollary</u>

Suppose $\operatorname{dom} f^{IC}(a,.)$ is nonempty and C is directionally stable. Then

$$f^C(a,v) = \liminf_{u \to v} f^{IC}(a,u) .$$

Although $T(E,e)$ is not convex in general, it enjoys a restricted convexity property. Namely

4-6 <u>Proposition</u>

$$T(E,e) + Q(E,e) \subset T(E,e)$$
$$T(E,e) + IQ(E,e) \subset IT(E,e) .$$

The proof of these inclusions is nothing but a direct application of the definitions. As above the following assertions follow :

if $IQ(E,e) \neq \emptyset$ then $T(E,e) = \operatorname{cl} IT(E,e)$;

if $\operatorname{dom} f^{IQ}(a,.) \neq \emptyset$ then $f^T(a,v) = \liminf_{u \to v} f^{IT}(a,u)$.

5 - TANGENTIAL CALCULUS AND SUBDIFFERENTIAL CALCULUS

In general the correspondance $E \mapsto C(E,e)$ is not isotone (i.e. does not respect inclusions). This strong defect is partly compensated by the following result in which E is said to be C-regular at e if

$$C(E,e) = T(E,e) \ .$$

5-1 Proposition

Let D and E be two subsets of X , $F = D \cap E$, $a \in \mathrm{cl}\, F$. Then

$$C(D,a) \cap IC(E,a) \subset C(F,a) \ .$$

If C is directionally stable and if $C(D,a) \cap IC(E,a) \neq \emptyset$ then

$$C(D,a) \cap C(E,a) \subset C(F,a) \ .$$

If moreover D and E are C-regular at a , then F is C-regular at a and

$$C(D,a) \cap C(E,a) = C(F,a) \ .$$

This result can be incorporated in the following property in which a mapping $f : D \to Y$ defined on some subset D of X , with values in some n.v.s. Y is said to be C-strictly differentiable at $a \in D$, if there is a linear continuous mapping $f'(a) : X \to Y$ such that for each sequence $((t_n,a_n)) \overset{\text{Ç}}{\to} (0,a)$ (with respect to D) and each $(v_n) \to v$ in X , with $v \in C(D,e)$, $a_n + t_n v_n \in D$ for each $n \in \mathbb{N}$ one has $((t_n,f(a_n)) \overset{\text{Ç}}{\to} (0,f(a))$ and

$$t^{-1}(f(a_n + t_n v_n) - f(a_n)) \to f'(a)(v) \ .$$

For $D = X$ and $C = T,P$ or Q this is just Hadamard-differentiability ; for $C = S$ this is exactly strict differentiability.

5-2 Proposition

Let F be a subset of Y and $E = f^{-1}(F)$ $(= D \cap f^{-1}(F))$, where $f : D \to Y$ is C-strictly differentiable at $a \in E$. Then

$$C(D,a) \cap f'(a)^{-1}(IC(F,f(a))) \subset C(E,a) \ .$$

If C is directionally stable and if $f'(a)(C(D,a)) \cap IC(F,f(a)) \neq \emptyset$ then

$$C(D,a) \cap f'(a)^{-1}(C(F,f(a))) \subset C(E,a) \ .$$

If moreover D and F are regular then equality holds and E is C-regular.

Similarly, if f is Q-strictly differentiable at a and if $f'(a)(Q(D,a)) \cap IQ(F,f(a)) \neq \emptyset$ then

$$T(D,a) \cap f'(a)^{-1}(T(F,f(a))) = T(E,a) .$$

One can derive chain rules from the preceding relations ; let us rather give two samples of rules for the addition (see also [6]).

5-3 Proposition

Let $h = f+g$. If there exists $\hat{v} \in X$ such that $f^Q(a,\hat{v}) < +\infty$, $g^{IQ}(a,\hat{v}) < +\infty$ then

$$h^T(a,x) \leqslant f^T(a,x) + g^T(a,x) \quad \text{for each} \quad x \in X .$$

If moreover $f^T(a,.)$ and $g^T(a,.)$ are convex then

$$\partial^T h(a) \subset \partial^T f(a) + \partial^T g(a) .$$

5-4 Proposition

Let $h = f+g$ where f and g are conically calm at a (i.e. for each $v \in X$ $f^K(a,v) > -\infty$, $g^K(a,v) > -\infty$) or such that $\text{dom } f^{IK}(a,.) = X = \text{dom } g^{IK}(a,.)$. Then if $\text{dom } f^P(a,.) \cap \text{dom } f^{IP}(a,.) \neq \emptyset$ then

$$h^P(a,x) \leqslant f^P(a,x) + g^P(a,x) \quad \text{for each} \quad x \in X \quad \text{and}$$

$$\partial^P h(a) \subset \partial^P f(a) + \partial^P g(a) .$$

6 - THE STAR DIFFERENCE

The following algebraic operation between two subsets of a vector space X will provide an interesting link between the cones we introduced ; it has been used by Pontrjagin [18], Psenicnyj [19] and Giner [7] who developped a subdifferential calculus using the star operation on various generalized derivatives and applied by Frankowska [6].

Given two subsets A and B of X their star-difference (or alternate difference) is the set

$$A \overset{*}{-} B = \{x \in X : x + B \subset A\} .$$

We set $A_* = A \overset{*}{-} A$; when A is a closed cone of a n.v.s. X, it has been shown in [4] and [7] that A_* is the intersection of the maximal convex subcones of A containing a boundary point of A . The two following lemmas give connections with a more functional point of view.

6-1 Lemma [6],[7]

The star of the epigraph E_h of a positively homogeneous functional

$h : X \to \overline{\mathbb{R}}$ is the epigraph of the sublinear functional $^{*}h$ given by

$$^{*}h(x) = \sup \{h(x+w) - r : (w,r) \in E_h\} := \sup \{h(x+w) \underset{\cdot}{+} (-h(w)) : w \in X\}$$

Proof

Let $(x,s) \in (E_h)_*$. As for each $(w,r) \in E_h$ we have $(w+x, r+s) \in E_h$ we get $s \geqslant \sup \{h(x+w) - r : (w,r) \in E_h\}$. Conversely if $(x,s) \in X \times \mathbb{R}$ is such that $s \geqslant \sup \{h(x+w) - r : (w,r) \in E_h\}$ then for each $(w,r) \in E_h$ we have $r+s \geqslant h(x+w)$ or $(x+w, r+s) \in E_h$ and $(x,s) \in (E_h)_*$. □

6-2 Lemma

If A and B are closed convex subsets of X , the support function h_C of $C = A \overset{*}{-} B$, given by $h_C(x^*) = \sup \langle x^*, C\rangle$ for $x^* \in X^*$, is the greatest of the weak-star lower-semicontinuous positively homogeneous functionals h on X^* such that $h + h_B \leqslant h_A$.

This follows from the fact that for a closed convex subset D of X one has $D + B \subset A$ iff $h_D + h_B \leqslant h_A$.

The star difference can be used in connection with Demyanov's theory of bidifferential calculus (or quasi-differential calculus [2]). Suppose $f : X \to \mathbb{R}$ has a directional derivative $h = f'(a,.)$ at $a \in X$ which is the difference of two sublinear mappings $p,q : h = p - q$. Let $\partial h(0) = \{x^* \in X^* : x^* \leqslant h\}$.

6-3 Proposition

One has $\partial h(0) = \partial p(0) \overset{*}{-} \partial q(0)$. In particular, if f attains a local minimum at 0 one has the following equivalent assertions :

$$0 \in \partial h(0) \iff 0 \in \partial p(0) \overset{*}{-} \partial q(0) \iff \partial q(0) \subset \partial p(0) .$$

Our interest in the star difference stems from the following fact

6-4 Proposition

For each subset E of X and $e \in \mathrm{cl}\, E$ one has $Q(E,e) = T(E,e)_*$ and

$$T(E,e) \overset{*}{-} K(E,e) \subset P(E,e) \subset K(E,d)_* ,$$

$$IK(E,e) \overset{*}{-} K(E,e) \subset IP(E,e) \subset IT(E,e) \overset{*}{-} T(E,e) = IQ(E,e)_* = IQ(E,e) .$$

It follows in particular that for any $f : X \to \overline{\mathbb{R}}$ finite at a one has

$$f^Q(a,.) = f^T(a,.)_* .$$

Thus, when $f^T(a,.)$ is convex, one has $f^Q(a,.) = f^T(a,.)$; in particular, when f is Hadamard-differentiable at a , one has

$$\partial^Q f(a) = \{f'(a)\} = \partial^P f(a) .$$

In [12] a more analytical (but simple) approach to subdifferential calculus is presented which in particular shares this enjoyable property which does not hold with the strict subdifferential $\partial^S f(a)$.

REFERENCES

[1] CLARKE F.H. : Optimization and Nonsmooth Analysis. Wiley, New-York (1983).

[2] CORNET B. : Contribution à la théorie mathématique des mécanismes dynamiques d'allocation des ressources. Thèse Univ. Paris 9 (1981).

[3] DEMYANOV V.F., RUBINOV A.M. : On quasidifferentiable functionals. Dokl. Akad. Nauk SSR 250 (1980) 21-25, Soviet Math. Dokl. 21(1) (1980) 14-17.

[4] DOLECKI S. : Hypertangent cones for a special class of sets. in "Optimization, theory and algorithms", J.B. Hiriart-Urruty et al. editors, Marcel Dekker, New-York (1983) pp. 3-11.

[5] DOLECKI S., PENOT J.P. : The Clarke's tangent cone and limits of tangent cones. Publ. Math. Pau (1983).

[6] FRANKOWSKA H. : The adjoint differential inclusions associated to a minimal trajectory of a differential inclusion. Cahiers de Math. de la Décision n° 8315, Univ. Paris IX (1983).

[7] GINER E. : Ensembles et fonctions étoilés ; application à l'optimisation et au calcul différentiel généralisé (manuscript, Toulouse) (1981).

[8] IOFFE A. : Approximate subdifferentials and applications I : the finite dimensional theory. Trans. Amer. Math. Soc. 281(1) (1984) 389-416.

[9] IOFFE A. : Calculus of Dini subdifferentials of functions and contingent coderivatives of set-valued maps. Nonlinear Anal. Th. Methods and Appl. 8(5) (1984) 517-539.

[10] KURATOWSKI K. : Topologie, I. Polish Scientific Publisher. P.W.N. Warzaw (1958), English translation PWN - Academic Press (1966).

[11] MICHEL P., PENOT J.P. : Calcul sous-différentiel pour des fonctions lipschitziennes et non lipschitziennes. C.R. Acad. Sc. Paris I 298(12) (1984) 269-272.

[12] MICHEL P., PENOT J.P. : A simple subdifferential calculus for locally lipschitzian functions (to appear).

[13] PENOT J.P. : Calcul sous-différentiel et optimisation, J. Funct. Anal. 27(2) (1978) 248-276.

[14] PENOT J.P. : On regularity conditions in mathematical programming. Math. Prog. Study 19 (1982) 167-199.

[15] PENOT J.P. : A characterization of tangential regularity. Nonlin. Anal. Theory, Methods and Appl. 5(6) (1981) 625-643.

[16] PENOT J.P. : Generalized higher order derivatives and higher order optimality conditions (to appear).

[17] PENOT J.P., TERPOLILLI P. : Cônes tangents et singularités. C.R. Acad. Sci. Paris 296 (1983), 721-724.

[18] PONTRJAGIN L.S. : Linear differential games II. Dokl. Akad. Nauk 175 (1967) 764-766.

[19] PSENICNYJ B.N. : Leçons sur les jeux différentiels. Cahier de l'IRIA n° 4 (1971) 145-226.

[20] ROCKAFELLAR R.T. : Directionally lipschitzian functions and subdifferential calculus. Proc. London Math. Soc. 39 (1979) 331-355.

[21] ROCKAFELLAR R.T. : Generalized directional derivatives and subgradients of nonconvex functions. Can. J. Math. 32(2) (1980) 257-280.

[22] ROCKAFELLAR R.T. : Generalized subgradients. in "Mathematical Programming : the State of the Art", Bonn 1982, A. Bachen, M. Grötschel, B. Korte, editors, Springer Verlag, Berlin (1983) 368-390.

[23] TREIMAN J. : Characterization of Clarke's tangent and normal cones in finite and infinite dimensions. Nonlinear Anal. Th., Methods and Appl. 7(7) (1983) 771-783.

[24] TREIMAN J. : Generalized gradients and paths of descent.Preprint, Univ. of Alaska (1984).

[25] WATKINS G.G. : Clarke's tangent vectors as tangents to Lipschitz continuous curves, J. Optim. Th. Appli. (to appear).

LIPSCHITZIAN STABILITY IN OPTIMIZATION: THE ROLE OF NONSMOOTH ANALYSIS

R.T. Rockafellar
Department of Mathematics, University of Washington, Seattle, WA 98195, USA

ABSTRACT

The motivations of nonsmooth analysis are discussed. Applications are given to the sensitivity of optimal values, the interpretation of Lagrange multipliers, and the stability of constraint systems under perturbation.

INTRODUCTION

It has been recognized for some time that the tools of classical analysis are not adequate for a satisfactory treatment of problems of optimization. These tools work for the characterization of locally optimal solutions to problems where a smooth (i.e. continuously differentiable) function is minimized or maximized subject to finitely many smooth equality constraints. They also serve in the study of perturbations of such constraints, namely through the implicit function theorem and its consequences. As soon as inequality constraints are encountered, however, they begin to fail. One-sided derivative conditions start to replace two-sided conditions. Tangent cones replace tangent subspaces. Convexity and convexification emerge as more natural than linearity and linearization.

In problems where inequality constraints actually predominate over equations, as is typical in most modern applications of optimization, a qualitative change occurs. No longer is there any simple way of recognizing which constraints are active in a neighborhood of a given point of the feasible set, such as there would be if the set were a cube or simplex, say. The boundary of the feasible set defies easy description and may best be thought of as a nonsmooth hypersurface. It does not take long to realize too that the graphs of many of the objective functions which naturally arise are nonsmooth in a similar way. This is the motivation for much of the effort that has gone into

* Research supported in part by a grant from the National Science Foundation at the University of Washington, Seattle.

introducing and developing various concepts of "tangent cone", "normal cone", "directional derivative" and "generalized gradient". These concepts have changed the face of optimization theory and given birth to a new subject, ***nonsmooth analysis***, which is affecting other areas of mathematics as well.

An important aim of nonsmooth analysis is the formulation of generalized necessary or sufficient conditions for optimality. This in turn receives impetus from research in numerical methods of optimization that involve nonsmooth functions generated by decomposition, exact penalty representations, and the like. The idea essentially is to provide tests that either establish (near) optimality (perhaps stationarity) of the point already attained or generate a feasible direction of improvement for moving to a better point.

Nonsmooth analysis also has other important aims, however, which should not be overlooked. These include the study of sensitivity and stability with respect to perturbations of objective and constraints. In an optimization problem that depends on a parameter vector v, how do variations in v affect the optimal value, the optimal solution set, and the feasible solution set? Can anything be said about rates of change?

This is where Lipschitzian properties take on special significance. They are intermediate between continuity and differentiability and correspond to ***bounds*** on possible rates of change, rather than rates themselves, which may not exist, at least in the classical sense. Like convexity properties they can be passed along through various constructions where true differentiability, even if one-sided, would be lost. Furthermore, they can be formulated in geometric terms that suit the study multifunctions (set-valued mappings), a subject of great importance in optimization theory but for which classical notions are almost entirely lacking.

It is in this light that the directional derivatives and subgradients introduced by F.H. Clarke [1] [2] should be judged. Clarke's theory emphasizes Lipschitzian properties and sturdily combines convex analysis and classical smooth analysis in a single framework. At the present stage of development, thanks to the efforts of many individuals, it has already had strong effects on almost every area of optimization, from nonlinear programming to the calculus of variations, and also on mathematical questions beyond the domain of optimization per se.

This is not to say, however, that Clarke's derivatives and subgradients are the only ones that henceforth need to be considered. Special situations certainly do require special insights. In particular, there are cases where special one-sided first and second derivatives that are more finely tuned than Clarke's are worth introducing. Significant and useful results can be obtained in such manner. But such results are likely to be relatively limited in scope.

The power and generality of the kind of nonsmooth analysis that is based on Clarke's ideas can be credited to the following features, in summary:

(a) Applicability to a huge class of functions and other objects, such as sets and multifunctions.

(b) Emphasis on geometric constructions and interpretations.

(c) Reduction to classical analysis in the presence of smoothness and to convex analysis in the presence of convexity.

(d) Unified formulation of optimality conditions for a wide variety of problems.

(e) Comprehensive calculus of subgradients and normal vectors which makes possible an effective specialization to particular cases.

(f) Coverage of sensitivity and stability questions and their relationship to Lagrange multipliers.

(g) Focus on local properties of a "uniform" character, which are less likely to be upset by slight perturbations, for instance in the study of directions of descent.

(h) Versatility in infinite as well as finite-dimensional spaces and in treating the integral functionals and differential inclusions that arise in optimal control, stochastic programming, and elsewhere.

In this paper we aim at putting this theory in a natural perspective, first by discussing its foundations in analysis and geometry and the way that Lipschitzian properties come to occupy the stage. Then we survey the results that have been obtained recently on sensitivity and stability. Such results are not yet familiar to many researchers who concentrate on optimality conditions and their use in algorithms. Nevertheless they say much that bears on numerical matters, and they demonstrate well the sort of challenge that nonsmooth analysis is now able to meet.

1. ORIGINS OF SUBGRADIENT IDEAS

In order to gain a foothold on this new territory, it is best to begin by thinking about functions $f: R^n \to R$ that are not necessarily smooth but have strong one-sided directional derivatives in the sense of

$$f'(x;h) = \lim_{\substack{t \downarrow 0 \\ h' \to h}} \frac{f(x+th') - f(x)}{t} \tag{1.1}$$

Examples are (finite) convex functions [3] and *subsmooth* functions, the latter being by definition representable locally as

$$f(x) = \max_{s \in S} f_s(x), \tag{1.2}$$

where S is a compact space (e.g., a finite, discrete index set) and $\{f_s \mid s \in S\}$ is a family of smooth functions whose values and derivatives depend continuously on s and x jointly. Subsmooth functions were introduced in [4]; all smooth functions and all finite convex functions on R^n are in particular subsmooth.

The formula given here for $f'(x;h)$ differs from the more common one in the literature, where the limit $h' \rightarrow h$ is omitted (weak one-sided directional derivative). It corresponds in spirit to true (strong) differentiability rather than weak differentiability. Indeed, under the assumption that $f'(x,h)$ exists for all h (as in (1.1)), one has f differentiable at x if and only if $f'(x;h)$ is linear in h. Then the one-sided limit $t \downarrow 0$ is actually realizable as a two-sided limit $t \rightarrow 0$.

The classical concept of ***gradient*** arises from the duality between linear functions on R^n and vectors in R^n. To say that $f'(x;h)$ is linear in h is to say that there is a vector $y \in R^n$ with

$$f'(x;h) = y \cdot h \quad \text{for all} \quad h. \tag{1.3}$$

This y is called the gradient of f at x and is denoted by $\nabla f(x)$.

In a similar way the modern concept of ***subgradient*** arises from the duality between sublinear functions on R^n and convex subsets in R^n. A function l is said to be ***sublinear*** if it satisfies

$$l(\lambda_1 h_1 + \ldots + \lambda_m h_m) \leq \lambda_1 l(h_1) + \ldots + \lambda_m l(h_m) \tag{1.4}$$

$$\text{when } \lambda_1 \geq 0, \cdots, \lambda_m \geq 0.$$

It is known from convex analysis [3, §13] that the finite sublinear functions l on R^n are precisely the support functions of the nonempty compact subsets Y of R^n: each l corresponds to a unique Y by the formula

$$l(h) = \max_{y \in Y} y \cdot h \quad \text{for all} \quad h. \tag{1.5}$$

Linearity can be identified with the case where Y consists of just a single vector y.

It turns out that when f is convex, and more generally when f is subsmooth [4], the derivative $f'(x,h)$ is always sublinear in h. Hence there is a nonempty compact subset Y of R^n, uniquely determined, such that

$$f'(x;h) = \max_{y \in Y} y \cdot h \quad \text{for all} \quad h. \tag{1.6}$$

This set Y is denoted by $\partial f(x)$, and its elements y are called subgradients of f at x. With respect to any local representation (1.4), one has

$$Y = \mathrm{co}\{\nabla f_s(x) \mid s \in S_x\}, \text{ where } S_x = \operatorname*{argmax}_{s \in S} f_s(x) \tag{1.7}$$

(co = convex hull), but the set $Y = \partial f(x)$ is of course by its definition independent of the representation used.

In the case of f convex [3, §23] one can define subgradients at x equivalently as the vectors y such that

$$f(x') \geq f(x) + y\cdot(x'-x) \text{ for all } x'. \tag{1.8}$$

For f subsmooth this generalizes to

$$f(x') \geq f(x) + y\cdot(x'-x) + o(|x'-x|), \tag{1.9}$$

but caution must be exercised here about further generalization to functions f that are not subsmooth. Although the vectors y satisfying (1.9) do always form a closed convex set Y at x, regardless of the nature of f, this set Y does not yield an extension of formula (1.6), nor does it correspond in general to a robust concept of directional derivative that can be used as a substitute for $f'(x;h)$ in (1.6). For a number of years, this is where subgradient theory came to a halt.

A way around the impasse was discovered by Clarke in his thesis in 1973. Clarke took up the study of functions $f : R^n \to R$ that are *locally Lipschitzian* in the sense of the difference quotient

$$|f(x'') - f(x')| \,/\, |x''-x'| \tag{1.10}$$

being bounded on some neighborhood of each point x. This class of functions is of intrinsic value for several reasons. First, it includes all subsmooth functions and consequently all smooth functions and all finite convex functions; it also includes all finite concave functions and all finite saddle functions (which are convex in one vector argument and concave in another; see [3, §35]). Second, it is preserved under taking linear combinations, pointwise maxima and minima of collections of functions (with certain mild assumptions), integration and other operations of obvious importance in optimization. Third, it exhibits properties that are closely related to differentiability. The local boundedness of the difference quotient (1.10) is such a property itself. In fact when f is locally Lipschitzian, the gradient $\nabla f(x)$ exists for all but a negligible set of points x in R^n (the classical theorem of Rademacher, see [5]).

Clarke discovered that when f is locally Lipschitzian, the special derivative expression

$$f^{\circ}(x;h) = \limsup_{\substack{t \downarrow 0 \\ h' \to h \\ x' \to x}} \frac{f(x'+th') - f(x')}{t} \tag{1.11}$$

is always a finite sublinear function of h. Hence there exists a unique nonempty compact convex set Y such that

$$f^{\circ}(x;h) = \max_{y \in Y} y \cdot h \quad \text{for all } h. \tag{1.12}$$

Moreover

$$f^{\circ}(x;h) = f'(x;h) \quad \text{for all } h \text{ when } f \text{ is subsmooth.} \tag{1.13}$$

Thus in denoting this set Y by $\partial f(x)$ and calling its elements subgradients, one arrives at a natural extension of nonsmooth analysis to the class of all locally Lipschitzian functions. Many powerful formulas and rules have been established for calculating or estimating $\partial f(x)$ in this broad context, but it is not our aim to go into them here; see [2] and [6], for instance.

It should be mentioned that Clarke himself did not incorporate the limit $h' \to h$ into the definition of $f^{\circ}(x;h)$, but because of the Lipschitzian property the value obtained for $f^{\circ}(x;h)$ is the same either way. By writing the formula with $h' \to h$ one is able to see more clearly the relationship between $f^{\circ}(x;h)$ and $f'(x;h)$ and also to prepare the ground for further extensions to functions f that are merely lower semicontinuous rather than Lipschitzian. (For such functions one writes $x' \to_f x$ in place of $x' \to x$ to indicate that x is to be approached by x' only in such a way that $f(x') \to f(x)$. More will be said about this later.)

Some people, having gone along with the developments up until this point, begin to balk at the "coarse" nature of the Clarke derivative $f^{\circ}(x;h)$ in certain cases where f is *not* subsmooth and nevertheless is being ***minimized***. For example, if $f(x) = -|x| + |x|^2$ one has $f^{\circ}(0;h) = |h|$, whereas $f'(0;h)$ exists too but $f'(0;h) = -|h|$. Thus f' reveals that every $h \neq 0$ gives a direction of descent from 0, in the sense of yielding $f'(0;h) < 0$, but f° reveals no such thing, inasmuch as $f^{\circ}(0;h) > 0$. Because of this it is feared that f° does not embody as much information as f' and therefore may not be entirely suitable for the statement of necessary conditions for a minimum, let alone for employment in algorithms of descent.

Clearly f° cannot replace f' in every situation where the two may differ, nor has this ever been suggested. But even in face of this caveat there are arguments to be made in favor of f° that may help to illuminate its nature and the supporting motivation. The Clarke derivative f° is oriented towards minimization problems, in contrast to f', which is neutral between minimization and maximization. In addition, it emphasizes a certain uniformity. A vector h with $f^\circ(x;h) < 0$ provides a descent direction in a strong *stable* sense: there is an $\varepsilon > 0$ such that for all x' near x, h' near h, and positive t near 0, one has

$$f(x' + th') < f(x') - t\varepsilon.$$

A vector h with $f'(x;h) < 0$, on the other hand, provides descent only from x; at points x' arbitrarily near to x it may give a direction of ascent instead. This instability is not without numerical consequences, since x might be replaced by x' due to round-off.

An algorithm that relied on finding an h with $f'(x;h) < 0$ in cases where $f^\circ(x;h) \geq 0$ for all h (such an x is said to be *substationary* point) seems unlikely to be very robust. Anyway, it must be realized that in executing a method of descent there is very little chance of actually arriving along the way at a point x that is substationary but not a local minimizer. One is easily convinced from examples that such a mishap can only be the consequence of an unfortunate choice of the starting point and disappears under the slightest perturbation. The situation resembles that of cycling in the simplex method.

Furthermore it must be understood that because of the orientation of the definition of f° towards minimization, there is no justice in holding the notion of substationarity up to any interpretation other than the following: a substationary point is either a point where a local *minimum* is attained or one where progress towards a local minimum is "confused". Sometimes, for instance, one hears cited as a failing of f° that f' is able to distinguish between a local minimum and a local maximum in having $f'(x;h) \geq 0$ for all h in the first case, but $f'(x;h) \leq 0$ for all h in the second, whereas $f^\circ(x;h) \geq 0$ for all h in both cases. But this is unfair. A one-sided orientation in nonsmooth analysis is merely a reflection of the fact that in virtually all applications of optimization, there is unambiguous interest in either maximization or minimization, but not both. For theoretical purposes it might as well be minimization.

Certainly the idea that a first-order concept of derivative, such as we are dealing with here, is obliged to provide conditions that distinguish effectively between a local minimum and a local maximum is out of line for other reasons. Classical analysis makes no attempt in that direction, without second derivatives. Presumably, second derivative concepts in nonsmooth analysis will eventually furnish the appropriate distinctions, cf. Chaney [7].

A final note on the question of f° versus f' is the reminder that $f^\circ(x;h)$ is defined for any locally Lipschitzian function f and even more generally, whereas $f'(x;h)$ is only defined for functions f in a narrower class.

An important goal of nonsmooth analysis is not only to make full use of Lipschitz continuity when it is present, but also to provide criteria for Lipschitz continuity in cases where it cannot be known *a priori*, along with corresponding estimates for the local Lipschitz constant. For this purpose, it is necessary to extend subgradient theory to functions that might not be locally Lipschitzian or even continuous everywhere, but merely lower semicontinuous. Fundamental examples of such functions in optimization are the so-called *marginal* functions, which give the minimum value in a parameterized problem as a function of the parameters. Such functions can even take on $\pm\infty$.

Experience with convex analysis and its applications shows further the desirability of being able to treat the indicator functions of sets, which play an essential role in the passage between analysis and geometry.

In fact, the ideas that have been described so far can be extended in a powerful, consistent manner to the class of all lower semicontinuous functions $f: R^n \longrightarrow \bar{R}$, where $\bar{R} = [-\infty, \infty]$ (extended real number system). There are two complementary ways of doing this, with the same result. In the continuation of the analytic approach we have been following until now, a more subtle directional derivative formula

$$f^\uparrow(x;h) = \lim_{\varepsilon\downarrow 0}\left[\limsup_{\substack{t\downarrow 0\\ x'\to_f x}}\left[\inf_{|h'-h|\le\varepsilon}\frac{f(x'+th')-f(x')}{t}\right]\right] \tag{1.14}$$

is introduced and shown to agree with $f^\circ(x;h)$ whenever f is locally Lipschitzian and indeed whenever $f^\circ(x;h)$ (in the extended definition with $x' \to_f x$, as mentioned earlier) is not $+\infty$. Moreover $f^\uparrow(x;h)$ is proved always to be a lower semicontinuous, sublinear function of h (extended-real-valued). From convex analysis, then, it follows that either $f^\uparrow(x;0) = -\infty$ or there is a nonempty closed convex set $Y \subset R^n$, uniquely determined, with

$$f^\uparrow(x;h) = \sup_{y\in Y} y\cdot h \quad \text{for all } h. \tag{1.15}$$

This is the approach followed in Rockafellar [8], [9]. One then arrives at the corresponding geometric concepts by taking f to be the indicator δ_C of a closed set C. For any $x \in C$, the function $h \mapsto \delta_C^\uparrow(x;h)$ is itself the indicator of a certain closed set

$T_C(x)$ which happens always to be a convex cone; this is the Clarke *tangent* cone to C at x. The subgradient set

$$N_C(x) = \partial\delta_C(x), \tag{1.16}$$

on the other hand, is a closed convex set too, the Clarke *normal* cone to C to x. The two cones are polar to each other:

$$N_C(x) = T_C(x)^\circ, \quad T_C(x) = N_C(x)^\circ. \tag{1.17}$$

In a more geometric approach to the desired extension, the tangent cone $T_C(x)$ and normal cone $N_C(x)$ can first be defined in a direct manner that accords with the polarity relations (1.16). Then for an arbitrary lower semicontinuous function $f: R^n \to \bar{R}$ and point x at which f is finite, one can focus on $T_E(x,f(x))$ and $N_E(x,f(x))$, where E is the epigraph of f (a closed subset of R^{n+1}). The cone $T_E(x,f(x))$ is itself the epigraph of a certain function, namely the subderivative $h \mapsto f^{\uparrow}(x;h)$, whereas the cone $N_E(x,f(x))$ provides the subgradients:

$$\partial f(x) = \{y \in R^n \mid (y,-1) \in N_E(x,f(x))\}. \tag{1.18}$$

The polarity between $T_E(x,f(x))$ and $N_E(x,f(x))$ yields the subderivative-subgradient relation (1.14). (Clarke's original extension of ∂f to lower semicontinuous functions [1] followed this geometric approach in defining normal cones directly and then invoking (1.17) as a definition for subgradients. He did not focus much on tangent cones, however, or pursue the idea that $T_E(x,f(x))$ might correspond to a related concept of directional derivative.)

The details of these equivalent forms of extension need not occupy us here. The main thing to understand is that they yield a basic criterion for Lipschitzian continuity, as follows.

THEOREM 1 (Rockafellar [10]). *For a lower semicontinuous function $f: R^n \to \bar{R}$ actually to be Lipschitzian on some neighborhood of the point x, it is sufficient (as well as necessary) that the subgradient set $\partial f(x)$ be nonempty and bounded. Then one has*

$$\limsup_{\substack{x' \to x \\ x'' \to x}} \frac{|f(x'')-f(x')|}{|x''-x'|} = \max_{y \in \partial f(x)} |y|. \tag{1.19}$$

This criterion can be applied without exact knowledge of $\partial f(x)$ but only an estimate that $\phi \neq \partial f(x) \subset Y$ for some set Y. If Y is bounded, one may conclude that f is locally Lipschitzian around x. If it is known that $|y| < \lambda$ for all $y \in Y$, one has from (1.19)

$$|f(x'') - f(x')| \leq \lambda |x'' - x'| \text{ for } x' \text{ and } x'' \text{ near } x.$$

2. LAGRANGE MULTIPLIERS AND SENSITIVITY

Many ways have been found for deriving optimality conditions for problems with constraints, but not all of them provide full information about the Lagrange multipliers that are obtained. The test of a good method is that it should lead to some sort of interpretation of the multiplier vectors in terms of sensitivity or generalized rates of change of the optimal value in the problem with respect to perturbations. Until quite recently, a satisfactory interpretation along such lines was available only for convex programming and special cases of smooth nonlinear programming. Now, however, there are general results that apply to all kinds of problems, at least in R^n. These results demonstrate well the power of the new nonsmooth analysis and are not matched by anything achieved by other techniques.

Let us first consider a nonlinear programming problem in its canonical parameterization:

$$(P_u) \qquad \text{minimize } g(x) \text{ subject to } x \in K \text{ and}$$
$$g_i(x) + u_i \leq 0 \text{ for } i = 1, \ldots, s,$$
$$= 0 \text{ for } i = s+1, \ldots, m,$$

where $g, g_1, \ldots, g_m$ are locally Lipschitzian functions on R^n and K is a closed subset of R^n; the u_i's are parameters and form a vector $u \in R^m$. By analogy with what is known in particular cases of (P_u), one can formulate the potential *optimality condition* on a feasible solution x, namely that

$$0 \in \partial g(x) + \textstyle\sum_{i=1}^m y_i \, \partial g_i(x) + N_K(x) \text{ with} \tag{2.1}$$

$$y_i \geq 0 \text{ and } y_i[g_i(x) + u_i] = 0 \text{ for } i = 1, \ldots, s,$$

and a corresponding *constraint qualification* at x:

the only vector $y=(y_1, \ldots, y_m)$ satisfying the version (2.2)

of (2.1) in which the term $\partial g(x)$ is omitted is $y=0$.

In *smooth programming*, where the functions $g, g_1, \ldots, g_m$ are all continuously differentiable and there is no abstract constraint $x \in K$, the first relation in (2.1) reduces to the gradient equation

$$0 = \nabla g(x) + \textstyle\sum_{i=1}^m y_i \nabla g_i(x),$$

and one gets the classical Kuhn-Tucker conditions. The constraint qualification is then equivalent (by duality) to the well known one of Mangasarian and Fromovitz.

In *convex programming*, where $g, g_1, \ldots, g_s$ are (finite) convex functions, $g_{s+1}, \ldots, g_m$ are affine, and K is a convex set, condition (2.1) is always sufficient for optimality. Under the constraint qualification (2.2), which in the absence of equality constraints reduces to the Slater condition, it is also necessary for optimality.

For the general case of (P_u) one has the following rule about necessity.

THEOREM 2 (Clarke [11]). *Suppose x is a locally optimal solution to (P_u) at which the constraint qualification* (2.2) *is satisfied. Then there is a multiplier vector y such that the optimality condition* (2.1) *is satisfied.*

This is not the sharpest result that may be stated, although it is perhaps the simplest. Clarke's paper [11] puts a potentially smaller set in place of $N_K(x)$ and provides along side of (2.2) a less stringent constraint qualification in terms of "calmness" of (P_u) with respect to perturbations of u. Hiriart-Urruty [12] and Rockafellar [13] contribute some alternative ways of writing the subgradient relations. For our purposes here, let it suffice to mention that Theorem 2 remains true when the optimality condition (2.1) is given in the slightly sharper and more elegant form:

$$0 \in \partial g(x) + y\,\partial G(x) + N_K(x) \text{ with } y \in N_C(G(x)+u), \tag{2.3}$$

where $G(x) = (g_1(x), \ldots, g_m(x))$ and

$$C = \{w \in R^m \mid w_i \leq 0 \text{ for } i=1,\ldots,s \text{ and } w_i=0 \text{ for } i=s+1,\ldots,m\}. \tag{2.4}$$

The notation $\partial G(x)$ refers to Clarke's generalized Jacobian [2] for the mapping G; one has

$$y\,\partial G(x) = \partial(\textstyle\sum_{i=1}^{m} y_i g_i)(x). \tag{2.5}$$

Theorem 2 has the shining virtue of combining the necessary conditions for smooth programming and the ones for convex programming into a single statement. Moreover it covers subsmooth programming and much more, and it allows for an abstract constraint in the form of $x \in K$ for an arbitrary closed set K. Formulas for calculating the normal cone $N_K(x)$ in particular cases can then be used to achieve additional specializations.

What Theorem 2 does *not* do is provide any interpretation for the multipliers y_i. In order to arrive at such an interpretation, it is necessary to look more closely at the properties of the marginal function

$$p(u) = \text{optimal value (infimum) in} (P_u). \tag{2.6}$$

This is an extended-real-valued function on R^m which is lower semicontinuous when the following mild *inf-boundedness condition* is fulfilled:

For each $\bar{u} \in R^m$, $\alpha \in R$ and $\varepsilon > 0$, the set of all $x \in K$ (2.7)

satisfying $g(x) \le \alpha$, $g_i(x) \le \bar{u}_i + \varepsilon$ for $i = 1,\ldots,s$, and

$\bar{u}_i - \varepsilon \le g_i(x) \le \bar{u}_i + \varepsilon$ for $i = s+1,\ldots,m$, is bounded in R^n.

This condition also implies that for each u with $p(u) < \infty$ (i.e. with the constraints of (P_u) consistent), the set of all (globally) optimal solutions to (P_u) is nonempty and compact.

In order to state the main general result, we let

$Y(u)$ = set of all multiplier vectors y that satisfy (2.1) (2.8)

for some optimal solution x to (P_u).

THEOREM 3 (Rockafellar [13]). *Suppose the inf-boundedness condition* (2.7) *is satisfied. Let u be such that the constraints of (P_u) are consistent and every optimal solution x to (P_u) satisfies the constraint qualification* (2.2). *Then $\partial p(u)$ is a nonempty compact set with*

$$\partial p(u) \subset \text{co}\, Y(u) \textit{ and } \text{ext}\, \partial p(u) \subset Y(u). \tag{2.9}$$

(where "ext" denotes extreme points). In particular p is locally Lipschitzian around u with

$$p^\circ(u;h) \leq \sup_{y \in Y(u)} y \cdot h \quad \textit{for all } h. \tag{2.10}$$

Indeed, any λ satisfying $|y| < \lambda$ for all $y \in Y(u)$ serves as a local Lipschitz constant:

$$|p(u'')-p(u')| \leq \lambda |u''-u'| \quad \textit{when } u' \textit{ and } u'' \textit{ are near } u. \tag{2.11}$$

For smooth programming, this result was first proved by Gauvin [14]. He demonstrated further that when (P_u) has a unique optimal solution x, for which there is a unique multiplier vector y, so that $Y(u) = \{y\}$, then actually p is differentiable at u with $\nabla p(u) = y$. For convex programming one knows (see [3]) that $\partial p(u) = Y(u)$ always (under our inf-boundedness assumption) and consequently

$$p'(u;h) = \max_{y \in Y(u)} y \cdot h. \tag{2.12}$$

Minimax formulas that give $p'(u;h)$ in certain cases of smooth programming where $Y(u)$ is not just a singleton can be for example found in Demyanov and Malozemov [15] and Rockafellar [16]. Aside from such special cases there are no formulas known for $p'(u;h)$. Nevertheless, Theorem 3 does provide an estimate, because $p'(u;h) \leq p^\circ(u;h)$ whenever $p'(u;h)$ exists. (It is interesting to note in this connection that because p is Lipschitzian around u by Theorem 3, it is actually differentiable almost everywhere around u by Rademacher's theorem.)

Theorem 3 has recently been broadened in [6] to include more general kinds of perturbations. Consider the parameterized problem

$$(Q_v) \qquad \text{minimize } f(v,x) \text{ over all } x \text{ satisfying}$$
$$F(v,x) \in C \text{ and } (v,x) \in D,$$

where v is a parameter vector in R^d, the functions $f: R^d \times R^n \rightarrow R$ and $F: R^d \times R^n \rightarrow R^m$ are locally Lipschitzian, and the sets $C \subset R^m$ and $D \subset R^d \times R^n$ are closed. Here C could be the cone in (2.4), in which event the constraint $F(v,x) \in C$ would reduce to

$$\begin{aligned} f_i(v,x) &\leq 0 \text{ for } i=1,\ldots,s, \\ &= 0 \text{ for } i=s+1,\ldots,m, \end{aligned}$$

but this choice of C is not required. The condition $(v,x) \in D$ may equivalently be written as $x \in \Gamma(v)$, where Γ is the closed multifunction whose graph is D. It represents therefore an abstract constraint that can vary with v. A fixed abstract constraint $x \in K$ corresponds to $\Gamma(v) \equiv K$, $D = R^d \times K$.

In this more general setting the appropriate optimality condition for a feasible solution x to (Q_v) is

$$(z,0) \in \partial f(v,x) + y\,\partial F(v,x) + N_D(v,x) \tag{2.13}$$

for some y and z with $y \in N_C(F(v,x))$,

and the constraint qualification is

the only vector pair (y,z) satisfying the version of (2.13) (2.14)

in which the term $\partial f(v,x)$ is omitted is $(y,z)=(0,0)$.

THEOREM 4 (Rockafellar [6, §8]). *Suppose that x is a locally optimal solution to (Q_v) at which the constraint qualification* (2.14) *is satisfied. Then there is a multiplier pair (y,z) such that the optimality condition* (2.13) *is satisfied.*

Theorem 4 reduces to the version of Theorem 2 having (2.3) in place of (2.1) when (Q_v) is taken to be of the form (P_u), namely when $f(v,x) \equiv g(x)$, $F(v,x) = G(x) + v$, $D = R^m \times K$ $(R^m = R^d)$, and C is the cone in (2.4).

For the corresponding version of Theorem 3 in terms of the marginal function

$$q(v) = \text{optimal value in } (Q_v), \tag{2.15}$$

we take inf-boundedness to mean:

For each $\bar{v} \in R^d$, $\alpha \in R$ and $\varepsilon > 0$, the set of all x (2.16)

satisfying for some v with $|v - \bar{v}| \le \varepsilon$

the constraints $F(v,x) \in C$, $(v,x) \in D$, and

having $f(v,x) \le \alpha$, is bounded in R^n.

Again, this property ensures that q is lower semicontinuous, and that for every v for which the constraints of (Q_v) are consistent, the set of optimal solutions to (Q_v) is nonempty and compact. Let

$Z(v)$ = set of all vectors z that satisfy the multiplier condition (2.13) for some optimal solution x to (Q_v) and vector y. (2.17)

THEOREM 5 (Rockafellar [6, §8]). *Suppose the inf-boundedness condition* (2.16) *is satisfied. Let v be such that the constraints of (Q_v) are consistent and every optimal solution x to (Q_v) satisfies the constraint qualification* (2.14). *Then $\partial q(v)$ is a nonempty compact set with*

$$\partial q(v) \subset \text{co}\, Z(v) \text{ and } \text{ext}\, \partial q(v) \subset Z(v). \tag{2.18}$$

In particular q is locally Lipschitzian around v with

$$q^{\circ}(v;h) \le \sup_{z \in Z(v)} z \cdot h \text{ for all } h. \tag{2.19}$$

Any λ satisfying $|z| < \lambda$ for all $z \in Z(v)$ serves as a local Lipschitz constant:

$$|q(v'') - q(v')| \le \lambda |v'' - v'| \text{ when } v' \text{ and } v'' \text{ are near } v. \tag{2.20}$$

The generality of the constraint structure in Theorem 5 will make possible in the next section an application to the study of multifunctions.

3. STABILITY OF CONSTRAINT SYSTEMS

The sensitivity results that have just been presented are concerned with what happens to the optimal value in a problem when parameters vary. It turns out, though, that they can be applied to the study of what happens to the feasible solution set and the optimal solution set. In order to explain this and indicate the main results, we must consider the kind of Lipschitzian property that pertains to multifunctions (set-valued mappings) and the way that this can be characterized in terms of an associated distance function.

Let $\Gamma: R^d \rightrightarrows R^n$ be a closed-valued multifunction, i.e. $\Gamma(v)$ is for each $v \in R^d$ a closed subset of R^n, possibly empty. The motivating examples are, first, $\Gamma(v)$ taken to be the set of all feasible solutions to the parameterized optimization problem (Q_v) above, and second, $\Gamma(v)$ taken to be the set of all optimal solutions to (Q_v).

One says that $\Gamma(v)$ is *locally Lipschitzian* around v if for all v' and v'' in some neighborhood of v one has $\Gamma(v')$ and $\Gamma(v'')$ nonempty and bounded with

$$\Gamma(v'') \subset \Gamma(v') + \lambda |v''-v'| B. \tag{3.1}$$

Here B denotes the closed unit ball in R^n and λ is a Lipschitz constant. This property can be expressed equivalently by means of the classical Hausdorff metric on the space of all nonempty compact subsets of R^n:

$$\text{haus}\,(\Gamma(v''),\Gamma(v')) \leq \lambda |v''-v'| \text{ when } v' \text{ and } v'' \text{ are near } v. \tag{3.2}$$

It is interesting to note that this is a "differential" property of sorts, inasmuch as it deals with rates of change, or at least bounds on such rates. Until recently, however, there has not been any viable proposal for "differentiation" of Γ that might be associated with it. A concept investigated by Aubin [17] now appears promising as a candidate; see the end of this section.

Two other definitions are needed. The multifunction Γ is *locally bounded* at v if there is a neighborhood V of v and a bounded set $S \subset R^n$ such that $\Gamma(v') \subset S$ for all $v' \in V$. It is *closed at* v if the existence of sequences $\{v_k\}$ and $\{x_k\}$ with $v_k \to v$, $x_k \in \Gamma(v_k)$ and $x_k \to x$ implies $x \in \Gamma(v)$. Finally, we introduce for Γ the *distance function*

$$d_\Gamma(v,w) = \text{dist}\,(\Gamma(v),w) = \min_{x \in \Gamma(v)} |x-w|. \tag{3.3}$$

The following general criterion for Lipschitz continuity can then be stated.

THEOREM 6 (Rockafellar [18]). *The multifunction Γ is locally Lipschitzian around v if and only if Γ is closed and locally bounded at v with $\Gamma(v) \neq \phi$, and its distance function d_Γ is locally Lipschitzian around (v,x) for each $x \in \Gamma(v)$.*

The crucial feature of this criterion is that it reduces the Lipschitz continuity of Γ to the Lipschitz continuity of a function d_Γ which is actually the marginal function for a certain optimization problem (3.3) parameterized by vectors v and w. This problem fits the mold of (Q_v), with v replaced by (v,w), and it therefore comes under the control of Theorem 5, in an adapted form. One is readily able by this route to derive the following.

THEOREM 7 (Rockafellar [18]). *Let Γ be the multifunction that assigns to each $v \in R^d$ the set of all feasible solutions to problem (Q_v):*

$$\Gamma(v) = \{x \mid F(v,x) \in C \ \textit{and} \ (v,x) \in D\}. \tag{3.4}$$

Suppose for a given v that Γ is locally bounded at v, and that $\Gamma(v)$ is nonempty with the constraint qualification (2.14) satisfied by every $x \in \Gamma(v)$. Then Γ is locally Lipschitzian around v.

COROLLARY. *Let $\Gamma: R^d \rightrightarrows R^n$ be any multifunction whose graph $D = \{(v,x) \mid x \in \Gamma(v)\}$ is closed. Suppose for a given v that Γ is locally bounded at v, and that $\Gamma(v)$ is nonempty with the following condition satisfied for every $x \in \Gamma(v)$:*

the only vector z with $(z,0) \in N_D(v,x)$ is $z = 0$. (3.5)

Then Γ is locally Lipschitzian around v.

The corollary is just the case of the theorem where the constraint $F(v,x) \in C$ is trivialized. It corresponds closely to a result of Aubin [17], according to which Γ is "pseudo-Lipschitzian" relative to the particular pair (v,x) with $x \in \Gamma(v)$ if

the projection of the tangent cone $T_D(v,x) \subset R^d \times R^n$ (3.6)
on R^d is all of R^d.

Conditions (3.5) and (3.6) are equivalent to each other by the duality between $N_D(v,x)$ and $T_D(v,x)$. The "pseudo-Lipschitzian" property of Aubin, which will not be defined here, is a suitable localization of Lipschitz continuity which facilitates the treatment of multifunctions Γ with $\Gamma(v)$ unbounded, as is highly desirable for other purposes in optimization theory (for instance the treatment of epigraphs dependent on a parameter vector v). As a matter of fact, the results in Rockafellar [18] build on this concept of Aubin and are not limited to locally bounded multifunctions. Only a special case has been presented in the present paper.

This topic is also connected with interesting ideas that Aubin has pursued towards a differential theory of multifunctions. Aubin defines the multifunction whose graph is the Clarke tangent cone $T_D(v,x)$, where D is the graph of Γ, to be the *derivative* of Γ at v relative to the point $x \in \Gamma(v)$. In denoting this derivative multifunction by $\Gamma'_{v,x}$, we have, because $T_D(v,x)$ is a closed convex cone, that $\Gamma'_{v,x}$ is a *closed convex process* from R^d to R^n in the sense of convex analysis [3, §39]. Convex processes are very much akin to linear transformations, and there is quite a *convex algebra* for them (see [3, §39], [19], and [20]). In particular, $\Gamma'_{v,x}$ has an *adjoint* $\Gamma'^{*}_{v,x}: R^n \rightrightarrows R^d$, which turns out in this case to be the closed convex process with

$$\text{gph } \Gamma'^{\bullet}_{v,x} = \{(w,z) \mid (z,-w) \in N_D(v,x)\}.$$

In these terms Aubin's condition (3.6) can be written as dom $\Gamma'_{v,x} = R^d$, whereas the dual condition (3.5) is $\Gamma'^{\bullet}_{v,x}(0) = \{0\}$. The latter is equivalent to $\Gamma'^{\bullet}_{v,x}$ being locally bounded at the origin.

There is too much in this vein for us to bring forth here, but the few facts we have cited may serve to indicate some new directions in which nonsmooth analysis is now going. We may soon have a highly developed apparatus that can be applied to the study of all kinds of multifunctions and thereby to subdifferential multifunctions in particular.

For example, as an aid in the analysis of the stability of optimal solutions and multiplier vectors in problem (Q_v), one can take up the study of the Lipschitzian properties of the multifunction

$\Gamma(v)$ = set of all (x,y,z) such that x is feasible in (Q_v)

and the optimality condition (2.13) is satisfied.

Some results on such lines are given in Aubin [17] and Rockafellar [21].

REFERENCES

[1] F.H. Clarke, "Generalized gradients and applications", Trans. Amer. Soc. 205 (1975), pp.247-262.

[2] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York, 1983.

[3] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton NJ, 1970.

[4] R.T. Rockafellar, "Favorable classes of Lipschitz continuous functions in subgradient optimization", *Processes in Nondifferentiable Optimization*, E. Nurminski (ed.), IIASA Collaborative Proceeding Series, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1982, pp. 125-143.

[5] S. Saks, *Theory of the Integral*, Monografie Matematyczne Ser., no. 7, 1937; 2nd rev.ed. Dover Press, New York, 1964.

[6] R.T. Rockafellar, "Extensions of subgradient calculus with applications to optimization", J. Nonlinear Anal., to appear in 1985.

[7] R.W. Chaney, Math. Oper. Res. 9 (1984).

[8] R.T. Rockafellar, "Generalized directional derivatives and subgradients of nonconvex functions", Canadian J. Math. *32* (1980), pp. 157-180.

[9] R.T. Rockafellar, *The Theory of Subgradients and its Applications to Problems of Optimization: Convex and Nonconvex Functions*, Heldermann Verlag, West Berlin, 1981.

[10] R.T. Rockafellar, "Clarke's tangent cones and the boundaries of closed sets in R^n", J. Nonlin. Anal. 3 (1970), pp.145-154.

[11] F.H. Clarke, "A new approach to Lagrange multipliers", Math. Oper. Res. 1 (1976), pp. 165-174.

[12] J-B Hiriart-Urruty, "Refinements of necessary optimality conditions in nondifferentiable programming, I," Appl. Math. Opt. 5 (1979), pp.63-82.

[13] R.T. Rockafellar, "Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming", Math. Prog. Study 17 (1982), 28-66.

[14] J. Gauvin, "The generalized gradient of a marginal function in mathematical programming problem", Math. Oper. Res. 4 (1979), pp.458-463.

[15] V.F. Demyanov and V.N Malozemov, "On the theory of nonlinear minimax problems", Russ. Math. Surv. 26 (1971), 57-115.

[16] R.T. Rockafellar,"Directional differentiability of the optimal value in a nonlinear programming problem", Math. Prog. Studies 21 (1984), pp. 213-226.

[17] J.P. Aubin, "Lipschitz behavior of solutions to convex minimization problems", Math. Oper. Res. 9 (1984), pp. 87-111.

[18] R.T. Rockafellar, "Lipschitzian properties of multifunctions", J. Nonlin. Anal., to appear in 1985.

[19] R.T. Rockafellar, "Convex algebra and duality in dynamic models of production", in Mathematical Models of Economic (J. Los', ed.), North-Holland, 1973, pp.351-378.

[20] R.T. Rockafellar, *Monotone Processes of Convex and Concave Type*, Memoir no.77, Amer. Math. Soc., Providence RI, 1967.

[21] R.T. Rockafellar, "Maximal monotone relations and the second derivatives of nonsmooth functions", Ann. Inst. H. Poincaré, Analyse Non Linéaire 2 (1985), pp.167-184.

UPPER-SEMICONTINUOUSLY DIRECTIONALLY DIFFERENTIABLE FUNCTIONS

A.M. Rubinov
Institute for Social and Economic Problems, USSR Academy of Sciences, Leningrad, USSR

1. INTRODUCTION

A generalized approximation of the subdifferential called the (ε,μ)-subdifferential is introduced for upper-semicontinuously directionally differentiable functions. The most attractive and important property of the (ε,μ)-subdifferential is that it can be taken to be a continuous mapping; this, in its turn, allows us to construct numerical methods for finding stationary points.

Let us consider the n-dimensional space $\mathbb{R}^n$ with some norm $\|\cdot\|$. Let X be an open set in this space, and a function f be defined, continuous and directionally differentiable on X. We say that the function f is *upper-semicontinuously directionally differentiable* (u.s.c.d.d.) at a point $x_0 \in X$ if for any fixed $g \in \mathbb{R}^n$ the function $x \longrightarrow f'(x,g)$ is upper-semicontinuous (in x) at this point and is bounded in some neighborhood of x_0. This last property means that there exists a number $C < \infty$ such that

$$|f'(x,g)| \leq C\|g\| \tag{1}$$

for all $g \in \mathbb{R}^n$ and every x in some neighborhood of x_0. Examples of u.s.c.d.d. functions include convex functions and maximum functions.

We say that a function f defined on X is *subdifferentiable* at a point $x \in X$ if it is directionally differentiable at x and if its directional derivative f'_x is a sublinear function (as a function of g).

Let $\underline{\partial} f(x)$ denote the subdifferential of f at x. By definition

$$f'_x(g) = \max_{v \in \underline{\partial} f(x)} (v,g) \qquad \forall x \in \mathbb{R}^n \quad .$$

Recall that the subdifferential is a convex compact set.

PROPOSITION 1. *If a function* f *is u.s.c.d.d. at a point* $x \in X$, *then it is also subdifferentiable at this point.*

Proof. The positive homogeneity of the function $f'_x(g) = f'(x,g)$ is obvious. Let us now check that it is subadditive. Take g_1, $g_2 \in \mathbb{R}^n$. Then there exist functions $\psi_1(\alpha)$ and $\psi_2(\alpha)$ such that

$$\psi_1(\alpha) \xrightarrow[\alpha\to+0]{} 0 \quad , \qquad \psi_2(\alpha) \xrightarrow[\alpha\to+0]{} 0$$

and

$$f'(x,g_1) = \frac{1}{\alpha} [f(x+\alpha g_1) - f(x)] + \psi_1(\alpha)$$

$$f'(x,g_1+g_2) = \frac{1}{\alpha} [f(x+\alpha g_1+\alpha g_2) - f(x)] + \psi_2(\alpha) \quad .$$

The above equalities imply that

$$f'(x,g_1+g_2) - f'(x,g_1) = \frac{1}{\alpha} [f(x+\alpha g_1+\alpha g_2) - f(x+\alpha g_1)] + \psi_3(\alpha) \quad ,$$

where

$$\psi_3(\alpha) = \psi_2(\alpha) - \psi_1(\alpha) \xrightarrow[\alpha\to+0]{} 0 \quad .$$

Fix some $\alpha > 0$, put $x_\alpha = x + \alpha g_1$, and define

$$M_\alpha = \sup_{0 \le \beta \le \alpha} f'(x_\alpha+\beta g_2, g_2) \quad .$$

It follows from the mean value theorem that

$$f(x_\alpha+\alpha g_2) - f(x_\alpha) \le M_\alpha \cdot \alpha \quad .$$

Therefore

$$f'(x,g_1+g_2) - f'(x,g_1) = \frac{1}{\alpha}[f(x_\alpha+\alpha g_2) - f(x_2)] + \psi_3(\alpha) \leq M_\alpha+\psi_3(\alpha) \quad .$$

Since f is an u.s.c.d.d. function, the derivative $f'(x,g_2)$ is u.s.c. (as a function of x). This means that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$f'(y,g_2) < f'(x,g_2) + \frac{\varepsilon}{2} \qquad \forall y \in B_\delta(x) \quad .$$

For α sufficiently small and $\beta \in (0,\alpha)$ we have

$$x_\alpha + \beta g_2 = x + \alpha g_1 + \beta g_2 \in B_\delta(x)$$

and therefore $M_\alpha < f'(x,g_2) + \varepsilon/2$. Assuming that $|\psi_3(\alpha)| < \varepsilon/2$ (which is the case if α is sufficiently small), we have

$$f'(x,g_1+g_2) - f'(x+g_1) \leq f'(x,g_2) + \varepsilon \quad ,$$

which implies (since ε is arbitrary) that the function $f'_x(g) = f'(x,g)$ is subadditive.

Let a function f defined on an open set $X \subset \mathbb{R}^n$ be u.s.c.d.d. on this set. It follows from Proposition 1 that f is subdifferentiable at every point $x \in X$ (and the subdifferential $\underline{\partial}f(x)$ is defined for every $x \in X$). Fix any $g \in \mathbb{R}^n$ and consider the function

$$q_g(x) = \max_{v \in \underline{\partial}f(x)} (v,g) = f'(x,g) \quad .$$

It follows from the definition that q_g is an u.s.c. function. Inequality (1) implies that the mapping $\underline{\partial}f$ is bounded in some neighborhood of every point $x \in X$. Thus the mapping $x \to \underline{\partial}f(x)$ is u.s.c.

Using methods from the topological theory of multivalued mappings (see, e.g., [1]) it is not difficult ~~t~~o show that every point $x_0 \in X$ has a neighborhood (in which the mapping $x \to \underline{\partial}f(x)$ is bounded) such that for any fixed $\varepsilon > 0$ we can find a continuous multivalued mapping b defined in this neighborhood which has convex compact sets as its images and for which

$$\underline{\partial} f(x) \subset b(x) \subset \underline{\partial} f(B_\varepsilon(x)) + B_\varepsilon \quad . \tag{2}$$

Here $B_\varepsilon(x) = x+B_\varepsilon$; $B_\varepsilon + B_\varepsilon(0)$.

For simplicity we assume that the mapping $x \to \underline{\partial} f(x)$ is bounded on all of set X. Then a continuous mapping b satisfying (2) can be defined on the entire set X.

Let ε and μ be positive numbers. It follows directly from (2) that there exists a continuous mapping b such that

$$\underline{\partial} f(x) \subset \underline{\partial} f(B_\varepsilon(x)) + B_\mu \qquad \forall x \in X \quad . \tag{3}$$

One example would be a mapping b which satisfies (2) for $\varepsilon' = \min\{\varepsilon,\mu\}$.

A continuous mapping b which satisfies (3) is called *a continuous* (ε,μ)*-subdifferential* of the function f and is denoted by $\underline{d}_{\varepsilon\mu} f$. Clearly, this mapping is not unique: if $0 < \varepsilon' \leq \varepsilon$, $0 < \mu' \leq \mu$ then every continuous (ε',μ')-subdifferential is also a continuous (ε,μ)-subdifferential.

The definition of a continuous (ε,μ)-subdifferential can be extended to the case in which one of the numbers ε and μ is zero. However, in this case we cannot guarantee the existence of a continuous (ε,μ)-subdifferential for an arbitrary u.s.c.d.d. function, although continuous $(\varepsilon,0)$-subdifferentials do exist for convex functions. We shall now describe one of these.

Let a function f be defined and convex on an open convex set X. By $\underline{\partial}_\varepsilon f(x)$ we denote the *conditional* ε*-subdifferential* of f at x with respect to the ball $B_\varepsilon(x)$ (see [2]):

$$\underline{\partial}_\varepsilon f(x) = \{v \in \mathbb{R}^n \mid f(z) - f(x) \geq (v,z-x) - \varepsilon \ \forall z \in B_\varepsilon(x)\}$$

PROPOSITION 2. *Let a function* f *be defined and convex on an open convex set* $X \in \mathbb{R}^n$. *Then the mapping* $\underline{\partial}_\varepsilon f$ *is a continuous* $(\varepsilon,0)$*-subdifferential of the function* f.

Proof. It follows from [3] that $\underline{\partial}_\varepsilon f(x)$ coincides with the closure of the set

$C_\varepsilon f(x) = \{v \in \mathbb{R}^n \mid \exists\, x' \in \text{int } B_\varepsilon(x) : v \in \underline{\partial} f(x'); f(x') - f(x) \geq (v, x'-x) - \varepsilon\}$.

From the definition, we have

$$C_\varepsilon f(x) \subset \underline{\partial} f(\text{int } B_\varepsilon(x)) \subset \underline{\partial} f(B_\varepsilon(x)) \quad .$$

In addition, $\underline{\partial} f(x) \subset C_\varepsilon f(x)$ and sets $\partial_\varepsilon f(x)$ are convex and compact (the latter follows from [3]). Thus

$$\underline{\partial} f(x) \subset \partial_\varepsilon f(x) \subset \underline{\partial} f(B_\varepsilon(x)) \quad . \tag{4}$$

It is now necessary to demonstrate the continuity of the mapping $\underline{\partial}_\varepsilon f(x)$. It follows from [3] that the support function $q_\varepsilon f(x,g)$ of the set $\underline{\partial}_\varepsilon f(x)$ is given by

$$q_\varepsilon f(x,g) = \inf_{0<\alpha \leq \frac{\varepsilon}{\|g\|}} \frac{1}{\alpha} [f(x+\alpha g) - f(x) + \varepsilon] \quad .$$

Fix any vector y and consider the function

$$h(x,\alpha) = \frac{1}{\alpha} [f(x+\alpha g) - f(x) + \varepsilon] \quad ,$$

which is jointly continuous in both variables on $X \times (0, \frac{\varepsilon}{\|g\|}]$. Fix $x_0 \in X$. Since

$$\lim_{\alpha \to +0} (hx_0, \alpha) = +\infty \quad ,$$

there exist numbers $\partial > 0$ and $\alpha_\infty > 0$ such that

$$\inf_{0<\alpha \leq \frac{\varepsilon}{\|g\|}} h(x,\alpha) = \min_{\alpha_0 \leq \alpha \leq \frac{\varepsilon}{\|g\|}} h(x,\alpha) \qquad \forall x \in B_\partial(x_0) \quad .$$

Since h is jointly continuous in both variables on the compact set $B_\partial(x_0) \times [\alpha_0, \frac{\varepsilon}{g}]$, the function

$$x \longrightarrow q_\varepsilon f(x,g) = \min_{\alpha_0 \leq \alpha \leq \frac{\varepsilon}{\|g\|}} h(x,\alpha)$$

is continuous at the point x_0. Also, from (4) and the boundedness of the subdifferential, the mapping $x \longrightarrow \underline{\partial}_\varepsilon f(x)$ is bounded in some neighborhood of x_0. Using results from [2], we then deduce that the mapping $x \longrightarrow \underline{\partial}_\varepsilon f(x)$ is continuous.

THEOREM 1. *(On the continuous* (ε,μ)*-subdifferential of a composition.)*

Let a function f *be defined, Lipschitzian and u.s.c.d.d. on an open set* $X_1 \subset \mathbb{R}^n$. *Suppose also that for any* $\varepsilon > 0$ *and* $\mu > 0$ *there exists a continuous* (ε,μ)*-subdifferential* $\underline{d}_{\varepsilon\mu}f$. *Let functions* $h_1,\dots,h_n$ *be defined and continuously differentiable on an open set* $X_2 \subset \mathbb{R}^m$, *where* $m \geq n$.

Consider a mapping $H(x) = (h_1(x),\dots,h_n(x))$ *such that*

(i) $H(X_2) \subset X_1$

(ii) The Jacobian matrix

$$H'_x = \begin{pmatrix} \dfrac{\partial h_1}{\partial x^{(1)}}, & \dots, & \dfrac{\partial h_1}{\partial x^{(n)}} \\ \\ \dfrac{\partial h_n}{\partial x^{(1)}}, & \dots, & \dfrac{\partial h_n}{\partial x^{(n)}} \end{pmatrix}$$

has a minor of n-th *order which does not vanish on the closure* cl X *of some bounded open subset* X *of the set* X_2.

Then the function $\phi(x) = f(H(x))$ *is u.s.c.d.d. and for any* $\delta > 0$, $\gamma > 0$ *there exist* $\varepsilon > 0$ *and* $\mu > 0$ *such that the mapping*

$$x \longrightarrow (H'_x)^* \, \underline{d}_{\varepsilon\mu} f(H(x))$$

is a continuous (δ,ν)*-subdifferential of the function* ϕ *on the set* X_1. *Here* * *denotes transposition.*

The proof is based on the following lemma.

LEMMA 1. *Under the assumptions of Theorem 1, for any* $\delta > 0$ *there must exist an* $\varepsilon > 0$ *such that*

$$H(x) + B_\varepsilon \subset H(x+B_\delta) \equiv H(B_\delta(x)) \qquad \forall x \in X \quad .$$

<u>Proof of Lemma 1</u>. Let us first show that the image of any neighborhood of a point $x \in X_2$ contains a ball centered at the point $H(x)$. Assume for the sake of argument that the minor which does not vanish (see condition (ii)) corresponds to the first n indices. Let

$$\bar{x} = (\bar{x}^{(1)},\ldots,\bar{x}^{(n)},\bar{x}^{(n+1)},\ldots,\bar{x}^{(m)}) \in X_2 \quad .$$

Consider the set

$$\tilde{X}_2 = \{y = (y^{(1)},\ldots,y^{(n)} \in \mathbb{R}^n \mid (y^{(1)},\ldots,y^{(n)},\bar{x}^{(n+1)},\ldots,\bar{x}^{(m)}) \in X_2\}$$

and the mapping $\tilde{H}$ defined on this set by the equality $\tilde{H}(y) = H(x)$, where $x = (y^{(1)},\ldots,y^{(n)},\bar{x}^{(n+1)},\ldots,\bar{x}^{(m)})$.

Since the Jacobian of this mapping does not vanish at the point $\bar{y} = (\bar{x}^{(1)},\ldots,\bar{x}^{(n)})$, it follows from the inverse function theorem that in some neighborhood of this point there exists a continuous mapping $\tilde{H}^{-1}$ which is the inverse of $\tilde{H}$. The continuity of $\tilde{H}^{-1}$ implies that the image of every sufficiently small neighborhood of $\bar{y}$ (under the mapping $\tilde{H}$) contains a ball centered at the point $\tilde{H}(\bar{y}) = H(\bar{x})$. Furthermore, the image of any neighborhood of the point $\bar{x}$ in the set X_2 (under the mapping H) contains a ball centered at the point $\bar{x}$.

Fix some $\delta > 0$. For any $x \in X_2$ let $\varepsilon(x)$ denote the supremum of the set of numbers $\varepsilon > 0$ such that

$$H(x) + \tilde{B}_\varepsilon \subset H(x+\tilde{B}_\delta) \qquad \forall x \in X_2 \ .$$

Here $\tilde{B}_\delta$ and $\tilde{B}_\varepsilon$ are open balls centered at zero with a radius of δ and ε, respectively. It follows from the above definitions that $\varepsilon(x) > 0$ for all x. Let us show that the function $\varepsilon(x)$ is l.s.c. Assuming the opposite, we should be able to find a sequence $\{x_k\}$ and numbers ε', $\varepsilon'' > 0$ such that

$$x_k \longrightarrow x,\ x_k \in X_2,\ \varepsilon(x) > \varepsilon'' > \varepsilon' > \varepsilon(x_k) \qquad \forall k \quad .$$

The inequality $\varepsilon' > \varepsilon(x_k)$ implies that there exist elements $\{y_k\}$ such that

$$\|H(x_k) - y_k\| < \varepsilon' , \quad y'_k \notin H(x_k + \tilde{B}_\delta) \quad . \tag{5}$$

Since the sequence $\{H(x_k)\}$ converges, the sequence $\{y_k\}$ is bounded. Without loss of generality we can assume that the limit $\lim y_k = y$ exists. Then since

$$\|H(x)-y\| = \lim \|H(x_k)-y_k\| \leq \varepsilon' < \varepsilon'' < \varepsilon(x) \quad ,$$

we have

$$y \in H(x) + \tilde{B}_{\varepsilon''} \subset H(x+\tilde{B}_\delta) \quad ,$$

i.e., for some $x' \in x + \tilde{B}_\delta$ the equality $y = H(x')$ holds. Let $\|x'-x\| = \delta' < \delta$, and take numbers γ and γ' such that $0 < 2\gamma < \gamma' < \delta - \delta'$. Since the image of a neighborhood contains a neighborhood and $y_k \rightarrow H(x')$, the inclusion $y_k \in H(x'+\tilde{B}_\gamma)$ holds for n sufficiently large. Let numbers k be such that

$$\|\tilde{x}-x_k\| \leq \|\tilde{x}-x'\| + \|x'-x\| + \|x-x_k\| < 2\gamma + \delta' < \delta \quad .$$

We conclude that $x' + \tilde{B}_\gamma \subset x_k + \tilde{B}_\delta$ and therefore that

$$y_k \in H(x'+\tilde{B}_\gamma) \subset H(x_k+\tilde{B}_\delta) \quad .$$

But this contradicts (5), showing that $\varepsilon(x)$ is l.s.c.

However, it is assumed that the set cl X is compact, and therefore $\varepsilon(x)$ achieves its minimum on cl X at some point x_0 and $\varepsilon(x) \geq \varepsilon(x_0) > 0$.

Proof of Theorem 1. Let $\phi(x) = f(H(x))$. Since f is Lipschitzian, we have

$$\phi'(x,g) = f'_H(H'_x(g)) = \max_{v\in\underline{\partial} f(H(x))} (v,H'_x(g)) =$$

$$= \max_{v\in\underline{\partial} f(H(x))} ((H'_x)^*v,g) = \max_{v'\in(H'_x)^*(\underline{\partial} f(H(x)))} (v',g) \quad .$$

We conclude that ϕ is an u.s.c.d.d. function and that $\underline{\partial}\phi(x) = (H'_x)^*(\underline{\partial}f(H(x))$. Let numbers $\delta > 0$, $\nu > 0$ be given. Find an $\varepsilon > 0$ which corresponds to δ (and whose existence is guaranteed by Lemma 1), and choose a μ such that $\mu\|(H'_x)^*\| \leqq \nu$. Take a continuous (ε,μ)-subdifferential $\underline{d}_{\varepsilon\mu}f$ of the function f. Then

$$\underline{\partial}f(H(x)) \subset \underline{d}_{\varepsilon\mu}f(H(x)) \subset \underline{\partial}f(Hx+B_\varepsilon) + \mu B^* \quad .$$

Applying the operator $(H'_x)^*$ to these inclusions we get

$$\begin{aligned}(H'_x)^*\underline{\partial}f(H(x)) &\subset (H'_x)^*\underline{d}_{\varepsilon\mu}f(H(x)) \\ &\subset (H'_x)^*\underline{\partial}f(H(x)+B_\varepsilon) + \mu(H'_x)^*B^* \\ &\subset (H'_x)^*\underline{\partial}f(H(x+B_\delta)) + \nu B^* \quad .\end{aligned}$$

Making use of the inequalities

$$\underline{\partial}\phi(x) = (H'_x)^*\underline{\partial}f(H(x))$$

$$\partial\phi(x+B_\delta) = \bigcup_{\|x'-x\|<\delta} (H'_x)^*\underline{\partial}f(H(x')) = (H'_x)^*\underline{\partial}f(H(x+B_\delta)) \quad ,$$

we finally arrive at

$$\underline{\partial}\phi(x) \subset (H'_x)^*\underline{d}_{\varepsilon\mu}f(H(x)) \subset \underline{\partial}\phi(x+B_\delta) + \nu B^* \quad .$$

Remark. If a function f has a continuous $(\varepsilon,0)$-subdifferential $\underline{d}_{\varepsilon 0}f$ for every $\varepsilon > 0$, then for any $\delta > 0$ there exists an $\varepsilon > 0$ such that the mapping $(H'_x)(\underline{d}_{\varepsilon 0}(H(x))$ is a continuous $(\varepsilon,0)$-subdifferential of the function $\phi = f(H)$ on the set X.

This result follows directly from the proof of the theorem.

Theorem 1 allows us to construct a continuous $(\varepsilon,0)$-subdifferential for one class of finite maximum functions.

THEOREM 2. *Let functions* $h_1,\ldots,h_n$ *be defined and continuously differentiable on an open set* $X \quad \mathbb{R}^m$ *(where* $m \geq n$*) and*

$$\phi(x) = \max_{i \in 1:n} h_i(x) \qquad \forall x \in \tilde{X} \quad .$$

Assume that the Jacobian matrix $\left\{\partial h_i/\partial x^{(j)}\right\}$ *has a minor of* n-th *order which does not vanish on the closure* cl X *of some bounded open subset* X *of the set* $\tilde{X}$. *Then for any* $\delta > 0$ *there exists an* $\varepsilon > 0$ *such that the mapping* $\underline{d}_{\varepsilon 0}(x)$ *defined below is a continuous* $(\delta,0)$*-subdifferential of* ϕ.

The mapping $\underline{d}_{\varepsilon 0}$ *is described by the relation*

$$\underline{d}_{\varepsilon 0}(x) = \left\{ y \in \mathbb{R}^m \,\middle|\, y = \sum_{i=1}^{n} v_i \frac{\partial h_i(x)}{\partial x^{(1)}}, \ldots, \sum v_i \frac{\partial h_i(x)}{\partial x^{(m)}} \right. ,$$

$$\left. v = (v_1, \ldots, v_n) \in V_\varepsilon(x) \right\} \quad ,$$

where

$$V_\varepsilon(x) = \left\{ v \in \mathbb{R}^m \,\middle|\, \sum_{i=1}^{n} v_i = 1 \; , \; v_i \geq 0 \; , \; v_i = 0 \text{ if } i \notin \tilde{R}_{2\varepsilon}(H(x)) \right. ,$$

$$\left. \phi(x) \leq \sum v_i h_i(x) + \varepsilon \right\}$$

$$\tilde{R}_{2\varepsilon}(H(x)) = \{ i \in 1:n \,|\, \phi(x) - h_i(x) < 2\varepsilon \} \quad .$$

2. A METHOD OF STEEPEST DESCENT

Let f be an u.s.c.d.d. function defined on $\mathbb{R}^n$. A point x is called an (ε,μ)-stationary point of f if

$$0 \in \underline{\partial} f(x+B_\varepsilon) + B_\mu \quad .$$

Observe that if a point x is (ε,μ)-stationary for all $\varepsilon > 0$, $\mu > 0$ then it is also stationary, i.e. $0 \in \underline{\partial} f(x)$.

Indeed, if $0 \in \underline{\partial} f(x+B_\varepsilon) + B_\mu$ then taking the limit as $\mu \to 0$ leads to $0 \in \underline{\partial} f(x+B_\varepsilon)$. But if $0 \in \underline{\partial} f(x+B_\varepsilon)$ $\forall \varepsilon$ then the upper-semicontinuity of the mapping $\underline{\partial} f$ implies that $0 \in \underline{\partial} f(x)$.

If $\underline{d}_{\varepsilon\mu} f$ is a continuous (ε,μ)-subdifferential of the function f and $0 \in \underline{d}_\varepsilon f(x)$, then x is an (ε,μ)-stationary point (by definition).

We shall now describe a steepest descent method based on the use of continuous (ε,μ)-subdifferentials.

Let

$$\phi(x,g) = \max_{v \in d_{\varepsilon\mu} f(x)} (v,g) \quad .$$

The function ϕ is the support function of the mapping $d_{\varepsilon\mu} f$. Consider the function

$$r(x) = \min_{\|g\| \le 1} \phi(x,g) \quad .$$

From the minimax theorem we have

$$r(x) = \min_{\|g\| \le 1} \max_{v \in \underline{d}_{\varepsilon\mu} f(x)} (v,g) = \max_{v \in \underline{d}_{\varepsilon\mu} f(x)} \min_{\|g\| \le 1} (v,g) =$$

$$= \max_{v \in \underline{d}_{\varepsilon\mu} f(x)} (-\|v\|) = - \min_{v \in \underline{d}_{\varepsilon\mu} f(x)} \|v\| \quad .$$

Thus $-r(x) = \rho(0, d_{\varepsilon\mu} f(x))$. If $r(x) = 0$ then $0 \in \underline{d}_{\varepsilon\mu} f(x)$, i.e., x is a stationary point.

Choose an arbitrary $x_0 \in \mathbb{R}^n$, and assume that the set $\{x \in \mathbb{R}^n \mid f(x) \le f|x_0\}$ is bounded.

Assume that a point x_k has already been found. If $r(x_k) = 0$ then x_k is an (ε,μ)-stationary point and the process terminates. Otherwise, if $r(x_k) < 0$, we find g_k such that

$$\|g_k\| = 1$$

$$r(x_k) = \min_{\|g\| \le 1} \phi(x_k,g) = \phi(x_k,g_k) \quad .$$

Now let us choose α_k such that

$$f(x_k+\alpha_k g_k) = \min_{\alpha \ge 0} f(x_k+\alpha g_k) \quad .$$

If the sequence $\{x_k\}$ thus constructed is finite then its last point is (ε,μ)-stationary by construction. Otherwise the following theorem is true.

<u>THEOREM 3</u>. *Any limit point of the sequence* $\{x_k\}$ *is an* (ε,μ)-*stationary point of the function* f.

<u>Proof</u>. We have

$$\begin{aligned} f(x_k+\alpha g_k) &= f(x_k) + \int_0^\alpha f'(x_k+\tau g_k,g_k)d\tau \\ &\leq f(x_k) + \int_0^\alpha \phi(x_k+\tau g_k,g_k)d\tau \quad . \end{aligned} \tag{6}$$

This inequality holds because $\underline{\partial} f(x) \subset \underline{d}_{\varepsilon\mu} f(x)$. Let us now prove that $\lim r(x_k) = 0$.

Assuming the opposite, we can find a subsequence $\left(x_{k_s}\right)$ such that

$$\lim_{s\to+\infty} r\left(x_{k_s}\right) = -a < 0 \quad .$$

Since the mapping $\underline{d}_{\varepsilon\mu}$ is continuous on the compact set $\{x|f(x) \leq f(x_0)\}$, it is also uniformly continuous, i.e., for any $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\rho(d_{\varepsilon\mu}f(x),d_{\varepsilon\mu}f(y)) < \varepsilon \quad \text{if} \quad \rho(x,y) < \delta \quad ,$$

where δ does not depend on points x and y. Take $\varepsilon = \frac{a}{2}$, and let $\alpha < \delta$. Then

$$\|(x_k+\varepsilon g_k) - x_k\| \leq \alpha < \delta \qquad \forall\tau \in (0,\alpha)$$

and therefore

$$\phi(x_k+\tau g_k,g_k) \leq \phi(x_k,g_k) + \frac{a}{2} \quad .$$

It now follows from (6) that

$$f\left(x_{k_s}+\alpha g_{k_s}\right) \leq f\left(x_{k_s}\right) + \int_0^{\alpha}\left[\phi\left(x_{k_s},g_{k_s}\right)+\frac{a}{2}\right]d\tau =$$

$$= f\left(x_{k_s}\right) + \alpha\left(r\left(x_{k_s}\right)+\frac{a}{2}\right) \quad .$$

But for s sufficiently large we have $r\left(x_{k_s}\right) < -\frac{3a}{4}$, and hence

$$f\left(x_{k_s}+\alpha g_{k_s}\right) \leq f\left(x_{k_s}\right) - \alpha\frac{a}{4} \quad .$$

Therefore

$$f\left(x_{k_s+1}\right) \leq f\left(x_{k_s}\right) - \alpha\frac{a}{4} \quad ,$$

which is impossible. It follows from this contradiction that we must have

$$\lim_{k\to\infty} r(x_k) = 0 \quad .$$

Since r is a continuous function the equality $r(x^*) = 0$ holds for any limit point x^* of the sequence $\{x_k\}$, i.e.,

$$0 \in \underline{d}_{\varepsilon\mu} f(x^*) \quad .$$

<u>Remark</u>. An analogous method can be used in the case $f = f_1+g$, where f is an u.s.c.d.d. function and d is a concave function, or to find a Clarke stationary point.

REFERENCES

1. Yu.G. Borisovich, B.G. Gelman, A.D. Myshkis and V.V. Obihovski. Multivalued mappings (in Russian). Achievements of Science and Engineering: Mathematical Analysis, 19(1982) 127-230.

2. V.F. Demyanov and V.K. Shomesova. Conditional subdifferentials of convex functions. Soviet Math. Dokl., 19(5) (1978)1181-1185.

3. V.F. Demyanov and L.V. Vasiliev. Nondifferentiable Optimization. Nauka, Moscow, 1981.

A NEW APPROACH TO CLARKE'S GRADIENTS IN INFINITE DIMENSIONS

Jay S. Treiman
Department of Mathematics, Lehigh University, Bethlehem, PA 18015, USA

1 Introduction:

One of the most useful tools developed for use in nonsmooth optimization is the generalized gradient set of Clarke. These gradients have been used on a variety of problems including necessary conditions for optimality, control theory and differential inclusions. Three different techniques can be used to define Clarke's gradients. They have characterizations in terms of directional derivatives [Clarke (1975), Rockafellar (1980)], the normal cone to the epigraph of a function [Clarke (1975)] and in terms of limits of proximal subgradients [Rockafellar (1981)]. Some of the strongest results involving Clarke's subgradients have been derived using the proximal subgradient formula [Rockafellar (1982)].

The characterization of Clarke's gradients in terms of proximal gradients is as follows. Let f be a l.s.c. function from $\mathbb{R}^n$ into $\bar{\mathbb{R}}$. A $v \in \mathbb{R}^n$ is a proximal subgradient to f at $\bar{x}$ if the function

$$f(x) - \langle v,x\rangle + r||x - \bar{x}||$$

has a minimum at $\bar{x}$ relative to some neighborhood of $\bar{x}$ for some r>0. Let

$$\hat{\partial}f(\bar{x}) = \begin{cases} v: \exists \text{ proximal subgradients } v^k \to v \text{ to } f \\ \quad \text{at } x^k \xrightarrow[f]{} \bar{x}. \end{cases}$$

and

$$\hat{\partial}^{\infty}f(\bar{x}) = \begin{cases} v: \exists \text{ proximal subgradients } v^k \text{ to } f \text{ at } x^k \xrightarrow[f]{} \bar{x} \\ \quad \text{with } \tau_k v^k \to v \text{ with } \tau_k \searrow 0. \end{cases}$$

The set of *Clarke subgradients* to f at $\bar{x}$ is given by

$$\partial f(\bar{x}) := \text{cl co } [\hat{\partial}f(\bar{x}) + \hat{\partial}^{\infty}f(\bar{x})]$$

Here the set $\hat{\partial}^{\infty}f(\bar{x})$ can be interpreted as the infinite subgradients.

There have several generalizations of this idea. They include the work of Thibault (1976), Kruger and Mordukhovich (1980) and Ioffe (1981).

In this paper a characterization of Clarke's gradients similar to the proximal subgradient formula is stated. This formula is valid in all reflexive Banach spaces. Several results proven using this characterization are also given.

2 The subgradient formula:

The main problem with proximal subgradients is that they may not exist in Banach spaces. They are replaced by ϵ-subgradients. Let E be a Banach space. A $v^* \in E^*$ is an ϵ-*subgradient* to a l.s.c. function f at $\bar{x}$ if

$$f(x) - \langle v^*,x\rangle + \epsilon||x - \bar{x}|| \qquad (1)$$

has a local minimum at $\bar{x}$.

It will be assumed throughout the rest of this paper that E has an equivalent norm that if Frechet differentiable off 0. This guarantees that ϵ-subgradients exist on a dense subset of the domain of f for any $\epsilon > 0$.

Theorem 1: [Treiman (1983)] *Let f be a lower semicontinuous function on E and $\bar{x}$ a point where f is finite. Take*

$$\hat{\partial} f(x) := \left\{ \begin{array}{l} v^* \ : \ \exists\ v^{*k} \xrightarrow{w^*} v^*,\ x^k \xrightarrow[f]{} x \textit{ and } \epsilon_k \searrow 0 \textit{ with} \\ \qquad v^{*k} \textit{ an } \epsilon_k\textit{-subgradient to } f \textit{ at } x^k \end{array} \right.$$

and

$$\hat{\partial}^{\infty} f(x) := \left\{ \begin{array}{l} v^* \ : \ \exists\ v^{*k} \xrightarrow{w^*} v^*,\ x^k \xrightarrow[f]{} x,\ \tau_k \searrow 0 \textit{ and } \epsilon_k \searrow 0 \textit{ with} \\ \qquad v^{*k} \textit{ an } \epsilon_k\textit{-subgradient to } \tau_k f \textit{ at } x^k. \end{array} \right.$$

Then

$$\partial f(x) = cl^{*} co\ [\hat{\partial} f(x) + \hat{\partial}^{\infty} f(x)].$$

A similar result holds in Banach spaces with an equivalent norm that is Gateaux differentiable off 0. These spaces include all separable spaces. The only differences are that the neighborhood in (1) is replaced by a set that absorbs a neighborhood of every element of E ⟨0⟩ and these absorbing sets must be uniform when taking the limits in Theorem 1.

This set of subgradients is differs from the broad cone of Ioffe (1981). In Ioffe's definition a similar ϵ-subgradient is used and is called the Dini γ-subdifferential. The major differences are that Ioffe's γ-subdifferentials are taken with respect to subspaces and he does not include infinite limits. This means that Ioffe's subgradient set can be much larger or smaller than Clarke's gradients.

The ϵ-subgradients described here are more closely related to the normals defined by Kruger and Mordukhovich (1980). A discussion of these relationships is contained in [Treiman (1983)].

3 Applications:

In this section we state several applications of Theorem 1. These are generalizations of Rockafellar's work

[Rockafellar (1985)] and will appear in [Treiman and Rockafellar (1985)]. The first of these results enables one to calculate Clarke's gradients in a special case.

Proposition 2: *Let E and X be Banach spaces with equivalent norms that are Frechet differentiable off 0 and $f\colon E \to \bar{\mathbb{R}}$ and $g\colon X \to \bar{\mathbb{R}}$ be lower semicontinuous functions. If $F(y,x) = f(y) + g(x)$ and $F(\bar{y},\bar{x})$ is finite then*

$$\partial F(\bar{y},\bar{x}) = (\partial f(\bar{y}),\partial g(\bar{x}))$$

If either $f(\bar{y})$ or $g(\bar{x})$ is empty then so is $\partial F(\bar{y},\bar{x})$.

The next result can be interpreted as a statemant about Lagrange multipliers. The proof of this result depends on a result similar to the result of Dolecki and Thera (1984) in this volume that does not require the existence of optimal solutions to perburbed problems.

In this theorem the concept of a tightly lipschitzian map is used. A map $F\colon X \to E$ is *tightly Lipschitzian* at $\bar{x}$ if F is Lipschitzian around $\bar{x}$ and for all h there is a compact set $H(h) \subset E$ such that for all $\delta > 0$ there is a $\mu > 0$ with

$$t^{-1}[F(x' + th) - F(x')] \in H(h) + \delta B$$

when $||x' - \bar{x}|| < \delta$ and $t \in (0,\delta)$.

Theorem 3: *Let $\bar{x}$ be a locally optimal solution to the problem*

$$\text{minimize } f(x) \quad \text{subject to} \quad F(x) + \bar{u} \in C,\ x \in D,$$

where $f\colon X \to \bar{\mathbb{R}}$ is lower semicontinuous with $f(\bar{x})$ finite, $F\colon X \to E$ is tightly Lipschitzian, E has an equivalent norm that is Frechet differentiable off 0 and $C \subset E$ and $D \subset X$ are closed sets. Suppose that the problem is calm in the sense that

$\nexists (x^k, u^k) \to (\bar{x}, \bar{u})$ with $F(x^k) + u^k \in C$, $x^k \in D$ such that

$$u^k \neq \bar{u} \quad \text{and} \quad \frac{f(x^k) - f(\bar{x})}{|u^k - \bar{u}|} \longrightarrow -\infty .$$

Then

$$\exists\ v^* \in N_C(F(\bar{x}) + \bar{u}) \quad \text{with} \quad 0 \in \partial(f + \delta_D)(\bar{x}) + \partial\langle v^*, F\rangle(\bar{x}).$$

Using this result several chain rules can be proven. In these chain rules the following concept is used. An element v^* of $\hat{\partial}^\infty f(x)$ is *nontrivial* if there is a sequence v^{*k} of ϵ_k-subgradients to $\tau_k f$ at $x^k \xrightarrow[f]{} x$ with $||v^{*k}|| > \delta > 0$ for some $\delta > 0$, $\epsilon_k \searrow 0$, $\tau_k \searrow 0$ and $v^{*k} \xrightarrow{w^*} v^*$. These elements give some information about the infinite behavior of the function around x.

Theorem 4: *Let* g: $X \to \mathbb{R}$ *be a directionally Lipschitzian lower semicontinuous function and* G: $E \to X$ *be tightly Lipschitzian where* X *has an equivalent norm that is Frechet differentiable off* 0 *and* $p(\bar{u}) := g(G(\bar{u}))$ *is finite. Assume that there are no nontrivial elements* $v^* \in \hat{\partial}^\infty p(\bar{u})$ *such that*

$$0 \in \partial\langle v^*, G\rangle(\bar{u})$$

Then for the sets

$$M(\bar{u}) := \bigcup_{y^* \in \hat{\partial} g(G(\bar{u}))} \partial\langle y^*, G\rangle(\bar{u}) \qquad \text{and}$$

$$M^\infty(\bar{u}) := \bigcup_{y^* \in \hat{\partial}^\infty g(G(\bar{u}))} \partial\langle y^*, G\rangle(\bar{u})$$

one has $\hat{\partial} p(\bar{u}) \subset M(\bar{u})$ *and* $\hat{\partial}^\infty p(\bar{u}) \subset M^\infty(\bar{u})$, *Thus*

$$\partial p(\bar{u}) \subset [M(\bar{u}) + M^{\infty}(\bar{u})].$$

If one assumes that the union over all nonzero elements of E of the H(h)'s in the definition of tightly Lischitzain is a separable subset of X one need only assume that g is l.s.c.

REFERENCES

Clarke, F. H. (1975) Generalized gradients and applications, Trans. Amer. Math. Soc. 205 pg. 247 - 262

Dolecki, S. and Thera, M. (1984) Upper bounds for Clarke's derivative of marginal functions in infinte dimensions, Proceeding of Sopron workshop on nondifferentiable optimization.

Ekeland, I. (1974) On the variational Principle. J. Math. Anal. Appl. 47 pg. 324 - 353

Ioffe, A. D. (1981a) Approximate subdifferentials of nonconvex functions. Chashiers Math. de la d Decision no.8120, Univ. Paris-Duaphine

Ioffe, A. D. (1981b) Nonsmooth analysis: differential calculus of nondifferentiable mappings, Trans. Amer. Math. Soc. 266 no. 1, pg 1 - 56

Kruger A. Ja. and Mordukhovich, B. Sh. (1980) Extreme points and Euler equations in nondifferentiable optimization. Dokl. Akad. Nauk BSSR 24 no. 8, pg 684 - 687

Rockafellar, R. T. (1979) Directionally Lipschitzain functions and subdifferential calculus Proc. London Math. Soc. (3) 39 pg 331 - 335

Rockafellar, R. T. (1980) Generalized directional derivatives and subgradients of nonconvex functions, Can. J. Math. 32 no. 2, pg. 257 - 280

Rockafellar, R. T. (1981) Proximal subgradients, marginal values and augmented Lagrangians in nonconvex optimization Math. Oper. Res. 6 no. 3, pg. 424 - 436

Rockafellar, R. T. (1982) Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming Math. Prog. Study 17 pg. 28 - 66

Rockafellar, R. T. (1985) Extensions of subgradient calculus with applications to optimization, Nonlinear Analysis, to appear.

Thibault, L. (1976) Quelgues propietes des sous-differentials de fonctions reeles localement lipschitziennes defines sur un espace de Banach separable C. R. Acad. Sci Paris, Ser. A-B 282 no. 10, Ai, A507 - A510

Treiman, J. S. (1983) A new characterization of Clarke's tangent cone and its applications to subgradietn analysis and optimization. Ph.D. Thesis, University of Washington

Treiman, J. S. and Rockafellar, R. T. (1985) Extensions of subgradient calculus in infinte dimensions, to appear.

II. MULTICRITERIA OPTIMIZATION AND CONTROL THEORY

A NONDIFFERENTIABLE APPROACH TO MULTICRITERIA OPTIMIZATION

Y. Evtushenko and M. Potapov
Computing Center, USSR Academy of Sciences, ul. Vavilova 40, Moscow, USSR

1. INTRODUCTION

Decision-making problems, the design of control systems, and the construction of multipurpose products all require the solution of multicriteria problems. These problems can be summarized in the following way. Let $x \in R^n$ be an n-dimensional vector of decisions (or construction parameters), and the constraint set $X \subset R^n$ to which the vectors x belong be given. The value of each decision (or the performance of the product) is estimated on the basis of m different scalar-valued criteria (objective functions): $F^i(x)$, $i \in [1:m]$. We shall denote these criteria by $F(x) = [F^1(x),\ldots,F^m(x)]$.

Decision makers would like to choose a feasible point $x \in X$ such that all the components of the vector $F(x)$ simultaneously take on the smallest possible values. However, this condition is usually unfulfillable: minimizing any one of the components will usually lead to an increase in the values of the others. Hence the term "solution of the multicriteria optimization problem" requires clarification. We will write the problem of multicriteria minimization of $F(x)$ on X as follows:

$$\min_{x \in X} F(x) \ . \tag{1}$$

Solving this problem means finding points from the Pareto set. We will say that the point x_* belongs to the Pareto set if $x_* \in X$, and there is no point x in X such that

(1) $F^i(x) \le F^i(x_*)$ for all $i \in [1:m]$ and

(2) $F^j(x) < F^j(x_*)$ for at least one $j \in [1:m]$.

The points which satisfy these conditions are also called Pareto optimal points, efficient points, or nondominated solutions. The collection of all points with the above properties is denoted X_* and called the *Pareto set*.

Introduce the images of the sets X and X_* under mapping $F(x)$:

$$Y = F(X) \ , \quad Y_* = F(X_*) \ .$$

In what follows, we will consider Y to be a nonempty set in R^m, and X to be a nonempty compact set in R^n.

The set Y is the Pareto set for the following elementary multicriteria problem:

$$\min_{y \in Y} y \ . \tag{2}$$

We will say that X_* is the Pareto set in decision (or parameter) space and its image Y_* is the Pareto set in criteria (or objective) space.

If the inequalities $y_1 = F(x_1) \le y_2 = F(x_2)$, $y_1 \ne y_2$, hold for two points $x_1, x_2 \in X$, then we will say that the point y_1 is more efficient than the point y_2, or that y_2 is less efficient than y_1.

We will assume that each component F satisfies the Lipschitz condition with the same constant L, i.e., for any x_1 and x_2 we have

$$|F^i(x_1) - F^i(x_2)| \le L \| x_1 - x_2 \| \ , \quad i \in [1:m] \ ,$$

which leads to the vector inequality

$$F(x_1) - eL \| x_1 - x_2 \| \le F(x_2) \ , \tag{3}$$

where $e \in R^n$ is the unit vector.

2. CONSTRUCTION OF THE NET

The structure of the Pareto set for even the simplest problems generally turns out to be very complex. It often happens that this set is nonconvex and nonconnected, so that it is difficult to approximate. Below we will attempt to construct a finite set A_k which resembles the usual notion of an E-net of the set Y_*. Take a set of points $A_k = [y_1, \ldots, y_k]$, where $y_i = F(x_i)$, $x_i \in X$, for all $i \in [1:k]$. We will assume that, in addition to A_k, the set of points x_i from the feasible set X is available or can easily be calculated.

Besides feasibility we impose two other conditions on the set of points A_k:

(1) for any $y_* \in Y_*$ there exists a vector $y_i \in A_k$ such that

$$y_i \leq y_* + Ee \; ; \tag{4}$$

(2) for any $y_j \in A_k$ there is no vector $y_i \in A_k$ such that $y_j \leq y_i$, $i \neq j$.

We will call the set of points A_k satisfying the above conditions an E-net of the Pareto set, and the conditions themselves the first and second net conditions, respectively.

For $y_i \in Y$ define the set

$$M_i = \{y \in R^m : y_i \leq y + Ee\} = \{y \in R^m : \min_{j \in [1:m]} (E + y^j - y_i^j) \geq 0\} \; .$$

This set contains the collection of all points which are less efficient than the point $y_i - Ee$.

Define $Z_k = \bigcup_{i=1}^k M_i$. This set can also be written in the form

$$Z_k = \{y \in R^m : \max_{i \in [1:k]} \min_{j \in [1:m]} [E + y^j - y_i^j] \geq 0\} \, .$$

The set A_k varies during the course of the calculations. If a point $\bar{y} \in Y$ is found such that $\bar{y} \leq y_i$, where $y_i \in A_k$, then y_i is taken out of A_k and replaced by $\bar{y}$. Several points can be removed simultaneously. Thanks to this, the second net condition of the Pareto set holds automatically. If the previous condition is not fulfilled and $\bar{y}$ does not belong to Z_k, then it is included in A_k, which is now written A_{k+1}.

If as a result of the construction of the set A_k it is found that

$$Y \subset Z_k \quad , \tag{5}$$

then A_k forms an E-net of the Pareto set. Indeed, for each $y_* \in Y_* \subset Y$ there is at least one point $y_i \in A_k$ such that (4) holds. The problem of constructing an E-net of the Pareto set has thus been reduced to constructing a set of points A_k satisfying (5).

The solution of the initial multicriteria optimization problem is therefore reduced to construction of the set A_k which satisfies condition (5). To do this we utilize the nonuniform space-covering technique proposed in Evtushenko (1971, 1974) for finding the global extremum of multivariable functions. This technique involves covering the set X with cubes inscribed in spheres of various radii. We present only the main formulae which differ from those described in Evtushenko (1971).

Let $y_i \in A_k$. Then the set M_i is of no interest from the viewpoint of E-net construction and can be omitted from consideration. To fulfill condition (5) we have to introduce the Lipschitz condition, or more precisely, the inequality (3).

Assume that the value $\bar{y} = F(\bar{x})$ is calculated at $\bar{x} \in X$, and suppose it turns out that $\bar{y} \in M_i$. From (3) it follows that

$$F(\bar{x}) - eL\|x - \bar{x}\| \leq F(x) \ .$$

If x is such that

$$F(x_i) - Ee \leq F(\bar{x}) - eL\|x - \bar{x}\| \ ,$$

then $y = F(x) \in M_i$. Hence all points in X which satisfy

$$eL\|x - \bar{x}\| \leq F(\bar{x}) - F(x_i) + Ee \qquad (6)$$

belong to the set M_i. The set defined by (6) contains a ball

$$B_i = \{x \in R^n : L\|x - \bar{x}\| \leq E + \min_{s \in [1:m]} |F^s(\bar{x}) - F^s(x_i)|\}$$

in the decision space. If $\bar{x} = x_i$ then the radius of the ball is at a minimum and is equal to E/L. In the case when A_k contains several points which are more efficient than $\bar{y}$, introduce the index set

$$I(\bar{y}) = \{i \in [1:k] : y_i \leq \bar{y}, \ y_i \in A_k\} \ .$$

This set contains the indices of vectors in A_k which are more efficient than $\bar{y}$. If $I(\bar{y})$ is nonempty then after determining $\bar{y} = F(\bar{x})$ one can eliminate all the points x for which (6) holds for at least one $i \in I(\bar{y})$. It is therefore optimal to choose an i such that the corresponding ball B_i has the largest radius. This radius is computed using

$$\rho = \frac{1}{L}[E + \max_{i \in I(\bar{y})} \min_{s \in [1:m]} (f^s(x) - F^s(x_i))] \ . \qquad (7)$$

Construction of the E-net of the Pareto set has thus been reduced to covering the set X with balls of the form (6). To implement this process one can use the approach described in Evtushenko (1971, 1974) and its extension. If X is bounded, then it can be covered in a finite number of steps, and the E-net will also be finite. Here, as in the search for global extrema, the computations can be speeded up by using local search methods. Such methods for determining the points in the Pareto set are now being successfully developed.

After the set A_k has been found, it is given to the engineer, who chooses his preferred set of design parameters. If the number of points in A_k turns out to be large, it can be reduced by discarding points which are close together. The distance between points can be defined in both criteria space and parameter space. The user gives a number N determining the smallest distance between points, and a special program "sifts" through the set A_k, leaving only the points which are separated by a distance greater than N.

We shall no w illustrate the application of the approach suggested above with a very simple example. Consider the case where $F^1(x) = x$, $F^2(x) = \sin \pi x$, $0 \leq x \leq 2$, $E = 0.001$. It is easy to show that in this case the Pareto set in decision space consists of the point $x = 0$ and the line segment (1,1.5]. In criteria space the Pareto set consists of the point $F^1 = F^2 = 0$ and the line $F^2 = \sin F^1$, where $1 < F^1 \leq 1.5$.

The sequence of points at which the vector function F was computed is shown in Figure 1. The suggested method allows us to more than halve the number of points at which vector function F must be calculated in order to guarantee the accuracy demanded in the problem, compared with the uniform covering technique. It can be seen from the figure that the covering steps are largest far from the Pareto set; when the Pareto set is being covered the step size is at a minimum and coincides with that required for uniform covering.

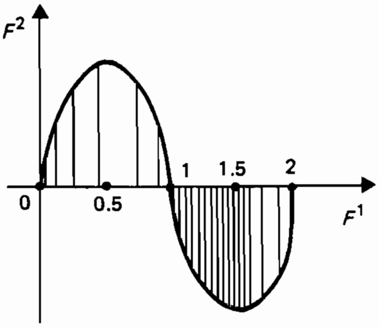


FIGURE 1 The sequence of points at which F was computed.

3. CONCLUSION

A numerical method for finding an E-approximation of a Pareto set is suggested. This method requires the feasible set to be covered with a nonuniform mesh only once. All other existing approaches involve global searches for multiple extrema. The approximate solution of the multicriteria problem is equivalent (in terms of labor) to the problem of finding the global minimum. There is, of course, some complication connected with the fact that here instead of calculating the value of $f(x)$ it is necessary to calculate m values of $F(x)$, and it is also necessary to remember the set of points A_k. However, the basic computations connected with the covering of X are roughly the same.

REFERENCES

Evtushenko, Y.G. (1971). *Zhurnal Vychislitelnoi Matematiki i Matimaticheskoi Fiziki*, 11(6):1390–1403.

Evtushenko, Y.G. (1974). Techniques for finding global extrema (in Russian). In *Operations Research*, Vol. 4, Computing Centre of the USSR Academy of Sciences, Moscow.

APPLICATION OF A SUBDIFFERENTIAL OF A CONVEX COMPOSITE FUNCTIONAL TO OPTIMAL CONTROL IN VARIATIONAL INEQUALITIES

B. Lemaire
University of Montpellier, Place E. Bataillon, 34060 Montpellier, France

INTRODUCTION

The chain rule for the subdifferential of a real convex functional composite with an affine operator and a real convex functional is well known (Ekeland-Temam, 1974). Various extensions of this classical case involving operators taking values in an ordered vector space have been considered by many people, for example Lescarret (1968), Levin (1970), Ioffe-Levin (1972), Valadier (1972), Zowe (1974), Penot (1976), Kutateladze (1977), Hiriart-Urruty (1980), Thera (1981) in a convex framework and Thibault (1980) in a non-convex situation.

§ 1 and § 2 are devoted to the chain rule for a real convex functional composite with a convex operator and a real non-decreasing convex functional. In § 3 , 4 , 5 we consider an optimal convex control problem with a non-differentiable cost function, in which the state of the system is defined as the (unique) solution of an elliptic variational inequality. The mapping between the control and the state is also non-differentiable but it is a convex operator. Applying the results of § 2 we can derive, by means of an adjoint state, necessary and sufficient optimality conditions improving the ones obtained by Mignot (1976). In § 6 these conditions are made explicit with an example.

1. DEFINITIONS AND NOTATION

All the vector spaces introduced in the sequel are real. X and Y denote topological vector spaces with respective topological duals X' and Y' . Y_+ is a convex cone in Y

which makes Y a partially ordered topological vector space (Perressini, 1967). We know that the ordering is defined as follows :

$$y^2 \leqq y^1 \Leftrightarrow y^1 \geqq y^2 \Leftrightarrow y^1 - y^2 \in Y_+$$

Y'_+ denotes the dual positive cone i.e. the cone of positive linear functionals on Y . $Y^{\bullet}$ stands for the set $Y \cup \{+\infty\}$ where $+\infty$ is a greatest element adjoined to Y . We extend in a natural way the addition and the scalar multiplication of Y to $Y^{\bullet}$. An operator f of X into $Y^{\bullet}$ is said to be convex if $f(\lambda x^1 + (1-\lambda)x^2) \leqq \lambda f(x^1) + (1-\lambda)f(x^2)$ for each x^1, x^2 in X and each real $\lambda \in [0,1]$. Its effective domain is the set $\text{dom } f = \{x \in X \mid f(x) \in Y\}$. As usual $L(X,Y)$ will denote the set of continuous linear operators of X into Y . By the subdifferential $\partial f(x)$ of f at $x \in \text{dom } f$, we mean the set of subgradients of f at x , i.e. the set

$$\partial f(x) = \{T \in L(X,Y) \mid f(x+h) \geqq f(x) + Th , \quad \forall h \in X\}$$

Given a functional φ of Y into $\mathbb{R}^{\bullet} = \mathbb{R} \cup \{+\infty\}$, φ is extended to $Y^{\bullet}$ by setting $\varphi(+\infty) = +\infty$. The effective domain of the composite real functional $\varphi \circ f$ of X into $\mathbb{R}^{\bullet}$ is then

$$\text{dom } \varphi \circ f = \text{dom } f \cap f^{-1}(\text{dom } \varphi)$$

2. THE CHAIN RULE

For an operator f of X into $Y^{\bullet}$ and a real functional φ of Y into $\mathbb{R}^{\bullet}$, we are going to give sufficient conditions for calculating the subdifferential of the composite $\varphi \circ f$ by the chain rule

$$\partial(\varphi \circ f)(x) = \partial\varphi(f(x)) \circ \partial f(x)$$

$$= \{y' \circ T \mid y' \in \partial\varphi(f(x)) , T \in \partial f(x)\}$$

The following results hold

Lemma 1. *If* φ *is non-decreasing then,* $\partial\varphi(y) \subset Y'_+$, $\forall\ y \in Y$.

Proof. Assume $\partial\varphi(y) \neq \phi$. Then $y \in \text{dom}\ \varphi$ and $\forall\ y' \in \partial\varphi(y)$ and $\forall\ z \in Y_+$, we have

$$\varphi(y) \geqslant \varphi(y-z) \geqslant \varphi(y) - <y',z>$$

i.e. $<y',z> \geqslant 0$.

Lemma 2.

$$\forall y' \in Y'_+ ,\ \forall x \in \textit{dom}\, f ,\ \partial(y' \circ f)(x) \supset y' \circ \partial f(x) .$$

Proof. Let $T \in \partial f(x)$ and $h \in X$. If $x+h \in \text{dom}\ f$,

$$f(x+h) - f(x) \geqq Th , \text{ and}$$

$$<y',f(x+h) - f(x)> \geqslant <y',Th>$$

If $x+h \notin \text{dom}\ f$, $<y',f(x+h)> = <y',+\infty> = +\infty$ (see § 1) , so $y' \circ T \in \partial(y' \circ f)(x)$.

Proposition 1. If φ is non-decreasing, then $\forall\ x \in \text{dom}\ \varphi \circ f$, $\partial(\varphi \circ f)(x) \supset \partial\varphi(f(x)) \circ \partial f(x)$.

Proof. By lemma 1 and lemma 2 ,

$$\bigcup_{y' \in \partial\varphi(f(x))} \partial(y' \circ f)(x) \supset \partial\varphi(f(x)) \circ \partial f(x)$$

Now, let $y' \in \partial\varphi(f(x))$, $x' \in \partial(y' \circ f)(x)$ and $h \in X$. If $x+h \in \text{dom}\ f$,

$$\varphi(f(x+h)) \geqslant \varphi(f(x)) + <y',f(x+h) - f(x)>$$

$$\geqslant \varphi(f(x)) + <x',h>$$

If $x+h \notin \text{dom}\ f$, $\varphi(f(x+h)) = \varphi(+\infty) = +\infty$ and the above inequality still holds, i.e. $x' \in \partial(\varphi \circ f)(x)$.

For the converse inclusion, we have the following intermediate result (see also Kutateladze, 1977).

Proposition 2. If φ is non-decreasing and convex, if f is a convex operator, if there exists $y \in R(f) \cap \operatorname{dom} \varphi$ where φ is continuous, then

$$\forall\ x \in \operatorname{dom} \varphi \circ f\ ,\ \partial(\varphi \circ f)(x) \subset \bigcup_{y' \in \partial\varphi(f(x))} \partial(y' \circ f)(x)$$

Proof. Let $x' \in \partial(\varphi \circ f)(x)$. The set

$$S = \{(f(x+h)+z, \varphi(f(x))+<x',h>) |\ x+h \in \operatorname{dom} f,\ z \in Y_+\}$$

is a convex subset of $Y \times \mathbb{R}$. As φ is non-decreasing, S and epi φ (the epigraph of φ) have only boundary points in common. Moreover epi φ has a non-empty interior. So, by the Hahn-Banach theorem, there exists $\tilde{y}' \in Y'$ and $\alpha \in \mathbb{R}$, such that $(\tilde{y}', \alpha) \neq 0$ and

$$\forall\ y \in \operatorname{dom} \varphi\ ,\ \forall\ \lambda \in \mathbb{R}\ ,\ \lambda \geqslant \varphi(y)\ ,\ \forall\ h \in \operatorname{dom} f{-}x\ ,$$

$$<\tilde{y}', y> +\ \alpha\lambda \geqslant <\tilde{y}', f(x+h)> +\ \alpha[\ \varphi(f(x) + <x',h>]$$

Taking $y = f(x)$ and $h = 0$, we get $\alpha \geqslant 0$. In fact $\alpha > 0$, otherwise

$$\forall\ y \in \operatorname{dom} \varphi\ \ ,\ \ <\tilde{y}', y> \ \geqslant\ <\tilde{y}', \bar{y}>\ ,$$

and $\tilde{y}' = 0$, because $\operatorname{dom} \varphi - \bar{y}$ is absorbing. Setting $y' = -\tilde{y}'/\alpha$, we get

(i) $\forall\ y \in \operatorname{dom} \varphi$, taking $\lambda = \varphi(y)$ and $h = 0$, $y' \in \partial\varphi(f(x))$

(ii) taking $y = f(x)$ and $\lambda = \varphi(f(x))$, $x' \in \partial(y' \circ f)(x)$.

Remark 1. In fact, by the proof of proposition 1 , the above proven inclusion is an equality.

Now, the question is : when the converse inclusion of lemma 2 does hold, that is to say (Valadier, 1972) when is f regularly subdifferentiable at x ? The answer is positive in the following cases.

Case 1 . f *is continuous affine with linear part* A .

Then f is convex and $\partial f(x) = \{A\}$. Moreover $\forall\ y' \in Y'_+$, $y' \circ f$ is continuous affine with linear part $y' \circ A$ and

$\partial(y' \circ f)(x) = \{y' \circ A\} = y' \circ \partial f(x)$. In fact, f is convex for any ordering on Y, and every φ is non decreasing for the particular ordering defined by $Y_+ = \{0\}$. Then we recover the case mentioned at the beginning of the introduction.

Case 2 . f *is Gateaux-differentiable at* x *with* G-*derivative* $f'(x)$, that is to say dom f - x is absorbing and $f'(x) \in L(X,Y)$ such that

$$\forall\ h \in X\ ,\ f'(x)h = \lim_{\lambda \to 0_+} \frac{f(x+\lambda h) - f(x)}{\lambda}$$

Then if $y' \in Y'_+$, $y' \circ f$ is G-differentiable at x with G-derivative $(y' \circ f)'(x) = y' \circ f'(x)$. Moreover, if Y_+ is *closed*, $\partial f(x) = \{f'(x)\}$. Then,

$$\partial(y' \circ f)(x) = \{(y' \circ f)'(x)\} = \{y' \circ f'(x)\} = y' \circ \partial f(x) .$$

Case 3 . f *is continuous at* x , Y *is a sequentially weakly complete Hausdorf locally convex space, which is an order complete vector lattice, normal, with order intervals relatively weakly compact, and* Y_+ *is closed.* Then (Valadier, 1972) $\partial f(x)$ is a non-empty compact and convex subset of $L_s(X,Y_\sigma)$ the space of linear operators of X into Y continuous for the weak topology $\sigma(Y,Y')$, equipped with the topology of simple convergence on X , and f is regularly subdifferentiable at x .

Examples of such a space Y are :

(i) the euclidean space $\mathbb{R}^m$ ordered by the order product of $\mathbb{R}$ or more generally by a cone generated by a set of m linearly independant vectors.

(ii) the space $L^p(\Omega,\Sigma,\mu)$, $1 \leqslant p < +\infty$ over a measured space Ω , ordered by the cone of μ-almost everywhere non negative functions.

Case 4 . X *is a reflexive Banach space,* f *is continuous at* x , Y *is a semi-reflexive Hausdorf locally convex space,* Y_+ *is closed and has a weakly compact base lying in a closed hyperplane not containing the origin.*

Then (Zowe, 1974), the same conclusion as in case 3 holds.

One example of such a Y is the space of Radon measures over a compact space, ordered by the cone of non-negative measures.

As a direct application of the chain rule we can recover the well-known formula of the subdifferential of the maximum of a finite family of convex functions. Namely, let f_i, $i = 1,\dots,m$, be m proper convex functions of the topological vector space X into $\mathbb{R}^{\bullet}$. Define the operator f of X into $Y^{\bullet} = \mathbb{R}^m \cup \{+\infty\}$, by

$$f(x) = \begin{cases} (f_1(x),\dots,f_m(x))^t & \text{if } x \in \bigcap_{i=1}^{m} \operatorname{dom} f_i \\ +\infty & \text{otherwise} \end{cases}$$

Then, for the order product defined by $Y_+ = \mathbb{R}^m_+$, f is convex. Now let φ of $\mathbb{R}^m$ into $\mathbb{R}$ defined by

$$\varphi(y) = \max_i y_i$$

Then φ is a continuous non-decreasing convex function. We have

$$\max_i f_i(x) = (\varphi \circ f)(x)$$

and

$$\partial f(x) = \prod_{i=1}^{m} \partial f_i(x)$$

Then the well-known result :

If, for each i, f_i *is continuous or* G-*differentiable at* $x \in \bigcap_i$ *dom* f_i, *then*

$$\partial(\max_i f_i)(x) = \operatorname{co}\{\partial f_i(x) \mid f_i(x) = \max_i f_i(x)\},$$

is an easy consequence of the above chain rule and the

Lemma 3 . $\forall\ y \in \mathbb{R}^m$, $\partial\varphi(y) = \operatorname{co}\{e^i \mid \varphi(y) = y_i\}$ where e^i denotes the i-th element of the canonical base of $\mathbb{R}^m$.

<u>Proof</u>. It is a particular case of lemma 4, § 6, hereafter.

3. VARIATIONAL INEQUALITIES AND ORDERING

Let V be a Hilbert space equipped with a continuous and coercive bilinear form a, and K a closed convex subset of V. Then (Lions-Stampacchia, 1967), for each $\ell \in V'$ topological dual of V, there exists a unique $y(\ell) \in K$ solution of the variational inequality :

$$a(y,\theta-y) \geqslant <\ell,\theta-y> \ , \ \forall\,\theta \in K \ , \quad y \in K \qquad (1)$$

or, with the notations of convex analysis,

$$\ell \in Ay + \partial\psi_K(y) \qquad (2)$$

where $A \in L(V,V')$ is the linear operator associated to the bilinear form a, and ψ_K denotes the indicatrice function of K. Moreover, the mapping $\ell \mapsto y(\ell)$ of V' (equipped with the dual norm) into V is Lipschitz continuous.

Now, introducing an ordering on V, we get the following abstract formulation of a well-known result of the classical theory of potential (Moreau, 1968).

Proposition 3. If V *is a vector lattice, the bilinear form* a *verifying* $a(y^+,y^-) \leqslant 0 \ \forall\, y \in V$ *; if* K *is hereditary :* $y + V_+ \subset K$, $\forall\, y \in K$, *and inf-stable : inf* $(y,z) \in K$, $\forall\, y,z \in K$, *then* $y(\ell)$ *is the least element of the set*

$$\mathcal{K}(\ell) = \{y \in K | a(y,\theta) \geqslant <\ell,\theta> \ , \forall\,\theta \in V_+\}$$

Proof. First, $y(\ell) \in \mathcal{K}(\ell)$. It is a trivial consequence of (1) and that K is hereditary. Then, let $y \in \mathcal{K}(\ell)$, and $z = y(\ell) - y$. We must prove $z^+ = 0$. Because K is inf-stable, $\inf(y(\ell),y) \in K$. But $\inf(y(\ell),y) = y(\ell) - z^+$. Putting in (1) as a θ, we get

$$- a(y(\ell),z^+) \geqslant - <\ell,z^+>$$

Moreover $a(y,z^+) \geqslant <\ell,z^+>$

Then $a(z,z^+) \leqslant 0$, and because $z = z^+ - z^-$,

$$a(z^+,z^+) \leqslant a(z^-,z^+) \leqslant 0 \ .$$

Finally the coercivity of a implies $z^+ = 0$.

Remark 2. This minimal property has been used by J.F. Durand (1972) in a finite dimensional context to prove the convergence of the Gauss-Seidel process for the inequality (2) where A is an M-matrix, with an argument of monotonicity.

Corollary 1 . Under the assumptions of proposition 3 , the mapping $\ell \mapsto y(\ell)$ *is a non-decreasing convex operator of* V' *into* V , V' *being ordered by the dual positive cone* V'_+ .

Proof. Let $\ell^1, \ell^2 \in V'$, $\lambda \in [0,1]$, $y^1 = y(\ell^1)$, $y^2 = y(\ell^2)$. We have $\lambda y^1 + (1-\lambda)y^2 \in K(\lambda\ell^1 + (1-\lambda)\ell^2)$. Therefore

$$y(\lambda\ell^1 + (1-\lambda)\ell^2) \leqq \lambda y^1 + (1-\lambda)y^2$$

If $\ell^1 \geqq \ell^2$, then $K(\ell^1) \subset K(\ell^2)$. So $y(\ell^1) \geqq y(\ell^2)$.

4. OPTIMAL CONTROL PROBLEM

Let us introduce the Hilbert space of *controls* U and the set of *admissible controls* U_{ad} which is a non-empty closed convex subset of U . We denote by b a continuous convex operator of U into V' ordered by the dual cone V'_+ . For $v \in U$, the *state* is defined as the solution $y(b(v))$ of the variational inequality (1) for $\ell = b(v)$. By corollary 1, the mapping between the *control* and the *state* is a continuous convex operator of U into V .

Then, let us consider the ordered Hausdorf locally convex space of *observations* Z . We assume that the mapping between the state and the *observation* $z(v)$ is a continuous non-decreasing convex operator c of V into Z :

$$z(v) = c(y(b(v)))$$

The cost function is defined by

$$J(v) = J_1(v) + \frac{1}{2} < Nv, v >$$

where $N \in L(U, U')$ is symmetric and coercive, and $J_1(v) = \Phi(z(v))$, with a lower-semi-continuous non-decreasing convex function Φ of Z into $\mathbb{R}$. Finally we consider the problem : find $u \in \mathcal{U}_{ad}$ (optimal control) such that

$$J(u) = \inf_{v \in \mathcal{U}_{ad}} J(v)$$

J is a lower-semi-continuous, strictly convex and coercive function of $\mathcal{U}$ into $\mathbb{R}$. So, by a classical argument (Ekeland-Temam, 1974) the optimal control u exists and is unique.

5. OPTIMALITY CONDITIONS

In fact J_1 is continuous because it is defined on the Banach space $\mathcal{U}$, and everywhere finite. So the optimal control u is characterized by : $\exists u' \in \partial J_1(u)$

$$< u' + Nu , v - u>_{\mathcal{U}'\mathcal{U}} \geqslant 0 , \quad \forall v \in \mathcal{U}_{ad} \qquad (3)$$

The problem is now to express u' by means of an adjoint state p. We have

$$J_1(v) = \Phi \circ c \circ y \circ b$$

We can apply the proposition 2 three times one after another. Then u is characterized by the existence of $z' \in \partial\Phi(z(u))$, $v' \in \partial(z' \circ c)(y(b(u)))$, $p \in \partial(v' \circ y)(b(u))$ and $u' \in \partial(p \circ b)(u)$ such that (3) holds. We can get more precise information if one of the four cases of § 2 holds for the operators b and (or) c. For instance, if c is affine with linear part C and, as a space Y, V' satisfies the conditions of case 3 or case 4, the characterization of the optimal control can be rewritten as :

$$\exists z' \in \partial\Phi(z(u)),\ \exists p \in \partial(z' \circ C \circ y)(b(u)),\ \exists B \in \partial b(u) \text{ , s.t.}$$
$$< B^{\star} p + Nu , v - u >_{\mathcal{U}'\mathcal{U}} \geqslant 0 , \quad \forall v \in \mathcal{U}_{ad} \qquad (4)$$

Let us assume now that (V,a) is a Dirichlet space on a locally compact space Ξ supplied with a Radon measure μ, ordered by the cone of μ-a.e. non-negative functions, and

$$K = \{v \in V | v \geqslant \xi \text{ quasi-everywhere on } \Xi\}$$

where $\xi : \Xi \to \overline{\mathbb{R}}$ is a quasi-upper-semi-continuous given function. Proposition 3 holds and (Mignot, 1976) the operator y has, at each $\ell \in V'$, a directional derivative $y'(\ell;h)$ in each direction $h \in V'$, which is the unique solution of the variational inequality :

$$a(y', \theta - y') \geqslant \langle h, \theta - y'\rangle , \forall \theta \in S_\ell , y' \in S_\ell \tag{6}$$

where S_ℓ is the closed convex cone of V defined by :

$$S_\ell = \{\theta \in V | \theta \geqslant 0 \text{ where } y(\ell) = \xi , \text{ and } a(y(\ell),\theta) = \langle \ell,\theta\rangle\}$$

Then, for $v' \in V'_+$, the real convex functional $v' \circ y$ has a directional derivative at ℓ given by

$$(v' \circ y)'(\ell;h) = \langle v', y'(\ell;h)\rangle_{V'V} , \forall h \in V'$$

and the subdifferential of $v' \circ y$ at ℓ is the set of $p \in V$ such that

$$\langle v', y'(\ell;h)\rangle \geqslant \langle h, p\rangle_{V'V} , \forall h \in V' \tag{7}$$

Now, using the techniques of Mignot (1976) we can derive the

Proposition 4 . $p \in V$ *satisfies* (7) *if and only if :*

$$a(\theta,p) \leqslant \langle v',\theta\rangle , \forall \theta \in S_\ell , p \in S_\ell$$

Proof. Let S be a closed convex cone of the Hilbert space V and $\theta \in V$. The a-projection $P_S(\theta)$ of θ onto S is defined as the unique solution of the variational inequality

$$a(\theta - q, w - q) \leqslant 0 , \forall w \in S , q \in S$$

$P^\star_S(\theta)$ denotes the $a^\star$-projection of θ onto S, where $a^\star$ is the adjoint bilinear form of $a^\star$. The a-polar cone of S is defined by

$$S^\circ_a = \{q \in V | a(q,\theta) \leqslant 0 , \forall \theta \in S\}$$

As a consequence of the bipolar theorem, we have

$$(S_a^\circ)_{a^\star}^\circ = S \quad \text{and} \quad P_S + P^\star_{S_a^\circ} = I \tag{8}$$

Because A, the linear operator associated to the bilinear form a, is an isomorphism of V onto V', (7) is equivalent to

$$< v', y'(\ell; A\theta) > \geqslant a(\theta, p) \ , \quad \forall\ \theta \in V \tag{9}$$

Taking $S = S_\ell$, we get $y'(\ell;A\theta) = P_S\theta$. So, by (8), (9) is equivalent to

$$< v', P_S\theta > \geqslant a(P_S\theta, p) + a(P^\star_{S_a^\circ}\theta, p) \ , \forall\ \theta \in V$$

or

$$\begin{cases} a(\theta,p) \leqslant < v', \theta > & , \ \forall\ \theta \in S \\ a(\theta,p) \leqslant 0 & , \ \forall\ \theta \in S_a^\circ \Leftrightarrow p \in (S_a^\circ)_{a^\star}^\circ = S \ . \end{cases}$$

6. EXAMPLE

Let $\Omega =]a,b[$ be an open bounded real interval. We choose as V, the sobolev space $H_o^1(\Omega)$. We know that V is included, with continuous injection, in $C(\overline{\Omega})$ the Banach space of continuous functions on $\overline{\Omega}$. We take

$$a(u,v) = \int_\Omega a_1 u'v'dx + \int_\Omega a_o\, uv\, dx \quad , \ \forall\ u,v \in V$$

where $a_o, a_1 \in L^\infty(\Omega)$, $a_o(x) \geqslant 0$, $a_1(x) \geqslant \alpha > 0$, a.e. in Ω. Then (V,a) is a Dirichlet space on Ω supplied with the Lebesgue measure. Let $\xi \in V$. We take

$$K = \{y \in V \mid y \geqslant \xi \ \text{ on } \ \Omega\} \ .$$

Introducing the differential operator A:

$$Ay = -\frac{d}{dx}(a_1 y') + a_o y \quad ,$$

we can interpret the variational inequality (1), for $\ell \in L^2(\Omega)$, as follows : $Ay - \ell$ is a positive measure on

Ω , concentrated on the closed subset of Ω :

$$\Omega^\circ = \{x \in \Omega \mid y(x) = \xi(x)\} .$$

We take now $\mathcal{U} = \mathcal{U}' = L^2(\Omega)$, and, for $u \in \mathcal{U}$, $b(u) = u^+$. We can easily prove that b is a continuous convex operator of $L^2(\Omega)$ into $L^2(\Omega)$ (then into V') . For each non-negative $p \in \mathcal{U}$, we have

$$(p \circ b)(u) = \int_\Omega p u^+ dx$$

and, as a consequence of the Lebesgue theorem of monotone convergence, the directional derivative of $p \circ b$ at u is given by

$$\forall\ v \in \mathcal{U} , \quad (p \circ b)'(u;v) = \int_{u=0} p v^+ dx + \int_{u>0} p v\, dx$$

Moreover, the set of βp where β is a measurable function on Ω verifying

$$\left.\begin{array}{lll} \beta(x) = 0 & \text{if} & u(x) < 0 \\ 0 \leqslant \beta(x) \leqslant 1 & \text{if} & u(x) = 0 \\ \beta(x) = 1 & \text{if} & u(x) > 0 \end{array}\right\} \qquad (10)$$

is a closed convex subset of $\mathcal{U}$ included in $\partial(p \circ b)(u)$, and, for each $v \in \mathcal{U}$, the measurable function β_v defined by

$$\beta_v(x) = 0 \quad \text{if} \quad u(x) < 0 \quad \text{or} \quad (u(x) = 0 \quad \text{and} \quad v(x) \leqslant 0)$$

$$\beta_v(x) = 1 \quad \text{if} \quad u(x) > 0 \quad \text{or} \quad (u(x) = 0 \quad \text{and} \quad v(x) > 0)$$

is such that

$$\int_\Omega \beta_v\, p\, v\, dx = (p \circ b)'(u;v)$$

Therefore,

$$\partial(p \circ b)(u) = \{\beta p \mid \beta \text{ measurable and } (10)\} .$$

Then, we take $Z = \mathcal{C}(\bar{\Omega})$ ordered by the cone of non negative functions on $\bar{\Omega}$. Let z_d given in Z . We take $z(v) = y(v^+) - z_d$. The operator of observation c is then continuous affine with linear part equal to the injection of

$H^1_o(\Omega)$ into $\mathcal{C}(\bar\Omega)$. Then, take

$$J_1(v) = |z(v)|_{\mathcal{C}(\bar\Omega)}$$

If $z_d \leqq y(v^+)$, $\forall\ v \in \mathcal{U}_{ad}$ (for instance $z_d \leqq \xi$) , then $J_1(v) = \Phi(z(v))$, with $\Phi(z) = \max\limits_{x \in \bar\Omega} z(x)$, which is a continuous non-decreasing convex function on Z .

Lemma 4 . For each $z \in Z$, $\partial\varphi(z)$ *is the subset of Radon probabilities on* $\bar\Omega$, *concentrated on*

$$\Omega(z) = \{x \in \bar\Omega \mid \varphi(z) = z(x)\} .$$

Proof. Because φ is non-decreasing we already know (see lemma 1) that, if $z' \in \partial\varphi(z)$, z' is a positive Radon measure on $\bar\Omega$. Then we have

$$\varphi(\zeta) \geqslant \varphi(z) + \langle z', \zeta - z\rangle , \quad \forall\ \zeta \in Z \qquad (11)$$

Taking $\zeta = z \pm \mathbb{1}$, we get $\langle z', \mathbb{1}\rangle = 1$. So $z' \in \mathcal{M}^1_+(\bar\Omega)$ i.e. is a Radon probability on $\bar\Omega$. Then (11) is equivalent to

$$\langle z', \varphi(z)\mathbb{1} - z\rangle = 0 \quad \text{or ,} \quad \text{as} \quad \varphi(z)\mathbb{1} - z \geqslant 0 ,$$

z' is concentrated on $\Omega(z)$.

Finally, the cost function being

$$J(v) = J_1(v) + \frac{N}{2}\,|v|^2_{\mathcal{U}} , \quad N > 0 ,$$

we can make explicit the general previous results in this particular situation.

There exists a unique optimal control $u \in \mathcal{U}_{ad}$ characterized by

$\exists$ z' Radon probability on $\bar\Omega$, concentrated on the subset $\Omega(z(u))$,
$\exists$ $p \in H^1_o(\Omega)$ s.t.

$$a(\theta,p) \leqslant \int_\Omega \theta\, dz' , \ \forall\ \theta \in S_u+ , \ p \in S_u+ , \ \text{where}$$

$S_u+ = \{\theta \in H^1_o(\Omega) \mid \theta \geqslant 0$ where $y(u^+) = \xi$ and $a(y(u^+),\theta) = \int_\Omega u^+\theta dx\}$,
$\exists\beta$ measurable function on Ω verifying (10),
such that

$$\int_\Omega (\beta p + Nu)(v-u)dx \geqslant 0 , \quad \forall\ v \in \mathcal{U}_{ad} .$$

Remark 3.

1. Taking $\mathcal{U}_{ad} = \mathcal{U}$, we get $u = -\frac{\beta p}{N}$. Because p is non-negative, $u^+ = 0$ and the optimal *state* is the least function of $H_0^1(\Omega)$ majorizing ξ and such that Ay is a positive measure on Ω.

2. We can interpret, unless formally, the adjoint inequality defining the adjoint state p as follows. Consider the partition of Ω between the three subsets

$$\Omega_1^\circ = \{x \in \Omega \mid y(u^+) = \xi \quad , \quad Ay(u^+) - u^+ > 0\}$$

$$\Omega_2^\circ = \{x \in \Omega \mid y(u^+) = \xi \quad , \quad Ay(u^+) - u^+ = 0\}$$

$$\Omega^+ = \{x \in \Omega \mid y(u^+) > \xi\}$$

Then

$$S_{u^+} = \{\theta \in H_0^1(\Omega) \mid \theta = 0 \text{ on } \Omega_1^\circ \text{ and } \geqslant 0 \text{ on } \Omega_2^\circ\}$$

and p is characterized by

$$\begin{cases} p = 0 & \text{on } \Omega_1^\circ \\ p \geqslant 0 \quad , \quad Ap \leqslant z' & \text{on } \Omega_2^\circ \\ Ap = z' & \text{on } \Omega^+ \end{cases}$$

REFERENCES

Durand J.F. (1972). L'algorithme de Gauss-Seidel appliqué à un problème unilatéral non symétrique. R.A.I.R.O., R-2, 23-30.

Ekeland I. and Temam R. (1974). Analyse convexe et problèmes variationnels. Dunod-Gauthier-Villars.

Hiriart-Urruty (1980). ε-subdifferential calculus, in convex analysis and optimization. Proc. coll. Imperial College London, 1-44.

Ioffe A.D. and Levin V.L. (1972). Subdifferential of convex functions. Trans. Moscow Math. Soc. 26, 1-72.

Kutateladze S.S. (1977). Formulas for computing subdifferentials. Soviet Math. Dokl. 18, n° 1, 146-148.

Lescarret C. (1968). Sous-différentiabilité de fonctions composées. Séminaire d'Analyse unilatérale, Montpellier.

Levin V.L. (1970). On the subdifferential of a composite functional. Soviet. Math. Dokl., vol. 11, n° 5.

Lions J.L. and Stampacchia G. (1967). Variational inequalities. Comm. pure Appl. Math. 20, 493-519.

Mignot F. (1976). Contrôle dans les inéquations variationnelles elliptiques. Journal of functional analysis 22, 130-185.

Moreau J.J. (1968). Majorantes sur-harmoniques minimales. Travaux du Séminaire d'Analyse unilatérale, Vol. 1, exposé n° 5.

Penot J.P. (1978). Calcul sous-différentiel et optimisation. Journal of functional analysis 27, 248-276.

Peressini A.L. (1967). Ordered topological vector spaces. Harper's series in modern mathematics.

Thibault L. (1980). Subdifferentials of compactly lipschitzian vector valued functions. Ann. Math. Pur. Appl. 125, 157-192.

Thera M. (1981). Subdifferential calculus for convex operators. Journal of Math. analysis and appl. 80, 78-91.

Valadier M. (1972). Sous-différentiabilité des fonctions convexes à valeurs dans un espace vectoriel ordonné. Math. scand. 30, 65-74.

Zowe J. (1974). Subdifferentiability of convex functions with values in an ordered vector space. Math. Scand. 34, 69-83.

ON SOME NONDIFFERENTIABLE PROBLEMS IN OPTIMAL CONTROL

J.V. Outrata and Z. Schindler
Institute of Information Theory and Automation, Czechoslovakian Academy of Sciences,
Pod vodårensku vě̌ží 4, 18208 Prague 8, Czechoslovakia

INTRODUCTION

Modern developments in nondifferentiable analysis have now made it possible to handle nondifferentiable optimal control problems. Maximum principles of considerable generality have been derived by Clarke (1976), and a number of effective numerical methods for minimizing nonsmooth objectives are available. Nevertheless, nondifferentiable optimal control problems are still difficult to solve. The reason lies in their structure, which in the most general case may involve compositions of nondifferentiable functionals and operators.

In this paper we study special types of such problems which can be solved with the help of a suitable bundle method. We have used two numerical codes by Lemaréchal: CONWOL for unconstrained minimization of convex objectives and BOREPS for minimization of weakly semismooth objectives with constraints in the form of upper and lower bounds, cf. Lemaréchal et al. (1980). We will use the following general model:

$$J\ (x,u) \rightarrow \inf$$

$$\text{subj. to} \qquad \qquad (\mathcal{P})$$

$$A\ (x,u) = \theta,$$

$$u \in \omega \subset U,$$

where $x \in X$ and $u \in U$ are the state and control variables, respectively. The spaces X and U are assumed to be Banach, $J[X \times U \rightarrow R]$, ω is a closed subset of U, $A[X \times U \rightarrow X]$. Moreover, we assume that the equality $A(x,u) = \theta$ defines a unique implicit function $x(u)$ which is locally Lipschitz. Finally, we denote $\phi(u) = J(x(u),u)$ and suppose that $\phi[U \to R]$ is locally Lipschitz over ω.

Section 1 explains how to solve some special types of (P) with the help of bundle methods, or, more precisely, how to compute elements of $\partial\phi$ (the Clarke's generalized gradient of ϕ) for any admissible control $u \in \omega$. Illustrations based on concrete practical problems are also provided. We have no state-space constraints in (P) since we assume that they have been included in the cost by a suitable penalty. In Section 2 a Sobolev type of Zangwill-Pietrzykowski penalty is studied and applied to a certain type of inequality state-space constraint.

We employ the standard notation in NDO; additionally, x^i is the i-th coordinate of a vector $x \in R^n$, B is the unit ball centered at the origin and $(x)^D$ denotes the projection of x onto D.

1. ESSENTIALLY NONSMOOTH PROBLEMS

We confine ourselves here to those problems in which the standard adjoint equation approach may be used to compute the desired elements of $\partial\phi$. Unfortunately, the structure of the problem only rarely enables us to obtain some inner approximation of $\partial\phi$ in this way. Regularity is crucial in considerations of this type.

Nondifferentiable objectives. In this part we will assume that A is continuously Fréchet differentiable over $X \times \tilde{\omega}$ with $A'_x(x,u)$ being continuous over $X \times \tilde{\omega}$ and utilize the chain rule II of Clarke (1983). $\tilde{\omega}$ is an open set containing ω.

Proposition 1.1. Let J *be locally Lipschitz in* u *for all* $x \in X$ *and Fréchet differentiable in* x *over* X *for all* $u \in \tilde{\omega}$ *with* $\nabla_x J(x,u)$ *being continuous over* $X \times \tilde{\omega}$. *Let* λ^* *be a solution of the equation*

$$A'_x(\bar{x},\bar{u})^* \lambda^* + \nabla_x J(\bar{x},\bar{u}) = \theta \tag{1.1}$$

at a fixed process $(\bar{x},\bar{u})$, $\bar{u} \in \omega$. *Then*

$$\partial\phi(\bar{u}) \supset \partial_u J(\bar{x},\bar{u}) + A'_u(\bar{x},\bar{u})^* \lambda^* \tag{1.2}$$

provided J *is regular at* $(\bar{x},\bar{u})$.

Proof. Due to Prop. 1 of Luenberger (1969) (Sect. 9.6) for $u \in \omega$, $h \in U$ and $\mu \in R_+$

$$\phi(u+\mu h) - \phi(u) = J(x_\mu,u+\mu h) - J(x,u) = J(x,u+\mu h) - J(x,u) + \\ + <\upsilon^*, A(x,u+\mu h) - A(x,u)> + o(\mu),$$

where trajectories x,x_μ correspond to controls $u,u+\mu h$, respectively, υ^* is a solution of (1.1) at the process (x,u) and $\lim_{\mu\to 0_+} o(\mu)/\mu = 0$. Hence, on denoting $\bar{x}_\mu$ the trajectory corresponding to $\bar{u} + \mu h$

$$\phi^o(\bar{u};h) = \overline{\lim_{\substack{\mu\to 0_+\\ u\to\bar{u}}}} (J(x_\mu,\ u+\mu h) - J(x(u),u))|\mu \geqslant$$

$$\geqslant \overline{\lim_{\mu\to 0_+}} (J(\bar{x}_\mu,\ \bar{u}+\mu h)-J(\bar{x},\bar{u}))|\mu = J'(\bar{x},\bar{u};\theta,h)+<\lambda^*,A_u'(\bar{x},\bar{u})h>=$$

$$= \overline{\lim_{\substack{\mu\to 0\\ u\to\bar{u}}}} (J(\bar{x},u+\mu h) - J(\bar{x},u))|\mu + <A_u'(\bar{x},\bar{u})^*\lambda^*,\ h>$$

by the regularity of J at $(\bar{x},\bar{u})$. □

As an example we may take the problem of operating an electric train between two stations with minimum energy losses:

$$\int_0^T x^2(t)(u(t))^+dt \to \inf \tag{1.3}$$

subj.to

$$\dot{x}(t) = f(x(t),u(t)) \text{ a.e. in } [0,T],$$
$$x(0) = a,\ x(T) = b,$$
$$u(t) \in \Omega(x^2(t)),$$

where $f[R^2 \times R \to R^2]$ is continuously differentiable, a, b are given vectors from R_+^2, Ω: $R \rightrightarrows R$ is a given nonempty compact measurable multifunction and $u \mapsto x$ is locally Lipschitz.
We set $X = C_o[0,T,R^2]$ and $A(x,u) = x(t)-a- \int_o^t f(x(\tau),u(\tau))d\tau$.
If $\bar{u}$ is admissible and the corresponding trajectory $\bar{x}$ satisfies $\bar{x}^2(t) \geq 0$ for $t \in [0,T]$, then $x^2(t)(u(t))^+$ is regular at $(\bar{x}(t), \bar{u}(t))$ for each t, and, consequently

$$\bar{x}^2(t)\xi(t) - \frac{\partial f(\bar{x}(t),\bar{u}(t))^T}{\partial u}\ p(t) \in \partial\phi(\bar{u}),$$

where $\xi(t)=1$ if $\bar{u}(t)>0$, $\xi(t)=0$ if $\bar{u}(t)<0$, $\xi(t) \in [0,1]$ for $\bar{u}(t)=0$ and p is the solution of the adjoint equation

$$\dot{p}(t) = - \frac{\partial f(\bar{x}(t),\bar{u}(t))^T}{\partial x} p(t) + \begin{bmatrix} 0 \\ (\bar{u}(t))^+ \end{bmatrix} \quad \text{a.e.}$$

backwards from a suitable terminal condition concerning the treatment of the terminal equality constraint $x(T) \rightleftarrows b$.

Proposition 1.2. Let $J=J_1(x)+J_2(u)$, *where* $J_1[X \to R]$, $J_2[U \to R]$ *are locally Lipschitz, and assume that* $\bar{\xi} \in \partial J_1(\bar{x})$, $\bar{\eta} \in \partial J_2(\bar{u})$ *at a fixed process* $(\bar{x},\bar{u})$. *Let the implicit function* $x(u)$ *be continuously Fréchet differentiable on a neighbourhood of* $\bar{u}$ *(which holds e.g. if* $A_x'(\bar{x},\bar{u})$ *is a linear homeomorphism of* X *onto* X*) and* λ^* *be a solution of the adjoint equation*

$$A'_x(\bar{x},\bar{u})^* \lambda^* + \bar{\xi} = \theta. \tag{1.4}$$

Then

$$A_u'(\bar{x},\bar{u})^* \lambda^* + \bar{\eta} \in \partial\Phi(\bar{u}) \tag{1.5}$$

provided any of the following conditions is satisfied:

(i) J_1, J_2 *are regular at* $\bar{x},\bar{u}$, *respectively;*

(ii) J_1 *is continuously Fréchet differentiable with* $\bar{\xi}$ *being its gradient at* $\bar{x}$.

(iii) J_2 *is continuously Fréchet differentiable with* $\bar{\eta}$ *being its gradient at* $\bar{u}$, *and either* $-J_1$ *is regular at* $\bar{x}$ *or* $x(u)$ *maps every neighbourhood of* $\bar{u}$ *to a set which is dense in a neighbourhood of* $\bar{x}$ *(e.g. if* $x'(\bar{u})$ *is onto).*

Proof. Under condition (ii) the statement is a direct consequence of the above mentioned result of Luenberger. Conditions (i) or (iii) imply due to the chain rule II that

$$(x'(\bar{u}))^* \bar{\xi} + \bar{\eta} \in \partial\phi(\bar{u})$$

taking into account the rule for generalized gradients of a finite sum of functions. To express the operator $(x'(\bar{u}))^*$ by means of the derivatives of A at $(\bar{x},\bar{u})$, observe that

$$A_x'(\bar{x},\bar{u})\, x'(\bar{u}) + A_u'(\bar{x},\bar{u}) = \theta.$$

Hence, for any $h \in U$

$$\langle\bar{\xi},x'(\bar{u})h\rangle = \langle\lambda^*,-A_x'(\bar{x},\bar{u})\, x'(\bar{u})h\rangle = \langle A_u'(\bar{x},\bar{u})^* \lambda^*,h\rangle$$

which completes the proof. □

Examples of this kind may be found e.g. in "production planning" problems cf. McMaster (1970). The following "minimum overshoot" problem also possesses an objective of the above form:

$$\max_{t \in [0,T]} (\langle c(t), x(t)\rangle - s)^+ \to \inf$$

subj.to

$$\dot{x}(t) = f(x(t),u(t)) \text{ a.e. in } [0,T],$$
$$x(0) = a,\ \langle c(T),\ x(T)\rangle = s,$$
$$u \in \omega \subset L_\infty[0,T,R^m],$$

where $f[R^n \times R^m \to R^n]$ is continuously differentiable, $s \in R$, $c \in C_0[0,T,R^n]$, $a \in R^n$, and $u \mapsto x$ is locally Lipschitz.

We set again $X = C_0[0,C,R^n]$ and introduce A as in the previous example. If $\bar{u}$ is admissible, $\bar{x}$ is the corresponding trajectory, and $\langle c(t),\bar{x}(t)\rangle > s$ for some $t \in [0,T]$, we denote

$$\theta = \{t \in [0,T] \mid \langle c(t),\bar{x}(t)\rangle = \max_{\tau \in [0,T]} \langle c(\tau),\bar{x}(\tau)\rangle\}.$$

According to Prop. 1.2 and Clarke (1983)

$$- \frac{\partial f(\bar{x}(t),\bar{u}(t))^T}{\partial u} p(t) \in \partial\phi(\bar{u}) \tag{1.6}$$

provided p is the solution of the adjoint equation

$$\dot{p}(t) = - \frac{\partial f(\bar{x}(t),\bar{u}(t))^T}{\partial x} p(t)$$

.backwards on the interval [0,T] from a terminal condition concerning the treatment of the terminal state condition and with the jump $c(t_1)$ at a time $t_1 \in \theta$. If $\langle c(t),\bar{x}(t)\rangle \leq s$ on [0,T], relation (1.6) is still true if p is the solution of the above adjoint equation without any jump.

Unfortunately, we are not able to provide any assertion of the type of Props. 1.1, 1.2 for a general objective J(x,u). However, its special structure may sometimes help us to obtain such statements - a problem of this sort has been investigated in Outrata (1983). In other cases the objective may be replaced by a regular one.

Nondifferentiable controlled systems. If U and X are Banach there is, to our knowledge, no available chain rule for computing generalized gradients of composite functionals J(x(u),u). Therefore, we have to confine ourselves to the finite-

-dimensional case and apply the Jacobian chain rule, cf. Clarke (1983). Nevertheless, the situation is still too complicated and we are forced to further restrictions. Namely, we will assume that $A = A_1(x)+A_2(u)$, where $A_1[X \to X]$ is continuously differentiable over $X=R^n$ and $A_2[U \to X]$ is locally Lipschitz over $\omega \subset U=R^m$. Furthermore, we require that $J = J_1(x)+J_2(u)$, where $J_1[X \to R]$, $J_2[U \to R]$ are continuously differentiable over X, $\tilde{\omega}$, respectively.

Proposition 1.3. Let $(\bar{x},\bar{u})$ *be a fixed process,* $A_1'(\bar{x})$ *be a linear homeomorphism of* X *onto* X *and* λ^* *be* ***the solution of the adjoint*** *equation*

$$(A_1'(\bar{x}))^T\lambda^* + \nabla J_1(\bar{x}) = \theta \quad . \tag{1.7}$$

Then

$$\partial\phi(\bar{u}) = \nabla J_2(\bar{u}) + (\partial A_2(\bar{u}))^T\lambda^*. \tag{1.8}$$

Proof. On denoting $v = A_2(u)$, $\bar{v} = A_2(\bar{u})$, Eq. $A_1(x)+v = \theta$ defines a unique implicit function $x=\mu(v)$ which is continuously differentiable on a neighbourhood of $\bar{v}$ with $\mu'(\bar{v}) = -(A_1'(\bar{x}))^{-1}$. According to the corollary of the Jacobian chain rule (Clarke, 1983)

$$\partial x(\bar{u}) = -(A_1'(\bar{x}))^{-1}\partial A_2(\bar{u}).$$

A direct application of the Jacobian chain rule gives now immediately

$$\partial\phi(\bar{u}) =-((A_1'(\bar{x}))^{-1}\partial A_2(\bar{u}))^T\nabla J_1(\bar{x}) + \nabla J_2(\bar{u}) =$$
$$= (\partial A_2(\bar{u}))^T\lambda^* + \nabla J_2(\bar{u}). \qquad \Box$$

An easy application of the above assertion is provided by the minimum-energy control of a linear plant with a dead band. After replacing the original control space $U = L_\infty[0,T]$ by R^m the problem may attain the following form

$$\frac{\Delta}{2}\sum_{i=0}^{m-1}(u^i)^2 + \frac{r}{2}\|y(T) - b\|^2_{R^n}$$

subj.to

$$\dot{y}(t) = A(t)y(t)+\psi_i(u^i) \text{ a.e. in } [i\Delta,(i+1)\Delta] \text{ ,} i=0,1,..m-1$$
$$y(0) = a,$$
$$u^i \in \omega^i \subset R, \; i=0,1,\dots,m-1,$$

where $m>1$ is a given integer, the stepsize $\Delta=T/m$, $r>0$ is a penalty parameter, a,b are given vectors from R^n, A is an $[n\times n]$ matrix of functions from $C_o[0,T]$ and $\psi_i(v) = (\psi_i^1(v),\ldots,\psi_i^n(v))^T$,

$$\psi_i^j(v) = \begin{cases} \beta_i^j(v-\varepsilon_i) & \text{for } v \geq \varepsilon_i \\ 0 & \text{for } |v| < \varepsilon_i \\ \beta_i^j(v+\varepsilon_i) & \text{for } v \leq \varepsilon_i,\ i=0,1,\ldots,m-1,\ j=1,2,\ldots,n. \end{cases}$$

Clearly,

$$\partial\psi_i(v) = \begin{cases} (\beta_i^1,\ \beta_i^2,\ldots,\ \beta_i^n)^T = \beta & \text{if } |v| > \varepsilon_i \\ 0 & \text{if } |v| < \varepsilon_i \\ \mathrm{co}(\theta,\ \beta) & \text{otherwise},\ i=0,1,\ldots,m-1. \end{cases}$$

To apply the preceding assertion, we set $X=R^n$ (the space of terminal states $y(T)$), $A_1=I$ (unit $[n\times n]$ matrix) and observe that

$$A_2(u^o,u^1,\ldots,u^{m-1}) = -r(T,0)a - \sum_{i=0}^{m-1} S_i\ \psi_i(u^i),$$

where $S_i = \int_{i\Delta}^{(i+1)\Delta} \Gamma(T,t)dt$ and Γ is the transition matrix, i.e. the solution of the matrix differential equation $\dot{\Gamma}(t,t_o) = A(t)\Gamma(t,t_o)$ on $[0,T]$ with the initial condition $\Gamma(t_o,t_o)=I$. We denote $\bar{u}=(\bar{u}^o,\bar{u}^1,\ldots,\bar{u}^{m-1})$, the elements of $\partial\phi(\bar{u})$ by $\upsilon=(\upsilon^o,\upsilon^1,\ldots,\upsilon^{m-1})$ and observe that the "modified" adjoint equation attains the form

$$\tilde{\lambda}_i^* = S_i^T r(\bar{y}(T)-b) = \int_{i\Delta}^{(i+1)\Delta} \Gamma^T(T,t)dt(r(\bar{y}(T)-b)),\ i=0,1,\ldots,m-1$$

$$\tilde{\lambda}^{*T} = (\tilde{\lambda}_o^{*T},\ \tilde{\lambda}_1^{*T},\ldots,\ \tilde{\lambda}_{m-1}^{*T}).\text{(Here } \tilde{\lambda}_i^* \sim S_i^T\lambda^* \text{ with } \lambda^* \text{ from (1.7))}$$

Using the properties of transition matrices we may rewrite it in the usual form

$$\tilde{\lambda}_i^* = \int_{i\Delta}^{(i+1)\Delta} p(t)dt, \qquad i=0,1,\ldots,m-1,$$

where p is the solution of the standard adjoint equation

$$\dot{p}(t) = -A^T(t)p(t)$$

backwards from the terminal condition $p(T) = r(\bar{y}(T)-b)$. Thus, by Prop. 1.3

$$\upsilon^i = \Delta\bar{u}^i + \langle\partial\psi_i(\bar{u}_i),\ \int_{i\Delta}^{(i+1)\Delta} p(t)dt\rangle,\ i=0,1,\ldots,m-1.$$

To be able to derive results of the type of Prop. 1.3 for more general cases, a deeper study of Lipschitz mappings is necessary. It is also possible that other generalized differentiability concepts with richer calculi will prove themselves to be more convenient with respect to different numerical methods, cf. e.g. Demyanov, Nikulina and Shablinskaya (1984).

2. NONSMOOTHNESS INTRODUCED BY THE TREATMENT

Various dual approaches have been developed for the numerical solution of optimal control problems. In this way we remove complicated state-space or mixed constraints by incorporating them in the objective - however, these new objectives may be nonsmooth. This is the case in Fenchel dualisation which proved itself to be very effective in the convex case (linear systems, convex objective and constraints). Such problems have been solved very rapidly with the help of CONWOL especially in those cases where the perturbation space was finite-dimensional (ordinary linear differential equations, terminal state constraints).

Here we turn our attention to Zangwill-Pietrzykowski exact penalties applied to inequality state-space constraints which are in the general case usually considered in the form

$$-q(x) \in D,$$

where $q[X \to Z]$, the "constraint" space Z is assumed to be Banach and D is a closed convex cone with the vertex at the origin. The exact penalty mentioned above takes the form

$$P_r(x) = r \operatorname{dist}(-q(x), D). \tag{2.1}$$

If Z is Hilbert, the penalty may be expressed in a more compact way by

$$P_r(x) = r \, \|(q(x))^{D^*}\|_Z, \tag{2.2}$$

where D^* is the positive dual cone to D. Sometimes there is a certain freedom in the choice of Z (and hence also D) so that we may use several different exact penalties of the type (2.1).

Let $X = H^1[0,T,R^n]$ and let the state-space constraint attain the form

$$q(x(t)) \leq 0 \quad \text{for} \quad t \in [0,T], \tag{2.3}$$

where $q[R^n \to R]$ is Lipschitz. Then we may choose the distance and the cone of nonnegative functions e.g. from spaces H^1, C_o, L_1. We already have sufficient numerical experience with the choice of C_o or L_1, cf. e.g. Outrata (1983). Therefore the rest of

Section 2 is devoted to the Sobolev case. The projection onto D^* in H^1 has been studied in Outrata and Schindler (1981) and the results enable us to compute it for piecewise affine functions of one variable very effectively. The objective ϕ of Section 1 is now given by

$$\phi(u) = J(x(u),u) + P_r(x(u)).$$

We will suppose that J is continuously Fréchet differentiable over $X \times \tilde{\omega}$, $(\bar{x},\bar{u})$ is a fixed process $(\bar{u} \in \omega)$, and the implicit function $x(u)$ is continuously Fréchet differentiable on a neighbourhood of $\bar{u}$.

Proposition 2.1. Let $\bar{x}$ *be nonfeasible with respect to the state--space constraint (2.3) and* λ^* *be a solution of the adjoint equation*

$$A_x'(\bar{x},\bar{u})^* \lambda^* + \nabla_x J(\bar{x},\bar{u}) + r^2(q'(\bar{x}))^*(q(\bar{x}))^{D^*}/P_r(\bar{x}) = \theta. \qquad (2.4)$$

Then ϕ *is Fréchet differentiable at* $\bar{u}$ *and*

$$\beta = \nabla_u J(\bar{x},\bar{u}) + A_u'(\bar{x},\bar{u})^* \lambda^* \qquad (2.5)$$

is its Fréchet derivative. If $\bar{x}$ *is feasible and* λ^* *is a solution of the adjoint equation (1.1), then* $\beta \in \partial\phi\ (\bar{u})$.

In the proof it suffices to combine a slightly modified assertion of Prop. 1.2 with the following lemma:

Lemma. Let Z *be Hilbert and* $z \in Z$. *Then the function* $g(z) = \|(z)^{D^*}\|$ *is Fréchet differentiable if* $-z \notin D$ *with*

$$\nabla g(z) = (z)^{D^*}/\|(z)^{D^*}\|. \qquad (2.6)$$

If $-z \in D$

$$\partial g(z) = B \cap D^* \cap \{z\}^{\perp}. \qquad (2.7)$$

Proof. Concerning Eq. (2.6), we refer to Zarantonello (1971). Eq. (2.7) can be proved by analysing the equivalence

$$\xi \in \partial g(\theta) \Longleftrightarrow \langle \xi, h\rangle \le \|(h)^{D^*}\| \text{ for all } h \in Z. \qquad \square$$

The investigated penalty characterizes the violation of the state space constraints in a very precise way. To realize it, note the right-hand side of the adjoint equation in the following example:

$$\sum_{i=0}^{k-1} \gamma_i(u^i) \to \inf$$

subj.to

$$x_{i+1} = f_i(x_i,u^i), \quad i=0,1...k-1, \quad x_o = a, \qquad (2.8)$$

$$x_i^1 \leq N, \ i = 1,2...k,$$

$$u^i \in \omega^i \subset R, \quad i=0,1,...,k-1,$$

where functions $\gamma_i[R^m \to R]$, $f_i[R^n * R^m \to R^n]$, $i=0,1,...,k-1$ are continuously differentiable and $a \in R^n$, $N \in R$ are given. The trajectories x are vector-valued piecewise affine functions given by sequences$(a,x_1,...,x_k)$. The penalty (2.2) attains for $Z = H^1[0,k]$ in this situation the form

$$P_r(x) = r \sqrt{s_0^2 + \sum_{i=0}^{k-1} (s_{i+1} - s_i)^2} \quad ,$$

where $s = (s_0,s_1...s_k) = (x^1-N)^{D*}$. Similarly, the adjoint variable $\lambda*$ may be expressed by a sequence $(p_o,p_1,...,p_k)$. On denoting $d_i = (s_i,0,...,0)^T$, Eq. (2.4) is equivalent to the difference scheme

$$p_{i-1} = \frac{\partial f_i(\bar{x}_i,\bar{u}^i)^T}{\partial x_i} p_i + r^2(-d_i+2d_{i-1}-d_{i-2})/P_r(\bar{x}), \quad i=1,2...k-1$$

which is to be solved backwards from the terminal condition

$$p_k = r^2(d_k - d_{k-1})/P_r(\bar{x}).$$

Thus, if $(a,\bar{x}_1,...,\bar{x}_k)$ is the trajectory corresponding to a control $(\bar{u}^o,\bar{u}^1,...\bar{u}^{k-1})$

$$\nabla_{u_i}\phi(\bar{u}^o,\bar{u}^1,...\bar{u}^{k-1}) = \nabla\gamma_i(\bar{u}^i) + \frac{\partial f_i(\bar{x}_i,\bar{u}^i)^T}{\partial u_i} p_{i+1}, \quad i=0,1,..,k-1$$

provided $x_i > N$ for some $i \in \{1,2,...,k\}$. Otherwise

$$(\nabla\gamma_o(\bar{u}^o),\nabla\gamma_1(\bar{u}^1),...,\nabla\gamma_{k-1}(\bar{u}^{k-1}))^T \in \partial\phi(\bar{u}^o,\bar{u}^1,...,\bar{u}^{k-1}).$$

3. NUMERICAL EXPERIENCE

We have performed a number of numerical experiments with the problem (1.3) solved by BOREPS. The mixed state-control constraints have been simplified to state constraints only by a simple transformation and they have been included in the cost by means of the exact C_o-penalty. For sloped railroads (where $x^2(t)$ could be negative), the objective has been regularized. The results are published in Outrata (1983).

The H^1-exact penalty has been tested on a rather complicated ecological problem of the type (2.8) with 3 state variables, 1 control variable and 360 steps of time-discretization again with the BOREPS routine. The results are encouraging.

REFERENCES

Clarke, F.H. (1976). The maximum principle under minimal hypotheses. SIAM J.Control Optim., 14:1078-1091.

Clarke, F.H. (1983). Optimization and Nonsmooth Analysis. Wiley, New York.

Dem´yanov, V.F., Nikulina, V.N., and Shablinskaya, I.R. (1984). Quasidifferentiable problems in optimal control. IIASA Working paper WP-84-2.

Lemaréchal, C., Strodiot, J.J., and Bihain, A. (1980). On a bundle algorithm for nonsmooth optimization. NPS4, Madison.

Luenberger, D.G. (1969). Optimization by Vector Space Methods. Wiley, New York.

McMaster, A.W. (1970). Optimal control in deterministic inventory model. WP-NPS-55MG0031A. U.S. Naval Postgraduate School, Monterey.

Outrata, J.V. (1983). On a class of nonsmooth optimal control problems. J.Appl.Math.Optim., 10:287-306.

Outrata, J.V., and Schindler, Z. (1981). An augmented Lagrangian method for a class of convex continuous optimal control problems. Problems of Control Inf.Theory 10(2):67-81.

Zarantonello, E.H. (1971). Projections on convex sets in Hilbert space and spectral theory. Contributions to Nonlinear Functional Analysis, ed. E.H.Zarantonello. Academic Press, New York.

ON SUFFICIENT CONDITIONS FOR OPTIMALITY OF LIPSCHITZ FUNCTIONS AND THEIR APPLICATIONS TO VECTOR OPTIMIZATION

S. Rolewicz
Institute of Mathematics, Polish Academy of Sciences, Sniadeckich 8, 00950 Warsaw, Poland

We shall start with the following numerical example concerning an optimization problem involving differentiable functions.

EXAMPLE 1. We consider the following optimization problem in three-dimensional space

$$f(x,y,z) = x + 2y - x^2 + y^2 - z^2 \to \inf$$

under conditions

$$(P) \quad g_1(x,y,z) = -(x+y) + z^2 \leq 0$$

$$g_2(x,y,z) = -y + z^4 \leq 0.$$

We want to show that (0,0,0) is a local minimum of problem (P). We shall first verify that the Kuhn-Tucker necessary conditions for optimality hold. Indeed, taking $\lambda_1 = \lambda_2 = 1$ and formulating the Lagrange function

$$L(x,y,z,\lambda_1,\lambda_2) = f(x,y,z) + \lambda_1 g_1(x,y,z)$$

$$+ \lambda_2 g_2(x,y,z) = x + 2y - x^2 + y^2 - z^2 +$$

$$+ (-(x+y) + z^2) + (-y + z^4) = -x^2 + y^2 + z^4$$

we trivially obtain that

$$\nabla L(x,y,z)\Big|_{(0,0,0)} = 0$$

Unfortunately the classical sufficient condition of optimality (Hesteness (1947), McShane (1942)) does not hold. The second differential of the Lagrangian at (0,0,0) is determined by the matrix

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and on the line orthogonal to the gradients $\nabla g_1\Big|_{(0,0,0)} =$ $= (-1,-1,0)$ and $\nabla g_2\Big|_{(0,0,0)} = (0,-1,0)$, i.e on the axis of z it simply vanishes.

This stimulates an approach to sufficient conditions which is different to the classical one proposed by McShane (1942) and Hesteness (1947).

The classical idea was based on direct approximation of the problem by approximations of linear and quadratic type.

Another approach is based on the idea of the implicit function theorem and in the simplest case can be expressed by the following:

THEOREM 1. (Rolewicz; 1980b)

Let D be a domain contained in n-dimensional real Euclidean space R^n. Let $f,g_1, \dots g_m$ be continuously differentiable functions defined on D.

We consider the following optimization problem

$$\text{(P)}\quad \begin{aligned} & f(x) \to \inf \\ & g_i(x) \leq 0, \quad i = 1,2, \dots ,m, \\ & x \in D. \end{aligned}$$

We assume that $x_o \in D$ and that

(i) all constraints are active at x_o, i.e. $g_i(x_o) = 0 \quad i = 1,2, \dots m$.

(ii) the gradients $\nabla g_i \big|_{x=x_o}$, $i = 1,2, \dots m$, are linearly independent

(iii) there are $\lambda_1, \dots \lambda_m$ strictly positive, $(\lambda_i > 0,\ i = 1,2, \dots m)$ such that the gradient at x_o of the Lagrange function is equal to 0,

$$\nabla (f + \sum_{i=1}^{m} \lambda_i g_i) \big|_{x=x_o} = 0.$$

Then x_o is a local minimum of the problem (P) if and only if it is a local minimum of the following equality problems:

$$f(x) \to \inf$$

$$\text{(Pe)} \qquad g_i(x) = 0 \ .$$

Having Theorem 1, we can easily show that $(0,0,0)$ is a local minimum in Example 1. Indeed, $g_1(x,y,z) = 0 = g_2(x,y,z)$ implies $y = z^4$ $x = z^4 - z^2$ and

$$f(x,y,z) = 2z^6.$$

Theorem 1 gives an algorithm reducing the problem of sufficient conditions for a problem with inequality constraints given by m functions of n variables, to the problem of sufficient conditions for a function of $(n-m)$ independent variables. The reduction procedure requires only the inversion of one matrix (Jacobian matrix at x_o) and for this reason is not computationally difficult.

Of course, a number of natural questions arise. How will the situation change if

(a) there are also equality constraints

(b) the Kuhn-Tucker necessary optimality conditions hold, but with certain $\lambda_i = 0$

or more generally

(c) the functions $f, g_1, \dots, g_m$ are defined on a Banach space

(d) the conditions $g_i(x) \leq 0$ are replaced by a condition $G(x) \leq 0$ where G maps a Banach space X into an ordered Banach space Y.

There is a possibility of extending Theorem 1 to the general case. This may be done using the following theorem:

THEOREM 2. (Rolewicz, 1981a). Let X, Y_1, Y_2, Z be Banach spaces over real numbers. We assume that Y_1, Y_2 are ordered. Let D be an open set in X. We assume that there are continuously Fréchet differentiable operators, F, G_1, G_2, H mapping D into real numbers (F), into $Y_1 (G_1)$, into $Y_2 (G_2)$, into $Z (H)$. Let $x_o \in D$. Suppose that

(i) $G_1(x_o) = G_2(x_o) = 0$

(ii) the differential ∇ of the mapping (G_1, G_2, H) taken at x_o, maps X into the product $Y_1 \times Y_2 \times Z$, (i.e. it is a surjection)

(iii) there is a uniformly positive linear functional φ_1 (i.e. such that there is $C > 0$ such that

$$\|y_1\| \leq C\varphi(y_1)$$

for $y_1 \in Y_1$, $y_1 \geq 0$)

and there are linear continuous functionals $\varphi_2 \in Y_2^*$, $\varphi_2 \geq 0$, $\Psi \in Z^*$ such that the gradient of the Lagrange function taken at the point x_o is equal to 0

$$\nabla (F(x) + \varphi_1(G_1(x)) + \varphi_2(G_2(x)) + \Psi(H(x))\Big|_{x_o} = 0.$$

Then x_o is a local minimum of the problem

$$F(x) \to \inf$$

$$(P) \qquad G_1(x) \leq 0, \quad G_2(x) \leq 0 \quad H(x) = 0$$

if and only if it is a local minimum of the following problem:

$$F(x) \to \inf$$

$$(Pe) \qquad G_1(x) = 0, \quad G_2(x) \leq 0, \quad H(x) = 0$$

Theorem 2 can be extended to the case of Lipschitz functions in the following way.

THEOREM 3. (Rolewicz, 1981a).

Let, X, Y_1, Y_2, Z be Banach spaces over reals. Let the spaces Y_1, Y_2 be ordered, Let D be an open domain in X. Let F, G_1, G_2, H be mappings defined on D with values being real numbers (F), having values in $Y_1 (G_1)$, in $Y_2 (G_2)$, in $Z(H)$. Let $x_o \in D$. We assume that

(i) $G_1(x_o) = G_2(x_o) = 0$

(ii) the multifunction $\Gamma(y_1, y_2, z) = \{x \in D: G_1 x = y_1, G_2 x = y_2, Hx = z\}$ is locally Lipschitzian at x_o i.e. there is a neighbourhood Q of x_o and a constant $K > 0$ such that

$$d(\Gamma(y_1, y_2, z) \cap Q, \Gamma(\bar{y}_1, \bar{y}_2, \bar{z}) \cap Q) \\ \leq K(\|y_1 - \bar{y}_1\| + \|y_2 - \bar{y}_2\| + \|z - \bar{z}\|)$$

where $d(A,B)$ denotes the Hausdorff distance of the sets A, B

(iii) there are odd functionals, φ_1 defined on Y_1, φ_2 defined on Y_2, Ψ defined on Z, where φ_2 is nonnegative, φ_1 is strictly positive (i.e. there is $C > 0$ such that

$$\|y_1\| \leq C\varphi_1(y_1) \quad \text{for} \quad y_1 \geq 0)$$

such that the Lagrange function

$$L(x) = F(x) + \varphi_1(G_1(x)) + \varphi_2(G_2(x)) + \Psi(H(x))$$

satisfies the Lipschitz condition with constant M.

(iv) $MKC < 1$.

Then x_o is a local minimum of the following problem:

$$\text{(P)} \quad \begin{aligned} & F(x) \to \inf \\ & G_1(x) \leq 0,\ G_2(x) \leq 0,\ H(x) \leq 0 \end{aligned}$$

if and only if it is a local minimum of the following problem:

$$\text{(Pe)} \quad \begin{aligned} & F(x) \to \inf \\ & G_1(x) = 0,\ G_2(x) \leq 0,\ H(x) = 0 \end{aligned}$$

Theorem 3 generalizes Theorem 2. If the hypotheses of Theorem 2 are satisfied, then by the Ljusternik theorem (Ljustèrnik, 1934) the multifunction $\Gamma(y_1,y_2,z)$ is pseudo--Lipschitzian with a certain constant K_o, i.e., there is a neighbourhood Q_o of x_o such that

$$\Gamma(y_1,y_2,z) \cap Q_o \subset \Gamma(\bar{y}_1,\bar{y}_2,\bar{z}) +$$

$$+ K_o(\|y_1 - \bar{y}_1\| + \|y_2 - \bar{y}_2\| + \|z - \bar{z}\|)B$$

where B is the unit ball in the space X.

It can be shown that $\Gamma(y_1,y_2,z)$ is in fact locally Lipschitzian with a Lipschitz constant K which is an arbitrary number greater than K_o (Rolewicz, 1980a). By (i.i.i) of Theorem 2 we can find a neighbourhood Q_1 such that $\Gamma(y_1,y_2,z)\cap Q_1$ satisfies the Lipschitz condition with constant K, and such that the Lipschitz constant of the Lagrange function is smaller then $\frac{1}{KC}$, i.e. (iv) of Theorem 3 holds.

Theorem 3 can be used for vector optimization in the following way.

THEOREM 4. (Rolewicz, 1983a).

Let P,X,Y_1,Y_2,Z be Banach spaces over real numbers. We assume that P,Y_1,Y_2 are ordered by cones.

Let D be a domain in the space X. Let F,G_1,G_2,H map D into $P(F)$, $Y_1(G_1)$, $Y_2(G_2)$, $Z(H)$. We assume that all mappings F,G_1,G_2,H are continuously Fréchet differentiable. Let $x_o \in D$. If

(i) the constraints G_1,G_2 are active at x_o (i.e. $G_1(x_o) = G_2(x_o) = 0$).

(ii) the gradient of (G_1,G_2,H) at x_o is a surjection of X onto the product $Y_1\times Y_2\times Z$

(iii) there are strictly positive linear functionals φ_1 defined on Y_1; α defined on P (i.e. such that there are constants C, $C_1>0$ such that

$$||p|| \leq C\alpha(p), \quad \text{for} \quad p\in P,\ p\geq 0$$

$$||y_1|| \leq C_1\varphi_1(y_1) \quad \text{for} \quad y_1\in Y,\ y_1\geq 0)$$

and a nonnegative linear continuous functional $\varphi_1 \in Y^*$ and a continuous linear functional $\Psi\in Z^*$, such that the gradient of the Lagrange function

$$\nabla(\alpha(F(x)) + \varphi_1(G_1(x)) + \varphi_2(G_2(x)) + \Psi(H(x))\Big|_{x=0} = 0$$

(iv) the space $L_1 = \ker \nabla F\big|_{x=0}$ and the space

$$L_2 = \ker \nabla G_1\big|_{x=0} \cap \ker \nabla H\big|_{x=0} \cap$$

$$\cap \neq \{x: \nabla G_2\big|_{x=0}(x) \leq 0\}$$

have a positive gap (i.e.

$$d = \max\ (\inf \{\|x - y\|,\ x \in L_1,\ y \in L_2,\ \|x\| = 1\},$$

$$\inf \{\|x - y\|,\ x \in L_1,\ y \in L_2\ \ \|y\| = 1\}) > 0$$

Then x_o is a local Pareto minimum of the following vector optimization problem:

$$\text{(VP)}\quad \begin{array}{l} F(x) \to \inf \\ G_1(x) \leq 0,\ G_2(x) \leq 0,\ H(x) = 0 \end{array}$$

The proof consists of three steps.

Step 1. We show that x_o is a local minimum of the following scalar problem with a Lipschitzian, but nondifferentiable goal function

$$\text{(SPe)}\quad \begin{array}{l} \alpha(F(x)) + \beta\|F(x) - F(x_o)\| \to \inf \\ G_1(x) = 0,\ G_2(x) \leq 0,\ H(x) = 0 \end{array}$$

for all $\beta > 0$.

Step 2. Using Theorem 3 we obtain that x_o is a local minimum of the following scalar problem

$$\text{(SP)}\quad \begin{array}{l} \alpha(F(x)) + \beta\|F(x) - F(x_o)\| \to \inf \\ G_1(x) \leq 0,\ G_2(x) \leq 0\ \ H(x) = 0 \end{array}$$

Step 3. Using the method of scalarisation (see for example Wierzbicki, 1979) we show that there is $\beta_o > 0$ such that if x_o is a local minimum of (SP) for β, $0 < \beta < \beta_o$, then it is a local Pareto minimum of problem (VP).

Theorem 4 above has a serious disadvantage, namely condition (iv).

In mathematical programming, in particular, (iv) implies that the number of coordinates in F must not be smaller than the difference between the dimension of the space and the number of conditions. For the case when this is not true we use the following theorem:

THEOREM 5. We consider the following vector optimization problem:

$$(VP)\qquad \begin{array}{l} f_1(x) \ \dots \ f_k(x) \to \inf. \\ g_1(x) \leq 0 \quad g_m(x) \leq 0. \end{array}$$

A point x_o is a Pareto minimum (local minimum) of problem (VP) if and only if it is a minimum (local minimum) of all the following scalar problems

$$(SP_i)\qquad \begin{array}{l} f_i(x) \to \inf \\ g_1(x) \leq 0, \ \dots \ , g_m(x) \leq 0 \\ f_1(x) \leq 0, \ \dots \ , f_{i-1}(x) \leq 0, \ f_{i+1}(x) \leq 0, \ \dots \ , f_K(x) \leq 0 \\ i = 1, 2, \ \dots \ , K. \end{array}$$

In Rolewicz (1984) a simple but nontrivial numerical problem is solved using Theorem 5.

REFERENCES

Hesteness, M.R. (1947), An indirect proof for the problem of Bolza in non-parametric form. Trans. Amer. Math. Soc. 62, 509-535.

Ljusternik, L.A. (1934), On conditional extrema of functionals (in Russian). Mat. Sb. 41, 390-401.

Mc Shane. (1942). Sufficient conditions for a weak relative minimum in the problem of Bolza. Trans. Amer. Math. Soc. 52, 344-379.

Rolewicz, S. (1980a). On intersection of multifunctions. Math. Operationsforschunᴦ und Stat. Ser. Optimization 11, 3-11.

Rolewicz, S. (1980b). On sufficient conditions of optimality in mathematical proqramming. Oper. Res. Verf. (Meth. of Oper. Res) 40, 149-152.

Rolewicz, S. (1981a). On sufficient conditions of optimality for Lipschitz functions. In Moeschlin, O, Pallaschke D (eds). Game Theory and Mathematical Economics, North-Holland, 351-355.

Rolewicz, S. (1981b). On sufficient conditions of vector optimization. Meth. of Oper. Res. 43, 151-157.

Rolewicz, S. (1983a). Sufficient conditions for Pareto optimization in Banach spaces. Stud. Math. 77, 111-114.

Rolewicz, S. (1983b). On sufficient conditions of optimality of second order. Ann. Pol. Math. 42, 297-300.

Rolewicz, S. (1984). Remarks on Sufficient Conditions of Optimality of Vector Optimization. Math. Operationsforschung und Stat. Ser. Optimization 15, 37-40.

Wierzbicki, A. (1979). On the use of penalty functions in multiobjective optimization. Oper. Res. Verf. (Methods of Oper. Res.) 31, 719-735.

OPTIMAL CONTROL OF HYPERBOLIC VARIATIONAL INEQUALITIES

Dan Tiba
INCREST, Department of Mathematics, Bd. Pacii 220, 79622 Bucharest, Romania

1. INTRODUCTION

Variational inequalities and free boundary problems arise in a natural way in a variety of physical phenomena. The study of their control, both from theoretical and numerical point of view, was initiated in the works of J.P. Yvon [12] and F.Mignot [7]. The literature is rich in results on elliptic and parabolic problems and we quote the recent book of Barbu [2] for a survey in this respect.

Our aim is to comment some new results on the control of hyperbolic variational inequalities based mainly on the recent works of the author [9], [10], [11]. In section 2 optimality conditions are obtained for the vibrating string with obstacle. In the next sections we study hyperbolic variational inequalities with unilateral conditions on the derivative of the state, in the domain or on the boundary. For the sake of brevity we shall give only outlines of proofs for the main results. More details can be obtained from the mentioned papers.

2. THE VIBRATING STRING WITH OBSTACLE

This is an example of a hyperbolic variational inequality with unilateral conditions on the unknown function:

$$y_{tt}-y_{xx}+w=u, \qquad w\in\beta(x,y) \tag{2.1}$$

$$y(0,x)=y_0(x), \qquad y_t(0,x)=v_0(x) \tag{2.2}$$

where $\beta(x,\cdot)$ is the maximal monotone graph

$$\beta(x,y)=\begin{cases} 0 & y>\varphi(x) \\ ]-\infty,\ 0] & y=\varphi(x) \\ \Phi & y<\varphi(x) \end{cases} \tag{2.3}$$

and φ is a continuous function on R.

Therefore the string is forced to vibrate above the given obstacle $y=\varphi(x)$. The physical meaning of the term $w\varepsilon\beta(x,y)$ is the unknown reaction of the obstacle and we formulate the control problem

(P) Minimize $\{|w|+|u|\}$

subject to $u\varepsilon L^2(B)$ and y, w satisfying (2.1), (2.2). Above $B=[0,T]xR$, $|\cdot|=|\cdot|_{L^2(B)}$ and $y_o\varepsilon H^1_{loc}(R)$, $v_o\varepsilon L^2_{loc}(R)$, $y_o\geq\varphi$ are given.

The equation (2.1), (2.2) was studied by Amerio and Prouse [1], Schatzman [8] by the method of the lines of influence of the obstacle.

This approach is difficult to follow here and we adopt the point of view from the unstable systems control theory as developed by J.L.Lions [6].

The control $u\varepsilon L^2(B)$ is called feasible if there are $y\varepsilon L^2(0,T;\ H^1_{loc}(R))$, $w\varepsilon L^2(B)$, $w\varepsilon\beta(y)$ a.e., such that $y(0,x)= =y_o(x)$ a.e. and

$$\int_B(\nabla y\nabla v+w\cdot v-y_t\cdot v_t)dxdt=\int_B u\cdot vdxdt+\int_R v_o(x)v(x,0)dx \tag{2.4}$$

for all $v\varepsilon H^1(B)$ with compact support and $v(T,x)=0$, $x\varepsilon R$. The pair $[y,w]$ is called a generalized solution of (2.1), (2.2) and the condition $w\varepsilon L^2(B)$ is a constraint on the set of admissible controls. However if u is an admissible control with $[y,w]$ the corresponding generalized solution, then u-w is also admissible with $[y,0]$ the corresponding generalized solution. Next any greater control from $L^2(B)$ is admissible and this shows that the feasible set is sufficiently rich for our problem to be well posed.

Proposition 2.1. *The existence of an admissible control implies the existence of at least one optimal pair* $[y^*,u^*]$ *for* *problem* (P).

We define the approximate problem

(P_ε) Minimize $\{|\beta^\varepsilon(y)|+|u|\}$

subject to

$$y_{tt}-y_{xx}+\beta^\varepsilon(x,y)=u \qquad (2.5)$$

and (2.2). Here β^ε is a regularization of β:

$$\beta^\varepsilon(x,y)=\int_{-\infty}^{\infty}\beta^\varepsilon(x,y+\varepsilon^2-\varepsilon^2\tau)\rho(\tau)d\tau \quad , \quad \varepsilon>0$$

where β_ε is the Yosida approximation of β and ρ is a Friedrichs mollifier, i.e. $\rho\geq 0$, $\rho(-\tau)=\rho(\tau)$, supp $\rho\subset[-1,1]$, $\rho\varepsilon C^\infty(R)$ and $\int_{-\infty}^{\infty}\rho(\tau)d\tau=1$.

Proposition 2.2. *If* $y_o\varepsilon H^1_{loc}(R)$, $v_o\varepsilon L^2_{loc}(R)$ *and* $u\varepsilon L^2(0,T;L^2_{loc}(R))$ *then the equation* (2.5) *has a unique generalized solution* $y\varepsilon L^\infty(0,T;H^1_{loc}(R))$ *and* $y_t\varepsilon L^\infty(0,T;\ L^2_{loc}(R))$.

Let J_ε and J be the cost functionals associated with (P_ε), (P).

Theorem 2.3. *Denote by* $[y^\varepsilon,u_\varepsilon]$ *an optimal pair for* (P_ε). *Then:*

i) $J_\varepsilon(u_\varepsilon)\leq J(u^*)$

ii) $\lim\limits_{\varepsilon\to 0} J_\varepsilon(u_\varepsilon)=J(u^*)$

iii) *on a subsequence*

$u_\varepsilon \to u^*$ *strongly in* $L^2(B)$

$\beta^\varepsilon(y^\varepsilon) \to w^*\varepsilon\beta(y^*)$ *strongly in* $L^2(B)$

$y^\varepsilon \to y^*$ *stronly in* $C(0,T;\ L^2_{loc}(R))$.

Corollary 2.4. *The problem* (P) *has a feasible control iff the sequence* $\{J_\varepsilon(u_\varepsilon)\}$ *is bounded.*

Now assume that $y_o\varepsilon H^1(R)$, $v_o\varepsilon L^2(R)$. Then $y^\varepsilon\varepsilon L^\infty(0,T;H^1(R))\cap W^{1,\infty}(0,T;\ L^2(R))$ and it is a strongly convergent sequence. Denote by $\psi:L^2(B)\to R$ the norm $\psi(u)=|u|$.

Theorem 2.5. *If* $\varphi\varepsilon C^1(R)$ *there is an optimal pair* $[y^*,u^*]$

in $L^2(0,T;\ H^1(R))xL^2(B)$, *an adjoint optimal state* $p^*\varepsilon L^2(B)$ *and a distribution* δ *on* B *with* $\text{supp}\ \delta\subset\{(t,x)\varepsilon B;\ y^*(t,x)=\varphi(x)\}$ *(the impact set) satisfying the optimality conditions:*

$$p^*_{tt}-p^*_{xx}=\delta \qquad \textit{in } \mathcal{D}'(B)\ , \qquad (2.6)$$

$$p^*\varepsilon\,\partial\psi(u^*) \qquad \text{a.e. } B\ , \qquad (2.7)$$

The proof is based on the following proposition

Proposition 2.6. *For every solution* $[y^\varepsilon,u_\varepsilon]$ *of* (P_ε) *there is* $p^\varepsilon\varepsilon L^\infty(0,T;\ H^1(R))\cap W^{1,\infty}(0,T;\ L^2(R))$ *such that:*

$$p^\varepsilon_{tt}-p^\varepsilon_{xx}+\beta^\varepsilon_y(y^\varepsilon)\cdot p^\varepsilon\varepsilon-\partial\psi(\beta^\varepsilon(y^\varepsilon))\cdot\beta^\varepsilon_y(y^\varepsilon)$$

$$p^\varepsilon(T,x)=p^\varepsilon_t(T,x)=0$$

$$p^\varepsilon\varepsilon\,\partial\psi(u_\varepsilon)$$

Proof of Theorem 2.5

Let y^*, u^*, p^* be such that on a subsequence $y^\varepsilon\to y^*$, $u_\varepsilon\to u^*$ strongly in $L^2(0,T;\ H^1(R))$, $L^2(B)$ and $p^\varepsilon\to p^*$ weakly in $L^2(B)$. Relation (2.7) is an obvious consequence of the demiclosedness of $\partial\psi$.

Concerning (2.6) we remark that y^* is continuous on B and $y^\varepsilon\to y^*$ uniformly on compact subsets of B.

Let $Q^n_\mu=\{(t,x)\varepsilon B;\ -n<x<n \text{ and } y^*(t,x)>\varphi(x)+\mu\}$ and $Q_o=\{(t,x)\varepsilon B;\ y^*(t,x)>\varphi(x)\}$ be open subsets of B.

There is $\varepsilon_o>0$ such that for $\varepsilon\leq\varepsilon_o$, $y^\varepsilon(t,x)\geq\varphi(x)+\frac{\mu}{2}$ on Q^n_μ , so $p^\varepsilon_{tt}-p^\varepsilon_{xx}=0$ on Q^n_μ for $\varepsilon\leq\varepsilon_o$. This follows from (2.3)and the definition of β^ε which imply $\beta^\varepsilon(x,y)=0$ for $y\geq\varphi(x)$.

Passing to the limit in $\mathcal{D}'(B)$ we see that the distribution $p^*_{tt}-p^*_{xx}$ vanishes on Q^n_μ. But $Q_o=\bigcup\limits_{n,\mu}Q^n_\mu$ and the proof is finished.

Remark. We underline that our results and methods apply also to higher dimensions or to finite domains. More general cost functionals including terms of the form $|y-y_d|$ or $|y(T)-y_d|_{L^2(R)}$ can be considered too.

3. UNILATERAL CONDITIONS ON THE DERIVATIVE

Let $\Omega \subset R^N$ be an open domain and $Q=]0,T[\times\Omega$ be a cylinder with lateral face $\Sigma=]0,T[\times\partial\Omega$. We analyse the control problem

$$\text{Minimize } \int_0^T L(y(t),\ u(t))dt \tag{3.1}$$

subject to:

$$y_{tt}-\Delta y+\beta(y_t)\ni Bu(t) \qquad \text{in } \Omega, \tag{3.2}$$

$$y(0,x)=y_o(x),\ y_t(0,x)=v_o(x) \qquad \text{in } \Omega, \tag{3.3}$$

$$y(t,x)=0 \qquad \text{in } \Sigma. \tag{3.4}$$

Here $\beta \subset R\times R$ is any maximal montone graph, $B:U \to H_o^1(\Omega)$ is a linear continuous operator with U a Hilbert space of control, $L:L^2(\Omega)\times U \to]-\infty,+\infty]$ is a convex, lower semicontinuous functional and $y_o \varepsilon H_o^1(\Omega) \cap H^2(\Omega)$, $v_o \varepsilon L^2(\Omega)$.

Equation (3.2)-(3.4) has a solution $y\varepsilon C(0,T;H_o^1(\Omega))$, $\partial y/\partial t\varepsilon C(0,T;L^2(\Omega)) \cap L^\infty(0,T;H_o^1(\Omega))$, $\partial^2y/\partial t^2\varepsilon L^2(0,T;L^2(\Omega))$ by a variant of a result from Barbu [3], p.279.

If some coercivity properties are assumed for L, then one may infer the existence of an optimal pair $[u^*,y^*]$.

Define the regularizations of β, L:

$$\beta^\varepsilon(y)=\int_{-\infty}^{\infty}\beta_\varepsilon(y-\varepsilon\tau)\rho(\tau)d\tau \tag{3.5}$$

$$L^\varepsilon(y,u)=\inf\left\{\frac{|y-z|^2_{L^2(\Omega)}+|u-v|^2_U}{2\delta(\varepsilon)}+L(z,v)\right\} \tag{3.6}$$

where $\delta(\varepsilon) \to 0$ when $\varepsilon \to 0$.

The approximate control problem is

$$\text{Minimize } \{\int_0^T L^\varepsilon(y,u)+\tfrac{1}{2}\int_0^T|u-u^*|^2_U\} \tag{3.7}$$

subject to (3.2)-(3.4) with β replaced by β^ε.

Problem (3.7) is a smooth control problem and one may obtain in quite a standard manner the necessary conditions:

Proposition 3.1. *For every approximate optimal pair* $[y^\varepsilon,u_\varepsilon]$ *there is* $m^\varepsilon\varepsilon C(0,T;L^2(\Omega))$ *such that:*

$$m^\varepsilon_{tt}-\Delta m^\varepsilon-\beta^\varepsilon_y(y^\varepsilon_t)\cdot m^\varepsilon_t=\int_t^T q_\varepsilon \qquad \text{in } Q, \tag{3.8}$$

$$m^{\varepsilon}(T,x)=m_t^{\varepsilon}(T,x)=0 \qquad \text{in } \Omega, \tag{3.9}$$

$$m^{\varepsilon}(t,x)=0 \qquad \text{in } \Sigma, \tag{3.10}$$

$$[q_{\varepsilon}(t),-B^*m_t^{\varepsilon}+u_{\varepsilon}(t)-u^*(t)]=\partial L^{\varepsilon}(y^{\varepsilon}(t),u_{\varepsilon}(t)). \tag{3.11}$$

Moreover, we have $y^{\varepsilon} \to y^*$ *strongly in* $C(0,T;H_o^1(\Omega))$, $y_t^{\varepsilon} \to$ $\to y_t^*$ *strongly in* $C(0,T;L^2(\Omega))$, $\beta^{\varepsilon}(y_t^{\varepsilon}) \to \beta(y_t^*)$ *weakly in* $L^2(\Omega)$, $u_{\varepsilon} \to u^*$ *strongly in* $L^2(0,T;U)$, $p^{\varepsilon}=-m_t^{\varepsilon} \to p$ *weakly* in* $L^{\infty}(0,T;$ $L^2(\Omega))$ *and* $q_{\varepsilon} \to q$ *weakly in* $L^1(0,T;L^2(\Omega))$ *where*

$$[q(t),\ B_p^*(t)]\varepsilon\partial L(y^*(t),u^*(t)) \quad \text{in } [0,T]. \tag{3.12}$$

To pass to the limit in the adjoint equation (3.8) the additional assumption that β is locally Lipshitz and satisfies

$$|\beta_y(y)\cdot y|\le C(|\beta(y)|+y^2+1) \qquad \text{a.e. } R \tag{3.13}$$

is made.

Theorem 3.2. *Let* $[y^*,u^*]\varepsilon W^{2,2}(0,T;L^2(\Omega))xL^2(0,T;U)$ *be an optimal pair for problem* (3.1)-(3.4). *There exist functions* $m\varepsilon L^{\infty}(0,T;H_o^1(\Omega))\cap W^{1,\infty}(0,T;L^2(\Omega))$, $q\varepsilon L^2(Q)$ *and* $h\varepsilon L^1(Q)$ *such that:*

$$m_{tt}-\Delta m-h=\int_t^T q \qquad \text{in } Q\ ,$$

$$m(T,x)=m_t(T,x)=0 \qquad \text{in } \Omega\ ,$$

Proof

Obviously $\{m^{\varepsilon}\}$ is bounded in $L^{\infty}(0,T;H_o^1(\Omega))$ and $\{m_t^{\varepsilon}\}$ is bounded in $L^{\infty}(0,T;L^2(\Omega))$. Fix n a natural number and consider

$$E_n^{\varepsilon}=\{(x,t)\varepsilon Q;\ |y_t^{\varepsilon}(x,t)|\le n\}\ , \qquad \varepsilon>0$$

then $|\beta_y^{\varepsilon}(y_t^{\varepsilon}(t,x))|\le C_n$ on E_n^{ε} with C_n independent of ε, as β is locally Lipschitz.

Denote by E a measurable subset of Q. We have

$$|\int_E m_t^{\varepsilon}\cdot\beta_y^{\varepsilon}(y_t^{\varepsilon})dxdt|\le C_n\int_E|m_t^{\varepsilon}|dxdt+C/n\int_{E-E_n^{\varepsilon}}|\beta^{\varepsilon}(y_t^{\varepsilon})|\cdot|m_t^{\varepsilon}|dxdt+$$

$$+C/n+C\int_{E-E_n^{\varepsilon}}|y_t^{\varepsilon}|\cdot|m_t^{\varepsilon}|dxdt\ .$$

Since $\{m_t^\varepsilon\}$, $\{\beta^\varepsilon(y_t^\varepsilon)\}$ are bounded in $L^2(Q)$, we obtain:

$$|\int_E \beta_y^\varepsilon(y_t^\varepsilon)\cdot m_t^\varepsilon dxdt| \le C\mu(E)^{1/2}\cdot C_n + C/n + C(\int_{E-E_n^\varepsilon} |y_t^\varepsilon|^2)^{1/2} .$$

But $\{y_t^\varepsilon\}$ is bounded in $L^\infty(0,T;H_o^1(\Omega))$ and by the Sobolev embedding theorem it yields $\{y_t^\varepsilon\}$ bounded in $L^s(Q)$ with some $s>2$. Then the last term is equicontinuous. The Dunford-Pettis criterion gives

$$\beta_y^\varepsilon(y_t^\varepsilon)\cdot m_t^\varepsilon \to h \quad \text{weakly in} \quad L^1(Q).$$

Combining with the results of Proposition 3.1 one can pass to the limit in (3.8)-(3.9) to finish the proof.

Remark. If β is a continuously differentiable function, it is easy to see that $h(t,x)=\beta_y(y_t^*(t,x))\cdot m_t(t,x)$ a.e. Q. In more general situations the Clarke [4] generalized gradient $\partial\beta$ will be used.

Assume that

$$\beta=\gamma-\lambda \tag{3.14}$$

where γ, λ are real, convex functions.

Theorem 3.3. *Under the above hypotheses, there are functions* $m\varepsilon L^\infty(0,T;H_o^1(\Omega))\cap W^{1,\infty}(0,T;L^2(\Omega))$ *and* $q\varepsilon L^2(Q)$ *satisfying*

$$m_{tt}-\Delta m-\partial\beta(y_t^*)\cdot m_t \int_t^T q \qquad \text{in } Q,$$

$$m(T,x)=m_t(T,x)=0 \qquad \text{in } \Omega,$$

$$[q(t), -B^*m_t(t)]\varepsilon\partial L(y^*(t),u^*(t)) \quad \text{in } [0,T].$$

Proof

For the sake of simplicity take β in (3.14) a real, convex function. Write

$$m_t^\varepsilon=m_+^\varepsilon-m_-^\varepsilon$$

where m_+^ε , m_-^ε are the positive and the negative part of m_t^ε up to a constant and are strictly positive. We can suppose

$m_+^\varepsilon \to v_+$, $m_-^\varepsilon \to v_-$ weakly in $L^2(\Omega)$,

$m_t = v_+ - v_-$.

By Proposition 3.1 and the Egorov theorem for every $\eta > 0$, there is $Q_\eta \subset Q$, $\operatorname{meas}(\Omega - \Omega_\eta) < \eta$ and $y_t^\varepsilon \to y_t^*$ uniformly on Q_η. We study first the weak convergence of $\beta_y^\varepsilon(y_t^\varepsilon) \cdot m_+^\varepsilon$ in $L^2(\Omega_\eta)$.

Since β is locally Lipschitz after a tedious computation involving (3.5) we reduce the problem to the study of the weak $L^2(Q_\eta)$ convergence for the sequence $m_+^\varepsilon \cdot \partial\beta((I+\varepsilon\beta)^{-1}(y_t^\varepsilon - \varepsilon\Theta))$, Θ fixed in $[-1,1]$.

Here $\partial\beta$ is just the subdifferential of the convex function β.

Consider the proper, closed saddle function

$$K(m,y) = \begin{cases} m\beta(y) & m \geq 0 \\ -\infty & m < 0 \ . \end{cases} \tag{3.15}$$

The maximal monotone operator ∂K in $R^2 x R^2$ is given by

$$\partial K(m,y) = [-\beta(y),\ m\partial\beta(y)] \tag{3.16}$$

Denote $\partial\tilde{K}$ the maximal monotone realization of ∂K in $L^2(Q_\eta) x L^2(Q_\eta)$. Then

$$[-\beta((I+\varepsilon\beta)^{-1}(y_t^\varepsilon - \varepsilon\Theta)),\ m_+^\varepsilon \cdot \partial\beta((I+\varepsilon\beta)^{-1}(y_t^\varepsilon - \varepsilon\Theta))] \varepsilon$$

$$\varepsilon\partial\tilde{K}(m_+^\varepsilon \ ,\ (I+\varepsilon\beta)^{-1}(y_t^\varepsilon - \varepsilon\Theta)) \qquad \text{a.e. } Q_\eta \ .$$

We remark that all the terms in the above relation are weakly convergent in $L^2(\Omega_\eta)$. Moreover the following condition is satisfies:

$$\lim_{\varepsilon,\mu\to 0} <[m_+^\varepsilon, (I+\varepsilon\beta)^{-1}(y_t^\varepsilon - \varepsilon\Theta)] - [m_+^\mu, (I+\mu\beta)^{-1}(y_t^\mu - \mu\Theta)] \ ,$$

$$[-\beta((I+\varepsilon\beta)^{-1}(y_t^\varepsilon - \varepsilon\Theta)),\ m_+^\varepsilon \partial\beta((I+\varepsilon\beta)^{-1}(y_t^\varepsilon - \varepsilon\Theta))] -$$

$$-[-\beta((I+\mu\beta)^{-1}(y_t^\mu - \mu\Theta)),\ m_+^\mu \partial\beta((I+\mu\beta)^{-1}(y_t^\mu - \mu\Theta))]>_{L^2(Q_\eta)^2} = 0$$

since $\beta((I+\varepsilon\beta)^{-1}(y_t^\varepsilon - \varepsilon\Theta))$, $(I+\varepsilon\beta)^{-1}(y_t^\varepsilon - \varepsilon\Theta)$ are uniformly convergent on Q_η .

Applying a wellknown property of monotone operators (Barbu [3], p.42) we get

$$[-\beta(y_t^*),\tilde{h}]\in\partial\tilde{K}(v_+,y_t^*)$$

where $\tilde{h}$ is the weak limit in $L^2(Q_\eta)$ of $m_+^\varepsilon\cdot\partial\beta((I+\varepsilon\beta)^{-1}(y_t^\varepsilon-\varepsilon\theta))$. Therefore

$$\tilde{h}(t,x)\in v_+(t,x)\cdot\partial\beta(y_t^*(t,x)) \qquad \text{a.e. } Q\,. \tag{3.17}$$

A similar treatment can be carried out for the sequence $m_-^\varepsilon\cdot\partial\beta((I+\varepsilon\beta)^{-1}(y_t^\varepsilon-\varepsilon\theta))$ and also in the case when (3.14) is assumed. Since the sections of $\partial\beta(y_t^*(t,x))$ which occur in (3.17) and in the other limits may differ then we can write $h(t,x)\in\varepsilon\partial\beta(y_t^*)(t,x)\cdot m_t(t,x)$ only by convention.

4. UNILATERAL CONDITIONS ON THE BOUNDARY

Now we study the distributed control problem:

$$\text{Minimize } \int_0^T L(y,u)dt \tag{4.1}$$

subject to:

$$y_{tt}-\Delta y=Bu \qquad \text{in } \Omega, \tag{4.2}$$

$$y(0,x)=y_0(x),\ y_t(0,x)=v_0(x) \qquad \text{in } \Omega, \tag{4.3}$$

$$-(\partial y/\partial n)\in\beta(y_t) \qquad \text{in } \Sigma. \tag{4.4}$$

Here $B:U\to H^1(\Omega)$ is a linear continuous operator and β is a strongly maximal monotone graph in RxR, that is $\beta=\alpha+\sigma I$, $\sigma>0$ and $\alpha\in$ RxR maximal monotone.

If $y_0\in H^2(\Omega)$, $v_0\in H^1(\Omega)$ and $-(\partial y_0/\partial n)\in\beta(v_0)$ a.e. $\partial\Omega$, there exists a unique solution y to (4.2)-(4.4) satisfying $y\in L^\infty(0,T;H^2(\Omega))\cap C(0,T;H^1(\Omega))$, $y_t\in L^\infty(0,T;H^1(\Omega))\cap C(0,T;L^2(\Omega))$, $y_{tt}\in L^\infty(0,T;L^2(\Omega))$.

Under some coercivity assumptions for L, one may infer the existence of an optimal pair $[u^*,y^*]$ in the problem (4.1)-(4.4).

The approximate control problem is defined by the cost functional (3.7) and the state system (4.2)-(4.4) with β replaced with β^ε given by:

$$\beta^{\varepsilon}(y)=\alpha^{\varepsilon}(y)+\sigma I \tag{4.5}$$

and α^{ε} obtained as in (3.5).

Due to the appropriate differentiability properties, we obtain the approximate optimality conditions:

Proposition 4.1. *For every approximate optimal pair* $[y^{\varepsilon},u_{\varepsilon}]$ *there is* $m^{\varepsilon}\varepsilon C(0,T;\ L^2(\Omega))$ *such that:*

$$m^{\varepsilon}_{tt}-\Delta m^{\varepsilon}=\int_t^T q_{\varepsilon} \qquad \text{in } Q, \tag{4.6}$$

$$m^{\varepsilon}(T,x)=m^{\varepsilon}_t(T,x)=0 \qquad \text{in } \Omega, \tag{4.7}$$

$$-(\partial m^{\varepsilon}/\partial n)=\beta^{\varepsilon}_y(y^{\varepsilon}_t)\cdot m^{\varepsilon}_t \qquad \text{in } \Sigma, \tag{4.8}$$

$$[q_{\varepsilon}(t),\ -B^*m^{\varepsilon}_t(t)+u_{\varepsilon}(t)-u^*(t)]=\partial L^{\varepsilon}(y^{\varepsilon}(t),u_{\varepsilon}(t)) \tag{4.9}$$

Moreover, $y^{\varepsilon}\to y^*$ *strongly in* $C(0,T;\ H^1(\Omega))$, $y^{\varepsilon}_t\to y^*_t$ *strongly in* $C(0,T;\ L^2(\Omega))$, $u_{\varepsilon}\to u^*$ *strongly in* $L^2(0,T;\ U)$, $p^{\varepsilon}=-m^{\varepsilon}_t\to p$ *weakly* in* $L^{\infty}(0,T;\ L^2(\Omega))$ *and* $q_{\varepsilon}\to q$ *weakly in* $L^1(0,T;\ L^2(\Omega))$ *where*

$$[q(t),\ B^*p(t)]\varepsilon\partial L(y^*(t),u^*(t)) \quad \text{in } [0,T]. \tag{4.10}$$

To pass to the limit in (4.6)-(4.8) one has to make again hypotheses (3.13), (3.14).

Theorem 4.2. *Let* $[u^*,y^*]\varepsilon W^{2,2}(0,T;\ L^2(\Omega))xL^2(0,T;\ U)$ *be an optimal pair for problem* (4.1)-(4.4). *There exist functions* $m\varepsilon L^{\infty}(0,T;\ H^1(\Omega))\cap W^{1,\infty}(0,T;\ L^2(\Omega))$, $q\varepsilon L^2(Q)$ *satisfying:*

$$m_{tt}-\Delta m=\int_t^T q \qquad \text{in } Q,$$

$$m(T,x)=m_t(T,x)=0 \qquad \text{in } \Omega,$$

$$-(\partial m/\partial n)\varepsilon\partial\beta(y^*_t)\cdot m_t \qquad \text{in } \Sigma.$$

Here $\partial\beta$ is the generalized gradient of the locally Lipschitz function β and the proof follows the same lines as in the previous section.

Remark. In the paper [9] an abstract scheme is built to obtain the results of sections 3 and 4. It allows other important applications to the parabolic case, to differential systems with delay.

Remark. Following the unstable systems approach, as in section 1, it is possible to obtain necessary conditions for hyperbolic control problems with strong nonlinearities, for instance exponentials. In this respect we quote the forthcoming paper [5].

REFERENCES

1. Amerio, L. and Prouse, G. (1975). Study of the motion of a string vibrating against an obstacle. Rend. di Matematica, 2.
2. Barbu, V. (1984). Optimal control of variational inequalities. Research Notes in Mathematics 100, Pitman, London.
3. Barbu, V. (1976). Nonlinear semigroups and differential equations in Banach spaces. Noordhoff, Leyden-Ed. Academiei, Bucureşti.
4. Clarke, F.H. (1975). Generalized gradients and applications. Trans. Amer. Math. Soc. 205.
5. Komornik, V. and Tiba, D. Optimal control of strongly nonlinear hyperbolic systems. In preparation.
6. Lions, J.L. (1983). Contrôle des systèmes distribués singuliers. Dunod, Paris.
7. Mignot, F. (1976). Contrôle dans les inequations variationelles elliptiques. Journal of Funct. Anal., 22.
8. Schatzman, M. (1979). Thèse, Univ. "Pierre et Marie Curie", Paris.
9. Tiba, D. (1985). Optimality conditions for distributed control problems with nonlinear state equaion. SIAM J. Control and Optimization, 23.
10. Tiba, D. (1984). Some remarks on the control of the vibrating string with obstacle. Rev. Roum. Math. Pures Appl., 10.
11. Tiba, D. (1984). Quelques remarques sur le contrôle de la corde vibrante avec obstacle. C.R.A.S. Paris. In print.
12. Yvon, J.P. (1974). Report INRIA, 53.

ON DUALITY THEORY RELATED TO APPROXIMATE SOLUTIONS OF VECTOR-VALUED OPTIMIZATION PROBLEMS

István Vályi
Bureau for Systems Analysis, State Office for Technical Development,
P.O. Box 565, 1374 Budapest, Hungary

1. INTRODUCTION

The notion of approximate solutions or ε-solutions emerged early in the development of modern convex analysis. An analogue of the well-known statement concerning the minimum of a convex function and its subgradient also holds in the approximate case: a convex function f has an ε-approximate minimum at x if and only if $0 \in \partial_\varepsilon f(x)$, where $\partial_\varepsilon f(x)$ is the ε-subdifferential of f at x. Particular attention has been paid to ε-subdifferentials (see Hiriart-Urruty, 1982; Demyanov, 1981). This has resulted in the construction of a new class of optimization procedures, the *ε-subgradient methods*. The virtually complete set of calculation rules derived for the ε-subdifferential has made possible the study and characterization of constrained convex optimization problems in both the real-valued and vector-valued cases, as in Strodiot et al. (1983), or for ordered vector spaces (Kutateladze, 1978).

Relatively little effort has been devoted to duality questions in this context (but see Strodiot et al., 1983, and the work of Loridan (1982), where duality is coupled with a technique based on Ekeland's maximum principle). Duality theory in the exact case has been thoroughly investigated even for vector-valued problems in terms of both strict optima (e.g., Ritter, 1969,1970; Zowe, 1976) and non-dominated optima (e.g., Tanino and Sawaragi, 1980; Corley, 1981). However, there is so far no corresponding theory for approximate solutions.

In this paper we intend to remedy this situation by stating some simple propositions on approximate optimal solutions; in addition we shall give some basic duality theorems for vector-valued situations, a number of which are also of interest in the scalar-valued case. In deriving the results we do not rely on the existence of ε-subgradients. This is important because until quite recently very little was known (especially in the vector-valued case) about the conditions under which the set of ε-subgradients is non-empty (Borwein et al., 1984). For these reasons we hope this paper may provide useful background information for a number of nondifferentiable optimization problems. Finally, we should mention that a vector-valued version of Ekeland's principle is also available (Vályi, 1985), but attempts to use it in the optimization context have so far failed.

Throughout this paper we shall consider only the algebraic case, although a parallel, topological version also exists. For details and proofs see Vályi (1984); other related issues are treated in Loridan (1984).

2. BASIC NOTIONS AND PRELIMINARIES

For basic definitions related to ordered vector spaces we refer the reader to Peressini (1967) or Akilov and Kutateladze (1978); for definitions related to convex analysis see Holmes (1975). All of the vector spaces considered here are real, all topologies are convex and Hausdorff, and the ordering cones are assumed to be convex, closed and pointed. A vector lattice in which every non-empty set with a lower bound possesses an infimum is said to be an *order complete space*. In order to ensure the existence of infima (or suprema) for all non-bounded sets, we add the elements ∞ and $-\infty$ to the order complete space Y and denote it by $\bar{Y}$. Here we suppose that the usual algebraic and ordering properties hold. Thus a set $H \subset Y$ which is not bounded from below has $\inf H = -\infty$, where $\inf \phi = \infty$.

The algebraic dual of the space Y will be denoted by Y', and the topological dual by Y^*. The cone of positive functionals with respect to the cone $C \subset Y$ or the dual of C is C^+, and C^* is the continuous dual. If both (X,K) and (Y,C) are ordered vector spaces, then $L^+(X,Y) \subset L(X,Y)$ denotes the cone of positive linear maps from X to Y, and $\alpha^+(X,Y) \subset \alpha(X,Y)$ the cone of continuous positive maps.

The sets of algebraic interior points and relative algebraic interior points of a set $H \subset Y$ are referred to as core (H) and rcore (H), respectively, and lina (H) denotes the set of linearly accessible points from H. The key tool in the theory that follows is the Hahn–Banach theorem for the scalar- and vector-valued cases. As shown in, e.g., Tuy (1972), there are more than 10 different but equivalent forms of this theorem, which has the following highly useful but little known corollary:

THEOREM 1. [Strict algebraic separation theorem, see Köthe (1976)]. *Let H be a convex subset of the real vector space Y, and let* rcore *$(H) \neq \phi$. If for some $y_0 \in Y$ we have $y_0 \notin$* lina *(H), then $y_0 \in Y$ can be strictly separated from H.*

In the vector case we have:

THEOREM 2 [Vector-valued separation theorem, see Zowe (1976)]. *Let X be a real vector space, (Y,C) be an order complete space, and S_1 and S_2 be convex subsets of the product space $X \times Y$. If*

$$y_1 \geq y_2 \quad \forall (x, y_1) \in S_1 \text{ and } (x, y_2) \in S_2$$

and

$$0 \in \operatorname{rcore} [P_X(S_1) - P_X(S_2)] \quad , \tag{1}$$

where $P_X: X \times Y \to X$ is the projection on X, then there exist a $T \in L(X, Y)$ and a $w \in Y$ such that

$$Tx_1 - y_1 \leq w \leq Tx_2 - y_2 \quad \forall (x_1, y_1) \in S_1 \text{ and } (x_2, y_2) \in S_2 \quad .$$

Using Theorem 2, Zowe proves a theorem concerning systems of convex inequalities, where the universal validity of the statement is shown to be equivalent to the Hahn–Banach extension theorem and in fact to the order completeness of the space. The definition of the inequality system, which is also used in defining the optimization problem, now follows.

Let Y, Y_i be real vector spaces ordered by the cones C, C_i, and X, Y_{m+j} be real vector spaces. We shall consider proper convex functions

$$f : X \to Y \cup \{\infty\}$$

$$f_i : X \to Y_i \cup \{\infty\}$$

and linear maps

$$f_{m+j} : X \to Y_{m+j} \quad , \; i \in I = \{1,2,\ldots,m\} \; , \; j \in J = \{1,2,\ldots,n\} \quad .$$

In addition, let

$$\mathbf{D} = \operatorname{dom} f \cap (\cap \{\operatorname{dom} f_i : i \in I\}) \neq \phi$$

be the common effective domain of the functions f and f_i, $i \in I$. For easy reference to the system, we shall use the following notation:

$$h = \mathbf{X}\{f_i : i \in I\} \; : \; X \to \mathbf{X}\{Y_i : i \in I\}$$

$$h : x \mapsto \{f_i(x) : i \in I\}$$

$$Z = \mathbf{X}\{Y_i : i \in I\} \text{ and } K = \mathbf{X}\{C_i : i \in I\} \quad .$$

Here (Z, K) is a vector space ordered by the pointed convex cone K, and

$$\operatorname{dom} h = \cap \{\operatorname{dom} f_i : i \in I\} \quad .$$

Similarly,

$$l = \mathbf{X}\{f_{m+j} : j \in J\} \; : \; X \to \mathbf{X}\{Y_{m+j} : j \in J\}$$

$$l\ :\ x \mapsto \{f_{m+j}(x) : j \in J\}$$

$$V = \mathbf{X}\{Y_{m+j} : j \in J\} \quad .$$

THEOREM 3 [Zowe (1976)]. *In addition to the notation and conditions given above, we shall assume that* (Y, C) *is an order complete space, and*

$$(0,0) \in \text{rcore}\ \{(h(x)+z\ , l(x)) : x\ \in \mathbf{D}\ ,\ z \in K\} \quad . \tag{2}$$

Then the following statements are equivalent:

$$f(x) \geq 0 \quad \forall x\ \in \{x \in \mathbf{D}: h(x) \leq 0\ ,\ l(x) = 0\} \tag{3}$$

$$\exists R \in L^{+}(Z, Y)\ ,\ S\ \in L(V, Y)\ , \text{ such that } f(x)\ + R\ \cdot h(x)\ + S\ \cdot l(x) \geq 0\ ,\ x\ \in \mathbf{D} \quad .(4)$$

We shall now consider different notions of approximate (order-) extremal points and their basic properties.

Definition 1. Let (Y, C) be an ordered vector space, $H \subset Y$, and $e\ \in C \subset Y$ be a positive element. Then an element $y\ \in H$ is said to be *strict e-minimal*, or $y\ \in S(e)$-min(H) if $H\ \subset y\ - e\ + C$. Conventionally, $S(e)$-min(H) = $-\infty$ if H is not bounded from below, and $S(e)$-min$(\phi) = \infty$.

The existence of strict optima, even of strict approximate optima, is very rare, and therefore the study of non-dominated optima is of major importance. As in the exact case, difficulties often arise when dealing with approximate non-dominated optima. This notion therefore has to be restricted to cases in which it can be characterized by linear functionals.

Definition 2. Let (Y, C) be an ordered vector space, $H\ \subset Y$, and $e\ \in C\ \subset Y$ be a positive element. Then a point $y\ \in H$ is said to be *P(e)-minimal* or $y\ \in P(e)$-min(H) if $(y\ -e\ -C) \cap H\ \subset \{y\ -e\}$. Conventionally, if this condition is not satisfied by any $y\ \in H$ then $P(e)$-min(H) = $-\infty$, and $P(e)$-min$(\phi) = \infty$. Further, let $y \in C^{+}$ and $\varepsilon \in \mathbb{R}^{+}$. Then $y\ \in H$ is $P(y^{*}, \varepsilon)$*-minimal* or $y\ \in P(y^{*}, \varepsilon)$-min$(H)$ if $<y^{*}, h> \geq <y^{*}, y> - \varepsilon\ \forall$ $h\ \in H$.

Now let core $(C) \neq \phi$. The element $y\ \in H$ is said to be *weakly P(e)-minimal (WP(e)-minimal)* or $y\ \in WP(e)$-min(H) if $(y\ -e\ -\text{core}(C)) \cap H\ = \phi$, with the same convention used earlier.

Now let us define the minimization problem (MP) and the corresponding vector-valued Lagrangian, which will then be studied from the point of view of the different notions of approximate optimality given in the last definition.

Definition 3. In addition to the notation and conditions given above, let us again assume that $e\ \geq 0$, $e\ \in Y$ is fixed. We define the *minimization problem* (MP) as follows:

Find elements $x_0 \in X$ such that

$$x_0 \in \{x \in \mathbf{D}: h(x) \leq 0 \, , \, l(x) = 0\} \tag{5}$$

$$f(x_0) \in \min \{f(x) \in Y: x \in \mathbf{D} \, , \, h(x) \leq 0 \, , \, l(x) = 0\} \quad . \tag{6}$$

The set

$$F = \{x \in X : x \in \mathbf{D} \, , \, h(x) \leq 0 \, , \, l(x) = 0\}$$

is called the set of ***feasible solutions***. Points $x_0 \in X$ which satisfy (5) and (6) with min replaced by $S(e)$-min will be called ***strict e-minimal solutions*** (or $S(e)$-solutions); if min is replaced by $P(e)$-min, then the points are called ***non-dominated e-minimal solutions*** or $P(e)$-solutions. ***$P(y^*, \varepsilon)$-solutions*** with $y^* \in C$ and $\varepsilon \in \mathbb{R}^+$, and ***weak $P(e)$-solutions*** (or $WP(e)$-solutions) can be defined in a corresponding manner.

It is important to note that the feasible set F and the set $f(F) + C$ of attainable points are convex, a fact which is essential for our results to be valid.

Definition 4. The (algebraic) vector-valued Lagrangian φ_L corresponding to the minimization problem is defined as follows:

$$\varphi_L : X \times L(Z, Y) \times L(V, Y) \rightarrow \bar{Y}$$

$$\varphi_L : (x, R, S) \mapsto \varphi_L(x, R, S) \quad ,$$

where

$$\varphi_L(x, R, S) = \begin{cases} \infty & \text{if } x \not\in \mathbf{D} \\ f(x) + R \cdot h(x) + S \cdot l(x) & \text{if } x \in \mathbf{D} \text{ and } R \in L^+(Z, Y) \\ -\infty & \text{if } x \in \mathbf{D} \text{ and } R \in L^+(Z, Y) \end{cases} \quad .$$

We shall call the set

$$\operatorname{dom} \varphi_L = \{(x, R, S) \in X \times L(Z, Y) \times L(V, Y): x \in \mathbf{D}, \, R \in L^+(Z, Y)\}$$

the ***effective domain of the Lagrangian*** φ_L.

3. APPROXIMATE DUALITY IN THE STRICT CASE

We shall now consider approximate solutions of the minimization problem (MP). First we shall formulate some simple relationships between approximate solutions corresponding to different $e \in Y - s$. Then we will turn to the strict e-approximate Kuhn–Tucker theorem, and finally describe some applications.

Proposition 1.

(a) The notions of strict optimum and $S(e)$-optimum coincide if $e = 0$.

(b) Let $e_1 \leq e_2$, $e_1, e_2 \in C$ and $x \in X$ be an $S(e_1)$-solution of (MP). Then x is also an $S(e_2)$-solution of (MP).

(c) Let (Y, C) be an ordered topological vector space and $\{e_\gamma \in C: \gamma \in \Gamma\}$ be a decreasing net with $\lim \{e_\gamma: \gamma \in \Gamma\} = e$. If $x \in X$ is an $S(e_\gamma)$-solution of (MP) for all $\gamma \in \Gamma$, then x is also an $S(e)$-solution of (MP).

(d) Let (Y, C) be an order complete space with a weakly sequentially complete topology, the ordering cone $C \subset Y$ be normal, and the sequence $\{e_n \in C: n \in \mathbb{N}\}$ be decreasing with

$$e = \inf \{e_n \in C : n \in \mathbb{N}\} \quad .$$

If $x \in X$ is an $S(e_n)$-solution of (MP) for every $n \in \mathbb{N}$, then x is also an $S(e)$-solution of (MP).

(e) Let (Y, C) be an ordered topological vector space and the set $\{f(x) \in Y: x \in F\} \subset Y$ be closed. Let us suppose in addition that there exist nets $\{x_\gamma \in X : \gamma \in \Gamma\}$ and $\{e_\gamma \in C : \gamma \in \Gamma\}$ with the following properties:

(i) $\{e_\gamma \in C: \gamma \in \Gamma\}$ is decreasing

(ii) $\lim e_\gamma = e$

(iii) x_γ is an $S(e_\gamma)$-solution of (MP)

(iv) there exists a $\gamma_0 \in \Gamma$ such that the set

$$S(e_{\gamma_0}) - \min \{f(x) \in Y: x \in F\}$$

is a compact subset of Y.

Then (MP) has an $S(e)$-minimal solution.

Definition 5. The element $(x_0, R_0, S_0) \in \operatorname{dom} \varphi_L$ is an $S(e)$-*saddle point* of the Lagrangian φ_L if

$$\varphi_L(x_0, R, S) - e \leq \varphi_L(x_0, R_0, S_0) \leq \varphi_L(x, R_0, S_0) + e \;\; \forall (x, R, S) \in \operatorname{dom} \varphi_L \quad .$$

We shall now establish an approximate Kuhn–Tucker theorem, i.e., a theorem which describes the implications of an element $x \in X$ being an approximate solution as compared with an approximate saddle point. In the special case $e = 0$, the theorems become identical with the results of Zowe (1976). As in that case, one implication is valid under fairly general conditions, while the other also requires a so-called constraint qualification. In this case one uses Theorem 3 (or Theorem 2), where condition

(2) (or (1) in the separation theorem) must be satisfied. The requirements formulated in the next definition are designed to do just that.

Definition 6 [Zowe (1976)]. We say that a problem satisfies the (algebraic) ***Slater–Uzawa constraint qualification*** if either

(i) there exists an $x_1 \in \operatorname{rcore}(\mathbf{D})$ such that

$$h(x_1) \in -\operatorname{rcore}(K) \quad , \quad l(x_1) = 0 \quad ,$$

or

(ii) $J = 0$ and there exists an $x_1 \in \mathbf{D}$ such that

$$h(x_1) \in -\operatorname{rcore}\{h(x) + k \in Z : x \in \mathbf{D} \, , \, k \in K\} \quad .$$

THEOREM 4. ***Let us suppose that, in addition to the conditions given earlier, the cone $K \subset Z$ is algebraically closed and*** core (K) ***is not empty. If*** $(x_0, R_0, S_0) \in \operatorname{dom} \varphi_L$ ***is an $S(e)$-saddle point for the Lagrangian φ_L, then $x_0 \in X$ is an $S(2e)$-minimal solution of*** (MP).

THEOREM 5. ***Let us suppose that, in addition to the conditions given earlier, (Y, C) is an order complete space and*** (MP) ***satisfies the Slater–Uzawa constraint qualification. If $x_0 \in X$ is an $S(e)$-minimal solution of*** (MP), ***then there exists an $(R_0, S_0) \in L^+(Z, Y) \times L(V, Y)$ such that $(x_0, R_0, S_0) \in$*** dom φ_L ***is an $S(e)$-saddle point of φ_L.***

We now use Theorems 1 and 2 to obtain a partial generalization of duality theorems by Golstein and Tuy for the vector-valued case (see Tuy, 1972 or Holmes, 1975).

Definition 7. Let us suppose that, in addition to the conditions given earlier, (Y, C) is an order complete space and $e \geq 0$. Consider the functions

$$P : X \rightarrow \bar{Y}$$

$$P : x \mapsto \sup \{\varphi_L(x, R, S) : R \in L(Z, Y) \, , \, S \in L(V, Y)\}$$

and

$$D : L(Z, Y) \times L(V, Y) \rightarrow \bar{Y}$$

$$D : (R, S) \rightarrow \inf \{\varphi_L(x, R, S) : x \in X\} \quad .$$

P and D are the ***strict (algebraic) primal*** and ***dual functions*** of the problem (MP).

Let

$$v = \inf \{P(x) \in \bar{Y} : x \in X\}$$

$$v^* = \sup \{D(R,S) \in Y: R \in L(Z,Y), S \in L(V,Y)\} \quad .$$

Then v and v^* are called the ***strict (algebraic) primal*** and ***dual values*** of (MP).

The problems

(P) Find elements $x \in X$ for which

$$P(x) \in S(e)\text{-min } \{P(x): x \in X\}$$

(D) Find elements $(R,S) \in L(Z,Y) \times L(V,Y)$ for which

$$D(R,S) \in S(e)\text{-max } \{D(R,S): R \in L(Z,Y), S \in L(V,Y)\}$$

are then the ***strict (algebraic) primal*** and ***dual problems***, respectively, for (MP).

Proposition 2. Let us suppose that, in addition to the conditions given earlier, $K \subset Z$ is algebraically closed, core $(K) \neq \phi$ and (Y, C) is an order complete space. Then the problem (P) is equivalent to (MP), or

$$P(x) = \begin{cases} f(x) & \text{if } x \in F \\ \infty & \text{if } x \notin F \end{cases} \quad .$$

Proposition 3 (Approximate weak duality). Let (Y, C) be an order complete space.

(i) The primal value of the minimization problem (MP) is greater than or equal to its dual value, i.e., $v \geq v^*$.

(ii) Let $x \in X$ be an $S(e)$-solution of the primal problem (P) and $(R,S) \in L(Z,Y) \times L(V,Y)$ be an $S(e)$-solution of the dual problem (D). Then

$$P(x) \geq D(R,S) \quad .$$

(iii) Suppose that for some $x \in X$, $(R,S) \in L(Z,Y) \times L(V,Y)$ we have

$$P(x) \leq D(R,S) + e \quad .$$

Then $x \in X$ is an $S(e)$-solution of the primal problem (P) and $(R,S) \in L(Z,Y) \times L(V,Y)$ is an $S(e)$-solution of the dual problem (D).

Definition 8. Let us suppose that, in addition to the conditions given earlier, (Y, C) is an order complete topological space, $\bar{e} = \{e_\gamma \in C: \gamma \in \Gamma\}$ is a decreasing net with $\lim \{e_\gamma: \gamma \in \Gamma\} = 0$, and $\bar{x} = \{x_\gamma \in X: \gamma \in \Gamma\}$, where x_γ is an $S(e_\gamma)$-solution of (MP). Then the net $\bar{x}$ is called a ***generalized strict solution*** of the problem (MP), and

$$v' = \inf \{f(x) \in Y: \gamma \in \Gamma \text{ and } \{x_\gamma \in X: \gamma \in \Gamma\} \text{ a generalized solution}\} \in Y$$

is the ***generalized strict value*** of (MP).

If $(x_\gamma, R_\gamma, S_\gamma) \in X \times L(Z, Y) \times L(V, Y)$ is an $S(e_\gamma)$-saddle point of the Lagrangian φ_L for all $\gamma \in \Gamma$, then the net $\{(x_\gamma, R_\gamma, S_\gamma): \gamma \in \Gamma\}$ is a *generalized strict (algebraic) saddle point.*

Here we should point out that this definition is more restricted than that given in Tuy (1972), as here we consider only feasible solutions while Tuy does not.

Definition 9. Let us suppose that, in addition to the conditions given earlier, (Y, C) is an order complete topological space. The problem (MP) is well-posed if

(i) its primal and dual values are equal, i.e., $v = v^*$

(ii) there is a net $\{(x_\gamma, R_\gamma, S_\gamma): \gamma \in \Gamma\}$ such that

$$\lim \{\varphi_L(x_\gamma, R_\gamma, S_\gamma): \gamma \in \Gamma\} = v \quad .$$

THEOREM 6. *Let us suppose that, in addition to the conditions given earlier, the cone $K \subset Z$ is algebraically closed,* core $(K) \neq \phi$, *(Y, C) is an order complete topological space, where $C \subset Y$ is a normal cone, and* (MP) *satisfies the Slater–Uzawa constraint qualification. If the problem* (MP) *has a generalized strict solution, then its generalized strict value equals its dual value.*

THEOREM 7. *Let us suppose that, in addition to the conditions given earlier, (Y, C) is an order complete topological space with the normal cone $C \subset Y$. Then if* (MP) *has a generalized strict saddle point, the problem is well-posed.*

COROLLARY 1. Let us suppose that, in addition to the conditions given earlier, $K \subset Z$ is algebraically closed, core $(K) \neq \phi$, (Y, C) is an order complete topological space with $C \subset Y$ normal, and (MP) satisfies the algebraic Slater–Uzawa constraint qualification. If the problem (MP) has a generalized solution, then it is well-posed.

It is worth noting that the reverse implication is trivial in the scalar case, which does not seem to be true here.

4. APPROXIMATE DUALITY IN THE NON-DOMINATED CASE

In this section we state propositions concerning the relations between the different types of non-dominated solutions of the problem (MP), and then give the corresponding Kuhn–Tucker theorems. The proof relies on the scalar version of Theorems 1 and 2.

Finally, in the case of $P(y^*, \varepsilon)$-solutions, we demonstrate the equivalence between primal–dual pairs of solutions and saddle points.

Proposition 2.

(a) Let $e_1 \leq e_2$ and $x \in X$ be a $P(e_1)$-minimal, $WP(e_1)$-minimal, or $P(y^*, \varepsilon)$-minimal solution of (MP). Then x is also a $P(e_2)$-solution, $WP(e_2)$-solution, or $P(y^*, \varepsilon)$-solution, respectively, of (MP).

(b) Let $y^* \in C^+$ be strictly positive. If $x \in X$ is a $P(y^*, 0)$-minimal solution, then it is also a P-solution.

(c) Let $y^* \in C^+$ be strictly positive and $\varepsilon \geq 0$. If $x \in X$ is $P(y^*, \varepsilon)$-minimal, then it is also $P(e \cdot \varepsilon / <y^*, e>)$-minimal.

(d) Let $e \geq 0$ and $x \in X$ be a $WP(e)$-solution. Then there exists a $y^* \in C^+$ such that $x \in X$ is a $P(y^*, <y^*, e>)$-solution.

(e) Assume that (Y, C) is an ordered topological vector space, the set

$$C + \{f(x) \in Y : x \in F\} \subset Y$$

is closed and $y^* \in C^*$. We also assume that

(i) $\{\varepsilon_n \in \mathbb{R}^+ : n \in \mathbb{N}\}$ is a decreasing sequence with $\lim \{\varepsilon_n : n \in \mathbb{N}\} = \varepsilon$

(ii) there is a $P(y^*, \varepsilon_n)$-solution for all $n \in \mathbb{N}$

(iii) the set $P(y^*, \varepsilon_1)$-min $\{f(x) : x \in F\}$ is compact. Then (MP) has a $P(y^*, \varepsilon)$-solution.

Definition 10. The element $(x_0, S_0, R_0) \in \operatorname{dom} \varphi_L$ is an (algebraic) $P(e)$-saddle point of the Lagrangian φ_L if

(i) $\varphi_L(x_0, R_0, S_0) \in P(e)\text{-min } \{\varphi_L(x, R_0, S_0) : x \in X\}$

(ii) $\varphi_L(x_0, R_0, S_0) \in P(e)\text{-max } \{\varphi_L(x_0, R, S) : (R, S) \in L(Z, Y) \times L(V, Y)\}$

and a $P(y^*, \varepsilon)$-*saddle point* (or a $WP(e)$-*saddle point*) if (i) and (ii) hold with $P(e)$-min, $P(e)$-max replaced by $P(y^*, \varepsilon)$-min, $P(y^*, \varepsilon)$-max (or $WP(e)$-min, $WP(e)$-max).

THEOREM 8. *Suppose that, in addition to the conditions given earlier,* $K \subset Z$ *is algebraically closed and* core$(K) \neq \phi$. *If* $(x_0, R_0, S_0) \in \operatorname{dom} \varphi_L$ *is a* $P(y^*, \varepsilon)$*-saddle point, then* $x_0 \in X$ *is a* $P(y^*, 2\varepsilon)$*-solution of* (MP).

THEOREM 9. *Suppose that, in addition to the conditions given earlier, the problem* (MP) *satisfies the algebraic Slater–Uzawa constraint qualification. If* $x_0 \in X$ *is a* $P(y^*, \varepsilon)$*-solution of* (MP), *then there exist* $(R_0, S_0) \in L^+(Z, Y) \times L(V, Y)$ *such that* $(x_0, R_0, S_0) \in \operatorname{dom} \varphi_L$ *is a* $P(y^*, \varepsilon)$*-saddle point of the Lagrangian* φ_L.

THEOREM 10. *Suppose that, in addition to the conditions given earlier, the problem* (MP) *satisfies the algebraic Slater–Uzawa constraint qualification, and* core $(C) = \phi$. *If* $x_0 \in X$ *is a WP(e)-solution of the problem* (MP) *then there exist elements* $(R_0, S_0) \in L^+(Z, Y) \times L(V, Y)$ *such that* $(x_0, R_0, S_0) \in \operatorname{dom} \varphi_L$ *is a WP(e)-saddle point of the Lagrangian* φ_L.

Theorems 8 and 9 reduce to the results of Corley (1981) and Tanino and Sawaragi (1980) in the exact case.

Definition 11. The $P(y^*, \varepsilon)$-*primal* and *dual functions* of the problem (MP) are defined by:

$$P(y^*, \varepsilon): X \to 2^{\bar{Y}}$$

$$P(y^*, \varepsilon)(x) = P(y^*, \varepsilon)\text{-max } \{\varphi_L(x, R, S): (R, S) \in L(Z, Y) \times L(V, Y)\}$$

and

$$D(y^*, \varepsilon): L(Z, Y) \times L(V, Y) \to 2^{\bar{Y}}$$

$$D(y^*, \varepsilon)(R, S) = P(y^*, \varepsilon)\text{-min } \{\varphi_L(x, R, S): x \in X\} \quad .$$

Using these functions, we define the $P(y^*, \varepsilon)$-*primal* and *dual problems* as follows:

$(P(y^*, \varepsilon))$ Find elements $x_0 \in X$ such that

$$P(y^*, \varepsilon)(x_0) \cap P(y^*, \varepsilon)\text{-min } \{ \cup \{P(y^*, \varepsilon)(x): x \in X\}\} \neq \phi$$

$(D(y^*, \varepsilon))$ Find elements $(R_0, S_0) \in L(Z, Y) \times L(V, Y)$ such that

$$D(y^*, \varepsilon)(R_0, S_0) \cap P(y^*, \varepsilon)\text{-max } \{ \cup \{D(y^*, \varepsilon)(R, S): (R, S) \in L(Z, Y) \times L(V, Y)\}\} \neq \phi.$$

Such elements $x_0 \in X$ and $(R_0, S_0) \in L(Z, Y) \times L(V, Y)$ are called the *solutions* of the problems $(P(y^*, \varepsilon))$ and $(D(y^*, \varepsilon))$, respectively.

Proposition 3. Suppose that, in addition to the conditions given above, the cone $K \subset Z$ is algebraically closed and core $(K) \neq \phi$. Then we have

(i) If $x_0 \in X$ is a $P(y^*, \varepsilon)$-solution of (MP), then it is a solution of the problem $(P(y^*, \varepsilon))$.

(ii) If $x_0 \in X$ is a solution of the problem $(P(y^*, \varepsilon))$ then it is a $P(y^*, 2\varepsilon)$-solution of (MP).

Definition 12. The point $(x_0, R_0, S_0) \in X \times L(Z, Y) \times L(V, Y)$ is a *primal–dual pair of solutions for* (y^*, ε), if

(i) $x_0 \in X$ is a solution to the problem $(P(y^*, \varepsilon))$, and

(ii) $f(x_0) \in D(y^*, \varepsilon)(R_0, S_0) \cap P(y^*, \varepsilon)\text{-max } \{ \cup \{D(y^*, \varepsilon)(R, S) : (R, S) \in L(Z, Y) \times$ $\times L(V, Y)\}\}$.

It is easy to see that (ii) implies that (R_0, S_0) is a solution to the problem $D(y^*, \varepsilon)$, and this has to be true for the element $f(x_0)$ (and perhaps also for others).

THEOREM 11. *Suppose that, in addition to the conditions given earlier, the cone* $K \subset Z$ *is algebraically closed and* core $(K) \neq \phi$. *Then we have*

(i) *If* $(x_0, R_0, S_0) \in \mathrm{dom}\ \varphi_L$ *is a* $P(y^*, \varepsilon)$*-saddle point of the Lagrangian* φ_L, *then it is a primal–dual pair of solutions for* $(y^*, 3\varepsilon)$.

(ii) *If* $(x_0, R_0, S_0) \in \mathrm{dom}\ \varphi_L$ *is a primal–dual pair of solutions for* (y^*, ε), *then it is a* $P(y^*, \varepsilon)$*-saddle point for* φ_L.

REFERENCES

Akilov, G.N. and S.S. Kutateladze (1978). *Ordered Vector Spaces.* Nauka, Novosibirsk (in Russian).

Borwein, J.M., J.P. Penot and M. Thera (1984). Conjugate convex operators. *Journal of Mathematical Analysis and Applications*, 102(2):399–414.

Demyanov, V.F. and L.V. Vasiliev (1981). *Nondifferentiable Optimization*. Nauka, Moscow (in Russian).

Corley, H.W. (1981). Duality theory with respect to cones. *Journal of Mathematical Analysis and Applications*, 84:560–568.

Hiriart-Urruty, J.B. (1982). ε-subdifferential calculus; convex analysis and optimization. *Research Notes in Mathematics*, Vol. 57, Pitman.

Holmes, R.B. (1975). *Geometric Functional Analysis.* Graduate Texts in Mathematics, Springer-Verlag, Berlin.

Köthe, G. (1966). *Topologische lineare Räume I.* Springer-Verlag, Berlin.

Kutateladze, S.S. (1978). Convex ε-programming (in Russian). *Dokladi Akademii Nauk SSSR*, Vol. 245

Loridan, P. (1982). Necessary conditions for ε-optimality. *Mathematical Programming Study*, 19:140–152.

Loridan, P. (1984). ε-solutions in vector minimization problems. *Journal of Optimization Theory*, 43(2):265–276.

Peressini, A.L. (1967). *Ordered Topological Vector Spaces.* Harper & Row, London.

Ritter, K. (1969, 1970). Optimization Theory in Linear Spaces, I, II, III, *Mathematische Annalen*, 182:189–206; 183:169–180; 184:133–154.

Strodiot, J.J., V.H. Nguyen and N. Heukemes (1983). ε-optimal solutions in nondifferentiable convex programming and some related questions. *Mathematical Programming*, 25:307–328.

Tanino, T. and Y. Sawaragi (1980). Duality theory in multiobjective programming. *Journal of Optimization Theory and Applications*, 31:509–529.

Tuy, H. (1972). Convex inequalities and the Hahn–Banach theorem. *Dissertationes Mathematicae*, 47.

Vályi, I. (1984). On duality theory related to approximate solutions of vector-valued optimization problems. Thesis, Eötvös Loránd University, Budapest, Hungary (in Hungarian).

Vályi, I. (1985). A general maximality principle and a fixed point theorem in uniform space. *Periodica Mathematica Hungarica*, 16(2).

Zowe, J. (1976). Konvexe Funktionen und Konvexe Dualitätstheorie in geordneten Vektorräumen. Habilitationsschrift dem Naturwissenschaftlichen Fachbereich IV der Bayerischen Julius-Maximilians-Universität Würzburg, Würzburg.

III. ALGORITHMS AND OPTIMIZATION METHODS

SEMINORMAL FUNCTIONS IN OPTIMIZATION THEORY

E.J. Balder
Mathematical Institute, University of Utrecht, 3508 TA Utrecht, The Netherlands
and Department of Statistics and Probability, Michigan State University,
East Lansing, MI 48824, USA

1. SEMINORMALITY OF FUNCTIONS

Let (X,d) be a metric space and let $(V,P,\langle\cdot,\cdot\rangle)$ be a pair of locally convex spaces, paired by a strict duality. A function $e\colon X \times V \to \overline{\mathbb{R}} \equiv [-\infty,+\infty]$ is defined to be *simple seminormal* (on $X \times V$) if there exist an l.s.c. (lower semicontinuous) function $f\colon X \to \overline{\mathbb{R}}$ and $p \in P$ with

$$e(x,v) = f(x) + \langle v,p\rangle.$$

A function $e\colon X \times V \to \overline{\mathbb{R}}$ is defined to be *seminormal* (on $X \times V$) if it is the pointwise supremum of a collection of simple seminormal functions on $X \times V$. In this way we extend a classical notion in the calculus of variations, due to Tonelli (1921), McShane (1934) and Cesari (1966). The *seminormal hull* $\tilde{\tilde{a}}$ of a function $a\colon X \times V \to \overline{\mathbb{R}}$ is defined to be the pointwise supremum of the collection of all (simple) seminormal functions e on $X \times V$ satisfying $e \le a$ (pointwise). We say that $a\colon X \times V \to \overline{\mathbb{R}}$ is *seminormal* at a point $(x,v) \in X \times V$ if $\tilde{\tilde{a}}(x,v) = a(x,v)$.

Example 1.1. Let $f\colon X \to \overline{\mathbb{R}}$ and $g\colon V \to \overline{\mathbb{R}}$ be given functions. Then for the functions $a_1, a_2\colon X \times V \to \overline{\mathbb{R}}$, defined by

$$a_1(x,v) \equiv f(x),\ a_2(x,v) \equiv g(v),$$

we have, denoting Fenchel conjugation in the usual way,

$$\tilde{\tilde{a}}_1(x,v) = \liminf_{y \to x} f(y),\ \tilde{\tilde{a}}_2(x,v) = g^{**}(v).$$

This shows that the seminormal hull concept straddles two important hull concepts in optimization theory.

Corresponding to a given function $a\colon X \times V \to \overline{\mathbb{R}}$ we define the function $b\colon X \times P \to \overline{\mathbb{R}}$ as follows:

$$b(x,p) \equiv a^*(x,p) \equiv \sup_{v \in V} [\langle v,p\rangle - a(x,v)].$$

Let $\bar{\bar{b}}\colon X \times P \to \overline{\mathbb{R}}$ be the u.s.c. hull of b with respect to the variable x; that is

$$\bar{\bar{b}}(x,p) \equiv \limsup_{y \to x} b(y,p).$$

It is easy to characterize the seminormal hull of a in terms of the function b (proofs of all statements to follow can be found in Balder (1983).

Proposition 1.2. The seminormal hull $\approx{a}$ of the function a is given by

$$\approx{a}(x,v) = \bar{\bar{b}}^{*}(x,v) \equiv \sup_{p\in P} [\langle v,p\rangle - \bar{\bar{b}}(x,p)]. \tag{1.1}$$

In optimal control theory the function b appears, under slightly modified circumstances, as the *Hamiltonian*, corresponding to a *Lagrangian* function a. A sufficient condition for seminormality, which can already be found in the work of Tonelli (1921), is as follows.

Theorem 1.3. If for $x \in X$ the following holds:

$$\text{a is sequentially l.s.c. at every point of } \{x\} \times V, \tag{1.2}$$

$$a(x,\cdot) \text{ is l.s.c. and convex on } V, \tag{1.3}$$

and if there exist a function $h: V \to (-\infty,+\infty]$ and $\delta > 0$ such that

h is inf-compact on V for every slope,

$a(y,v) \geq h(v)$ for every $y \in X$, $d(y,x) < \delta$, and every $v \in V$,

then

a is seminormal at every point of $\{x\} \times V$.

Roughly speaking, the above "superlinear" growth condition allows the interchange of monotone limit and Fenchel conjugation in (1.1). A more subtle result of this kind is given next, where we consider seminormality of a function on $X \times V \times \mathbb{R}$ with respect to the framework consisting of (X,d) and $(V \times \mathbb{R}, P \times \mathbb{R})$ paired by the duality

$$\langle\langle(v,\lambda), (p,q)\rangle\rangle \equiv \langle v,p\rangle + \lambda q.$$

This function is as follows. Let $h: V \to [0,+\infty]$ and $h': [0,+\infty) \to [0,+\infty]$ be given functions, and define the function $a_{1,\varepsilon}: X \times V \times \mathbb{R} \to \bar{\mathbb{R}}$, $\varepsilon > 0$, by

$$a_{1,\varepsilon}(x,v,\lambda) \equiv \max(a(x,v),\lambda) + \varepsilon h(v) + \varepsilon h'(\max(-\lambda,0)).$$

Theorem 1.4. If for $x \in X$ (1.2)-(1.3) hold and if

h is convex and inf-compact on V for every slope,

h' is nondecreasing, l.s.c. and convex on $[0,+\infty)$ with $\lim_{\gamma\to\infty} h'(\gamma)/\gamma = +\infty$,

then

$a_{1,\varepsilon}$ is seminormal at every point of $\{x\} \times V \times \mathbb{R}$.

2. SEMINORMALITY OF MULTIFUNCTIONS

Following Cesari (1966), we say that a multifunction $Q: X \rightrightarrows V$ (which may have empty values) has *property* (Q) *at* a point $x \in X$ if

$$Q(x) = \bigcap_{\delta>0} \text{cl co} \cup\{Q(y) : y \in X,\ d(y,x) < \delta\}. \tag{2.1}$$

Let χ_Q: $X \times V \to \{0,+\infty\}$ be the indicator function of Q. The next result - which is new - is a direct consequence of (1.1).

Theorem 2.1. For every $x \in X$ the following are equivalent:

Q has property (Q) at x,

χ_Q is seminormal at every point of $\{x\} \times V$

In fact, the proof of this result reveals that the seminormal hull of χ_Q is precisely the indicator function of the multifunction defined by the right-hand side of (2.1). Let us illustrate the usefulness of Theorem 2.1 by an example:

Example 2.2. Suppose that P is a Banach space. Let f: $P \to \mathbb{R}$ be a function which is locally Lipschitz near $x \in P$. Then the generalized gradient multifunction $\partial f(\cdot)$ in the sense of Clarke (1975), defined in a neighborhood N of x, has property (Q) at x. To see this, we take X = N, V = dual of P (with weak star topology), a = indicator function of $\partial f(\cdot)$, b = generalized directional derivative in the sense of Clarke (1975). The desired result then follows from Proposition 1.2 and Theorem 2.1, since b is u.s.c. on X in the variable y (by definition) and convex and continuous in the variable p (by the Lipschitz condition).

Our next result complements Theorem 2.1; in a more rudimentary form it can be found in Cesari (1970).

Theorem 2.3. For every $x \in X$ the following are equivalent:

a is seminormal at every point of $\{x\} \times V$,

the epigraphic multifunction Q_a: $X \overset{\rightarrow}{\rightarrow} V \times \mathbb{R}$ of a has property (Q) at x (here $Q_a(y) \equiv$ epigraph of $a(y,\cdot)$).

3. SEMINORMALITY OF INTEGRAL FUNCTIONALS

We suppose now in addition that X, V and P are Suslin spaces for their respective topologies. Let $(T,\mathcal{T},\mu)$ be an abstract σ-finite measure space. Let $(\mathcal{X},d)$ be the space of all $(\mathcal{T},\mathcal{B}(X))$-measurable functions from T into X, equipped with the essential supremum metric d, and let $(\mathcal{V},\mathcal{P},\langle\cdot,\cdot\rangle)$ be a pair of decomposable vector spaces of equivalence classes of scalarly μ-integrable functions going from T into V and P respectively, such that for every $v \in \mathcal{V}$, $p \in \mathcal{P}$ the integral in

$$\langle v,p\rangle \equiv \int_T \langle v(t),p(t)\rangle\ \mu(dt)$$

is well-defined and finite (cf. Castaing-Valadier (1977), Ch. VII for some details). Let ℓ: $T \times X \times V \to \overline{\mathbb{R}}$ be a given function. By outer integration we define the integral functional I_ℓ: $\mathcal{X} \times \mathcal{V} \to \overline{\mathbb{R}}$:

$$I_\ell(x,v) \equiv \tilde{\int}_T \ell(t,x(t),v(t))\ \mu(dt).$$

Seminormality of I_ℓ is defined with respect to the framework consisting of

(X,d) and $(V,P,<\cdot,\cdot>)$. The main result of Balder (1983) is as follows.

Theorem 3.1. If for $x \in X$ the following holds: there exist $p_0 \in P$, $\phi_0 \in P$, $\phi_0 \in L_1(T,T,\mu)$ and $\delta > 0$ such that for μ-a.e. $t \in T$

$$\ell(t,y,v) \geq <v,p_0(t)> + \phi_0(t) \text{ for every } y \in X,\ d(y,x(t)) < \delta, \text{ and every } v \in V, \tag{3.1}$$

$$\ell(t,\cdot,\cdot) \text{ is seminormal at every point of } \{x(t)\} \times V, \tag{3.2}$$

then

$$I_\ell \text{ is seminormal at every point of } \{x\} \times V. \tag{3.3}$$

Conversely, if (3.1) and (3.3) hold and if

ℓ is $T \times B(X \times V)$-measurable,

$I_\ell(x,\cdot)$ is not identically equal to $+\infty$ on V,

then (3.2) holds.

In the terminology of Balder (1983), Theorem 3.1 shows that seminormality *in the small* (3.2) and seminormality *in the large* (3.3) are equivalent under broad conditions. We can use this result to shed new light on the (sequential) lower semicontinuity properties of I_ℓ. First, in the spirit of Balder (1984), we define a subset V_0 of V to be *almost Nagumo tight* if there exist a sequence $\{B_i\}$ in T, monotonically decreasing to a μ-null set, and a sequence of $T \times B(V)$-measurable functions $h_i : T \times V \to [0,+\infty]$ such that for every $i \in \mathbb{N}$

$$\sup_{v \in V_0} \int_{T \setminus B_i} h_i(t,v(t))\, \mu(dt) < +\infty ,$$

and for μ-a.e. $t \in T$

$h_i(t,\cdot)$ is convex and inf-compact on V for every slope.

Examples of almost Nagumo tight subsets of V include weakly converging or merely uniformly L_1-bounded sequences in $L_1(T,T,\mu;V)$, in case V is a separable reflexive Banach space (cf. Brooks - Chacon (1980)).

We arrive at lower semicontinuity of I_ℓ via a stronger seminormality property of the integral functional $I_{\ell_1} : X \times V \times L_1(T,T,\mu) \to \overline{\mathbb{R}}$, defined by

$$I_{\ell_1}(x,v,\lambda) \equiv \tilde{\int}_T \max\, (\ell(t,x(t),v(t)),\lambda(t))\, \mu(dt);$$

here seminormality is defined with respect to the framework composed of (X,d) and $(V \times L_1(T,T,\mu), P \times L_\infty(T,T,\mu), <<\cdot,\cdot>>)$, where

$$<<(v,\lambda),(p,q)>> \equiv \int_T [<v(t),p(t)> + \lambda(t)q(t)]\, \mu(dt).$$

Theorem 3.2. If for $x \in X$, $V_0 \subset V$ and $L_0 \subset L_1(T,T,\mu)$ the following holds: for μ-a.e. $t \in T$

$\ell(t,\cdot,\cdot)$ is sequentially l.s.c. at every point of $\{x(t)\} \times V$,

$\ell(t,x(t),\cdot)$ is l.s.c. and convex on V,

and also

$\mathcal{V}_0$ is almost Nagumo tight,

$\{\max(-\lambda,0): \lambda \in \mathcal{L}_0\}$ is uniformly μ-integrable,

then there exists a function $J: \mathcal{X} \times \mathcal{V} \times L_1(T,\mathcal{T},\mu) \to \overline{\mathbb{R}}$ such that

J is seminormal at every point of $\{x\} \times \mathcal{V} \times L_1(T,\mathcal{T},\mu)$,

$J(y,v,\lambda) = I_{\ell_1}(y,v,\lambda)$ for every $y \in \mathcal{X}$, $v \in \mathcal{V}_0$, $\lambda \in \mathcal{L}_0$.

This *coincident seminormality* result follows from Theorem 1.4 and the implication (3.2)⇒(3.3) in Theorem 3.1; it immediately implies a well-known semicontinuity result for the integral functional I_ℓ. Conversely, using the implication (3.3) ⇒ (3.2) of Theorem 3.1, one can derive necessary conditions for such lower semicontinuity. We refer to Balder (1983) for details.

REFERENCES

Balder, E.J. (1983). On Seminormality of Integral Functionals and Their Integrands, Preprint No. 302, Mathematical Institute, Utrecht. To appear in SIAM J. Control Optim.

Balder, E.J. (1984). A general approach to lower semicontinuity and lower closure in optimal control theory. SIAM J. Control Optim. 22:570-598.

Brooks, J.K. and Chacon, R.V. (1980). Continuity and compactness of measures. Adv. Math. 37:16-26.

Castaing, C., and Valadier, M. (1977). Convex Analysis and Measurable Multifunctions. Springer-Verlag, Berlin.

Cesari, L. (1966). Existence theorems for weak and usual optimal solutions in Lagrange problems with unilateral constraints. I. Trans. Amer. Math. Soc. 124:369-412.

Cesari, L. (1970). Seminormality and upper semicontinuity in optimal control. J. Optim. Theory Appl. 6:114-137.

Cesari, L. (1983). Optimization-Theory and Applications. Springer-Verlag, Berlin.

Clarke, F.H. (1975). Generalized gradients and applications. Trans. Amer. Math. Soc. 205:247-262.

McShane, E.J. (1934). Existence theorems for ordinary problems of the calculus of variations. Ann. Scuola Norm. Sup. Pisa (2) 3:181-211, 287-315.

Tonelli, L. (1921). Fondamenti di Calcolo delle Variazioni. Zanichelli, Bologna.

THE GENERAL CONCEPT OF CONE APPROXIMATIONS IN NONDIFFERENTIABLE OPTIMIZATION

K.-H. Elster and J. Thierfelder
Technical University of Ilmenau, Am Ehrenberg, 6300 Ilmenau, GDR

1. INTRODUCTION

General optimization problems connected with necessary conditions for optimality have been studied by many authors in recent years. Since Clarke (1975) introduced the notion of a generalized gradient and the corresponding tangent cone, numerous papers havebeen published which extend standard smooth and convex optimization results to the general case.

In this paper we show how necessary optimality conditions may be constructed for local solutions of nonsmooth nonconvex optimization problems involving inequality constraints.

We shall use the approach developed by Dubovitskij and Miljutin (1965), which is closely connected with appropriate cone approximations of sets and differentiability concepts (to obtain multiplier conditions). Having studied the properties of numerous published cone approximations (see Thierfelder 1984), we propose a general definition of a local cone approximation K and introduce the corresponding K-directional derivative and K-subdifferential of a functional $f:X \rightarrow \overline{R}$. Using these notions it is possible to derive general multiplier conditions which turn out to be true generalizations of the Kuhn-Tucker theory for smooth and convex optimization problems.

2. LOCAL CONE APPROXIMATIONS

Let $[X,\tau]$ be a locally convex Hausdorff space and $[X^*,\sigma^*]$ be the topological dual space of X endowed with the weak $*$ (star) topology. We consider the problem

$$(P): f_o(x) \rightarrow \min,\ x \in S := \{\bar{x} \in X \mid f_i(\bar{x}) \leqq 0,\ i \in I := \{1,\dots,m\}\} \quad ,$$

where the $f_i: X \rightarrow \bar{R}$, $i \in \{0\} \cup I$, are extended real-valued functionals.

The definition of an abstract local cone approximation is fundamental to the following considerations since it can be used to replace an arbitrary set by a simple structured set. Moreover, the K-directional derivative leads to generalized differentiability for an extended real-valued functional.

Definition 2.1. The mapping $K: 2^X \times X \rightarrow 2^X$ is called a local cone approximation if a cone $K(M,x)$ is associated with each set $M \subset X$ and each point $x \in X$ such that

(i) $K(M-x,0) = K(M,x)$,

(ii) $K(M \cap U,x) = K(M,x)\ \forall\ U \in \mathcal{U}(x)$

(iii) $K(M,x) = X$ if $x \in \operatorname{int} M$,

(iv) $K(M,x) = \emptyset$ if $x \notin \bar{M}$,

(v) $K(\phi(M),\phi(x)) = \phi(K(M,x))$

(vi) $0^+M \subset 0^+K(M,x)$.

Here $\mathcal{U}(x)$ is the system of neighborhoods of x, $\phi: X \rightarrow X$ is any linear homomorphism, and the recession cone 0^+M of a set $M \subset X$ is defined by

$$0^+M := \{y \in X \mid M + ty \subset M \quad \forall t > 0\} \quad , \qquad 0^+\emptyset = X \quad .$$

Condition (i) represents the invariance of the cone approximation with respect to simultaneous translations of the set M and the point x. Without loss of generality it can be assumed that the vertices of the approximation cones are located at the origin.

Conditions (ii)-(iv) express local properties of the cone approximation. Hence, the cones are determined completely by the behavior of the set M on an (arbitrarily small) neighborhood of x. In particular $K(X,x) = X$ and $K(\emptyset,x) = \emptyset$ for each $x \in X$.

Condition (v) requires invariance of the cone approximation with respect to any linear homeomorphism (such as rotation and reflection).

Condition (vi) gives a relation between the recession cones of the set M and the cone K(M,x). This property is used to prove certain propositions concerning K-directional derivatives (cf. the proof of Theorem 3.1).

It can easily be shown that well-known cone approximations such as the cone of feasible directions, the cone of interior displacements, the cone of adherent displacements, Clarke's tangent cone and others (see Clarke, 1975; Dubovitskij and Milutin, 1965; Rockafellar, 1980; Thierfelder, 1984) satisfy conditions (i),..., (vi) above. The set of local cone approximations defined by Definition 2.1 is therefore nonempty.

Additional local cone approximations can be constructed using the following lemma:

Lemma 2.1. *Let* $K(.,.)$ *and* $K_i(.,.)$, $i = 1,\dots,\ell$ *be local cone approximations. Then*

$$\operatorname{int} K(.,.),\ \overline{K}(.,.),\ \operatorname{conv} K(.,.),\ X\setminus K(X\setminus .,.)\ ,$$

$$\bigcap_{i=1}^{\ell} K_i(.,.),\quad \bigcup_{i=1}^{\ell} K_i(.,.),\quad \sum_{i=1}^{\ell} K_i(.,.)$$

are also local cone approximations.

Proof.

1. Let $K(.,.)$ be a local cone approximation as specified in Definition 2.1. To prove that $\operatorname{int} K(.,.)$ is also a local cone approximation it suffices to prove (v) and (vi). Since ϕ is continuous we have on the one hand

$$\operatorname{int} K(\phi(M),\phi(x)) = \operatorname{int} \phi(K(M,x)) \subset \phi(\operatorname{int} K(M,x))\ ,$$

while on the other we conclude from the continuity of ϕ^{-1} that

$$\operatorname{int} K(M,x) = \operatorname{int} \phi^{-1}(K(\phi(M),\phi(x))) \subset \phi^{-1}(\operatorname{int}(K(\phi(M),\phi(x))))\ ,$$

and hence

$$\phi(\operatorname{int} K(M,x)) \subset \operatorname{int} K(\phi(M),\phi(x))\ .$$

Condition (vi) is true because

$$0^+K(M,x) \subset 0^+(\text{int } K(M,x)) \quad .$$

The propositions concerning $\overline{K}(.,.)$ and conv $K(.,.)$ can be proved in an analogous way.

2. To prove that $X\setminus K(X\setminus .,.)$ is a local cone approximation we consider only (vi). From $0^+(X\setminus M) = -0^+M$ we immediately obtain

$$0^+M = -0^+(X\setminus M) \subset -0^+K(X\setminus M,x) = 0^+(X\setminus K(X\setminus M,x)) \quad .$$

3. The proof of the other propositions is trivial. □

From Lemma 2.1 the set of all local cone approximations is algebraically closed with respect to set operations such as union, intersection and the sum of a finite family of cones, and taking the interior, the closed hull and the convex hull, and the double complement due to $X\setminus K(X\setminus M,x)$.

The algebraic structure of this set will not be considered here since the aim of the present paper is to demonstrate the usefulness of local cone approximations in deriving general optimality propositions for nonlinear optimization problems.

3. K-DIRECTIONAL DERIVATIVES AND K-SUBDIFFERENTIALS

Let $f:X\to\overline{R}$, $x\in X$, $|f(x)|<\infty$, and let $K:2^{X\times R}\times(X\times R)\to 2^{X\times R}$ be a local cone approximation as specified in Definition 2.1. Using the fact that traditional directional derivatives are positively homogeneous and that their epigraphs can be considered to be cone approximations of the epigraphs of the original functions, we introduce a general directional derivative of a functional f.

Definition 3.1. The mapping $f^K(x,.)$, $X\to\overline{R}$ with

$$f^K(x,y) := \inf\{\xi\in R \mid (y,\xi)\in K(\text{epi } f, (x,f(x)))\}$$

is called a K-directional derivative of f at x.

It is known from convex analysis that the subdifferential of a convex or a locally convex function at a point x is represented by the set of all linear continuous supporting functionals of the (one-sided) directional derivative

$$f'(x,y) := \lim_{t\downarrow 0} \frac{f(x+ty)-f(x)}{t}, \; y\in X$$

(see Ioffe and Tikhomirov, 1979). Using the K-directional derivative we introduce the K-subdifferential of a functional f.

Definition 3.2. The set

$$\partial_K f(x) := \{x^*\in X^* \mid x^*(y) \leqq f^K(x,y) \;\; \forall y\in X\}$$

is called the K-subdifferential of f at x, and the elements of $\partial_K f(x)$ are called K-subgradients of f at x.

If $f:X\to\overline{R}$ is convex and the cone of feasible directions

$$Z(M,x) := \{y\in X \mid \exists \lambda > 0 \;\; \forall t\in (0,\lambda) : x+ty\in M\}$$

is used for K(.,.), then we obtain

$$f^K(x,y) = \lim_{t\downarrow 0} \frac{f(x+ty)-f(x)}{t} = f'(x,y) \;\; \forall y\in X \quad ,$$

$$\partial_K f(x) = \{x^*\in X^* \mid x^*(y) \leqq f'(x,y) \;\; \forall y\in X\} \quad .$$

This example shows that the notions introduced above are proper generalizations of the corresponding notions from convex analysis.

Now we shall derive some basic propositions.

Theorem 3.1. *Let* $f:X\to\overline{R}$, $x\in X$, $|f(x)| < \infty$. *Then*

$$(1) \quad \operatorname{epi} f^K(x,.) = \{(y,\xi) \mid \forall\varepsilon > 0 \; \exists \overline{\xi}\in R : |\xi - \overline{\xi}| < \varepsilon \textit{ and } (y,\overline{\xi})\in K(\operatorname{epi} f, (x,f(x)))\} \quad ,$$

$$(2) \quad \partial_K f(x) = \{x^*\in X^* \mid (x^*,-1)\in K^*(\operatorname{epi} f, (x,f(x)))\} \quad .$$

Here the polar cone K^* of K is defined as

$$K^* := \{x^*\in X^* \mid x^*(y) \leqq 0 \;\; \forall y\in K\} \quad .$$

Proof.

1. Let $(y,\xi) \in \text{epi } f^K(x,.)$. Then

$$\inf \{\overline{\xi} \in R \mid (y,\overline{\xi}) \in K \;(\text{epi } f,(x,f(x)))\} \leqq \xi \quad .$$

Hence for each $\varepsilon > 0$ there exists a $\overline{\xi} < \xi + \varepsilon$ such that

$$(y,\overline{\xi}) \in K(\text{epi } f,\; (x,f(x))) \quad . \tag{3.1}$$

Since

$$(0,1) \in 0^+(\text{epi } f) \subset 0^+K(\text{epi } f,\; (x,f(x)))$$

from Definition 2.1 (vi), we deduce that for each $\varepsilon > 0$ there exists a $\overline{\xi} \in R$ which satisfies (3.1) and $|\overline{\xi} - \xi| < \varepsilon$. Thus one inclusion is true. The reverse inclusion is trivial.

2. Using the first proposition of this theorem we obtain

$$\begin{aligned}\partial_K f(x) &= \{x^* \in X^* \mid (x^*,-1)(y,f^K(x,y)) \leqq 0 \;\forall y \in X\} \\ &= \{x^* \in X^* \mid (x^*,-1)(y,\xi) \leqq 0 \;\forall (y,\xi) \in \text{epi } f^K(x,.)\} \\ &= \{x^* \in X^* \mid (x^*,-1) \in K^*(\text{epi } f,(x,f(x)))\} \;. \;\square\end{aligned}$$

From the second proposition of Theorem 3.1 we conclude that the K-subdifferential $\partial_K f(x)$ is convex and closed. Moreover, we have

$$\partial_K f(x) = \partial_{\overline{\text{conv}}\; K}\; f(x) \quad .$$

Hence, without loss of generality, we shall assume in the following that $K(.,.)$ is convex and closed, i.e., $f^K(x,.)$ is a l.s.c. convex functional.

Theorem 3.2. *Let* $f^K(x,.)$ *be convex and l.s.c. Then*

$$(1) \quad f^K(x,0) \geqq 0 \iff \partial_K f(x) \neq \emptyset \quad ,$$

$$(2) \quad f^K(x,0) = 0 \Rightarrow f^K(x,y) = \sup \{x^*(y) \mid x^* \in \partial_K f(x)\} \;\forall y \in X,$$

(3) $\exists\, U(0): f^K(x,y) \leqq 1 \;\; \forall y \in U(0) \;\Rightarrow\; \partial_K f(x)$ *compact and*

$$f^K(x,y) = \max\,\{x^*(y) \,|\, x^* \in \partial_K f(x)\} \;\; \forall y \in X.$$

Proof.

1. Let $f^K(x,0) \geqq 0$. In the case when $f^K(x,0) = +\infty$ we have $f^K(x,.) \equiv +\infty$ from the assumptions of the theorem. Hence $\partial_K f(x) = X^*$. In the case when $f^K(x,0) = 0$ the functions $f^K(x,.)$ and thus $(f^K(x,.))^*(.)$ are proper and, moreover,

$$\partial_K f(x) = \{x^* \in X^* \,|\, x^*(y) - f^K(x,y) < \infty \;\; \forall y \in X\}$$

$$= \operatorname{dom}\, (f^K(x,.))^* \neq \emptyset \quad ,$$

where $(f^K(x,.))^*$ is the Fenchel conjugate function of $f^K(x,.)$. If, conversely, $\partial_K f(x) \neq \emptyset$, then from $f^K(x,.) \geqq x^*(.)$, $x^* \in \partial_K f(x)$, we obtain $f^K(x,0) \geqq 0$.

2. Let $f^K(x,0) = 0$, i.e., $f^K(x,.)$ is proper. Now we have

$$(f^K(x,.))^*(x^*) = \sup\,\{x^*(y) - f^K(x,y) \,|\, y \in X\}$$

$$= \begin{cases} 0 \text{ if } x^* \in \partial_K f(x) & , \\ \infty \text{ otherwise} , & \end{cases}$$

$$= \chi_{\partial_K f(x)}(x^*)$$

and

$$(f^K(x,.))^{**}(y) = (\chi_{\partial_K f(x)})^*(y) = \sup\,\{x^*(y) \,|\, x^* \in \partial_K f(x)\}.$$

Use of the Fenchel-Moreau theorem leads to assertion (2) of the theorem.

3. If $A := \{y \in X \,|\, f^K(x,y) \leqq 1\} \supset U(0)$, then

$$\partial_K f(x) \subset \{x^* \in X^* \,|\, x^*(y) \leqq 1 \;\; \forall\, y \in A\}$$

$$\subset \{x^* \in X^* \,|\, x^*(y) \leqq 1 \;\; \forall\, y \in U(0)\} \quad .$$

From the Alaoglu-Bourbaki theorem and the fact that $\partial_K f(x)$ is closed, we deduce that $\partial_K f(x)$ is compact. Using proposition (2) we obtain assertion (3). □

We shall now look at the connection between the generalized directional derivatives $f^K(x,.)$ and $f^{\text{int } K}(x,.)$, and between the generalized subdifferentials $\partial_K f(x)$ and $\partial_{\text{int } K} f(x)$.

Theorem 3.3. *Let* $f^K(x,.)$ *be convex and l.s.c. Then the following properties hold:*

(1) *If* dom $f^{\text{int } K}(x,.) \neq \emptyset$ *then*

$$f^K(x,y) = \liminf_{\bar{y} \to y} f^{\text{int } K}(x,\bar{y}) \ \forall\, y \in X \quad ,$$

$$f^{\text{int } K}(x,y) = \limsup_{\bar{y} \to y} f^K(x,\bar{y}) \ \forall\, y \in X \quad ,$$

$$\partial_{\text{int } K} f(x) = \partial_K f(x) \quad .$$

(2) *If* $y \in \text{dom } f^{\text{int } K}(x,.)$ *then*

$$f^{\text{int } K}(x,y) = f^K(x,y) \quad .$$

Proof.

1. From the assumptions of the theorem we have int $K(\text{epi } f, (x,f(x))) \neq \emptyset$. Using

$$K(\text{epi } f,\ (x,f(x))) = \overline{\text{int } K}(\text{epi } f,\ (x,f(x)))$$

we obtain assertion (1) by Theorem 3.1.

2. If $y \in \text{dom } f^{\text{int } K}(x,.)$ then, since $f^{\text{int } K}(x,.)$ is u.s.c., there exists a neighborhood of y such that $f^{\text{int } K}(x,.)$ is bounded above on that neighborhood. Hence $f^{\text{int } K}(x,.)$ is continuous at y, and using assertion (1) we obtain assertion (2). □

To formulate multiplier conditions for problem (P) in terms of the K-subdifferentials and the K-directional derivatives of the functionals concerned, we need a relation linking linear combinations of the K-subdifferentials and the corresponding linear combinations of K-directional derivatives for a finite family of functionals. In convex analysis we have such a theorem (the Moreau-Rockafellar theorem) but this is not applicable here. As a first step in developing an appropriate theorem we set

$$\infty + \alpha = \infty\ , \quad 0 \cdot \alpha = 0 \quad \forall \alpha \in \overline{R}\ ,$$

$$\emptyset + M = \emptyset,\ 0 \cdot M = \{0\}\ \forall M \subset X \ \text{(or } M \subset X^*) \ .$$

Theorem 3.4.

(1) $\partial_K(\lambda f)(x) = \lambda \cdot \partial_K f(x) \quad \forall \lambda > 0$.

(2) $0 \in \sum_{i=1}^{m} \partial_{K_i} f_i(x) \Rightarrow \sum_{i=1}^{m} f_i^{K_i}(x,y) \geqq 0 \quad \forall y \in X$.

(3) *Let* $f_i^{K_i}(x,.)$, $i = 1,\dots,m$, *be convex. If there is a*

$y_1 \in \bigcap_{i=1}^{m} \operatorname{dom} f_i^{K_i}(x,.)$

such that all the $f_i^{K_i}$ *except one are continuous at* y_1, *then*

$$\sum_{i=1}^{m} f_i^{K_i}(x,y) \geqq 0 \quad \forall y \in X \Rightarrow 0 \in \sum_{i=1}^{m} \partial_{K_i} f_i(x) \ .$$

4. OPTIMALITY CONDITIONS

To prove necessary conditions for optimality in problem (P), we have to approximate sets which can be described in terms of the level sets of a finite number of extended real-valued functionals by cones. Since the cone approximations defined by Definition 2.1 are determined only by the geometrical form of the corresponding sets, we may introduce certain cones in the same way as in smooth (or convex) optimization, where the growth behavior of the functionals describing the sets is taken into account by the K-directional derivatives.

Let $f: X \to \overline{R}$, $f_i: X \to \overline{R}$, $i \in \Omega$, where Ω is a finite index set. We assume all functionals to be finite at the point $x \in X$.

Definition 4.1.

(1) The set

$$A_f^K(x) := \{y \in X \mid f^K(x,y) < 0\}$$

is called the *cone of descending directions* of f at x.

(2) The set

$$C_f^K(x) := \{y \in X \mid f^K(x,y) \leqq 0\}$$

is called the *linearizing cone* of f at x.

(3) $$A_\Omega^K(x) := \bigcap_{i\in\Omega} A_{f_i}^K(x)\,,\quad C_\Omega^K(x) := \bigcap_{i\in\Omega} C_{f_i}^K(x)$$

$$A_\emptyset^K(x) = C_\emptyset^K(x) = X \quad .$$

We shall now give inclusions relating these sets, making especial use of the cone int K(.,.).

Lemma 4.1. *Let* $f^K(x,.)$ *be convex and l.s.c., and let* $A_f^{\text{int } K}(x) \neq \emptyset$. *Then*

(1) $$\text{int } C_f^{\text{int } K}(x) = A_f^{\text{int } K}(x) \subset C_f^{\text{int } K}(x) \subset \overline{A_f^{\text{int } K}(x)} \quad ,$$

$$\text{int } C_f^K(x) \subset A_f^K(x) \subset C_f^K(x) = \overline{A_f^K(x)} \quad ,$$

$$\text{int } C_f^{\text{int } K}(x) = \text{int } C_f^K(x)\,,\ A_f^{\text{int } K}(x) \subset A_f^K(x) \quad ,$$

$$C_f^{\text{int } K}(x) \subset C_f^K(x)\,,\ \overline{A_f^{\text{int } K}(x)} = \overline{A_f^K(x)} \quad .$$

(2) *All of the above sets are convex cones.*

We also introduce a cone for which a dual relation holds with respect to the linearizing cone.

Definition 4.2. The set

$$B_\Omega^K(x) := \{x^* \in X^* \mid x^* \in \sum_{i\in\Omega} \lambda_i \partial_K f_i(x)\,,\ \lambda_i \geqq 0,\ i \in \Omega\}$$

is called the *cone of K-subgradients* of the functionals f_i,

$i \in \Omega$, at x.
Using the polar cone of B_{Ω}^{K} we obtain

$$(B_{\Omega}^{K}(x))^{*} \subset C_{\Omega}^{K}(x) \quad .$$

This inclusion can be sharpened by assumptions concerning the convexity and closure of the cone K.

Lemma 4.2. *Let* $f_i^K(x,.)$ *be convex and l.s.c., and let* $f_i^K(x,0) = 0 \quad \forall i \in \Omega$. *Then*

$$(1) \quad (C_{\Omega}^{K}(x))^{*} = \overline{B_{\Omega}^{K}(x)} \quad .$$

$$(2) \quad (B_{\Omega}^{K}(x))^{*} = C_{\Omega}^{K}(x) \quad .$$

We shall assume that the following conditions are satisfied at $x \in X$:

$$|f_i(x)| < \infty \quad \forall i \in \{0\} \cup I \quad , \tag{4.1}$$

$$f_i \text{ is u.s.c. at } x \quad \forall i \in I \setminus I(x) \quad , \tag{4.2}$$

where $I(x) := \{i \in I | f_i(x) = 0\}$ is the index set of active constraints at x. Condition (4.1) ensures the K-directional differentiability of f at x and (4.2) implies that only active constraints have to be taken into account in the local description of the feasible set S.

Lemma 4.3. *If* $x \in S$ *is a local solution of* (P), *then there exists a neighborhood* V(x) *such that*

$$N(f_o,x) \cap (S \cap V(x)) = \emptyset \quad ,$$

where

$$N(f_o,x) := \{\overline{x} \in X | f_o(\overline{x}) < f_o(x)\} \quad .$$

In the approximation of the sets $N(f_o,x)$ and $S \cap V(x)$ we use the classical tangent cone

$$T(M,x) := \{y \in X \mid \forall U(y)\ \forall \lambda > 0\ \exists\, t \in (0,\lambda)\ \exists\, \bar{y} \in U(y) : x + t\bar{y} \in M\}$$

and the cone of interior displacements

$$D(M,x) := \{y \in X \mid \exists\, U(y)\ \exists\, \lambda > 0\ \ \forall t \in (0,\lambda)\ \forall \bar{y} \in U(y) : x + t\bar{y} \in M\} \quad .$$

Theorem 4.4. *Let* $x \in S$ *be a local solution of* (P). *Then*

(1) $A^D_{f_o}(x) \cap T(S,x) = \emptyset$,

(2) $A^T_{f_o}(x) \cap D(S,x) = \emptyset$.

Proof. We shall only prove (1), since (2) may be proved in an analogous way. Assume $y \in A^D_{f_o}(x) \cap T(S,x)$. Then on the one hand we have

$$\forall U(y)\ \forall\ \lambda > 0\ \exists\, t \in (0,\lambda)\, \exists\, \bar{y} \in U(y) : x + t\bar{y} \in S \quad , \qquad (4.3)$$

while on the other, by Theorem 3.1, there is a real $\xi < 0$ such that $(y,\xi) \in D(\text{epi } f_o, (x,f_o(x)))$, i.e.,

$$\exists\, U_1(y)\ \exists\, \lambda_1 > 0\ \forall t \in (0,\lambda_1)\ \forall \bar{y} \in U_1(y) :\ (x,f_o(x)) + t(\bar{y},\xi) \in \text{epi } f_o$$

and hence

$$f_o(x + t\bar{y}) \leqq f_o(x) + t\,\xi < f_o(x) \quad . \qquad (4.4)$$

From (4.3) and (4.4) we conclude that for each neighborhood $V(x)$ of x there exists a point $\bar{x} := x + ty \in S \cap V(x)$ such that $f_o(\bar{x}) < f_o(x)$ and thus $\bar{x} \in N(f_o,x)$. Then by Lemma 4.3 x is not a local solution of (P). □

A disadvantage of the optimality conditions given in Theorem 4.4 is that the cones which occur are in general not convex and hence the assumptions regarding their separability are not satisfied.

We therefore assume that the cone approximations have the following additional properties:

(V1) $K(.,.)$ convex and closed,

(V2) $x \in \bar{M} \Leftrightarrow 0 \in K(M,x)$,

(V3) $K(.,.) \subset T(.,.)$,

(V4) $\text{int } K(.,.) \subset D(.,.)$,

(V5) $A^{K}_{I(x)}(x) \subset K(S,x)$.

Now Theorem 4.4 leads immediately to the following result:

Theorem 4.5. *Let* $x \in S$ *be a local solution of* (P). *Then*

$$(1) \quad A^{\text{int } K}_{f_o}(x) \cap K(S,x) = \emptyset \quad ,$$

$$(2) \quad A^{K}_{f_o}(x) \cap \text{int } K(S,x) = \emptyset \quad .$$

Since the cones under consideration are convex, we can formulate an optimality condition in the dual space X^*.

Theorem 4.6. *Let* $x \in S$ *be a local solution of* (P). *If one of the two conditions*

$$(\text{B1}) \quad \text{dom } f_o^{\text{int } K}(x,.) \cap K(S,x) \neq \emptyset \quad ,$$

$$(\text{B2}) \quad \text{dom } f_o^{K}(x,.) \cap \text{int } K(S,x) \neq \emptyset$$

is satisfied, then $0 \in \partial_K f_o(x) + K^*(S,x)$.

Proof.

1. Let (B1) be satisfied. Then from Theorem 4.5 (1) we have

$$f_o^{\text{int } K}(x,y) \geqq 0 \geqq \forall y \in K(S,x) \quad . \qquad (4.5)$$

Obviously $\partial_K f_o(x) \neq \emptyset$, since otherwise by Theorem 3.2 we would have $f_o^K(x,0) = -\infty$ and hence, using the lower-semicontinuity property, $f_o^K(x,0) \equiv \pm\infty$. (Here $f^K(x,.) \equiv \pm\infty$ means that $f^K(x,.)$ has no finite values.) It follows from Theorem 3.2 that $f_o^{\text{int } K} \equiv \pm\infty$ and hence by (4.5) we obtain

$$f_o^{\text{int } K}(x,y) = +\infty \;\forall y \in K(S,x) \quad ,$$

in contradiction to (B1).

Now we construct a set $M \subset X \times R$ defined as follows:

$$M := (K(S,x) \times R) \cap \operatorname{epi} f_o^{\operatorname{int} K}(x,.) \quad .$$

From (B1) we have $M \neq \emptyset$ and by Theorem 3.1

$$M^* = (K^*(S,x) \times \{0\}) + K^*(\operatorname{epi} f_o, (x, f_o(x))) \quad . \qquad (4.6)$$

From (4.5) we obtain, for $(0,-1) \in X^* \times R$, that

$$(0,-1)(y,\xi) = -\xi \leqq 0 \quad \forall (y,\xi) \in M \quad ,$$

i.e., $(0,-1) \in M^*$.

Making use of (4.6), we can deduce the existence of an $x^* \in K^*(S,x)$ such that

$$(-x^*,-1) \in K^*(\operatorname{epi} f_o, (x, f_o(x))) \quad ,$$

i.e., $-x^* \in \partial_K f_o(x)$. This proves the assertion of the theorem under assumption (B1).

An analogous proof can be developed taking (B2) instead of (B1). □

Remark. Theorem 4.6 is stated for certain special cases (K is Clarke's tangent cone; int K is the cone of epi Lipschitzian directions) in Hiriart-Urruty (1979) and Rockafellar (1981).

Assuming an appropriate regularity condition

(RB1) $(K^*(S,x) \subset B^K_{I(x)}(x))$ and (B1) or (B2),

we can deduce the existence of an optimality condition of the Kuhn-Tucker type.

Theorem 4.7. *Let* $x \in S$ *be a local solution of* (P). *If* (RB1) *is satisfied, then there exist multipliers* $\lambda_i \geqq 0$, $i \in I(x)$, *such that*

$$0 \in \partial_K f_o(x) + \sum_{i \in I(x)} \lambda_i \partial_K f_i(x) \quad ,$$

$$f_o^K(x,y) + \sum_{i \in I(x)} \lambda_i f_i^K(x,y) \geqq 0 \quad \forall y \in X \quad .$$

The proof follows immediately from Theorem 4.6 and Theorem 3.4.

We shall now give some other regularity conditions which are also sufficient for (RB1). Writing

$$(R)\quad \begin{cases} (B1) \text{ or } (B2) \text{ is satisfied} , \\ \partial_K f_i(x) \neq \emptyset \quad \forall\, i \in I(x) , \\ B^K_{I(x)}(x) = \overline{B^K_{I(x)}(x)} \end{cases}$$

we formulate the regularity conditions

$$\begin{array}{lll} (RB2) & K^*(S,x) \subset (C^K_{I(x)}(x))^*, & (R) \text{ is satisfied,} \\ (RB3) & C^K_{I(x)}(x) \subset K(S,x) , & (R) \text{ is satisfied,} \\ (RB4) & A^{\operatorname{int} K}_{I(x)}(x) \neq \emptyset , & (R) \text{ is satisfied,} \\ (RB5) & \operatorname{dom} f^K_o(x,.) \cap A^{\operatorname{int} K}_{I(x)}(x) \neq \emptyset , & \\ & \partial_K f_i(x) \neq \emptyset \quad \forall\, i \in I(x) , B^K_{I(x)}(x) = \overline{B^K_{I(x)}(x)} . & \end{array}$$

Note that

(RB2) is a generalized Gould-Tolle condition (see Gould and Tolle, 1971)

(RB3) is a generalized Abadie condition (see Abadie, 1967)

(RB4) and (RB5) are generalized Slater conditions.

The proof of the following theorem is given in Elster and Thierfelder (1985).

Theorem 4.8.

$$(RB5) \Rightarrow (RB4) \Rightarrow (RB3) \Leftrightarrow (RB2) \Rightarrow (RB1) .$$

Using Theorem 4.5 and assumption (V5) we obtain optimality conditions of the Kuhn-Tucker type.

Theorem 4.9. *Let* $x \in S$ *be a local solution of* (P). *Then*

$$(1)\quad A^{\operatorname{int} K}_{f_o}(x) \cap A^K_{I(x)}(x) = \emptyset ,$$

$$(2) \quad A^K_{f_o}(x) \cap A^{\text{int } K}_{I(x)}(x) = \emptyset \quad .$$

We can now deduce a proposition of the John type.

Theorem 4.10. *Let* $x \in S$ *be a local solution of* (P). *Then*

(1) *There exist multipliers*

$\lambda_i \geqq 0$, $i \in \{0\} \cup I(x)$, *not all of which vanish, such that*

$$\lambda_o f_o^{\text{int } K}(x,y) + \sum_{i \in I(x)} \lambda_i f_i^K(x,y) \geqq 0$$

$$\forall\, y \in \operatorname{dom} f_o^{\text{int } K}(x,.) \cap \Big(\bigcap_{i \in I(x)} \operatorname{dom} f_i^K(x,.)\Big) \quad .$$

(2) *There exist multipliers*

$\lambda_i \geqq 0$, $i \in \{0\} \cup I(x)$, *not all of which vanish, such that*

$$\lambda_o f_o^K(x,y) + \sum_{i \in I(x)} \lambda_i f_i^{\text{int } K}(x,y) \geqq 0$$

$$\forall\, y \in \operatorname{dom} f_o^K(x,.) \cap \Big(\bigcap_{i \in I(x)} \operatorname{dom} f_i^{\text{int } K}(x,.)\Big) \quad .$$

An optimality condition can be derived using the condition

(B3) $\exists\, i_o \in \{0\} \cup I(x)$:

$$\operatorname{dom} f_{i_o}^K(x,.) \cap \Big(\bigcap_{i \in \{0\} \cup I(x) \setminus \{i_o\}} \operatorname{dom} f_i^{\text{int } K}(x,.)\Big) = X \quad .$$

Theorem 4.11. *Let* $x \in S$ *be a local solution of* (P). *If* (B3) *is satisfied, then there exist multipliers* $\lambda_i \geqq 0$, $i \in \{0\} \cup I(x)$, *not all of which vanish, such that*

$$(1) \quad 0 \in \sum_{i \in \{0\} \cup I(x)} \lambda_i \partial_K f_i(x) \quad ,$$

$$(2) \quad \sum_{i \in \{0\} \cup I(x)} \lambda_i f_i^K(x,y) \geqq 0 \quad \forall\, y \in X \quad .$$

Proof. Let $i_o \neq 0$. By the first assertion of Theorem 4.10 there exist multipliers $\lambda_i \geqq 0$, $i \in \{0\} \cup I(x)$, not all of which vanish, such that

$$\lambda_o f_o^{\text{int } K}(x,y) + \sum_{i\in I(x)} \lambda_i f_i^K(x,y) \geqq 0 \quad \forall\, y \in X \quad .$$

Since (B3) is satisfied, assumption (3) of Theorem 3.4 is satisfied and assertion (1) follows from Theorem 3.1.

Assertion (1) and Theorem 3.4 (2) immediately lead to assertion (2).

If $i_o = 0$ then the assertion can be proved in an analogous way using Theorem 4.10 (2). □

If the regularity condition

(RB6) $(A^K_{I(x)}(x) \neq \emptyset)$ and (B3)

is satisfied, then we obtain an optimality condition of the Kuhn-Tucker type from Theorem 4.11.

Theorem 4.12. *Let* $x \in S$ *be a local solution of* (P). *If* (RB6) *is satisfied then* $\lambda_o \neq 0$ *in Theorem* 4.11.

Proof. Let us assume that $\lambda_o = 0$. Then it follows that

$$\sum_{i\in I(x)} \lambda_i f_i^K(x,y) \geqq 0 \quad \forall\, y \in X \quad ,$$

where the multipliers $\lambda_i \geqq 0$, $i \in I(x)$, do not all vanish. This contradicts $A^K_{I(x)}(x) \neq \emptyset$ and thus (RB6). □

5. CONCLUDING REMARKS

In this paper we give certain optimality conditions which are true generalizations of well-known results derived for smooth, convex and Lipschitzian optimization problems. We obtain the same results if concrete cone approximations are used.

Let (P) be a convex optimization problem: we assume that the functionals $f_i, i \in \{0\} \cup I$, are convex and continuous at the point $x \in S$.

If $K(.,.)$ is the classical tangent cone $T(.,.)$ and if int $K(.,.)$ is replaced by the cone of interior displacements $D(.,.)$, then we can prove

$$f_i^T(x,y) = f_i^D(x,y) = \lim_{t\downarrow 0} \frac{f_i(x+ty) - f_i(x)}{t} = f_i'(x,y) \quad \forall\, y \in X$$

and hence

$$\partial_T f_i(x) = \partial_D f_i(x) = \partial f_i(x) \quad ,$$

where the $\partial f_i(x)$ are subdifferentials of the type used in convex analysis.
[Note: if $M \subset X$ is convex then $D(M,x)$ is an open convex cone, $T(M,x)$ is a closed convex cone and $D(M,x) \subset \text{int}\ T(M,x)$. In the case $D(M,x) \neq \emptyset$ the equality holds.]

It is clear that (B1) is always satisfied due to $0 \in T(S,x)$ and dom $f_o'(x,.) = X$. Since all the functionals are subdifferentiable the regularity conditions take the following form:

(RB1') $$T^*(S,x) \subset B_{I(x)}(x) := \{x^* \in X^* \,|\, x^* \in \sum_{i \in I(x)} \lambda_i \partial f_i(x) \quad , \quad \lambda_i \geqq 0,\ i \in I(x)\} \quad .$$

(RB2') $$T^*(S,x) \subset \{y \in X \,|\, f_i'(x,y) \leqq 0 \quad \forall\, i \in I(x)\}^*$$

and $B_{I(x)}(x)$ is closed.

(RB3') $$T(S,x) \supset \{y \in X \,|\, f_i'(x,y) \leqq 0 \quad \forall\, i \in I(x)\}$$

and $B_{I(x)}(x)$ is closed.

In the special case when

$$A^D_{I(x)}(x) = A^T_{I(x)}(x) = \{y \,|\, f_i'(x,y) < 0 \quad \forall\, i \in I(x)\} \neq \emptyset \ ,$$

we have (see Lemma 4.1)

$$\begin{aligned}(C_{I(x)}(x))^* &:= (\bigcap_{i \in I(x)} y \in X \,|\, f_i'(x,y) \leqq 0\})^* \\ &= (\bigcap_{i \in I(x)} (\partial f_i(x))^*)^* \\ &= \sum_{i \in I(x)} (\partial f_i(x))^{**} \\ &= \sum_{i \in I(x)} \text{cone}\ \partial f_i(x) \quad = B_{I(x)}(x) \quad .\end{aligned}$$

(Note that $0 \in \partial f_i(x) \quad \forall i \in I(x)$ from our assumption of convexity and the sets cone $(\partial f_i(x))$ are closed due to the compactness of the subdifferentials.) Hence $B_{I(x)}(x)$ is closed.

Since dom $f_i'(x,.) = X$ holds for all $i \in \{0\} \cup I(x)$, condition (B3) is satisfied. Moreover, the regularity conditions (RB4), (RB5) and (RB6) take the form of the well-known Slater condition

$$\exists y \in X: \; f_i'(x,y) < 0 \quad \forall i \in I(x) \quad .$$

Then Theorems 4.10 and 4.11,and Theorems 4.7 and 4.12, are the theorems given by John and Kuhn and Tucker, respectively.

Similar results can be obtained in the smooth case and, furthermore, in the Lipschitzian case if Clarke's tangent cone is used for K(.,.).

REFERENCES

Abadie, J. (1967). On the Kuhn/Tucker theorem. In: J. Abadie (ed.): *Nonlinear Programming*. North-Holland, Amsterdam, pp. 17-36.

Clarke, F. (1975). Generalized gradients and applications. *Trans. Americ. Math. Soc.*, 205:247-262.

Clarke, F. (1983). *Optimization and Nonsmooth Analysis*. John Wiley, New York.

Dubovitskij, A.J. and Miljutin, A.A. (1965). Extremum problems under constraints (in Russian). *Vychisl. Mat. i Mat. Fiz.*, 5: 395-453.

Elster, K.-H. and Thierfelder, J. Abstract cone approximations and generalized differentiability in nonsmooth optimization. (Forthcoming).

Gould, F. and Tolle, W. (1971). Geometry of optimality conditions and constraint qualifications. *Math. Programming*, 2(1): 1-18.

Hiriart-Urruty, J.B. (1979). Tangent cones, generalized gradients and mathematical programming in Banach spaces. *Math. Op. Res.*, 4(1): 79-97.

Ioffe, A.D. and Tikhomirov, V.M. (1979). *Theorie der Extremalaufgaben*. DVW Berlin.

Rockafellar, R.T. (1980). Generalized directional derivatives and subgradients of nonconvex functions. *Canad. J. Math.*, 32 (2): 257-280.

Rockafellar, R.T. (1981). *The Theory of Subgradients and its Applications in Problems of Optimization.* Heldermann-Verlag, Berlin.

Thierfelder, J. (1984). *Beiträge zur Theorie der nichtglatten Optimierung.* Diss. A, TH Ilmenau.

AN ALGORITHM FOR CONVEX NDO BASED ON PROPERTIES OF THE CONTOUR LINES OF CONVEX QUADRATIC FUNCTIONS

Manlio Gaudioso
CRAI, Via Bernini 5, 87036 Quattromiglia de Rende, Italy

1. INTRODUCTION

The objective of the paper is to suggest a model algorithm for the unconstrained minimization of a Lipschitz convex function of several variables, not necessarily differentiable.

The proposed algorithm stems from a property of the contour lines of the convex quadratic differentiable functions which allows us to represent the ordinary Newton's direction in terms of information about the gradient and the objective function values.

This idea is extended to the nondifferentiable case by means of some recent results on the approximate (or perturbed) first order directional derivatives (Hiriart-Urruty 1982, Lemarechal and Zowe 1983).

Nevertheless, in order to attain to an implementable method, a number of simplifying assumptions are to be introduced. Consequently the resulting numerical algorithm can be considered as belonging to the family of the well known bundle methods (Lemarechal 1977, Lemarechal Strodiot and Bihain 1981, Gaudioso and Monaco 1982).

In section 2 the basic ideas underlying the approach are presented and in section 3 a model algorithm, together with its convergence properties, is outlined.

2. THE APPROACH

The following proposition provides a simple characterization of Newton's direction for convex quadratic functions.

Proposition 1. Given a convex quadratic function $f:R^n \longrightarrow R$, any point $x \in R^n$ and the gradient $g \triangleq \nabla f(x)$, the solution d^* of the problem

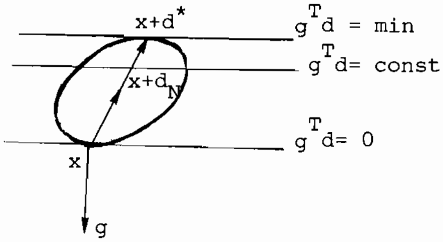
$$\min_d \ g^T d \tag{1}$$

$$\text{s.t. } f(x+d) = f(x)$$

is a scalar multiple of Newton's direction d_N at the point x(in fact $d^* = 2d_N$)

Proof. Straightforward application of first order optimality conditions.

A pictorial representation of the proposition is given in fig. 1



In order to explore the potential use of the property in the framework of convex non smooth optimization the following proposition, proved by Hiriart-Urruty (1982) and Lemarechal and Zowe (1983) is particularly helpful:

Proposition 2. Given $f:R^n \longrightarrow R$, f Lipschitz and convex, then, for any $x \in R^n$, $d \in R^n$, the following holds:

$$f(x+d) = f(x) + \max_{\varepsilon \geq 0} \{f'_\varepsilon(x,d) - \varepsilon\}$$

where $f'_\varepsilon(x,d)$ is the approximate (or perturbed) directional derivative of f at the point x along the direction d and is defined as

$$f'_\varepsilon(x,d) = \inf_{t>0} \frac{f(x+td) - f(x) + \varepsilon}{t}$$

It is important to note that $f'_\varepsilon(x,d)$ is the support function of $\partial_\varepsilon f(x)$, the ε-subdifferential of f at the point x, i.e.

$$f'_\varepsilon(x,d) = \max_{v \in \partial_\varepsilon f(x)} v^T d$$

On the basis of proposition 2, problem (1) may be formally rewritten as

$$\begin{aligned} &\min_d f'(x,d) \\ \text{s.t.}\ &\max_{\varepsilon \geq 0} \left[f'_\varepsilon(x,d) - \varepsilon\right] = 0 \end{aligned} \qquad (2)$$

As a result of this reformulation, problem (2) appears suitable, at least theoretically, fordefining a direction finding step in an algorithmic context for convex non smooth optimization. On the other hand it provides Newton's direction if applied to a quadratic function.

Nevertheless, in order to devise an implementable algorithm, modifications are to be introduced in the definition of the problem (in fact it requires complete information about the ε-subdifferentials).

In this aim, consider the point x, $g \in \partial f(x)$ and a bundle of points and subgradients $x^{(i)}$, $g^{(i)} \in \partial f(x^{(i)})$, $i \in I$ (x may be the current estimate of the minimum and the $x^{(i)}$'s are points previously obtained in some iterative descent process).

It is well known that

$$g^{(i)} \in \partial_{\alpha_i} f(x) \qquad \text{where}$$

$$\alpha_i \triangleq f(x) - \left[f(x^{(i)}) + g^{(i)T}(x-x^{(i)}) \right] \geq 0$$

It is easy to show that the following problem

$$\begin{aligned} &\min_{d} \; f'(x,d) \\ \text{s.t.} \quad & g^{(i)T} d - \alpha_i \leq 0 \qquad \forall i \in I \end{aligned} \tag{3}$$

is a relaxation, in the usual sense, of problem (2). (In fact every d feasible for problem (2) is also feasible for problem (3) as consequence of the property of the approximate (perturbed) directional derivative of being the support function of the ε-subdifferential).

Moreover, taking in consideration the properties of the ordinary derivative, problem (3) may be further modified:

$$\begin{aligned} &\min_{d,v} \; v \\ \text{s.t.} \quad & g^{(i)T} d \leq v \qquad i \in C \\ & g^{(i)T} d - \alpha_i \leq 0 \qquad i \in F \end{aligned} \tag{4}$$

where C and F are respectively the set of indices of subgradients related to points "close" to x and "far" from x in the sense that will be defined later. Obviously C U F = I.

Bounded solution of problem (4) requires dual feasibility, which implies the existence of multipliers λ_i, $i \in C$ and μ_i, $i \in F$ such that:

$$\sum_{i \in C} \lambda_i g^{(i)} + \sum_{i \in F} \mu_i g^{(i)} = 0$$

$$\sum_{i \in C} \lambda_i = 1$$

$$\lambda_i \geq 0, \quad \mu_i \geq 0$$

Therefore, as usual in bundle methods, some limitation on the variable d needs to be introduced. A possible way is the following:

$$\begin{aligned} &\min_{d,v} \; v + \frac{1}{2} d^T d \\ \text{s.t.} \quad & g^{(i)T} d \leq v \qquad i \in C \\ & g^{(i)T} d - \alpha_i \leq 0 \qquad i \in F \end{aligned} \tag{P}$$

The dual (D) of problem (P) is obtained as

$$\min \left\| \sum_{i \in C} \lambda_i g^{(i)} + \sum_{i \in F} \mu_i g^{(i)} \right\|_2^2 + \sum_{i \in F} \mu_i \alpha_i$$

$$\sum_{i \in C} \lambda_i = 1 \qquad \text{(D)}$$

$$\lambda_i \geq 0 \qquad i \in C$$

$$\mu_i \geq 0 \qquad i \in F$$

3. THE ALGORITHM

Before the description of the possible use of the solution of problem (P) (or, equivalently, (D)) in the direction finding step of an algorithm for the minimization of convex non smooth functions, some properties of the optimal solutions of (P) and (D) are listed.

Primal and dual optimal solutions are related in the following way:

$$d^* = -\left(\sum_{i \in C} \lambda_i^* g^{(i)} + \sum_{\mu_i \in F} \mu_i^* g^{(i)} \right)$$

$$v^* = - \| d^* \|_2^2 - \sum_{i \in F} \mu_i^* \alpha_i$$

Note also that v^* is non positive and $v^*=0$ implies that $\| \sum_{i \in C} \lambda_i^* g^{(i)} \| = 0$, i.e. that some approximate optimality condition is satisfied.

Moreover the following proposition can be easily proved

Proposition 4. If $v^* \geq -\eta$, η being any positive number, then

$$\| \sum_{i \in C} \lambda_i^* g^{(i)} \| \leq \sqrt{\eta} \left(1 + \frac{k\sqrt{\eta}}{\alpha_{\min}} \right)$$

where k is the upper bound on the norm of the subgradient and $\alpha_{\min}$ is defined as $\min_{i \in F} \alpha_i$

The properties of the solution of the problems defined above are useful in order to define the direction finding step in a descent algorithm for the minimization of a Lipschitz convex function $f: R^n \longrightarrow R$ which in addition is supposed to be not unbounded from below.

One iteration of the algorithm is summarized by the following steps, where x is assumed to be the current estimate of the minimum, $g \in \partial f(x)$ and a bundle (eventually empty) of subgradients $g^{(i)}$, $i \in F$ is available, together with the corresponding scalars α_i defined in the previous section.

The positive parameters $\bar{t}$, m_1 and m_2 are given, $0 < m_2 < m_1 < 1$; initially $C= \{1\}$ and the subgradient g is conventionally indicated by $g^{(1)}$.

STEP 1. Solve the quadratic programming problem

$$\min_{v,d} \; v + \frac{1}{2} d^T d$$

$$\text{s.t.} \quad g^{(i)^T} d \le v \qquad i \in C$$

$$g^{(i)^T} d - \alpha_i \le 0 \qquad i \in F$$

and obtain d^* and v^*.
Perform a termination test on the value of v^*.

<u>STEP 2.</u> Line search. Perform a line search along d^* finding $t > 0$ and $g^+ \in \partial f(x + td^*)$ such that

$$g^{+^T} d^* \ge m_1 v^*$$

and either

$$\text{a)} \quad f(x + td^*) - f(x) \le m_2 t v^*$$

or

$$\text{b)} \quad t \; \|d^*\|_2 \le \bar{t}$$

In case a) move to the new point $x^+ = x + td^*$, update the set F, and iterate.
In case b) consider the point x^+ as a point "close" to x, update the set C, create accordingly the new quadratic programming problem and return to Step 1.

The following propositions hold; they are similar to propositions holding in classical bundle methods.

<u>Proposition 5.</u> After a finite number of "serious steps" (case a) of the line search) the quantity $- v^*$ is reduced below any positive fixed value, provided that f is not unbounded from below.

<u>Proof.</u> Suppose that $\{x^{(k)}\}$ is the sequence of points obtained as results of successfully line searches, correspondent to the sequences $\{v_k^*\}$ and $\{t_k\}$.
For any integer n the following holds

$$f(x^{(x+1)}) - f(x^{(0)}) \le m_2 \sum_{k=0}^{n} t_k v_k^* \le m_2 \bar{t} \sum_{k=0}^{n} \frac{v_k^*}{\|d_k^*\|} \le 0$$

Since f is bounded from below, it follows that

$$\left\{ \frac{v_k^+}{\|d_k^*\|} \right\} \longrightarrow 0 \text{ which in turn implies that } \frac{|v_k^*|^2}{\|d_k^*\|^2} \longrightarrow 0, \text{ but}$$

$$|v_k^*| \ge \|d_k^*\|^2 \text{ hence } v_k^* \longrightarrow 0.$$

<u>Proposition 6.</u> At any point x which does not satisfy some prefixed stopping criterion on the value v^*, a descent direction is found in a finite number of steps.

<u>Proof.</u> It is easy to verify that, as consequence of the condition $g^{+^T} d^* \ge m_1 v^* > v^*$, in case of repeated failures of the line

search (case b)) an increasing sequence of values $\{v_k^*\}$ is obtained. Moreover this sequence is bounded from above by zero.

To prove that $v_k^* \longrightarrow 0$, note that, being $v^* = -\|d^*\|^2 - \sum_{i \in F} \mu_i^* \alpha$, also the sequence $\{\|d_k^*\|\}$ is bounded.

Thus consider a convergent subsequence of $\{v_k^*\}$ and $\{d_k^*\}$ and let v_s^* and d_s^* be the successor of v_k^* and d_k^* in such subsequence. Assuming that g^+ is the subgradient evaluated along the direction d_k^*, the following hold:

$$g^{+T} d_k^* \geq m_1 v_k^*$$

$$g^{+T} d_s^* \leq v_s^*$$

hence

$$g^{+T} (d_s^* - d_k^*) \leq v_s^* - m_1 v_k^*$$

and, passing to the limit, the result follows.

The proposition above ensures that after a finite number of failed line searches either a successful one is performed or the value $-v^*$ is reduced below any positive prefixed value.

The following proposition clarifiės the meaning of "point close to x" and justifies the termination test based on the value of v^* as an ε-optimality condition.

Proposition 7. Any point obtained as result of case b) of the line search provides an ε-subgradient at the current point, for $\varepsilon = 2\bar{t}k$ (k is the upper bound on the norm of the subgradient).

Proof. Consider a point $x^+ = x + td^*$, obtained as result of a line search performed along the direction d^* starting from point x. Let $g \in \partial f(x)$ and that case b) of the line search occours $\left(t \|d^*\|_2 \leq \bar{t}\right)$.

Any subgradient g^+ at the point x^+ belongs to the α-subdifferential of f at point x for α defined as

$$\alpha = f(x) - f(x^+) + tg^{+T} d^* \geq 0$$

On the other hand the following inequality holds:

$$f(x^+) - f(x) \geq tg^T d^*$$

then

$$\alpha \leq td^{*T}(g^+ - g) \leq 2k\bar{t}$$

4. CONCLUSIONS

The paper presents some ideas to modify the bundle methods for convex optimization. Guidelines for definition of numerical algorithms are discussed as well, although a number of open questions (deletion rules, possible restricted step approach, appropriate methods for solving the quadratic programming subproblem) deserve some research effort in order to guarantee numerical effectiveness.

REFERENCES

Gaudioso, M., and Monaco, M.F. (1982). A bundle type approach to the unconstrained minimization of convex nonsmooth functions. Math. Prog., 23: 216-226.

Hiriart-Urruty, J.-B.(1982). Limiting behaviour of the approximate first order and second order directional derivatives for a convex function. Nonlinear Analysis, Theory, Methods & Applications, 6(12): 1309-1326.

Lemarechal, C. (1977). Bundle methods in nonsmooth optimization. In C. Lemarechal and R. Mifflin (Eds.), Nonsmooth Optimization. IIASA Proc. Series 3, Pergamon Press, Oxford.

Lemarechal, C., Strodiot, J.J., and Bihain, A. (1981). On a bundle algorithm for nonsmooth optimization. In O.L. Mangasarian, R.R. Meyer and S.M. Robinson (Eds.), Nonlinear Programming 4. Academic Press, New York.

Lemarechal, C., and Zowe, J. (1983). Some remarks on the construction of higher order algorithms in convex optimization. Appl. Math. Opt., 10: 51-68.

A NOTE ON THE COMPLEXITY OF AN ALGORITHM FOR TCHEBYCHEFF APPROXIMATION

A.A. Goldstein
Department of Mathematics, University of Washington, Seattle, WA 98195, USA

ABSTRACT

Some remarks are given concerning the complexity of an exchange algorithm for Tchebycheff Approximation. We consider an "exchange" algorithm that constructs the best polynomial of uniform approximation to a continuous function defined on a closed interval or a finite point set of real numbers. The first, and still popular, class of methods for this problem have been called "exchange algorithms". We shall consider the simplest method of this class, a blood relative of the dual simplex method of linear programming, and a special case of the cutting plane method. The the idea of the method was initiated by Remes, [1] and [2]. See also Cheney [3], for further developments. Klee and Minty [4], (1972) showed by example that the number of steps in a Simplex method can be exponential in the dimension of the problem. Since then considerable effort has been expended trying to explain the efficiency experienced in practice. Recently, probabilistic models have been assumed that yield expected values for the number of steps with low order monomial behaviour. See for example, Borgwardt [5], and Smale [6]. Alternatively, one might ask can one somehow classify the good problems from the bad ones. We believe that this may be possible for the exchange algorithm.

Let T = [0,1], or a finite subset of distinct points of [0,1] with card T > n+1. Let A(t) = $(1, t, \ldots, t^{n-1})$. Assume that f is in C^1(T). There exists an n-tuple x^* minimizing the function F(x) = max{|[A(t), x]| - f(t) : t ε T}, where [,] denotes the dot product. Given $\varepsilon > 0$ we seek x^k to minimize F within a tolerance of ε. Needed in exchange algorithms is the maximization of |[A(t),x] - f(t)| for fixed x. A novelty of the formulation below is that this maximization can have an error ≦ η, where η depends on ε. Most of the arguments however are borrowed from [1], [2] and [3]. The number of steps k to ensure that $F(x^k) - F(x^*) < \varepsilon$ will be shown to be proportional to log(1/ε) and to $1/\vartheta$, where $\vartheta > 0$ is a number that depends on f and n. Some remarks about the behavior of ϑ will be made. At k=1 in the algorithm that follows we take $t_i^1 = .5(1 - \cos(i\pi/n))$, $0 \leqq i \leqq n$. See II below.

ALGORITHM

1) At the kth iteration a positive number η and a set of n+1 points $0 \leqq t_0^k \leqq t_1^k, \ldots, \leqq t_n^k \leqq 1$ is given. Solve the equations

$$(-1)^i M^k = \sum_{j=1}^{j=n} A_j(t_i^k) x_j^k - f(t_i^k), \qquad 0 \leqq i \leqq n$$

for (x^k, M^k), where $M^k > 0$. (If $M^k < 0$, replace $(-1)^i$ by $(-1)^{i+1}$).

2) Calculate $\overline{t^k}$ such that $|R(\overline{t^k})| = |[A(\overline{t^k}), x^k] - f(\overline{t^k})| \geqq F(x^k) - \eta$. If $|R(\overline{t^k})| = M^k$, stop.

3) If $|R(\overline{t_k})| > M^k$, replace one of the points t_i^j, $0 \leqq i \leqq n$, by $\overline{t_k}$ in such a way that R(t) alternates signs on the points $t_0^k < t_1^k, \ldots, < t_n^k$.

4) Return to 1) with k+1 replacing k.

Supported in part by NIH Grant RR01243 at the University of Washington.

CLAIM: There exists a positive number ϑ such that given $\varepsilon > 0$ and a non-negative number $\eta < \varepsilon\vartheta/(1+\vartheta)$, and a positive integer $k > \vartheta^{-1}[\log\vartheta^{-1}+log(R(\overline{t}_1)-M^1)+\log\varepsilon^{-1}]+log(1+\vartheta)$ then for some s, $1\leq s\leq k+1$, $F(x^s)-F(x^*)<\varepsilon$.

PROOF The inequalities $M^k\leq F(x^*)\leq F(x)$ will be used below. If $|R(\overline{t_k})|=M^k$, $F(x^k)-F(x^*) \leq \eta$. Hence we stop. By [2, p.43] we may write:

$$M^k=|\sum_{i=0}^{i=n}R(t_i^k)d_i^k(-1)^i|/(\sum_{i=0}^{i=n}d_i^k)=\sum_{i=0}^{i=n}\lambda_i^k|R(t_i^k)|$$

where $\lambda_i^k>0$, and$\sum_{i=0}^{i=n}\lambda_i^k=1$.

Since $|R(\overline{t_k})|>M^k$,it follows that $M^{k+1}>M^k$, because n values of $|R(t_{k_i})|=M^k$, while the remaining one exceeds M^k. Hence for some λ_j^k, $0\leq j\leq n$, $M^{k+1} = \lambda_j^k(|R(\overline{t_k})|-M^k) + M^k$. Let $\overline{\vartheta} = \inf\{\lambda_i^k:0\leq i\leq n;1\leq k\leq\infty\}$ Assume temporarily that $\overline{\vartheta}>0$. Deny the claim. Then for every s = 1,2,...,k+1, $F(x^s)-M^s> F(x^s)-F(x^*) \geq \varepsilon$. Then $F(x^*)-M^k > M^{k+1}-M^k \geq(|R(\overline{t_k})|-M^k)\vartheta \geq (F(x^k)-\eta-M^k)\vartheta \geq (F(x^*)-M^k)\vartheta\geq (\varepsilon-\eta)\vartheta \geq (\varepsilon-\frac{\varepsilon\vartheta}{1+\vartheta})\vartheta = \frac{\varepsilon\vartheta}{1+\vartheta} > \eta$.

Since $M^k-M^{k+1}\leq-(F(x^*)-M^k)\vartheta$,

$$F(x^*)-M^k+M^k-M^{k+1} \leq(F(x^*)-M^k)(1-\vartheta) \leq (|R(\overline{t_k})|-M^k)(1-\vartheta)$$

Whence $F(x^*)-M^{k+1} \leq (|R(\overline{t}_1)|-M^1)(1-\vartheta)^k$ and $F(x^{k+1}-\eta-M^{k+1})\vartheta \leq F(x^*)-M^{k+1} \leq(|R(\overline{t}_1)|-M^1)(1-\vartheta)^k$. If $F(x^{k+1})-\eta-M^{k+1} < \varepsilon-\eta$, we have our contradiction. Choose k so that $(1-\vartheta)^k(R(\overline{t}_1)-M^1) <\frac{\varepsilon\vartheta}{1+\vartheta}$. Then, using $-\log(1-\vartheta)>\vartheta$, we get

$$k > \vartheta^{-1}[\log\frac{1}{\varepsilon}+\log\frac{1}{\vartheta}+\log(R(\overline{t}_1-M^1)+\log(1+\vartheta)]$$

It remains to show that $\overline{\vartheta} > 0$. Let $\underline{t}=(t_0,...,t_n)$ and set:

$$M(\underline{t})=\min\{\max\{|[A(t_i,x]-f(t_i)|:0\leq i\leq n\}:x\varepsilon R^n\}.$$

If $\underline{t} = (t_0^k,t_1^k,...,t_n^k)$ then $M(\underline{t})=M^k$. Let $\underline{T} = \{\underline{t}=(t_0,t_1,...,t_n):0\leq t_0\leq t_1\leq,...,t_n\leq 1\}$. $\underline{T}$ is a compact subset of R^{n+1}. We claim that $M(\underline{t})$ is continuous on $\underline{T}$. This follows by the continuity of f and the Vandermonde matrix if the components of $\underline{t}$ are non-coalescing, i.e., if $t_{i+1}\neq t_i$ for $0\leq i\leq n-1$. If some components coalesce then $M(\underline{t}) = 0$, since in this case x can be chosen so that the polynomial [A(t), x] interpolates f at $(t_0,...,t_n)$. Suppose then that $\{\underline{t}^s\}$ is any sequence with non-coalescing components converging to $\overline{\underline{t}}$ and assume that for some index i, $\overline{t}_1=\overline{t}_{i+1}$. There are at least 1 and at most n distinct components of $\overline{\underline{t}}$. Choose $\overline{x}$ closest to the origin such that:

$$R(\overline{t_i},\overline{x})=[A(\overline{t_i}),\overline{x}]-f(\overline{t_i})=0, \quad 0\leq i\leq n.$$

Since $\max\{|[A(t_i^s),\overline{x}]-f(t_i^s)|:0\leq i\leq n\} \geq M(\underline{t}^s)$ we have that $\lim M(\underline{t}^s) = 0 = M(\overline{\underline{t}})$. If $M^1 > 0$ we define the compact set

$$S=\{\underline{t}\varepsilon T:M^1\leq M(\underline{t})\leq\min\{\max\{|[A(t),x]-f(t)|:t\varepsilon[0,1]\}:x\varepsilon R^n\}.$$

If $M^1 = 0$, replace M^1 by M^2. (If $M^2 = 0$, we have a solution to our problem.) Let $G(\underline{t}) = \min\{(t_{i+1}-t_i):1\leq i\leq n\}$ and $\lambda(\underline{t})=\min\{(d_i(\underline{t})/\sum_{i=0}^{i=n}d_i(\underline{t})):0\leq i\leq n\}$. Since $G(\underline{t})$ is continuous on S, it achieves a minimum, say at $\underline{t}^{\bullet}$. Because $M(\underline{t}^{\bullet}) \geqq M^1$, it follows that $G(\underline{t}^{\bullet}) = \gamma > 0$. Since $\gamma > 0$, $\lambda(\underline{t})$ is continuous; it achieves a minimum ϑ on S. Clearly $\vartheta \leqq \vartheta$. By the formula for d_i in Example 1 below, it is seen that $\vartheta > 0$.

REMARK 1 If card T = m > n, the traditional algorithm (η = 0) may be employed and the maximization of |R(t)| has a cost proportional to m.

REMARK 2 Assume R $\varepsilon C^2[0,1]$, R alternates on t_i^k, $0 \leqq i \leqq n$, and $R'(t)$ vanishes no more than n+1 times on [0,1]. Given $\mu > 0$ such that $|R''(t)| \leqq \mu$ for all t $\varepsilon[\overline{t}, t^{\bullet}]$ where $t^{\bullet}$ is a local maximum of |R(t)| and $\overline{t}$ is the closest point of $\{t_0^k, \ldots, t_n^k\}$ to $t^{\bullet}$. Thus $|t^{\bullet} - \overline{t}| < 1/n$. By Taylor's theorem $\eta = |R(t) - R(t^{\bullet})| = |R''(\xi)|(t - t^{\bullet})^2$. Thus if $|t - t^{\bullet}| \leqq \sqrt{\frac{\eta}{\mu}}$ then t is a satisfactory maximizer. Using the bisection method to find t requires k steps where $2^{-k}/n < \sqrt{\frac{\eta}{\mu}}$ whence

$$k > [.5(log(1/n) + \log \mu) + \log \frac{1}{\eta}] / \log 2$$

For each cycle this process would be applied n+1 times. Thus it is plausible that the exchange algorithm can be effective.

II Some Remarks about ϑ.

The weights $\{d_i^k : 0 \leqq i \leqq n\}$, and hence ϑ depend on the distribution of the points $T_k^n = \{t_0^k \cdots, t_n^k\}$. Let $T^n = \{t_i = \frac{1}{2}(1 - \cos i\frac{\pi}{n}) : 0 \leqq i \leqq n\}$. For this distribution (Example 1), ϑ^{-1} = Kn < 2n . Thus if $T_k^n - T^n$ is sufficiently small, $\vartheta_k^{-1} <$ 4n, a pleasant complexity. Let P_{n-1} be the polynomial of degree n-1 that best approximates f on [0,1]. The critical points of P_{n-1} are points of [0,1] where the magnitude of the difference of f(t) and $P_{n-1}(t)$ is maximal. By Remark 3) below, for every n there are continuous functions for which the critical points induce $\vartheta^{-1} \leqq$ 4n. By a remarkable theorem of Kadic [7], for any f belonging to C[0,1], the critical points of P_{n-1} are asymptotically equal to T_n. Unfortunately, it is not established whether ϑ^{-1}/Kn tends to 1 as n goes to infinity. Moreover,Remark 3 and Example 3 show that for every n there is a continuous function f such that the corresponding polynomial P_{n-1} of best approximation has values of the weights $\lambda_i \leqq 2^{-(n+1)}$ for all but 2 values of i.

PROJECT. Given a natural number N find a family of functions F_N with the property that if n > N and f belongs to F_N then P_{n-1} has critical points near T_n. Likely candidates would be power series whose coefficients C_k for k > N converged at a sufficiently high speed. Are there others?

CLAIM 2. Suppose P_n is a polynomial of best approximation to f on [0,1] Let $E_n = ||f - P_n||_\infty$. Given $\varepsilon > 0$ assume that $\sqrt{E_n / E_{n-1}} = 1/n^{2+\varepsilon}$. Then $\lim_{n \to \infty} \frac{\vartheta}{Kn} = 1$.

Proof. Consider approximating f(x/π) on [0, π] by Q_{n-1} a polynomial in cos x of degree n-1. The points corresponding to T^n above are now simply $\{x_k = (\pi k/n) : 0 \leqq k \leqq n\}$. Let $\overline{x}_k$ be the critical points of Q_{n-1}. Kadic [7] proves that the following inequality holds for each n, every α, 0≦α≦.5 and every k, k=0,1,2,...,n

$$|\overline{x}_k - x_k| \leqq (\pi\alpha/n) + (n\alpha)^{-1/2} \, arcosh \frac{E_{n-1} + E_n}{E_{n-1} - E_n} \qquad (A)$$

Let $E_n = q(n)E_{n-1}$ and assume that 0 < q(n) < 1/2. Using arcosh u = log(u + $(u^2-1)^{1/2}$) we find setting u = $\frac{1+q_n}{1-q_n}$ and $u^2-1 = 4q_n/(1-q_n)^2$ that arcosh $\frac{1+q_n}{1-q_n}$ = log $\left(1 + \frac{2q_n}{1-q_n} + 2\frac{\sqrt{q_n}}{1-q_n} \right) \leq 8\sqrt{q_n}$. Set $\alpha = n^{-(1+2\varepsilon/3)}$ and $\sqrt{q(n)} = 1/(n^{2+\varepsilon})$. Then

$$|\overline{x}_k - x_k| \leqq (\pi+8)/n^{2+(2/3)\varepsilon} = (\pi/n)(s(n)) \quad 0 \leqq i \leqq n \qquad (B)$$

Here $s(n) = (\pi+8)/\pi n^{2+2\varepsilon/3}$. Assume that s(n) < 1 and k > i.

Let $\overline{t_i} = .5(1-\cos \overline{x}_i)$ and $t_i = .5(1-cos\ x_i)$. Then

$$|\overline{t_i} - t_i| \leqq |(sin(\xi))(\pi\ s(n)/n)|$$

with $((i-1)\pi/n) \leq \xi(n) \leq (i+1)\pi/n.$ Also $t_{i+1} - t_i = -(sin(i+.5)\pi/n)(2sin\,\pi/n)$, Thus

$$|\overline{t_i} - t_i|/|t_{i+1} - t_i| \leqq r(n)s(n)$$

$$|\overline{t_k} - t_k|/|t_k - t_{k+1}| \leqq g(n)s(n)$$

where r(n) and g(n) tend to 1 as n goes to infinity.

Since

$$\overline{t_k} - \overline{t_i} = \overline{t_k} - t_k + t_k - t_i - (\overline{t_i} - t_i)$$

$$\left|\frac{\overline{t_k} - \overline{t_i}}{(t_k - t_i)} - 1\right| \leqq \left\{ \frac{(t_k - t_{k-1})}{(t_k - t_i)} g(n)s(n) + \frac{(t_{i+1} - t_i)}{(t_k - t_i)} r(n)s(n) \right\}$$

$$\leqq \{ (r(n)+g(n))s(n) \} = c(n)$$

For each d_i^{-1} there are n products of the form $t_k - t_i$, and since (g(n)+r(n)) is bounded and s(n) goes to 0 faster than 1/n, $\lim_{n\to\infty}(1 \pm c(n))^n = 1$. Thus

$$(1-c(n))^n \leqq d_i/\overline{d_i} \leqq (1+c(n))^n.$$

Similarly

$$(1-c(n))^n \leqq \frac{\sum d_i}{\sum \overline{d_i}} \leqq (1+c(n))^n$$

Whence

$$\left(\frac{1-c(n)}{1+c(n)}\right)^n ,= \frac{\lambda_i}{\overline{\lambda_i}} \leqq \left(\frac{1+c(n)}{1-c(n)}\right)^n$$

Thus if the estimate of the above claim is realistic, we see that the class of functions for which $\vartheta/2n$ tends to 1 is quite limited.

EXAMPLE 1. Assume n is even. If the points $t_i \varepsilon T^n$ then $\max(d_i)/\min(d_i) = 2$, and $d_i / \sum_{i=0}^{i=n} d_i > 1/2n$, $0 \leqq i \leqq n$.

PROOF. The points t_i are symmetrically spaced with respect to $t_{n/2}$. By a formula due to de la Valle Poussin (see [2,p.25])

$$d_i^{-1} = (t_i - t_0)(t_i - t_1),\ldots,(t_i - t_{i-1})(t_{i+1} - t_i),\ldots,(t_n - t_i)$$

we see that $d_0=d_n$ $d_1=d_{n-1},\ldots,$ $d_{(n-2)/2}=d_{(n+2)/2}$; and $d_0\leqq d_1\leqq,\ldots,d_{n/2}$. Let $t_0=0$ and $t_n = 1$. Then

$$d_0^{-1}=(1)(.5-.5cos[(n-1)\pi/n])(.5-.5cos[(n-2)\pi/n]),\ldots,(.5),\ldots,(.5-.5cos\,\pi/n)$$

Using $cos(n-k)\pi/n=(-1)cos(k\pi/n)$ we get:
d_0^{-1} $=(.5)^{n-1}(sin^2(\pi/n))(sin^2(2\pi/n)),\ldots,(sin^2((.5n-1)/\pi))$. And $d_{n/2}^{-1}$ =
$.5(-.5\cos(n-1)\pi/n),\ldots,.5cos(.5n-1)(\pi)/n,\ldots,(5cos(\pi/n))(.5)$
$=(.5)^n(\sin\pi/n)^2(sin\,2\pi/n)^2,\ldots,(sin(.5n-1)\pi/n)^2$. Since d_n = $.5d_{n/2}=d_0$ and $nd_{n/2} \geqq \sum_{i=0}^{i=n} d_i$, we get that $d_i/\sum_{i=0}^{i=n} d_i > 1/2n$.

The set T^n is not optimal, that is there are distributions for which induce larger values of ϑ than T^n.

EXAMPLE 2. If the points t_i are equally spaced, the numbers d_i are proportional to the binomial coefficients and $\vartheta^{-1}\leqq 2^n$, the value 2^n being achieved at t_0 and t_n.

EXAMPLE 3. Things can get worse. In the following example, all but 2 of the weights tend exponentially to 0 as n goes to infinity. Assume n is odd and all points are equally spaced except at the middle of the interval , that is: $t_{i+1}-t_i$ = h, if i ≠ (n-1)/2 and $t_{(n+1)/2}-t_{(n-1)/2}=\Delta$, with h = $(t_n-t_0\ -\Delta)/n$ = (1 - Δ)/n. The numbers $d_{(n-1)/2}$ and $d_{(n+1)/2}$ are equal to say $d^\bullet$ and the number Δ appears as a factor only in $d^\bullet$. The form of d_i^{-1} is $K_ih^r(\alpha h+\delta)((\alpha+1)h+\delta)\ldots$. Let K_0=minK_i. Then $d_i(h,\Delta) < d_i(h,\alpha) < 1/K_0h^{n+1}$. Let $d^\bullet(h,\Delta) = (1/D^\bullet(h,\Delta)\Delta) \geqq(1/D^\bullet(h,h)\Delta) \geqq(1/K^\bullet h^{n)}\Delta$.

CLAIM. Let $\beta=2^{-n}K_0/nK^\bullet$. Assume that $\Delta/(1-\Delta) \leqq \beta$. Then $\vartheta^{-1} \geqq 2^{n+1.}$

PROOF. If $d_i\neq d^\bullet$ then $\max_{i=n} d_i/d^\bullet$ $\leqq(K^\bullet h^n\Delta)/K_0h^{n+1}$ = $nK^\bullet\Delta/K_0(1-\Delta)$ = $2^{-n}\Delta/\beta(1-\Delta) \leqq 2^{-n}$. Since $\sum_{i=0}^{i=n} d_i\geqq 2d^\bullet$, we get that $d_i\ /\sum_{i=0}^{i=n} d_i\leqq 2^{-(n+1)}$.

REMARK 3. Given a number σ > 0 and the set { a = $t_1 < t_2 <,\ldots,t_{n+2}$ = b } there exists a function f belonging to C[a,b] such that if P_n is the best Tchebycheff approximation of f, then max { |P_n(t) - f(t)| : t ε [a,b] } = |P_n(t) - f(t) | = σ, $1\leqq i\leqq n+2$.

PROOF. Let g(x) = σcos{(n+1)(x-a)π/(b-a)} . Then g alternates sign on a, a+(b-a)/(n+1),..., a + n(b-a)/(n+1), b. Let x(t) be the monotone piece-wise linear function through the points: (a,a), (t_2, a + (b-a)/(n+1)),...,(t_{n+1}, a + n(b-a)/(n+1)), (b,b). The function h(t) = g(x(t)) alternates on $t_1, t_2, \ldots, t_{n+2}$, with amplitude $\sigma=|h(t_i)|$, $1\leqq i\leqq n+2$. Let Q_n be any fixed polynomial of degree n and set f = Q_n - h . Let R_n be any polynomial of degree n. Then max {|R(t) - f(t): t ε[a,b]} achieves a minimum at the polynomial Q_n, because Q_n - f = h, and h has the equi-oscillation property. Hence $Q_n=P_n$.

ACKNOWLEDGEMENTS

We are grateful to Professor E.W. Cheney for helpful criticisms and advice and to the referee who kindly pointed out numerous errors, and suggested several clarifications.

BIBLIOGRAPHY

1. Remes, E. Ya. Sur le calcul effectif des polynomes d'approximation de Tchebychef. Compte Rendus 199, ps 337-340, 1934.

2. Remes, E. Ya. General Computational Methods of Tchebychef Approximation. AEC Translations #4491, 1957.

3. Cheney, E.W. Introduction to Approximation Theory. McGraw Hill, 1966.

4. Klee,V., and Minty, G.J. How good is the Simplex Algorithm? in Shish, O. Ed., Proceeding of the 3rd Symp. on Inequalities, A.P. N.Y., pp.159-175, 1972.

5. Borgwardt, K.H. The Average Number of Pivot Steps Required by the Simplex Method is Polynomial. Zeitschrift *für* OR vol.26, pp. 157-177, 1982.

6. Smale, S. The Problem of the Average Speed of the Simplex Method. Proc. of the 11th Int. Symp. on Math. Prog. U. Bonn, pp. 530-539, Aug. 1982.

7. Kadec, M.I. On the Distribution of Points of Maximum Deviation in the Approximation of Continuous Functions by Polynomials. AMS Translations (2) 26 (1963) 231-234, 1960.

DESCENT METHODS FOR NONSMOOTH CONVEX CONSTRAINED MINIMIZATION

K.C. Kiwiel
*Systems Research Institute, Polish Academy of Sciences, Newelska 6,
01447 Warsaw, Poland*

1. INTRODUCTION

We are concerned with methods for solving the problem

$$\text{minimize} \quad f(x) \quad \text{over all} \quad x \in R^N \tag{1.1a}$$

$$\text{satisfying} \quad F(x) \le 0, \tag{1.1b}$$

$$h_i(x) \le 0 \quad \text{for each} \quad i \in I, \tag{1.1c}$$

where the (possibly nonsmooth) functions f and F are real-valued and convex on R^N, h_i are affine and $|I| < \infty$. We assume that the feasible set $S=S_h \cap S_F$ is nonempty, where $S_h = \{x : h_i(x) \le 0, i \in I\}$ and $S_F=\{x : F(x) \le 0\}$, and that $F(\tilde{x}) < 0$ for some $\tilde{x}$ in S_h (the Slater condition). We suppose that for each $x \in S_h$ one can compute $f(x)$, $F(x)$ and two arbitrary subgradients $g_f(x) \in \partial f(x)$ and $g_F(x) \in \partial F(x)$; these evaluations are not required for $x \notin S_h$.

We shall present two algorithms for problem (1.1). Their convergence analysis will appear elsewhere (see the ref. list). Here we wish to concentrate on the following two basic ideas.

First, we show that nondifferentiabilities of f and F can be tackled by employing their polyhedral models with at most $N+3$ linear pieces. This eliminates the difficulties with increasing storage and work of earlier methods (Kelley, 1960; Mifflin, 1982; Strodiot et al. 1983), which use k pieces at the k-th iteration. A uniform bound on storage and work per iteration is obtained by following the subgradient selection strategy of Kiwiel (1983). This strategy drops irrelevant linear pieces by exploiting properties of quadratic programming subproblems that generate search directions.

Secondly, our treatment of constraints differs from that employed in existing feasible point methods (Mifflin, 1982; Strodiot et al. 1983). Our algorithms may approach the boundary of S more rapidly than do the latter methods, thus attaining faster convergence. To this end, we use exact penalty functions, whereas Mifflin (1983) used another penalty technique. Moreover, our algorithms find a solution in a finite number of iterations whenever f and F happen to be polyhedral and some mild regularity conditions are satisfied. This attractive property is not possessed by the existing feasible point methods. In effect, our algorithms seem to be natural extensions to the nonsmooth case of the widely used method of successive quadratic approximations (see, e.g., Pshenichny, 1983). We hope, therefore, that they will inherit the efficiency of its predecessor.

From lack of space, we shall report elsewhere extensions to nonconvex locally Lipschitzian problems done in the spirit of Kiwiel (1984d).

2. LINEARLY CONSTRAINED PROBLEMS

For simplicity, we start with the reduced version of (1.1)

$$\text{minimize} \quad f(x) \quad \text{over all} \quad x \in S_h. \tag{2.1}$$

Our method for solving (2.1) generates a sequence of points $\{x^k\}_1^\infty \subset S_h$ with nonincreasing $\{f(x^k)\}$, which is intended to converge to the required solution, and a sequence of trial points $\{y^k\} \subset S_h$. The starting point $x^1=y^1 \in S_h$ is provided by the user. Each y^j defines the linearization of f

$$f_j(x) = f(y^j) + \langle g_f(y^j), x-y^j \rangle \text{ for all } x. \tag{2.2}$$

At the k-th iteration, f is approximated around x^k by

$$\hat{f}^k(x) = \max\{f_j(x) : j \in J_f^k\}, \tag{2.3}$$

where $J_f^k \subset \{1,\ldots,k\}$ and $|J_f^k| \le N+2$. By convexity, $f(x) \ge \hat{f}^k(x)$ for all x, and $f(y^j)=\hat{f}^k(y^j)$ for all $j \in J_f^k$. Since we want to minimize f on S_h, we may minimize its approximation $\hat{f}^k$. Therefore, the next trial point y^{k+1} is chosen to

$$\text{minimize} \quad \hat{f}^k(y) + \tfrac{1}{2}|y-x^k|^2 \quad \text{over all} \quad y \in S_h, \tag{2.4}$$

where the stabilizing term $|y-x^k|^2/2$ keeps $y^{k+1} \in S_h$ in the region where $\hat{f}^k$ should be close to f. Without this term, subproblem (2.4) would be closer to (2.1) globally (as in cutting plane methods), but need not have a solution. If $y^{k+1}=x^k$, the method may stop because x^k is optimal.

The algorithm makes a serious step from x^k to $x^{k+1}=y^{k+1}$ only if the objective is significantly reduced, as measured by the test

$$f(y^{k+1}) \le f(x^k) + m\, v^k, \tag{2.5}$$

where $m \in (0,1)$ is a fixed parameter and

$$v^k = \hat{f}^k(y^{k+1}) - f(x^k) \tag{2.6}$$

is the predicted decrease $(v^k < 0)$. Otherwise, a null step $x^{k+1}=x^k$ occurs, but y^{k+1} will enrich the next approximation $\hat{f}^{k+1}$ with the piece f_{k+1} $(k+1 \in J^{k+1})$, thus increasing the chance of finding a better y^{k+2}.

It remains to choose J^{k+1}. In practice, we find a search direction $d^k=y^{k+1}-x^k$ by solving for (d^k,u^k) the quadratic programming subproblem

$$\begin{aligned} &\text{minimize} \quad u + \tfrac{1}{2}|d|^2 \quad \text{over all} \quad (d,u) \in R^{N+1} \\ &\text{satisfying} \quad f_j(x^k) + \langle g_f(y^j), d \rangle \le u \quad \text{for} \quad j \in J_f^k, \\ &\qquad\qquad h_i(x^k) + \langle \nabla h_i, d \rangle \le 0 \quad \text{for} \quad i \in I \end{aligned} \tag{2.7}$$

and find its Lagrange multipliers λ_j^k, $j \in J_f^k$, ν_i^k, $i \in I$, such that

$$\hat{J}_f^k = \{ j \in J_f^k : \lambda_j^k \neq 0 \} \tag{2.8}$$

satisfies $|J_f^k| \le N+1$. Then $\hat{f}^{k+1}$ (cf. (2.3)) defined by

$$J_f^{k+1} = \hat{J}_f^k \cup \{k+1\}$$

contains all the pieces f_j contributing to d^k and $y^{k+1}=x^k+d^k$, since replacing J_f^k by $\hat{J}_f^k$ in (2.7) does not change its solution. The remaining "inactive" pieces are dropped. This <u>subgradient selection strategy</u> ensures that $|J_f^k| \le N+2$ for all k.

We may add that typical quadratic programming routines for solving (2.7) will automatically produce at most N+1 nonzero Lagrange multipliers λ_j^k, since (2.7) involves N+1 variables. In practice, it is more efficient to solve the dual of (2.7) (see Kiwiel, 1984g).

<u>Theorem 2.1</u>. The algorithm described above minimizes f on S_h, i.e. $\{x^k\} \subset S_h$ and $f(x^k) \downarrow \inf \{f(x) : x \in S_h\}$. Moreover, $\{x^k\}$ converges to a solution of problem (2.1) whenever this problem has any solution.

It is worth adding that if f is polyhedral and problem (2.1) satisfies some regularity condition (Kiwiel, 1983), which is weaker than the Haar condition, then the method stops with an optimal x^k after a finite number of iterations.

In practice one may use a stopping criterion of the form $|v^k| \le \varepsilon_s$ with small positive ε_s (e.g. $\varepsilon_s=10^{-6}$), since we have the estimate

$$f(x^k) \le f(x) + |v^k|+|v^k|^{1/2}|x-x^k| \quad \text{for all} \quad x \in S_h.$$

Then,. for bounded S_h, termination occurs with

$$f(x^k) \le \min_{S_h} f + \varepsilon_s + \varepsilon_s^{1/2} \max \{|x-x^k| : x \in S_h,\ f(x) \le f(x^k)\}.$$

3. METHOD OF LINEARIZATION

We shall now extend the method of Section 2 to the nonlinearly constrained problem (1.1).

In order to treat the nonlinear constraint (1.1b) in the preceding algorithm, it suffices to use the linearizations of F

$$F_j(x) = F(y^j) + \langle g_F(y^j), x-y^j \rangle \quad \text{for all} \quad x$$

for defining the k-th polyhedral lower approximation to F

$$\hat{F}^k(x) = \max\{F_j(x) : j \in J_F^k\}$$

with $J_F^k \subset \{1,...,k\}$ and $|J_F^k| \le N+2$. Then (2.4) is extended to the subproblem

$$\text{minimize} \quad \hat{f}^k(y) + \tfrac{1}{2}|y-x^k|^2 \quad \text{over all} \quad y \in R^N \tag{3.1a}$$

$$\text{satisfying} \quad F^k(y) \le 0, \tag{3.1b}$$

$$h_i(y) \le 0 \quad \text{for} \quad i \in I. \tag{3.1c}$$

This is a local approximation to problem (1.1). It differs from the corresponding subproblem of the cutting plane method (Kelley, 1960) in that the presence of the stabilizing quadratic term $|y-x^k|^2/2$ enables one to select $J_f^k \cup J_F^k$ not necessarily equal to $\{1,...,k\}$ without impairing convergence.

Since y^{k+1} or x^k may not lie in S_F, for assessing whether y^{k+1} is better than x^k we need a certain merit function that combines the objective value $f(x)$ with the (nonlinear) constraint violation $F(x)_+=\max\{F(x),0\}$. To this end, we shall use the exact penalty function

$$e(x;c) = f(x) + c\,F(x)_+ \quad \text{for all} \quad x\,,$$

where $c=c^k > 0$ will be the penalty coefficient of the k-th iteration. We shall choose c^k large enough to ensure that $e(\cdot;c^k)$ has minima only at solutions to problem (1.1). Moreover, c^k will be such that the following approximate derivative of $e(\cdot;c^k)$ at x^k in the direction $d^k=y^{k+1}-x^k$

$$v^k = \hat{f}^k(x^k+d^k) + c^k\hat{F}^k(x^k+d^k)_+ - e(x^k;c^k)$$

is negative, so that d^k is approximately a direction of descent for $e(\cdot;c^k)$ at x^k. The algorithm will take a serious step from x^k to $x^{k+1}=y^{k+1}=x^k+d^k$ if y^{k+1} is better than x^k in the sense that

$$e(y^{k+1};c^k) \le e(x^k;c^k) + m\,v^k\,, \tag{3.2}$$

where $m \in (0,1)$ is a parameter. Otherwise, a null step $x^{k+1}=$

x^k will occur. In this case the new subgradient information collected at y^{k+1} will enable the method to generate a better next search direction d^{k+1}.

Algorithm 3.1

Step 0 (Initialization). Select the starting point $x^1 \in S_h$, a final accuracy tolerance $\varepsilon_s \geq 0$, a line search parameter $m \in (0,1)$ and an initial penalty coefficient $c^o > 0$. Set $y^1 = x^1$ and $J_f^1 = J_F^1 = \{1\}$. Set $k=1$.

Step 1 (Direction finding). Find the solution (d^k, u^k) to the quadratic programming subproblem

$$\text{minimize} \quad u + \tfrac{1}{2}|d|^2 \quad \text{over all} \quad (d,u) \in R^{N+1} \tag{3.3a}$$

$$\text{satisfying} \quad f_j(x^k) + \langle g_f(y^j), d \rangle \leq u \quad \text{for} \quad j \in J_f^k, \tag{3.3b}$$

$$F_j(x^k) + \langle g_F(y^j), d \rangle \leq 0 \quad \text{for} \quad j \in J_F^k, \tag{3.3c}$$

$$h_i(x^k) + \langle \nabla h_i, d \rangle \leq 0 \qquad \text{for} \quad i \in I \tag{3.3d}$$

and corresponding Lagrange multipliers λ_j^k, $j \in J_f^k$, μ_j^k, $j \in J_F^k$, and ν_i^k, $i \in I$, such that the sets

$$\hat{J}_f^k = \{j \in J_f^k : \lambda_j^k \neq 0\} \quad \text{and} \quad \hat{J}_F^k = \{j \in J_F^k : \mu_j^k \neq 0\}$$

satisfy $|J_f^k \cup J_F^k| \leq N+1$. Set

$$\tilde{c}^k = \sum_{j \in J_F^k} \mu_j^k .$$

Step 2 (Penalty updating). If $\tilde{c}^k < c^{k-1}/2$, set $c^k = c^{k-1}$; otherwise, set $c^k = 2\max\{c^{k-1}, \tilde{c}^k\}$.

Step 3 (Stopping criterion). Set $v^k = u^k - e(x^k; c^k)$. If $v^k \geq -\varepsilon_s$, terminate; otherwise, continue.

Step 4 (Line search). Set $y^{k+1} = x^k + d^k$. If (3.2) holds, set $x^{k+1} = y^{k+1}$; otherwise, set $x^{k+1} = x^k$.

Step 5 (Linearization updating). Set $J_f^{k+1} = \hat{J}_f^k \cup \{k+1\}$, $J_F^{k+1} = \hat{J}_F^k \cup \{k+1\}$ and compute

$$f_{k+1}(x^{k+1}) = f(y^{k+1}) + \langle g_f(y^{k+1}), x^{k+1}-y^{k+1} \rangle ,$$

$$F_{k+1}(x^{k+1}) = F(y^{k+1}) + \langle g_F(y^{k+1}), x^{k+1}-y^{k+1} \rangle ,$$

$$f_j(x^{k+1}) = f_j(x^k) + \langle g_f(y^j), x^{k+1}-x^k \rangle \quad \text{for} \quad j \in \hat{J}_f^k ,$$

$$F_j(x^{k+1}) = F_j(x^k) + \langle g_F(y^j), x^{k+1}-x^k \rangle \quad \text{for} \quad j \in \hat{J}_F^k .$$

Step 6. Increase k by 1 and go to Step 1.

A few remarks on the algorithm are in order.

Note that the sequence of penalty coefficients $\{c^k\}$ is nondecreasing. The property $\tilde{c}^k < c^k$ ensures that $v^k < 0$ at Step 4. Our penalty updating rules make c^k eventually constant if $\{y^k\}$ stays bounded. Such an automatic limitation of penalty growth is important in practice, since large values of c^k may force the algorithm to follow closely the boundary of S_F, thus preventing fast convergence.

If the algorithm terminates at Step 3 then

$$\begin{aligned} &f(x^k) \le f(x) + \varepsilon_s + \varepsilon_s^{1/2}\, |x-x^k| \quad \text{for all} \quad x \in S, \\ &F(x^k)_+ \le \varepsilon_s / c^k . \end{aligned} \tag{3.4}$$

The above estimates show that x^k is approximately optimal.

Observe that replacing J_f^k and J_F^k by $\hat{J}_f^k$ and $\hat{J}_F^k$ in (3.3) yields an equivalent subproblem. Thus, once again, subgradient selection on the basis of Lagrange multipliers ensures uniformly bounded storage and work per iteration, since $|J_f^k \cup J_F^k| \le N+3$ for all k.

Theorem 3.2. Suppose that Algorithm 3.1 generates infinite sequences $\{x^k\}$ and $\{y^k\}$ such that $\{y^k\}$ is bounded. Then $\{x^k\}$ converges to a solution of problem (1.1). Moreover, the penalty coefficient c^k stays constant after a finite number of iterations, and $v^k \to 0$.

Observe that the assumption of Theorem 3.2 is satisfied if S_h is bounded, since $\{y^k\} \subset S_h$ by construction. Also for bounded S_h Theorem 3.2 implies finite termination (with x^k satisfying (3.4)) if the final accuracy tolerance ε_s is positive.

We may add that, under mild conditions, even with $\varepsilon_s=0$ the algorithm will terminate at an optimal x^k after finitely many iterations if f and F happen to be piecewise linear.

4. EXACT PENALTY FUNCTION METHOD

Another way of solving problem (1.1) is to

$$\text{minimize} \quad e(x;\bar{c}) = f(x) + \bar{c}F(x)_+ \quad \text{over all} \quad x \in S_h \tag{4.1}$$

with $\bar{c} > 0$ large enough (see, e.g. Demyanov and Vasiliev,1985). Since the above problem is a special case of (2.1), we may use the method of Section 2 and choose suitable $\bar{c}$ in the course of calculations.

Thus let the k-th approximation to $e(\cdot;c^k)$ be

$$\hat{e}^k(x;c^k) = \max\{e_j(x) : j \in J^k\} \quad \text{for all} \quad x,$$

where $J^k \subset \{1,\ldots,k\}$ satisfies $|J^k| \leq N+1$, whereas

$$e_j(x) = f_j(x) + c^kF_j^+(x),$$

$$F_j^+(x) = \begin{cases} F(y^j) + \langle g_F(y^j), x-y^j \rangle & \text{if} \quad F(y^j) > 0, \\ 0 & \text{if} \quad F(y^j) \leq 0 \end{cases}$$

are linearizations at y^j of the convex functions $e(\cdot;c^k)$ and $F(\cdot)_+$, respectively, and f_j is given by (2.2). Introducing $J_+^k=\{j : F(y^j) > 0,\ 1 \leq j \leq k\}$ and $J_0^k=\{j : F(y^j) \leq 0,\ 1 \leq j \leq k\}$, we see that $F_j^+(\cdot)=F_j(\cdot)$ if $j \in J_+^k$, and $F_j^+(\cdot)=0$ if $j \in J_0^k$. We may now proceed as in Section 2 to motivate the subproblem

$$\text{minimize} \quad \hat{e}^k(y;c^k) + \tfrac{1}{2}|y-x^k|^2 \quad \text{over all} \quad y \in S_h,$$

which gives rise to the following method.

Algorithm 4.1.

Step 0 (Initialization). Select the starting point $x^1 \in S_h$ and a final accuracy tolerance $\varepsilon_s \geq 0$. Choose a line search parameter $m \in (0,1)$, an initial penalty coefficient $c^1 > 0$ and an initial unconstrained minimization tolerance $\delta^1 > 0$. Set $y^1=x^1$, $J_+^1=\{1\}$ and $J_0^1=\emptyset$ if $F(y^1) > 0$, $J_+^1=\emptyset$ and $J_0^1=\{1\}$ if $F(y^1) \leq 0$, and $J=\{1\}$. Set $k=1$.

Step 1 (Direction finding). Find the solution (d^k,u^k) to the subproblem

$$\begin{aligned}&\text{minimize}\quad u+\tfrac{1}{2}|d|^2 \quad \text{over all} \quad (d,u)\in R^{N+1}\\ &\text{satisfying}\quad f_j(x^k)+\langle g_f(y^j),d\rangle \le u \quad \text{for} \quad j\in J_0^k,\\ &f_j(x^k)+c^kF_j(x^k)+\langle g_f(y^j)+c^kg_F(y^j),d\rangle\le u \quad \text{for} \quad j\in J_+^k,\\ &h_i(x^k)+\langle \nabla h_i,d\rangle \le 0 \quad \text{for} \quad i\in I\end{aligned} \tag{4.2}$$

and corresponding Lagrange multipliers λ_j^k, $j\in J_0^k\cup J_+^k$, and ν_i^k, $i\in I$, such that the sets

$$\hat{J}_0^k=\{j\in J_0^k:\lambda_j^k\neq 0\} \quad \text{and} \quad \hat{J}_+^k=\{j\in J_+^k:\lambda_j^k\neq 0\}$$

satisfy $|\hat{J}_0^k\cup\hat{J}_+^k|\le N+1$. Set

$$v^k=\hat{e}^k(x^k+d^k;c^k)-e(x^k;c^k).$$

Step 2 (Penalty updating). If $|v^k|\le\delta^k$ and $F(x^k)>|v^k|$, set $c^{k+1}=2c^k$ and $\delta^{k+1}=\delta^k/2$; otherwise, set $c^{k+1}=c^k$ and $\delta^{k+1}=\delta^k$.

Step 3 (Stopping criterion). If $|v^k|\le\varepsilon_s$ and $F(x^k)\le\varepsilon_s$, terminate. Otherwise, continue.

Step 4 (Line search). Set $y^{k+1}=x^k+d^k$. If

$$e(y^{k+1};c^{k+1})\le e(x^k;c^{k+1})+mv^k,$$

set $x^{k+1}=y^{k+1}$ (serious step); otherwise, set $x^{k+1}=x^k$ (null step).

Step 5 (Linearization updating). Set

$$J_+^{k+1}=\hat{J}_+^k\cup\{k+1\} \quad \text{and} \quad J_0^{k+1}=\hat{J}_0^k \quad \text{if} \quad F(y^{k+1})>0,$$

$$J_+^{k+1}=\hat{J}_+^k \quad \text{and} \quad J_0^{k+1}=\hat{J}_0^k\cup\{k+1\} \quad \text{if} \quad F(y^{k+1})\le 0.$$

Set $J^{k+1}=J_+^{k+1}\cup J_0^{k+1}$. Compute

$$f_{k+1}(x^{k+1})=f(y^{k+1})+\langle g_f(y^{k+1}),x^{k+1}-y^{k+1}\rangle,$$

$$f_j(x^{k+1})=f_j(x^k)+\langle g_f(y^j),x^{k+1}-x^k\rangle \quad \text{for} \quad j\in\hat{J}_+^k\cup\hat{J}_0^k,$$

$$F_{k+1}(x^{k+1}) = F(y^{k+1}) + \langle g_F(y^{k+1}), x^{k+1}-y^{k+1} \rangle \quad \text{if} \quad F(y^{k+1})>0,$$

$$F_j(x^{k+1}) = F_j(x^k) + \langle g_F(y^j), x^{k+1}-x^k \rangle \quad \text{for} \quad j \in \hat{J}^k_+.$$

Step 6. Increase k by 1 and go to Step 1.

The penalty updating scheme of Step 2 is based on the relation

$$e(x^k;c^k) \le e(x;c^k)+|v^k|+|v^k|^{1/2}|x-x^k| \quad \text{for all} \quad x \in S_h. \quad (4.3)$$

Thus $|v^k|$ indicates how much x^k differs from being optimal in (4.1) with $\bar{c}=c^k$. Moreover, (4.3) implies

$$f(x^k) \le f(x) + |v^k|+|v^k|^{1/2}|x-x^k| \quad \text{for all} \quad x \in S,$$

so x^k is an approximate solution to problem (1.1) if both $|v^k|$ and $F(x^k)_+$ are small. The penalty coefficient is increased only if $e(\cdot;c^k)$ has been approximately minimized, as indicated by relations (4.3) and $|v^k| \le \delta^k$ (with progressively smaller minimization tolerances $\{\delta^k\}$), but x^k is significantly infeasible $(F(x^k) > |v^k|)$. This penalty scheme is due to Kiwiel (1984e).

If the algorithm terminates then

$$f(x^k) \le f(x) + \varepsilon_s + \varepsilon_s^{1/2}|x-x^k| \quad \text{for all} \quad x \in S$$

and $F(x^k) \le \varepsilon_s$, so x^k is an approximate solution to problem (1.1). Of course, x^k is optimal if additionally $\varepsilon_s=0$.

Observe that the algorithm does not in fact require computation of $F(y)$ and $g_F(y)$ if $y \in S_F$. This is useful in certain applications.

Theorem 4.2. If Algorithm 4.1 generates a bounded infinite sequence $\{x^k\}$ (e.g. if S_h is bounded), then $\{x^k\}$ converges to a solution of problem (1.1). Moreover, the penalty coefficients $\{c^k\}$ stay constant for all large k, and $v^k \to 0$.

It is worth adding that, under mild conditions, Algorithm 4.1 also has the finite termination property in the polyhedral case.

Summing up, we observe that global convergence properties of Algorithms 3.1 and 4.1 are essentially the same. However, Algorithm 3.1 exploits the structure of problem (1.1) more fully by using the natural constraints (3.1b) and employing $e(\cdot;c^k)$ as a merit function only. These advantages have to be weighed against additional effort involved in quadratic programming when $I=\emptyset$.

5. CONCLUDING REMARKS

We have extended the widely used constraint linearization technique to the nonsmooth case. In particular, this technique ensures finite convergence in the polyhedral case, an important property not possessed by the existing feasible point methods.

Let us now comment on possible modifications and extensions.

For large N, we may replace subgradient selection with subgradient aggregation (Kiwiel, 1983, 1984a,1984c) to reduce the number of constraints of the form (3.3b,c) to as few as four without impairing global convergence. This will save storage and work per iteration. However, convergence may be slow if too few constraints (linear pieces) are used. Also it is easy to include more efficient line searches in the methods (Kiwiel, 1984a, 1984d, 1984f).

Additional information about the problem function structure can be used for modifying subproblems (2.7),(3.3) and (4.2) so as to increase the efficiency of the algorithms. Suitable techniques may be found in (Kiwiel, 1984f) for max-type functions, and in (Kiwiel, 1984b) for large-scale linearly constrained problems.

We shall report elsewhere extensions of the algorithms to the nonconvex case of locally Lipschitzian problem functions satisfying the semismoothness condition of Kiwiel (1984a,1984d).

REFERENCES

Demyanov V.F. and L.V. Vasiliev (1984). Nondifferentiable Optimization. Springer, Heidelberg.

Kelley J.E. (1960). The cutting plane method for solving convex programs. J. SIAM, 8, 703-712.

Kiwiel K.C. (1983). An aggregate subgradient method for nonsmooth convex minimization. Math.Programming, 27, 320-341.

Kiwiel K.C. (1984a). A linearization algorithm for constrained nonsmooth minimization. System Modelling and Optimization, P. Thoft-Christensen, ed., Lect. Notes Control Inform. Sci. 59, Springer, Berlin, pp. 311-320.

Kiwiel K.C. (1984b). A descent algorithm for large-scale linearly constrained problems. CP-84-15, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Kiwiel K.C. (1984c). An algorithm for linearly constrained convex nondifferentiable minimization problems. J. Math. Anal. Appl. (to appear).

Kiwiel K.C. (1984d). A linearization algorithm for nonsmooth minimization. Math. Oper. Res. (to appear).

Kiwiel K.C. (1984e). An exact penalty function algorithm for nonsmooth constrained convex minimization problems. IMA J. Num. Anal. (to appear).

Kiwiel K.C. (1984f). A method for minimizing the sum of a convex function and a continuously differentiable function. J. Optim. Theory Appl. (to appear).

Kiwiel K.C. (1984g). A method for solving certain quadratic programming problems arising in nonsmooth optimization. ZTSW-84, Systems Research Institute, Warsaw.

Mifflin R. 1982 . A modification and an extension of Lemarechal's algorithm for nonsmooth minimization. Nondifferential and Variational Techniques in Optimization, D.C. Sorensen and R.J.-B. Wets, eds., Math. Programming Study 17, pp. 77-90.

Mifflin R. 1983 . A superlinearly convergent algorithm for one-dimensional constrained minimization with convex functions. Math. Oper. Res., 8, 185-195.

Pshenichny B.N. (1983). Method of Linearizations. Nauka, Moscow (in Russian).

Strodiot J.-J., V.H. Nguyen and N. Heukemes (1983). ε-Optimal solutions in nondifferentiable convex programming and related questions. Math. Programming, 25, 307-328.

STABILITY PROPERTIES OF INFIMA AND OPTIMAL SOLUTIONS OF PARAMETRIC OPTIMIZATION PROBLEMS

Diethard Klatte and Bernd Kummer
Department of Mathematics, Humboldt University, 1086 Berlin, GDR

1. INTRODUCTION

In the analysis of parametric optimization problems it is of great interest to explore certain stability properties of the optimal value function and of the optimal set mapping (or some selection function of this mapping): continuity, smoothness, directional differentiability, Lipschitz continuity and the like. For a survey of this field we refer to comprehensive treatments of various aspects of such questions in the recent works of Fiacco (1983), Bank et al. (1982) and Rockafellar (1982).

In the present paper we consider an optimization problem that depends on a parameter vector $t \in T \subset R^m$:

$$P(t): \qquad \min \left\{ f_o(x,t) \;/\; x \in M(t) \right\}, \quad t \in T,$$

where T is nonempty, M: $T \longrightarrow 2^{R^n}$ is a closed-valued multifunction, and f_o is a real-valued function defined on $R^n \times T$. We define the <u>infimum function</u> φ and the <u>optimal set map</u> ψ by

$$\varphi(t) := \quad \inf \left\{ f_o(x,t) \;/\; x \in M(t) \right\}, \quad t \in T,$$

$$\psi(t) := \quad \left\{ x \in M(t) \;/\; f_o(x,t) = \varphi(t) \right\}, \quad t \in T.$$

Let $\psi_{loc}(t)$ denote the <u>set of all local minimizers</u> for $f_o(\cdot,t)$ w.r. to M(t). For $\varepsilon > 0$, the <u>set of ε-optimal solutions</u> is $\psi_\varepsilon(t) := \left\{ x \in M(t) \;/\; f_o(x,t) \leq \varphi(t) + \varepsilon \right\}$. Given $Q \subset R^n$ we set

$$M_Q(t) := \quad M(t) \cap \text{cl } Q ,$$

$$\varphi_Q(t) := \inf \{f_o(x,t) \,/\, x \in M_Q(t)\},$$
$$\psi_Q(t) := \{x \in M_Q(t) \,/\, f_o(x,t) = \varphi_Q(t)\},$$

where "cl" stands for closure. The symbol int X will be used to denote the interior of a set $X \subset R^n$. Further, $\|\cdot\|$ denotes the Euclidean norm, $U_\varepsilon(t) :=$ ε-neighborhood of t, $d(x,Z) := \inf_z \{\|x-z\| \,/\, z \in Z\}$ $(x \in R^n,\ Z \subset R^n)$, $d_H(Y,Z) := \inf_k \{k \,/\, d(y,Z) \leq k\ (\forall y \in Y),\ d(z,Y) \leq k\ (\forall z \in Z)\}$ (Hausdorff-distance of $Y,Z \subset R^n$). The closed unit ball in R^n will have the standard symbol B_n.

Adapting Rockafellar's definitions of Lipschitzian functions, we shall say that a multifunction F from $T \subset R^m$ to R^n is Lipschitzian on $D \subset T$ if there is some constant $L > 0$ such that $d_H(F(s),F(t)) \leq L\|s-t\|$ $(\forall s,t \in D)$. F is Lipschitzian around $t' \in T$ if there are real numbers $\varepsilon > 0$ and $L > 0$ such that $d_H(F(s),F(t)) \leq L\|s-t\|$ $(\forall s,t \in U_\varepsilon(t') \cap T)$. F is upper Lipschitzian at $t' \in T$ if there are real numbers $\varepsilon > 0$ and $L > 0$ such that $d(x,F(t')) \leq L\|t-t'\|$ $(\forall t \in U_\varepsilon(t') \cap T,\ \forall x \in F(t))$. A single-valued function g is said to be Lipschitzian on D (resp. around t') if $t \longrightarrow F(t) = \{g(t)\}$ has this property.

In the present paper we shall discuss the Lipschitz stability of P(t). Above all, our attention is focused on standard problems in parametric convex or quadratic optimization and thereby on the derivation of conditions under which the map or some "portion" of ψ_{loc} exhibit a certain Lipschitz behavior. In the literature, there are two approaches to these studies. The first one has been applied in parametric linear and quadratic programming; it makes use of the fact that a polyhedral multifunction F from R^m to R^n is upper Lipschitzian on R^m (cf. Walkup and Wets 1969, Robinson 1979,1981, Klatte 1983). The second approach is based on the application of implicit-function theorems (for systems of nonlinear equations and inequalities) to the parameterized Kuhn-Tucker system of the optimization problem considered; it requires restrictive smoothness and regularity assumptions on the objective function and on the constraints; in particular, second-order optimality conditions play an important role (cf. Fiacco 1983, Robinson

1982, Hager 1979). With respect to special classes of parametric programs the question arises whether some Lipschitz behavior of Ψ or Ψ_{loc} can be "saved" also in the absence of second-order regularity assumptions. One aim of our paper is to help clarifying this question by some constructive results and simple but instructive examples of ill-behaved parametric programs. A particular answer will be that if second-order conditions are dropped then, even for the class of parametric convex programs with right-hand side perturbations only, upper Lipschitz continuity of Ψ or the existence of a (Lipschitz-) continuous selection of Ψ cannot be expected, in general.

In contrast to this situation, the Lipschitz continuity of φ holds under rather natural assumptions. We mention here the following very simple but useful result (cf., e.g., Cornet 1983).

Lemma 1. Consider problem P(t). Let $T' \subset T$, and suppose that for some $Q \subset R^n$ and each $t \in T'$, we have $M(t) \subset Q$. If f_o is Lipschitzian on $Q \times T'$ with modulus β_f, and if M is Lipschitzian on T' with modulus β_M, then φ is Lipschitzian on T' (with modulus $\beta_f(\beta_M+1)$).

When M is defined as the solution set mapping of a system $f(x,t) \leqq 0$, where f is a locally Lipschitzian vector function, then certain constraint qualifications (for example, the Slater condition in the convex case, and the Mangasarian-Fromovitz condition in the smooth case) ensure that M is Lipschitzian in some sense; a detailed discussion of this question can be found in Rockafellar's (1984) paper which also covers results of Robinson, Levitin, Aubin and other authors concerning implicit multifunction theorems.

2. CONVEX PROBLEMS

Consider the parametric optimization problem P(t) under the following additional requirements:

(1) $M(t) := \{x \in R^n \,/\, f_i(x,t) \leqq 0 \ (i=1,\dots,s);\ f_j(x,t) = 0 \ (j=s+1,\dots,s+r)\}$,

(2) $f_i\colon R^n \times T \longrightarrow R$ is continuous on $R^n \times T$ $(\forall i \in \{0,1,\dots,s+r\})$,

(3) $f_i(\cdot,t)$ is convex on R^n $(\forall t \in T,\ \forall i \in \{0,1,\dots,s\})$,
$f_j(\cdot,t)$ is affine-linear $(\forall t \in T,\ \forall j \in \{s+1,\dots,s+r\})$;

we denote this parametric problem by $P_1(t)$. If (3) is replaced by (3)' then we have the special case of convex programs with right-hand side perturbations only:

(3)' $f_o(x,t) = h_o(x)$, $f_i(x,t) = h_i(x) - t_i$ $(t \in T;\ i=1,\dots,s)$,
where h_i $(i=0,1,\dots,s)$ is convex on R^n and $r=0$, $s=m$.

This special parametric program will be symbolized by $P_2(t)$.

First we state a theorem which is, in fact, a simple consequence of Robinson's (1976) inversion theorem for convex multifunctions. Using other methods of proof, Eremin and Astafiev (1976)§27 and Blatt (1980) presented similar results.

Theorem 1. Consider the parametric convex problem $P_1(t)$. Suppose that for some $t' \in T$,

(i) $\psi(t')$ is a nonempty, bounded set,

(ii) the Slater condition is satisfied w.r. to $M(t')$, i.e., there is a point $x' \in M(t')$ with $f_i(x',t') < 0$ $(i=1,\dots,s)$ such that the gradients $\nabla_x f_{s+1}(\cdot,t'),\dots, \nabla_x f_{s+r}(\cdot,t')$ are linearly independent,

(iii) there are an open convex set $W \supset \psi(t')$ and a neighborhood U of t' such that f_o is Lipschitzian on $W \times U$,

(iv) for each $x \in W$ and each $i \in \{1,2,\dots,s+r\}$, $f_i(x,\cdot)$ is Lipschitzian around t' with some modulus independent of x.

Then φ is Lipschitzian around t', and there is a number $\tilde{\varepsilon} > 0$ such that for all $0 < \varepsilon < \tilde{\varepsilon}$, ψ_ε is Lipschitzian around t'.

Proof: Set $Q := (\psi(t')+B_n) \cap W$. Taking (i), (ii) and (iv) into account and applying Corollary 2 in Robinson (1976), we have that M_Q is Lipschitzian around t'. Note that ψ is upper semicontinuous at t' (cf. Bank et al. 1982, Th. 4.3.3), hence for t near t', $\psi(t) = \psi_Q(t)$. Lemma 1 then yields the Lipschitz continuity of φ around t'. The assumptions (2), (3), (i) and (ii) ensure that the map $(t,\varepsilon) \longrightarrow \psi_\varepsilon(t)$ is upper semicontinuous at $(t',0)$ (cf. Bank et al. 1982, Cor. 4.3.3.2), and so if $\|t-t'\|$ and ε are sufficiently small, say $\|t-t'\| < \tilde{\varepsilon}$, $0 < \varepsilon < \tilde{\varepsilon}$, then $\psi_\varepsilon(t) \subset Q$. Let $0 < \varepsilon < \tilde{\varepsilon}$. Apply now Corollary 2 in Robinson (1976) to the map $t \longrightarrow M_\varepsilon(t) := \{x \in M_Q(t) \ / \ f_o(x,t) - \varphi(t) \leq \varepsilon\}$;

we only note that (a) $M_\varepsilon(t')$ contains a Slater point, (b) all functions describing M_ε are Lipschitzian w.r. to t, and (c) for all t near t' it holds $M_\varepsilon(t) = \Psi_\varepsilon(t) \subset Q$. //

In the Examples 1 and 2 we shall point out that, under the assumptions of Theorem 1, a Lipschitz behavior of ψ cannot be expected, in general, not even for the special problem $P_2(t)$. Example 1 is due to B. Schwartz (private communication).

Example 1. The optimal set map of the parametric program

$$\min_{(x,y)} \{ y \; / \; y \geqslant x^2 \, , \; y \geqslant t \}, \; t \in R,$$

is not upper Lipschitzian at t=0. Obviously, the optimal sets are $\psi(t) = \{(x,y) \in R^2 \; / \; -\sqrt{t} \leq x \leq \sqrt{t} \, , \; y = t \}$, if $t \geq 0$.

Example 2. (ψ is single-valued) Let G be the function defined by

$$G(x,y) := \begin{cases} |y| \exp\,(-x/|y|) & \text{if } x \geq 0, \; y \neq 0 \\ 0 & \text{if } x \geq 0, \; y = 0 \\ |y| - x & \text{if } x \leq 0. \end{cases}$$

G is convex (cf. Bank et al. 1982, p.52). Consider the problem

$$\min \{ G(x,y) \; / \; x^2 + (y+1)^2 \leq 1 \, , \; y \leq t \} \, , \; t \in R.$$

It is easy to check that $\psi(t) = \{((1 - (1+t)^2)^{1/2}, \; t)\}$ for $-1 \leq t \leq 0$.

When the constraints are given by more complicated convex functions it may even happen that there is no continuous (let alone Lipschitzian) selection of ψ, cf. §4. However, for parametric problems in which the objective function as well as the constraint functions are convex and quadratic (see Example 1 above), there exists for ψ a selection function which satisfies a certain kind of Lipschitz condition (for the proof we refer to Klatte and Kummer 1984):

Theorem 2. Consider the parametric convex problem $P_2(t)$. For each $i \in \{0,1,\dots,m\}$, let h_i be defined as

$$h_i(x) = x^T C^i x + p^{i\,T} x + q_i \, ,$$

where C^i is a symmetric, positive semidefinite (n,n)-matrix, $p^i \in R^n$ and $q_i \in R$. If $\psi(0) \neq \emptyset$, and if the Slater condition is satisfied w.r. to M(0), then for every $x \in \psi(0)$, there are a constant L and a neighborhood U of 0 such that

$$d(x, \psi(t)) \leq L \, \| t \| \quad (\forall t \in U).$$

Remark: If $P_2(t)$ has the special form $\min \{ x^T C^o x + p^{o\,T} x \,/\, Ax \leq t \}$, $t \in R^m$, with fixed vector $p^o \in R^n$ and fixed matrices A and C^o of suitable order (C^o symmetric, positive semidefinite), then ψ is even Lipschitzian on its effective domain $\mathrm{dom}\, \psi := \{ t \,/\, \psi(t) \neq \emptyset \}$, cf. Klatte 1984 a.

3. NON-CONVEX QUADRATIC PROBLEMS

In this paragraph we restrict our considerations to the study of stability of local optimal solutions to the parametric quadratic program

$$P_3(t): \qquad \min \{ f(x,t) \,/\, x \in M(t) \},$$

with the parameter tuple $t = (C,p,A,b)$, where

$$f(x,t) := \frac{1}{2} x^T C x + p^T x \quad , \quad M(t) := \{ x \in R^n \,/\, Ax \leq b \},$$

and C varies over all symmetric (n,n)-matrices, A varies over all (m,n)-matrices, and the parameters p and b are vectors in R^n and R^m, respectively. The set of all such parameter tuples is denoted by T. As for more general classes of parametric problems we only refer to a few publications in which various aspects of current research in our subject are treated. Concerning Lipschitz properties of the infimum function: Rockafellar (1982, 1984), Gauvin and Dubeau (1982), Fiacco (1983). Concerning Lipschitz properties of local minimizers and stationary points (under second-order conditions): Robinson (1982), Fiacco (1983). Concerning continuity properties of local minimizers (in the absence of second-order informations): Robinson (1983), Klatte (1984a,b).

Following Robinson (1983) we shall say that a nonempty set $X \subset R^n$ is a strict local minimizing set for $f(\cdot,t)$ w.r. to $M(t)$, if there is an open set $Q \supset X$ such that $X = \psi_Q(t)$. We recall that $\psi_Q(t) = \{ x \in M(t) \cap \mathrm{cl}\, Q \,/\, f(x,t) = \varphi_Q(t) \}$. Obviously, such a strict local minimizing set is a subset of $\psi_{loc}(t)$, and it is always closed. Typical examples of strict local minimizing sets are the following:

(i) $X=\{z\}$ if z is a strict local minimizer for $f(\cdot,t)$ w.r. to $M(t)$;

(ii) $X=\Psi(t)$ if $\Psi(t)\neq\emptyset$.

Let $KT(t)$ denote the set of Kuhn-Tucker points of the program $P_3(t)$ (for fixed t):

$$(4)\quad KT(t) := \left\{ (x,u)\in R^n\times R^m \;/\; \begin{array}{l} Cx + A^Tu + p = 0, \\ Ax \qquad\quad - b \leq 0, \\ u\geq 0,\ u^T(Ax-b)=0 \end{array} \right\} .$$

The set of stationary points, denoted $SP(t)$, is

$$(5)\quad SP(t) = \pi_n(KT(t)) \qquad (\pi_n := \text{canonical projection to } R^n),$$

the set of Lagrange multipliers at $x\in SP(t)$ is given by

$$LM(x,t) = \{u\in R^m \;/\; (x,u)\in KT(t)\} .$$

As usual, $KT(\cdot)$, $SP(\cdot)$ and $LM(\cdot,\cdot)$ are considered to be multifunctions. The norm in the parameter space T is defined by

$$\| t\|_T := \max\{\|C\|,\ \|p\|,\ \|A\|,\ \|b\|\},\ t=(C,p,A,b),$$

where $\|\cdot\|$ is always the Euclidean norm of the corresponding linear space.

The next theorem covers results by Robinson (1979), who assumes convexity of the initial problem at $t=t^o$, and Hager (1979), who assumes that for all t near t^o the multifunction KT is single-valued. A detailed proof of Theorem 3 is in Klatte (1984a,b).

<u>Theorem 3.</u> Consider problem $P_3(t)$. Let $t^o=(C^o,p^o,A^o,b^o)$ be a given parameter tuple, and let X be a (nonempty) bounded, strict local minimizing set for $f(\cdot,t^o)$ w.r. to $M(t^o)$. Suppose that the Slater condition is satisfied w.r. to $M(t^o)$.

Then $K := KT(t^o)\cap(X\times R^m)$ is nonempty and compact, and there are a bounded, open set $D'\supset K$ and a constant $L>0$ such that the following is true:

(a) If D is any open set with $K\subset D\subset D'$, then one has, for some neighborhood U_D of t^o,

$$\emptyset \neq D\cap KT(t) \subset K + L\,\|t-t^o\|_T\, B_{n+m} \quad (\forall t\in U_D).\ ^{+)}$$

+) $X+Y := \{x+y \,/\, x\in X,\ y\in Y\}$; $\beta X := \{\beta x \,/\, x\in X\}$ $(\beta\in R)$.

(b) If Q is any open set with $X \subset Q \subset \Pi_n(D')$, then one has, for some neighborhood U_Q of t^o,
$\emptyset \neq Q \cap \Psi_{loc}(t) \subset Q \cap SP(t) \subset X + L \|t - t^o\|_T B_n \ (\forall t \in U_Q)$.
Further, the infimum function φ_Q is Lipschitzian around t^o.

In Klatte (1984a,b) there is an example which shows that in Theorem 3 the assumption "X is a bounded, strict local minimizing set" cannot be replaced by the weaker assumption "X is a nonempty, bounded subset of $\Psi_{loc}(t^o)$":

Example 3. It is not difficult to verify that for the parametric program

$$\min \{xy - x^2 \ / \ x \geq t \ , \ y \leq 1\} \ , \ t \in R,$$

we have

$$\Psi_{loc}(0) = \{(0,a) \in R^2 \ / \ 0 < a \leq 1\} \ , \text{ but } \ SP(t) = \emptyset \ \text{ if } t > 0.$$

A further example illustrates the fact that there is no analogy to Theorem 3 with respect to the (global) optimal set mapping Ψ:

Example 4. $\min \{x(1 - tx) \ / \ x \geq 0 \ , \ 1 - tx \geq -t\} \ , \ t \geq -1.$

Obviously,

$$\Psi(t) = \begin{cases} \{0\} & \text{if } -1 \leq t \leq 0, \\ \left\{\frac{1+t}{t}\right\} & \text{if } t > 0 \end{cases}$$

We note that all assumptions of Theorem 3 are fulfilled and, really, $\Psi_{loc}(t) \cap Q \equiv \{0\} \ (\forall t \geq -1)$ with $Q := \{x \ / \ -1 < x < 1\}$.

Outline of proof of Theorem 3.

1^o First we note that for each $x \in X$, the set $LM(x,t^o)$ is nonempty and bounded, since the Slater condition is satisfied w.r. to $M(t^o)$. By Robinson (1982, Th. 2.3), the multifunction $LM(\cdot,t^o)$ is upper semicontinuous on X. This, together with the compactness of X, implies that $K := KT(t^o) \cap (X \times R^m) = X \times \bigcup_{x \in X} LM(x,t^o) \subset X \times Y$, where Y is a compact subset of R^m. With no loss of generality let Y be a polyhedral convex set satisfying $K \subset \text{int } Y$. Since K is obviously closed, K is a compact set.

2^o The representations (4) and (5) tell us that $SP(t^o)$ is a union of finitely many polyhedral convex sets $X_1,\dots,X_N$ ($X_k \neq \emptyset \ \ \forall k$). Define

$$I(X) := \{ i \in \{1,\dots,N\} \ / \ X \cap X_i \neq \emptyset \} .$$

If x is an arbitrary point of $X \cap X_i$, $i \in I(X)$, then for each $y \in X_i$, we obtain $f'(x,t^o;y-x) \geq 0$ and $f'(y,t^o;x-y) \geq 0$, where $f'(z,t^o;w)$ is the directional derivative of $f(\cdot,t^o)$ at z in the direction w (note that $x,y \in SP(t^o)$ and that the vectors $y-x$ and $x-y$ are feasible directions for $M(t^o)$ at x resp. y). Hence,

$$f(x,t^o) = f(y,t^o) \quad (\forall x \in X \cap X_i \quad \forall y \in X_i \quad \forall i \in I(X)).$$

Since X is a strict local minimizing set, this implies that $X_i \subset X$ ($\forall i \in I(X)$). Because of the compactness of X there is a number $\varepsilon > 0$ such that $(X + \varepsilon B_\infty) \cap X_j = \emptyset$ ($\forall j \notin I(X)$), where B_∞ is the unit cube in R^n. Setting $Q' := X + \text{int}\, \varepsilon B_\infty$, we thus have

$$X = \bigcup_{i \in I(X)} X_i = SP(t^o) \cap \text{cl } Q'.$$

3^o(Lipschitz property) Let A_I and $A_{\bar{I}}$ (or b_I, $b_{\bar{I}}$) denote the submatrix of A (or the subvector of b) which is built, for $i \in I$ or $i \in \bar{I} := \{1,\dots,m\} \setminus I$, by the rows a^i of A (or the components b_i of b). Because of the special structure of $KT(t)$ we can split $KT(t)$ into components $F^{I,J}(t)$ as follows:

$$KT(t) = \bigcup_{(I(t),J(t)) \in Z} F^{I(t),J(t)}(t),$$

where, for $t=(C,p,A,b)$ and $I,J \subset \{1,\dots,m\}$,

$$F^{I,J}(t) := \left\{ (x,u) \ / \ \begin{array}{l} Cx + A^T u + p = 0 , \ A_I x = b_I \\ A_{\bar{I}} x \leqq b_{\bar{I}} , \ u_J = 0, \ u_{\bar{J}} \geqslant 0 \end{array} \right\}$$

and

$$Z := \{ (I,J) \in \{1,\dots,m\} \times \{1,\dots,m\} \ / \ I \cup J = \{1,\dots,m\} \} .$$

Set $D_i' := (X_i + \text{int}\, \varepsilon B_\infty) \times \text{int } Y \quad (i \in I(X))$ and define

$$Z_i := \{ (I,J) \in Z \ / \ F^{I,J}(t^o) \cap \text{cl } D_i' \neq \emptyset \} \quad (i \in I(X)).$$

By 1^o and 2^o, $KT(t^o) \cap D_i' \neq \emptyset$ and so $Z_i \neq \emptyset$ for all $i \in I(X)$, thus $KT(t^o) \cap \text{cl } D_i'$ has the representation

$$KT(t^o) \cap \text{cl } D_i' = \bigcup_{(I,J) \in Z_i} (F^{I,J}(t^o) \cap \text{cl } D_i') \quad (\forall i \in I(X)).$$

Taking the compactness of $\text{cl } D_i'$ into account and using the fact that the multifunctions $t \longrightarrow KT(t) \cap \text{cl } D_i'$ and $t \longrightarrow F^{I,J}(t) \cap \text{cl } D_i'$ are closed (cf. Bank et al. 1982, Th. 3.1.1), it is easy to show that, for some neighborhood U_i of t^o,

$$KT(t)\cap \mathrm{cl}\, D_i' = \bigcup_{(I,J)\in Z_i} (F^{I,J}(t)\cap \mathrm{cl}\, D_i') \quad (\forall i\in I(X)\ \forall t\in U_i).$$

By Daniel (1973), the multifunctions $F^{I,J}(\cdot)\cap \mathrm{cl}\, D_i'$ are upper Lipschitzian (note that $\mathrm{cl}\, D_i'$ are convex polyhedra, by construction). Then it follows that $KT(\cdot)\cap D'$ is also upper Lipschitzian at t^o, where $D' := \bigcup_{i\in I(X)} D_i'$.
Because of $K = KT(t^o)\cap((SP(t^o)\cap \mathrm{cl}\, Q')\times Y)$ (by 1° and 2°) and hence $K = KT(t^o)\cap \mathrm{cl}\, D'$ we have obtained the Lipschitz property of part (a) (which obviously also holds for any open set D with $K\subset D\subset D'$). The Lipschitz property of assertion (b) follows by standard arguments from the fact that $SP(t) = \pi_n(KT(t))$.

4° (solvability) Let Q be any open set satisfying $X\subset Q\subset Q'$. Then there is a point $x_Q\in Q$ such that $A^o x_Q \angle b^o$, and hence we can find a neighborhood V of (A^o,b^o) such that $Ax_Q < b$ $(\forall (A,b)\in V)$. Thus, the sets $\{x\in \mathrm{cl}\, Q \;/\; Ax\leq b\}$ are nonempty and compact for all $(A,b)\in V$. For all $t=(C,p,A,b)$ with $(A,b)\in V$, we have, by the Weierstraß theorem,

$$\psi_Q(t) \neq \emptyset .$$

Further, Berge's (1963) stability results provide that ψ_Q is upper semicontinuous at t^o. Hence, $\psi_Q(t)\subset Q$ if $\|t - t^o\|_T$ is sufficiently small, and so there is a neighborhood U_Q of t^o such that

$$\emptyset \neq \psi_Q(t)\subset \psi_{loc}(t)\cap Q \qquad (\forall t\in U_Q).$$

The Lipschitz continuity of φ_Q easily follows from the compactness of X and the Slater condition (by application of Lemma 1. Hence (b) is shown.
Concerning the remaining assertion of part (a) we only mention that if D is any open set with $K\subset D\subset D'$, then it is not difficult to derive that $KT(t)\cap D$ is nonempty if $\|t-t^o\|_T$ is sufficiently small; one has to apply part (b) which is already shown and to take into account the upper semicontinuity of the multifunction $LM(\cdot)$ on $X\times\{t^o\}$ (cf. again Robinson 1982, Th. 2.3), the details are omitted here. //

Remark: In the case of fixed matrices $C=C^o$ and $A=A^o$ the (global) optimal set map ψ is upper Lipschitzian on $R^n \times R^m$, and the infimum function φ is Lipschitzian on each bounded convex subset of dom ψ := $\{(p,b) \in R^n \times R^m \ / \ \psi(p,b) \neq \emptyset\}$, cf. Klatte (1983). The set dom ψ is, in this case, a union of finitely many polyhedral convex sets.
We further note that (for the parametric program $P_3(t)$) the inclusion $Q \cap \psi_{loc}(t) \subset Q \cap SP(t)$ in part (b) of Theorem 3 may be strict (see Robinson (1982, p.213)).

4. OPTIMAL AND ε-OPTIMAL SELECTIONS

In this last section we consider the existence of a continuous or Lipschitzian function s which assigns to each $t \in T$ a single point $s(t) \in \psi(t)$ (or $s(t) \in \psi_\varepsilon(t)$); such a function s will be called an optimal selection (or ε-optimal selection). Obviously, this question is closely related to the more general theory of continuous selections for arbitrarily given multifunctions $F: T \longrightarrow 2^{R^n}$, where the basic results are wellknown from Michael's famous papers (cf. Michael 1956). In particular, a continuous selection for F exists if F is lower semicontinuous on T, and F(t) is nonempty and convex for all $t \in T$. As it concerns Lipschitzian selections we mention here

Theorem 4. Let T be compact and $F: T \longrightarrow 2^{R^n}$ be a Lipschitzian multifunction with modulus L, and suppose that the sets F(t), $t \in T$, are nonempty, convex and compact. Then there is a Lipschitzian selection s for F with modulus $n \cdot L$.

Two independent and different proofs have been given by Dommisch (1983) and, for a slightly modified version of the preceding theorem, by Aubin and Cellina (1982). Note that Dommisch's Lipschitz modulus $n \cdot L$ for s (provided that F has the modulus L) is better than the one obtained by Aubin and Cellina.

However, the existence of a Lipschitzian selection is not a privilege of Lipschitzian multifunctions only:

Theorem 5. Let T be compact and $F: T \longrightarrow 2^{R^n}$ be a multifunction with nonempty and convex images F(t) for all $t \in T$. Suppose further all sets $F^-(x) := \{t \in T \ / \ x \in F(t)\}$ $(x \in R^n)$

to be open (w.r. to the induced topology).
Then there is a Lipschitzian selection s for F.

Proof: We adapt the well-known idea of the partition of unity. Obviously,

$$T = \bigcup_{x \in R^n} F^-(x) .$$

Since T is compact and the sets $F^-(x)$ are open, there are finitely many points x^k $(k=1,\dots,N)$ such that

$$T = \bigcup_{k=1}^{N} F^-(x^k).$$

The closed sets $A_k := T \setminus F^-(x^k)$ $(k=1,\dots,N)$ then fulfil

$$\bigcap_{k=1}^{N} A_k = \emptyset .$$

Let $d_k\colon T \longrightarrow R$ be the distance functions $d_k(t) := d(t,A_k)$ $(\forall k)$, therefore

$$d(t) := \sum_{k=1}^{N} d_k(t) > 0 \quad (\forall t \in T).$$

Moreover, each d_k is Lipschitzian (with modulus 1). Since T is compact, we observe that $l := \inf_{t \in T} d(t) > 0$, and the function s defined by

$$s(t) := \sum_{k=1}^{N} d_k(t) \cdot d(t)^{-1} \, x^k$$

is therefore again Lipschitzian with a modulus depending on N, l and $\max_k \|x^k\|$. Because of

$$s(t) \in \operatorname{conv} \left\{ x^k \,/\, x^k \in F(t) \right\} \subset F(t)$$

("conv":= convex hull) the proposition is true. //

In the case $F = \psi$, the application of the Theorems 4 and 5 is difficult, because its hypotheses are usually too strong. However, if we put $F(t) = \psi_\varepsilon(t)$ both theorems allow immediate proof of the following corollaries.

Corollary 1. Consider the parametric convex problem $P_1(t)$ and suppose the assumptions of Theorem 1 to be satisfied for all $t' \in T'$, where T' is a compact convex subset of T. Then there is a number $\tilde{\varepsilon} > 0$ such that for all $0 < \varepsilon < \tilde{\varepsilon}$ there is a Lipschitzian ε-optimal selection on T'.

Proof: Apply Theorem 1 and Theorem 4. //

Corollary 2. Consider the parametric convex problem $P_1(t)$ in the case $r=0$ (without equality constraints) and suppose that for all elements t of a compact subset T' of T, $\psi(t)$ is nonempty and bounded, and the Slater condition is satisfied w.r. to $M(t)$.
Then, for each $\varepsilon > 0$, there is a Lipschitzian ε-optimal selection on T'.
Proof: Apply Theorem 5 to the map $F(t):=\{x \,/\, f(x,t)<\varphi(t)+\varepsilon,\ g(x,t)<0\}$. //

Even if we have right-hand side perturbations only, the suppositions of Theorem 1 (or Corollary 1) do not guarantee the existence of a continuous optimal selection:

Example 5. Consider the parametric convex program

$$\min_{(x,y,z)} \left\{ G(x,y)+z \;/\; \begin{array}{l} G(1-x,y) \leq t_1 + z \\ \quad\quad\quad\ y \geq t_2 + z \\ 0 \leq x,y,z \leq 1 \end{array} \right\},$$

where G is defined as in Example 2. For $t=(0,0)$ there is a Slater point (with $x=\frac{1}{2}$), but no selection of ψ is continuous at $t=(0,0)$. Indeed, setting $t_1=t_2=q$ $(q \longrightarrow +0)$ one easily verifies that the only solutions are

$$x_q = 1, \quad y_q = q, \quad z_q = 0.$$

In the case $t_1 = q\exp(-(2q)^{-1})$, $t_2 = q$ $(q \longrightarrow +0)$, however, the only solutions are

$$x_q = \frac{1}{2}, \quad y_q = q, \quad z_q = 0.$$

Thus, a selection of ψ which is continuous at (0,0) cannot exist.

Finally, we give an example which shows that in Theorem 4 the convexity assumption cannot be dropped, in general. This is an example of a closed Lipschitzian multifunction F with nonempty and compact images, but without any continuous selection.

Example 6. Let $T = B_2$ be the unit ball of R^2. For $t \neq 0$ we put

$$u(t) := t \cdot \|t\|^{-1} \quad \text{and} \quad Q(t) := \left\{ x \in R^2 \,/\, \|x-u(t)\| \geq \|t\| - \tfrac{1}{2} \right\}.$$

Now, define

$$F(t) := \text{bd } B_2 \cap Q(t) \quad \text{with} \quad \text{bd } B_2 = \{t \,/\, \|t\| = 1\}.$$

Then F is Lipschitzian with modulus 3π and, since $t \notin F(t)$ for all $t \in T$, there is no continuous selection s for F; otherwise the function s would have a fixed point $t = s(t) \in F(t)$.

5. REFERENCES

Aubin, J.P. and A. Cellina (1982). Differential Inclusions. Manuscript, CEREMADE, Université de Paris-Dauphine, (Book edition in Springer-Verlag, Berlin-Heidelberg-New York).

Bank, B., J. Guddat, D. Klatte, B. Kummer and K. Tammer (1982). Non-Linear Parametric Optimization. Akademie-Verlag, Berlin.

Berge, C. (1963). Topological Spaces. Macmillan, New York.

Blatt, H.-P. (1980). Lipschitz-Stabilität von Optimierungs- und Approximationsaufgaben. Numerical Methods of Approximation Theory, ISNM 52, Birkhäuser Verlag, Basel, pp. 9-28.

Cornet, B. (1983). Sensitivity analysis in optimization. CORE Discussion Paper No. 8322, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Daniel, J.W. (1973). On perturbations in systems of linear inequalities. SIAM J. Numer. Anal., Vol. 10, 299-307.

Dommisch, G. (1983). Zur Existenz von Lipschitz-stetigen, differenzierbaren bzw. meßbaren Auswahlfunktionen für mengenwertige Abbildungen. Forschungsstudie, Humboldt-Universität Berlin, Sektion Mathematik.

Eremin, I.I. and N.N. Astafiev (19??). Introduction to the Theory of Linear and Convex Programming. Nauka, Moscow (in Russian).

Fiacco, A.V. (1983). Introduction to Sensitivity and Stability Analysis in Nonlinear Programming. Academic Press, New York.

Gauvin, J. and F. Dubeau (1982). Differential properties of the marginal function in mathematical programming. Math. Programming Study 19, 101-119.

Hager, W.W. (1979). Lipschitz continuity for constrained processes. SIAM J. Control Optim., Vol. 17, 321-338.

Klatte, D. (1983). On the Lipschitz behavior of optimal solutions in parametric problems of quadratic optimization and linear complementarity. IIASA-WP-83-121, International Institute for Applied Systems Analysis, Laxenburg/Austria.

Klatte, D. (1984a). Beiträge zur Stabilitätsanalyse nichtlinearer Optimierungsprobleme. Dissertation B, Humboldt-Universität Berlin, Sektion Mathematik.

Klatte, D. (1984b). On stability of local and global optimal solutions in parametric problems of nonlinear programming, Part II. (forthcoming in: Seminarberichte der Sektion Mathematik der Humboldt-Universität Berlin).

Klatte, D. and B. Kummer (1984). On the (Lipschitz-) continuity of solutions of parametric optimization problems. In: K. Lommatzsch (ed.): Proceedings der 16. Jahrestagung "Mathematische Optimierung" Sellin 1984, Seminarbericht, Sektion Mathematik der Humboldt-Universität Berlin.

Michael, E. (1956). Continuous selections I-III. Ann. of Math., (Part I:) Vol. 63 (1956), 361-382; (Part II:) Vol. 64 (1956), 562-580; (Part III:) Vol. 65 (1957), 375-390.

Robinson, S.M. (1976). Regularity and stability for convex multivalued functions. Math. of Operations Research, Vol. 1, 130-143.

Robinson, S.M. (1979). Generalized equations and their solutions, Part I: Basic theory. Math. Programming Study 10, 128-141.

Robinson, S.M. (1981). Some continuity properties of polyhedral multifunctions. Math. Programming Study 14, 206-214.

Robinson, S.M. (1982). Generalized equations and their solutions, Part II: Applications to nonlinear programming. Math. Programming Study 17, 200-221.

Robinson, S.M. (1983). Persistence and continuity of local minimizers. Preprint, University of Wisconsin-Madison. (Also Collaborative Paper CP-84-5, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Rockafellar, R.T. (1982). Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming. Math. Programming Study 17, 28-66.

Rockafellar, R.T. (1984). Lipschitzian properties of multifunctions. (to appear)

Walkup, D. and R. Wets (1969). A Lipschitzian characterization of convex polyhedra. Proceed. Amer. Math. Soc., Vol. 23, 167-173.

ON METHODS FOR SOLVING OPTIMIZATION PROBLEMS WITHOUT USING DERIVATIVES

K. Lommatzsch and Nguyen Van Thoai
Department of Mathematics, Humboldt University, 1086 Berlin, GDR

INTRODUCTION

'Smooth' methods have been developed and used because under the assumption of smoothness it is possible to use the methods of differential calculus. For example, there are a great number of methods for solving convex optimization problems in which both the minimized objective and the set of feasible points can be expressed with the aid of differentiable convex functions. In some cases, however, the problems connected with the calculation of gradients have led to the development of algorithms which do not use derivatives. (Nevertheless, differentiability is still necessary to prove optimality, convergence assertions, etc.) The most successful optimization method - the well-known simplex method of linear programming - does not use derivatives. On the other hand, there are methods which make partial use of gradients, linearization etc., but which do not depend on differentiability assertions to prove their convergence.

In Section 1 of this note we consider two such methods and in Section 2 we present an algorithm for concave programming problems which is based on a branch-and-bound technique.

1. METHODS OF CENTERS AND OF POINTS OF GRAVITY

The problem can be formulated as follows:

(P1) $\min\{f(x) \mid x \in M\}$,

where $f(x)$ is a convex function defined on R_n and M is an (n-dimensional) convex compact subset of R_n.
The main idea of Huard's method of centers (cf. [1]) consists, roughly speaking, in calculating the centers of the sets $M(t) = \{x \in M \mid f(x) \leq t\}$ by using certain distance functions $d(x,t)$ defined on $M(t)$. If $M = \{x \in R_n \mid g_i(x) \leq 0,\ i=1,\dots,m\}$, then the distance function can be defined as follows:

$$d(x,t) = \max\{g_1(x),\dots,g_m(x),\ f(x) - t\} .$$

Then the algorithm is of the following general form:

step 1: t_o given, set $k \leftarrow 0$;

step 2: Compute x^{k+1} as a solution of

$$\min\{d(x,t_k) \mid x \in M(t_k)\} ;$$

step 3: $t_{k+1} = \varrho f(x^{k+1}) + (1-\varrho)t_k, \quad \varrho \in (0,1]$;

step 4: Set $k \longleftarrow k+1$ and go to step 2.

Under certain assumptions the convergence of this algorithm can be proved. As the solution of step 2 is connected with considerable difficulties, P. Huard and others suggested to replace the problem of step 2 by some other problem (e.g. linearization of functions occurring in the description of the set M by using gradients, cf. [1]).

The idea of the method of points of gravity is based on computing the points of gravity in the sets M(t) mentioned above, cf. [2]. In the algorithm described above we have to replace only step 2 by

step 2': Compute the points of gravity x^{k+1} of the set $M(t_k)$.

Under certain assumptions the algorithm converges to one of the points of solution of problem (P1). Similarly to the preceding algorithm, the subproblems contained in step 2' are very difficult. Nevertheless, these subproblems can be replaced by computing the points of gravity of finitely many boundary points of the sets $M(t_k)$, e.g. if $x^k \in \operatorname{int} M(t_k)$ and $d^1, \ldots, d^n$ is a given system of orthogonal directions on R_n, then

step 2": $$x^{k+1} = \frac{1}{2n} \sum_{s=1}^{n} (\underline{r}^s + \bar{r}^s),$$

where for $s = 1, \ldots, n$

$$\underline{r}^s = x^k + \underline{\alpha}_s d^s, \quad \bar{r}^s = x^k + \bar{\alpha}_s d^s \ ,$$

$$\underline{\alpha}_s = \min\{\alpha \in R_1 \mid x^k + \alpha d^s \in M(t_k)\} \ ,$$

$$\bar{\alpha}_s = \max\{\alpha \in R_1 \mid x^k + \alpha d^s \in M(t_k)\} \ .$$

Of course, if step 2" is used in the algorithm, the rate of convergence and the numerical properties of the algorithm depend to a high degree on the geometrical properties of the sets $M(t_k)$ and on the position of the points x^k in $M(t_k)$. On the other hand, the algorithm needs only very simple calculations.

2. AN ALGORITHM FOR SOLVING CONCAVE OPTIMIZATION PROBLEMS

We consider the problem

(P2) $$\min\{f(x) \mid x \in M\} \ ,$$

where $f(x)$ is a concave function defined on R_n and M is an (n-dimensional) compact convex subset of R_n. It is well-known that

a) there always exists an extremal point $e \in M$ such that $f(e) \leq f(x)$ for all $x \in M$;

b) if $f(x)$ is concave on the halfline $H(x^0)$ with the initial point x^0 and if there exists a point $x^1 \in H(x^0)$ where $f(x^1) < f(x^0)$, then the function $f(x)$ decreases unboundedly along $H(x^0)$;

c) if the concave function $f(x)$ is bounded from below along the halflines $H^1(x^0), \ldots, H^r(x^0)$ with common initial point x^0, then $f(x)$ is bounded also on the convex hull of these halflines and $f(x^o) \leq f(x)$ for all $x \in co(H^1(x^o),\dots,H^r(x^o))$.

The main idea of the algorithm proposed by Hoang Tuy and Nguyen Van Thoai (cf. [3], varied and implemented for a polyhedral set M by N.V. Thoai in [4]) consists in covering the constraint set M by a system of polyhedral cones K^i, $i=1,2,\dots$, in computing lower bounds of the objective function f(x) on the sets $K^i \cap M$ (bounding) and in bisecting a cone K^i which belongs to one of the smallest lower bounds (branching) and so on. In the algorithm, polyhedral cones K having a common vertex w^o, $w^o \in \text{int}M$, are used. Each of these cones has exactly n edges $H^j = \{x \in R_n \mid x = w^o + \tau(u^j - w^o), \tau \geq 0\}$, $j \in J(K) = \{j_1,\dots,j_n\}$, where w^o, $u^{j_1},\dots,u^{j_n}$ is a system of linearly independent points in R_n.

A. Computation of lower bounds of the objective function f(x) on $K \cap M$.

For $j \in J(K)$ and for a parameter γ, which is characteristic of the algorithm, we determine:

a) $w^j = w^o + {}_j(u^j - w^o)$,
where $\tau_j = \max\{\tau \geq 0 \mid w^o + \tau(u^j - w^o) \in M\}$;

b) $\beta(K,\gamma) = \min\{\gamma; f(w^o); f(w^j),\ j \in J(K)\}$;

c) $\eta_j(\gamma) = \sup\{\eta \geq 0 \mid f(w^o + \eta(w^j - w^o)) \geq \beta(K,\gamma)\}$;

d) $G(K,\gamma) = \{j \in J(K) \mid \eta_j(\gamma) < \infty\}$;

e) $\bar{\eta}_j(\gamma) = \min\{\eta_j(\gamma), c\}$,
where c is a given, sufficiently large number;

f) $y^j(\gamma) = w^o + \bar{\eta}_j(\gamma)\ (u^j - w^o)$,
obviously $f(y^j(\gamma)) \geq \beta(K,\gamma)$;

g) $z^j(\gamma) = w^o + \bar{\alpha}(K,\gamma)\ (y^j(\gamma) - w^o)$,
where $\bar{\alpha}(K,\gamma)$ is the optimal value of the optimization problem :

h) $\max\left\{\sum_{j \in J(K)} \lambda_j \mid w^o + \sum_{j \in J(K)} \lambda_j (y^j(\gamma) - w^o) \in M,\ \lambda_j \geq 0,\ j \in J(K)\right\}$;

i) $$\varrho(K,\gamma) = \begin{cases} \beta(K,\gamma) \text{ if } G(K,\gamma) = \emptyset \text{ or } \bar{\alpha}(K,\gamma) \leq 1 , \\ \min\{\beta(K,\gamma); f(z^j(\gamma)),\ j \in J(K)\} \text{ otherwise .} \end{cases}$$

Obviously $\varrho(K,\gamma) \leq f(x)$ for all $x \in K \cap M$.

B. Bisection of the convex cone K.

We determine one of the longest edges of the (n-1)-dimensional simplex which is generated by the points $u^{j_1},\dots,u^{j_n}$. Let it have the endpoints u^{j_r} and u^{j_s}, $j_r, j_s \in J(K)$. With the aid of the bisection point $u^{j_{n+1}} = 1/2(u^{j_r} + u^{j_s})$ and the edges of K we define two new cones K^1 and K^2 with vertex w^o: K^1 has the edges $H^j(w^o)$, $j \in J(K^1) = \{j_1,\dots,j_{r-1},j_{r+1},\dots,j_{n+1}\}$, and K^2 has the edges $H^j(w^o)$, $j \in J(K^2) = \{j_1,\dots,j_{s-1},j_{s+1},\dots,j_{n+1}\}$. In [3] it was shown that a sequence of cones $\{K^i\}_{i=1}^{\infty}$, where K^{i+1} is constructed from K^i by the bisection process described above, converges to a halfline with the initial point w^o.

Algorithm. (Step 0): Let $w^o \in \operatorname{int} M$ be given and n+1 linearly independent points $v^1,\dots,v^{n+1}$, where $w^o \in \operatorname{int} co(v^1,\dots,v^{n+1})$. Further, let $L^o = \{K^1,\dots,K^{n+1}\}$, where K^i, $i \in I^o = \{1,\dots,n+1\}$, is a cone with vertex w^o and edges H^j, $j \in J(K^i) = \{1,\dots, i-1, i+1,\dots,n+1\}$.

step 1: For $i \in I^o$ compute the points $w^i = w^o + \tau_i(v^i - w^o)$ according to formula a) above,
construct the set
$W^o = \{w^o, w^1,\dots,w^{n+1}\}$,
compute the number
$\gamma_o = \min\{f(w^i),\ i=0,1,\dots,n+1\}$,
and determine a point $x^o \in W^o$ with $f(x^o) = \gamma_o$;

step 2: For $i \in I^o$ compute the lower bounds $\varrho(K^i,\gamma_o)$ defined in i) above and set $\mu_o = \min\{\varrho(K^i,\gamma_o\},i \in I^o\}$;

step 3: $k \leftarrow 0$;

step 4: If $\mu_k = \gamma_k$, then stop;

step 5: Otherwise, for an index $i \in I^k$ with $\varrho(K^i,\gamma_k)\ \mu_k$
bisect the cone K^i into the cones K^{n+2+2k} and K^{n+3+2k}, (v^{n+2+k} be the bisection point) and set
$I^{k+1} = (I^k \setminus \{i\}) \cup \{n+2+2k\} \cup \{n+3+2k\}$;

step 6: Compute the point

$$w^{n+2+k} = w^o + \tau_{n+2+k}(v_{n+2+k} - w^o)$$

according to formula a) above,
construct the set $W^{k+1} = W^k \cup \{w^{n+2+k}\}$
and compute $\gamma_{k+1} = \min\{\gamma_k, f(w^{n+2+k})\}$,
if $\gamma_{k+1} < \gamma_k$, then set $x^{k+1} \leftarrow w^{n+2+k}$,
otherwise $x^{k+1} \leftarrow x^k$;

step 7: For r=2,3 compute the lower bounds $\varrho(K^{n+r+2k}, \gamma_{k+1})$ and set

$$\mu_{k+1} = \min\{\varrho(K^i, \gamma_{k+1}), i \in I^{k+1}\};$$

step 8: $k+1 \leftarrow k$ and go to step 4.

Remarks:

1.) This algorithm either yields an optimal solution after finitely many cycles or it generates an infinite sequence of points $\{x^k\}$ which converges to an optimal point of problem (P2) (cf. [3],[4]). In each cycle we have to solve a convex optimization problem (compare step 7 and k) above) with a linear objective function (for this purpose we can use the method of points of gravity from section 1).

2.) If in problem (P2) the set M of feasible points is polyhedral, then the steps 0,1 and 2 of the algorithm can be shortened: A nondegenerated vertex of M may serve as initial point w^o, the points $v^1, \dots, v^{n+1}$ can be dropped and the points $w^1, \dots, w^n$ (cf. a) and step 6 above) can be computed immediately as the vertices of M adjacent to w^o, the start set L^o contains one cone only. The optimization problem of step 7 is linear. For this case, in [4] an implemented algorithm which is written in FORTRAN and tested on a computer ESER 1022 is presented; some smaller examples are also given there.

REFERENCES

[1] Huard, P.: Programmation mathematique convexe, RIRO 2 (1968) 7, pp. 43-59.

[2] Lommatzsch, K.: Ein Gradienten- und Schwerpunktverfahren der linearen und nichtlinearen Optimierung, Aplikace Mat. 11 (1966), pp. 303-313.

[3] N.V. Thoai, H. Tuy: Convergent Algorithms for Minimizing a Concave Function, Math. of O.R 5 (1980), pp. 556-566.

[4] N.V. Thoai: Verfahren zur Lösung konkaver Optimierungsaufgaben auf der Basis eines verallgemeinerten Erweiterungsprinzips, Diss. (B), Humboldt-Universität Berlin, 1984.

AN ACCELERATED METHOD FOR MINIMIZING A CONVEX FUNCTION OF TWO VARIABLES

F.A. Paizerova
Department of Applied Mathematics, Leningrad State University,
Universiteskaya Nab. 7/9, Leningrad 199164, USSR

A method for minimizing a convex continuously differentiable function of two variables was proposed in [1], where it was shown that its rate of convergence is geometric with coefficient 0.9543. We shall describe two modifications of this method with improved convergence rates.

Let $Z \in E_2$, a function f be convex and continuously differentiable on E_2. Assume that we know that a minimum point of f is contained in a convex quadrilateral ABCD. The area of this quadrilateral is called the uncertainty area. Let R be the point of intersection of the diagonals of the quadrilateral. Let us choose four points M,N,Q,P on intervals AC and BC which are all at the same distance ε from R (where $\varepsilon > 0$ is fixed).

Now let us compute the function f at these points and at the point R (see Figure 1).

Case 1

$$f(Q) > f(R), \quad f(P) > f(R) \quad , \tag{1}$$

$$f(M) > f(R), \quad f(N) > f(R) \quad . \tag{2}$$

In this case R is (within ε-accuracy) a minimum point of f on AC and BD, and then by the properties of continuously differentiable functions the point R is a minimum point of f on ABCD (to within the given accuracy ε) and the process terminates.

Case 2. If inequality (1) is satisfied but inequality (2) is not, then R is a minimum point of f on BD. If $f(M) < f(R)$ then

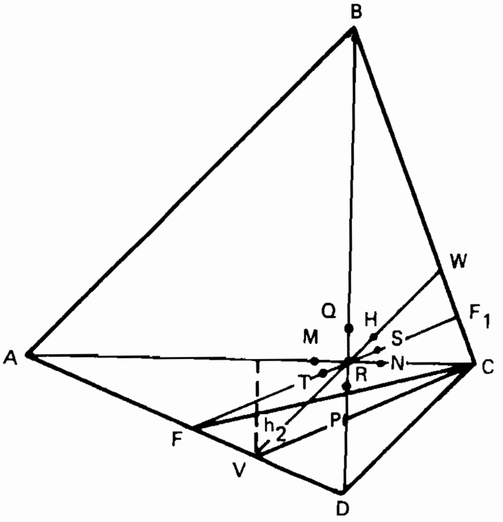


Fig. 1

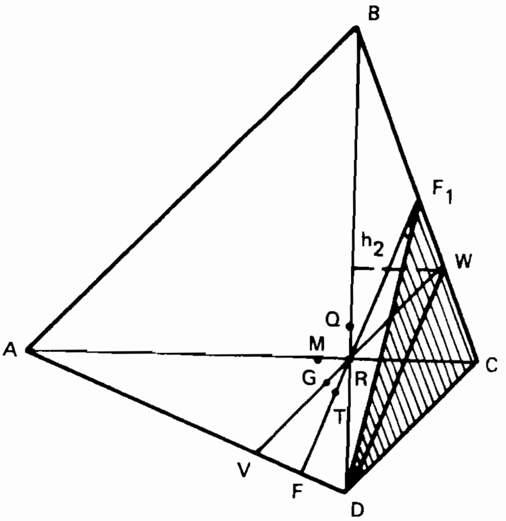


Fig. 2

$$f(Z) > f(R) \qquad \forall Z \in BDC$$

and therefore a minimum point of f lies within the triangle ABD. If $f(N) < f(R)$ then

$$f(Z) > f(R) \qquad \forall Z \in ABD$$

and a minimum point of f lies within the triangle BDC.

Case 3. If inequality (2) is satisfied but (1) is not then we argue analogously.

These three cases were discussed in [1] and are treated in the same way here. The difference between our method and that of [1] is demonstrated in the following case 4.

Case 4. Suppose that both inequalities (1) and (2) are satisfied. Then there exist two points (say, M and Q) such that

$$f(M) < f(R), \ f(Q) < f(R) \quad .$$

It follows from the convexity of f that

$$f(Z) > f(R) \qquad \forall Z \in DRC \quad .$$

Let us draw the line VW which passes through the point R and is parallel to the line DC. On the interval VW let us choose two points G and H at a distance ε from R. If $f(H) \geq f(R)$ and

$f(G) \geq f(R)$ then R is (within ε-accuracy) a minimum point of the function $f(Z)$ on the line VW (see [2]) and since $f(M) < f(R)$ then

$$f(Z) > f(R) \qquad \forall\, Z \in VWCD \quad .$$

This case was also discussed in [1]. The case left to be discussed is the one where either $f(H) < f(R)$ or $f(G) < f(R)$. At this point our method diverges from the method described in [1]. We will suggest two modifications of this method. For the sake of argument assume that $f(H) < f(R)$.

1. First modification. It is assumed that

$$f(H) < f(R) \quad .$$

Then (see Figure 1)

$$f(Z) > f(R) \qquad \forall\, Z \in VRCD \quad .$$

Moreover,

$$f(Z) > f(R) \qquad \forall\, Z \in VCD \quad .$$

Let us draw the line FF_1 which passes through the point R and is parallel to the line VC. On the interval FF_1 let us choose two points T and S at a distance ε from R.

If

$$f(T) \geq f(R) \quad \text{and} \quad f(S) \geq f(R)$$

then R is (within ε-accuracy) a minimum point of f on FF_1 and

$$f(Z) > f(R) \qquad \forall\, Z \in FF_1CD \quad .$$

If

$$f(S) < f(R) \text{ then}$$

$$f(Z) > f(R) \qquad \forall\, Z \in FRCD$$

and furthermore,

$$f(Z) > f(R) \qquad \forall Z \in FCD \quad .$$

As a result we get the quadrilateral ABCF which contains a minimum point of the function f. Let us compute the ratio of the areas of the quadrilaterals ABCF and ABCD.

Assume that

$$\frac{RD}{BR} = \alpha, \ \frac{AR}{RC} \geq \alpha, \ \frac{RC}{AR} = \alpha_1 \geq \alpha \quad .$$

Let h be the height of the triangle ABC. Then

$$S_{ABCD} = \frac{1}{2}\,(1+\alpha)AC\cdot h; \quad S_{ACD} = \frac{1}{2}\,\alpha\, AC\cdot h \quad ,$$

$$RC = \frac{\alpha_1}{(1+\alpha_1)}\,AC \quad .$$

Here S_{ABC} is the area of the triangle ABC. We have

$$S_{VCD} = S_{DRC} = \frac{1}{2}\,\alpha\cdot h\cdot RC = \frac{\alpha}{2}\,\frac{\alpha_1}{(1+\alpha_1)}\,AC\cdot h \quad .$$

Let us define h_2. Since

$$S_{AVC} = \frac{1}{2}\,AC\cdot h_2 \quad \text{and } S_{AVC} = S_{ACD} - S_{VCD} =$$

$$= \frac{1}{2}\,\alpha\,\cdot AC\cdot h - \frac{\alpha\cdot\alpha_1}{2(1+\alpha_1)}\,AC\cdot h = \frac{\alpha}{2(1+\alpha_1)}\,AC\cdot h$$

we have

$$h_2 = \frac{S_{AVC}}{\frac{1}{2}AC} = \frac{\alpha}{1+\alpha_1}\quad h \quad .$$

This leads to

$$S_{FVC} = S_{VRC} = \frac{1}{2}\,RC\cdot h_2 = \frac{\alpha\cdot\alpha_1}{2(1+\alpha_1)^2}\,AC\cdot h \quad ,$$

$$S_{FCD} = S_{VCD} + S_{FVC} = \frac{\alpha\cdot\alpha_1}{2(1+\alpha_1)}\,AC\cdot h + \frac{\alpha\cdot\alpha_1}{2(1+\alpha_1)^2}\,AC\cdot h =$$

$$= \frac{\alpha\cdot\alpha_1(2+\alpha_1)}{2(1+\alpha_1)^2} = AC\cdot h \quad .$$

Hence, the ratio of the area of the quadrilateral ABCF to the area of the quadrilateral ABCD is

$$1 - \frac{\alpha\cdot\alpha_1(2+\alpha_1)}{(1+\alpha)(1+\alpha_1)^2} \quad . \tag{3}$$

Since

$$\frac{\alpha_1(2+\alpha_1)}{(1+\alpha_1)^2} \geq \frac{\alpha(2+\alpha)}{(1+\alpha)^2} \text{ if } \alpha_1 \geq \alpha \text{ this result implies}$$

$$1 - \frac{\alpha\cdot\alpha_1(2+\alpha_1)}{(1+\alpha)(1+\alpha_1)^2} \leq 1 - \frac{\alpha^2(2+\alpha)}{(1+\alpha)^3} \quad . \tag{4}$$

If we decrease the uncertainty area as shown in Figure 2, similar arguments lead us again to (4).

If at some step it turns out that $\frac{RD}{BR} = \alpha \leq \alpha_0$ (where α_0 will be defined later) then we draw a line passing through D and parallel to AC, and then extend AB and BD until they intersect this line (see Figure 3). Instead of the quadrilateral ABCD let us take the triangle A_1BC_1. In the case of a quadrilateral we had four lines passing through R. In the case of a triangle we take the point of intersection of its medians (the point R_1) instead of R.

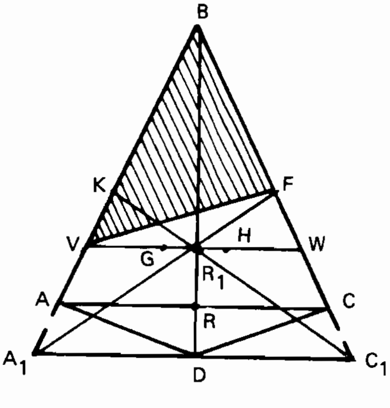


Fig. 3

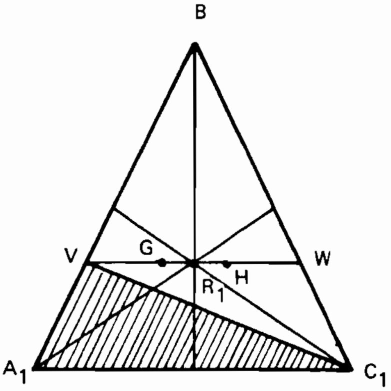


Fig. 4

If a minimum point of f is not contained in the quadrilateral $KBFR_1$ (Fig. 3) then we draw the line VW passing through R_1 and parallel to the line A_1C_1. On the interval VW let us choose two points G and H at a distance ε from R_1.

If

$$f(G) \geq f(R_1) \quad \text{and} \quad f(H) \geq f(R_1)$$

then R_1 is (within ε-accuracy) a minimum point of f on VW and

$$f(Z) > f(R_1) \qquad \forall\, Z \in VBW \quad .$$

Consider the case $f(H) < f(R_1)$. Then we conclude that

$$f(Z) > f(R_1) \qquad \forall\, Z \in VBFR_1$$

and furthermore,

$$f(Z) > f(R_1) \qquad \forall\, Z \in VBF \quad .$$

Thus, we have a new quadrilateral A_1VFC_1 which contains a minimum point.

Let us define the ratio of the area of the quadrilateral A_1VFC_1 and the quadrilateral ABCD. Let h be the height of the triangle ABC. We have

$$S_{ABCD} = \frac{1}{2}\, A_1C_1 \cdot h, \quad S_{A_1BC_1} = \frac{1}{2}\,(1+\alpha)\, A_1C_1 \cdot h \quad ,$$

$$S_{VBF} = \frac{1}{6}\,(1+\alpha)\, A_1C_1 \cdot h \quad .$$

Hence,

$$S_{A_1VFC_1} = \frac{1}{3}\,(1+\alpha)\, A_1C_1 \cdot h$$

and

$$\frac{S_{A_1VFC_1}}{S_{ABCD}} = \frac{2}{3}\,(1-\alpha) \quad . \tag{5}$$

Let us consider the case where the triangle $A_1R_1C_1$ (see Fig. 4) does not contain a minimum point of f. Let us draw the line VW passing through the point R_1 and parallel to the line A_1C_1, and argue as above. Let VBC_1 be a triangle which contains a minimum point of f. We get

$$S_{A_1VC_1} = \frac{1}{6}\ (1+\alpha)\ A_1C_1h$$

and the ratio of the area of the new triangle VBC_1 and the quadrilateral ABCD is $\frac{2}{3}$ $(1+\alpha)$, i.e. (5) holds again.

If $\alpha \leq \alpha_0 \approx 0.335$, then we must construct a triangle since it guarantees a greater decrease in the uncertainty area. The quantity α_0 is then a solution of the equation

$$1 - \frac{\alpha^2(2+\alpha)}{(1+\alpha)^3} = \frac{2}{3}\ (1+\alpha) \quad .$$

The convergence of this modification of the method from [1] is geometric with the rate

$$q = \frac{2}{3}\ (1+\alpha_0) \sim 0.89 \quad .$$

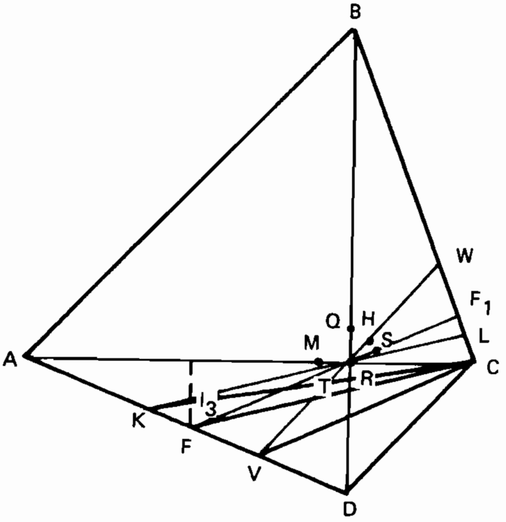


Fig. 5

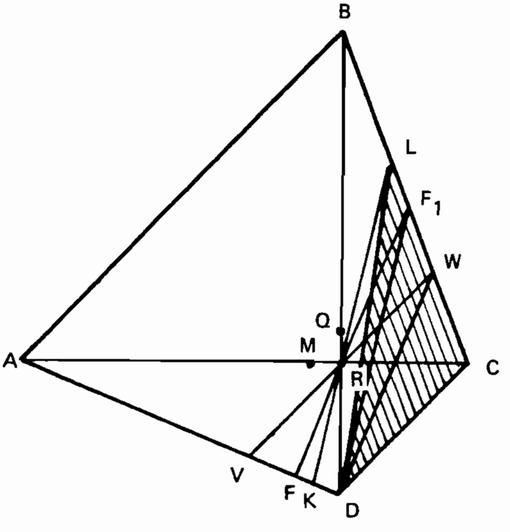


Fig. 6

2. Second modification. Let us again (see Fig. 5) assume that

$$f(M) < f(R) \quad .$$

Then

$$f(Z) > f(R) \qquad \forall Z \in ORCD \quad .$$

Furthermore,

$$f(Z) > f(R) \qquad \forall Z \in VCD \quad .$$

Let us draw the line FF_1 passing through R and parallel to the line VC. On the interval FF_1 let us choose two points T and S at a distance ε from R.

If

$$f(T) \geq f(R) \quad \text{and} \quad f(S) \geq f(R)$$

then R is (within ε-accuracy) a minimum point of f on FF_1 and

$$f(Z) > f(R) \qquad \forall Z \in FF_1CD \quad .$$

Let

$$f(S) < f(R) \quad .$$

Then

$$f(Z) > f(R) \qquad \forall Z \in FRCD$$

and furthermore

$$f(Z) > f(R) \qquad \forall Z \in FCD \quad .$$

Now let us again draw the line KL passing through R and parallel to FC and proceed as above.

As a result we get the new quadrilateral ABCK which contains a minimum point of f. Now let us compute the ratio of the areas of the new quadrilateral ABCK and the quadrilateral ABCD.

Assume that

$$\frac{RD}{BR} = \alpha, \ \frac{AR}{RC} \geq \alpha, \ \frac{RC}{AR} = \alpha_1 \geq \alpha \quad .$$

Let h be the height of the triangle ABC. It follows from the computations above that

$$S_{ABCD} = \frac{1}{2}\,(1+\alpha)AC\cdot h, \ S_{ACD} = \frac{1}{2}\,\alpha\cdot AC\cdot h \quad ,$$

$$RC = \frac{\alpha_1}{1+\alpha_1}\,AC, \quad S_{FCD} = \frac{\alpha\cdot\alpha_1(2+\alpha_1)}{2(1+\alpha_1)^2}\,AC\cdot h \quad .$$

Let us find h_3. Since $S_{AFC} = \frac{1}{2}\,AC\cdot h_3$ and

$$S_{AFC} = S_{ACD} - S_{FCD} = \frac{1}{2}\,\alpha\cdot AC\cdot h - \frac{\alpha\cdot\alpha_1(2+\alpha_1)}{2(1+\alpha_1)^2}\,AC\cdot h =$$

$$= \frac{1}{2}\,\alpha\cdot AC\cdot h\left(1 - \frac{\alpha_1(2+\alpha_1)}{(1+\alpha_1)^2}\right) = \frac{\alpha}{2(1+\alpha_1)^2}\,AC\cdot h$$

we have

$$h_3 = \frac{\alpha}{(1+\alpha_1)^2}\,AC\cdot h, \ S_{FKC} = S_{FRC} = \frac{1}{2}\,RC\cdot h_3 = \frac{\alpha\cdot\alpha_1}{2(1+\alpha_1)^3}\,AC\cdot h \ .$$

Therefore

$$S_{KCD} = S_{FCD} + S_{FKC} = \frac{\alpha\cdot\alpha_1(1+\alpha_1)}{2(1+\alpha_1)^2}\,AC\cdot h +$$

$$+ \frac{\alpha\cdot\alpha_1}{2(1+\alpha_1)^3}\,AC\cdot h = \frac{\alpha\cdot\alpha_1}{2(1+\alpha_1)^2}\,AC\cdot h\ \left(2+\alpha_1+\frac{1}{1+\alpha_1}\right) =$$

$$= \frac{\alpha\cdot\alpha_1(\alpha_1^2+3\alpha_1+3)}{2(1+\alpha_1)^3}\,AC\cdot h \quad .$$

The ratio of the areas of the new quadrilateral **ABCK** and the quadrilateral ABCD is

$$1 - \frac{\alpha\alpha_1(\alpha_1^2+3\alpha_1+3)}{(1+\alpha)(1+\alpha_1)^3} \quad . \tag{6}$$

Since

$$\frac{\alpha_1(\alpha_1^2+3\alpha_1+3)}{(1+\alpha_1)^3} \geq \frac{\alpha(\alpha^2+3\alpha+3)}{(1+\alpha)^3} \qquad \forall\, \alpha_1 \geq \alpha \quad ,$$

it follows from (6) that

$$1 - \frac{\alpha\alpha_1(\alpha_1^2+3\alpha_1+3)}{(1+\alpha)(1+\alpha_1)^3} \leq 1 - \frac{\alpha^2(\alpha^2+3\alpha+3)}{(1+\alpha)^4} \quad . \tag{7}$$

If we decrease the uncertainty area as shown in Fig. 6, we again obtain the same relation (7).

Let (see Fig. 7)

$$f(H) < f(R) \quad .$$

Then

$$f(Z) > f(R) \qquad \forall\, Z \in VRCD$$

and furthermore

$$f(Z) > f(R) \qquad \forall\, Z \in VCD \quad .$$

Let us draw the line FF_1 passing through the point R and parallel to the line VC. On the interval FF_1 let us choose two points T and S at a distance ε from R. If

$$f(T) \geq f(R) \quad \text{and} \quad f(S) \geq f(R)$$

then R is (within ε-accuracy) a minimum point of f on FF_1 and

$$f(Z) > f(R) \qquad \forall\, Z \in FF_1\,CD \quad .$$

Let

$$f(T) < f(R).$$

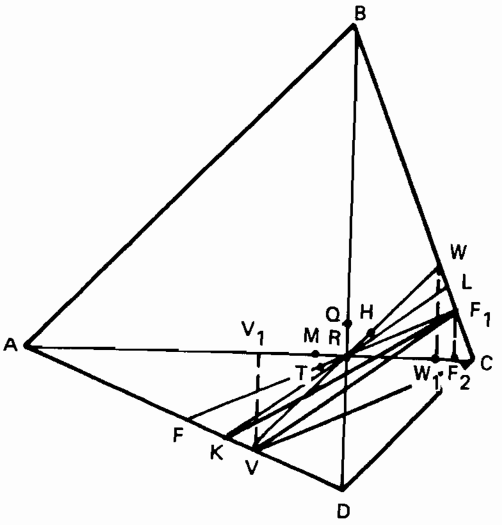


Fig. 7

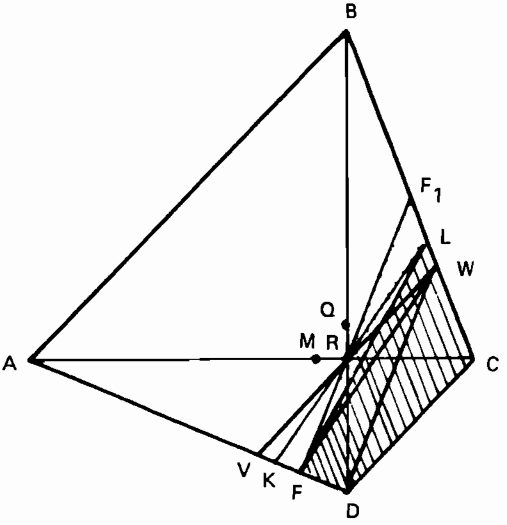


Fig. 8

Then

$$f(Z) > f(R) \qquad \forall\, Z \in V R F_1 CD$$

and furthermore

$$f(Z) > f(R) \qquad \forall\, Z \in V F_1 CD \quad .$$

Let us again draw the line KL passing through R and parallel to the line VF_1 and argue as above. As a result we get a new quadrilateral ABF_1K which contains a minimum point of f. Find the ratio of the areas of the quadrilaterals ABF_1K and ABCD.

Assume that

$$\frac{RD}{BR} = \alpha, \ \frac{RD}{AR} = \alpha_1 = \alpha, \ \frac{AR}{RC} \geq \alpha \quad .$$

The triangles DRC and ABR are similar since

$$\frac{RD}{BR} = \frac{RC}{AR} = \alpha, \ \angle\, DRC = \angle\, ARB \quad .$$

We have $\frac{DC}{AB} = \alpha$ and DC is parallel to AB.
The line VW is parallel to the line DC by construction. Thus, $VW \| AB$. The triangles ABD and VRD are also similar since the

corresponding angles are equal. Therefore

$$\frac{BD}{RD} = \frac{AB}{VR} \quad .$$

Analogously the fact that the triangles BCD and BWR are similar implies that

$$\frac{BD}{RB} = \frac{DC}{WR} \quad .$$

Therefore VR = WR and $\angle$ ARV = $\angle$ CRW. We have $VV_1 = WW_1$. The line FF_1 is parallel to the line VC by construction. Since the triangles VWC and RWF_1 are similar, we have

$$\frac{VW}{WR} = \frac{WC}{WF_1} = 2 \quad .$$

Hence,

$$WF_1 = F_1C, \ F_1F_2 = \frac{1}{2} WW_1 = \frac{1}{2} VV_1 \quad .$$

We have

$$S_{KF_1CD} = S_{VCD} + S_{VF_1C} + S_{KF_1V} = S_{VCD} + S_{VF_1C} + S_{VRF_1} =$$

$$= S_{VCD} + S_{VRC} + S_{RF_1C} \quad .$$

From the computations above it follows that

$$RC = \frac{\alpha_1}{1+\alpha_1} AC, \ VV_1 \equiv h_2 = \frac{\alpha}{1+\alpha_1} h, \ S_{VCD} = \frac{\alpha\alpha_1}{2(1+\alpha_1)} AC \cdot h \ ,$$

$$S_{VRC} = \frac{\alpha\alpha_1}{2(1+\alpha_1)^2} AC \cdot h, \ S_{ABCD} = \frac{1}{2} (1+\alpha) AC \cdot h \quad .$$

Thus,

$$S_{RF_1C} = \frac{1}{2} RC \cdot FF_1 = \frac{\alpha\alpha_1}{4(1+\alpha_1)^2} AC \cdot h \quad .$$

Then

$$S_{KF_1CD} = \frac{\alpha\alpha_1(2\alpha_1+5)}{4(1+\alpha_1)^2} = AC\cdot h \quad .$$

The ratio of the areas of the new quadrilateral ABF_1K and the quadrilateral ABCD is

$$1 - \frac{\alpha\alpha_1(2\alpha_1+5)}{2(1+\alpha_1)^2(1+\alpha)} = 1 - \frac{\alpha^2(2\alpha+5)}{2(1+\alpha)^3} \qquad (8)$$

(since $\alpha_1 = \alpha$).

If we decrease the uncertainty area as shown in Fig. 8 then we again have (8). The estimate (8) is worse than (7).

In the case

$$\frac{RD}{AR} = \alpha_1 > \alpha$$

we always have an estimate better than (8). If at some step

$$\frac{RD}{BR} = \alpha \leq \alpha_0$$

then we enlarge the quadrilateral to a triangle and instead of the quadrilateral ABCD we take the triangle A_1BC_1 (Fig. 9).

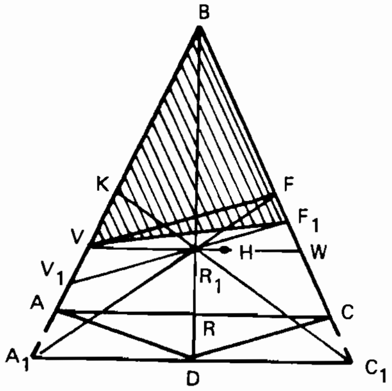


Fig. 9

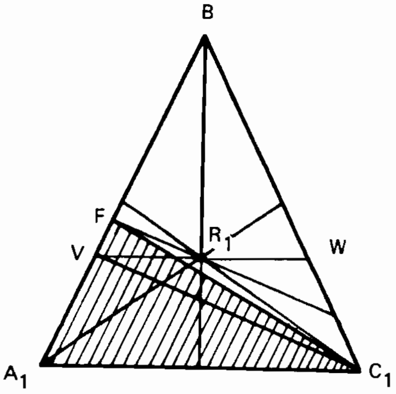


Fig. 10

Let R_1 be the point of intersection of the medians of triangle A_1BC_1. Let there be no minimum point of f in the quadrilateral $KBFR_1$. Then let us draw the line VW passing through the point R_1 and parallel to the line A_1C_1. On the interval VW choose two points G and H at a distance ε from R_1. If

$$f(G) \geq f(R_1) \quad \text{and} \quad f(H) \geq f(R_1)$$

then R_1 is (within ε-accuracy) a minimum point of f on VW and

$$f(Z) > f(R_1) \qquad \forall Z \in VBW \quad .$$

In the case $f(H) < f(R_1)$ we have

$$f(Z) > f(R_1) \qquad \forall Z \in VBFR_1$$

and moreover

$$f(Z) > f(R_1) \qquad \forall Z \in VBF \quad .$$

Let us draw the line V_1F_1 passing through the point R_1 and parallel to the line VF, and argue analogously. Let a quadrilateral $A_1VF_1C_1$ be obtained which contains a minimum point of f. Let h be the height of the triangle ABC. We have

$$S_{ABCD} = \frac{1}{2} A_1C_1 \cdot h , \; S_{A_1BC_1} = \frac{1}{2} (1+\alpha) A_1C_1 \cdot h \quad ,$$

$$S_{VBF} = \frac{1}{6} (1+\alpha) A_1C_1 \cdot h , \; S_{VFF_1} = S_{VFR_1} = \frac{1}{36} (1+\alpha) A_1C_1 \cdot h \; ,$$

$$S_{VBF_1} = \frac{1}{37} (1+\alpha) \; A_1C_1 \cdot h \quad .$$

The ratio of the new quadrilateral $A_1VF_1C_1$ and the quadrilateral ABCD is

$$\frac{11}{18} (1+\alpha) \quad . \tag{9}$$

If we decrease the triangle as shown in Fig. 10, then the ratio of the areas of the new triangle FBC_1 and the quadrilateral ABCD is

$$\frac{5}{9}\,(1+\alpha) \quad . \tag{10}$$

The estimate (9) is worse than the estimate (10).

If

$$\alpha \leq \alpha_0 \approx 0.3787$$

then it is necessary to construct a triangle. The quantity α_0 is a solution of the equation

$$1 - \frac{\alpha^2(2\alpha+5)}{2(1+\alpha)^3} = \frac{11}{18}\,(1+\alpha) \quad .$$

This modification of the method displays geometric convergence with a rate $q \approx 0.8425$.

REFERENCES

1. V.F. Demyanov. "On minimizing a convex function on a plane", Zh. Vychisl. Mat. Mat. Fiz. 16(1) (1976) 247-251.

2. D.J. Wilde. Optimum Seeking Methods. Prentice-Hall Intern. Series in the Physical and Chemical Engineering Sciences, Prentice-Hall, Englewood Cliffs, N.J., 1964.

ON THE STEEPEST-DESCENT METHOD FOR A CLASS OF QUASI-DIFFERENTIABLE OPTIMIZATION PROBLEMS

D. Pallaschke and P. Recht
Institute of Statistics and Mathematical Economics, University of Karlsruhe,
P.O. Box 6308, 7500 Karlsruhe 1, FRG

INTRODUCTION

In a recent paper V.F.Demyanov, S.Gamidov and T.J.Sivelina presented an algorithm for solving a certain type of quasidifferentiable optimization problems [3].
More precisely, they considered the class $\mathcal{F}$ of all functions given by

$$\mathcal{F} = \{f:\mathbb{R}^n \longrightarrow \mathbb{R} \mid f(x) = F(x,y_1(x),\dots,y_m(x))\} \quad ,$$

where

$$y_i:\mathbb{R}^n \longrightarrow \mathbb{R} \quad \text{is defined by}$$

$$y_i(x) = \max_{j\in I_i} \phi_{ij}(x) \qquad I_i = 1,\dots,N_i;\ i=1,\dots,m$$

and

$$\phi_{ij}:\mathbb{R}^n \longrightarrow \mathbb{R} \quad \text{for all } i\in\{1,\dots,m\} \text{ and all } j\in I_i .$$

The functions F and ϕ_{ij} under consideraticn are assumed to belong to the classes $C_1(\mathbb{R}^{n+m})$ and $C_1(\mathbb{R}^n)$ respectively.
The optimization problem consists in minimizing a function $f \in \mathcal{F}$ under constraints.

In this paper we will apply the minimization algorithm of [3] to another class of quasidifferentiable functions.
We are able to prove for this type of optimization problems a convergence theorem similar to that in [3].

1. STEEPEST-DESCENT METHOD

We will briefly recall the steepest descent algorithm for minimizing a quasidifferentiable function in the unconstrained case.

Let $f:\mathbb{R}^n \longrightarrow \mathbb{R}$ be a quasidifferentiable function.

Then for every $\tilde{x}\in\mathbb{R}^n$ there exist two compact, convex sets $\overline{\partial} f|_{\tilde{x}}$ and $\underline{\partial} f|_{\tilde{x}}$, such, that for every $g\in\mathbb{R}^n$, $\|g\|_2 = 1$, the directional derivative is given by:

$$\left.\frac{df}{dg}\right|_{\tilde{x}} = \max_{v\in\underline{\partial} f|_{\tilde{x}}} \langle v,g\rangle + \min_{w\in\overline{\partial} f|_{\tilde{x}}} \langle w,g\rangle \quad .$$

Here $\langle,\rangle$ denotes the canonical inner product in $\mathbb{R}^n$.

In terms of these two sets, a steepest descent direction for f at $\tilde{x}$ is given by

$$g|_{\tilde{x}} := g(\tilde{x}) = -\frac{v_o+w_o}{\|v_o+w_o\|_2}$$

with

$$\|v_o+w_o\|_2 = \max_{w\in\overline{\partial} f|_{\tilde{x}}} \left(\min_{v\in\underline{\partial} f|_{\tilde{x}}} \|v+w\|_2\right).$$

Now, in the steepest descent algorithm, we start with an arbitrary point $x_o\in\mathbb{R}^n$.

Let us assume that for $k\geq 0$ the point $x_k\in\mathbb{R}^n$ has already been defined, then define

$$x_{k+1} := x_k + \alpha_k\cdot g(x_k) \quad ,$$

where $g(x_k)$ is a steepest descent direction of f at x_k and the real number $\alpha_k \geq 0$ is choosen in such a way that

$$\min_{\alpha\geq 0} f(x_k+\alpha g(x_k)) = f(x_k+\alpha_k g(x_k)) .$$

Obviously, the sequence $(x_k)_{k\in\mathbb{N}}$ induces a monotonously decreasing sequence $(f(x_k))_{k\in\mathbb{N}}$ of values of the function f.

A modification of the steepest descent algorithm is proposed in [3]. Therefore we define:

Definition: Let ε,μ be positive real numbers and $f\colon \mathbb{R}^n \to \mathbb{R}$ be quasidifferentiable. Let N be a neighbourhood of all points $x_o \in \mathbb{R}^n$, where f is not differentiable. Then for $x_o \in N$ we define:

$$\underline{\partial}_\varepsilon f|_{x_o} := \operatorname{conv}\Big(\bigcup_{\substack{s \in \mathbb{R}^n \\ \|s\|_2 \leq \varepsilon}} \underline{\partial} f|_{x_o+s} \Big)$$

$$\bar{\partial}_\mu f|_{x_o} := \operatorname{conv}\Big(\bigcup_{\substack{s \in \mathbb{R}^n \\ \|s\|_2 \leq \mu}} \bar{\partial} f|_{x_o+s} \Big)$$

If $x_o \notin N$, then $\underline{\partial}_\varepsilon f|_{x_o} := \underline{\partial} f|_{x_o}$ and $\bar{\partial}_\mu f|_{x_o} := \bar{\partial} f|_{x_o}$.

If $\underline{\partial}_\varepsilon f|_{x_o}$ and $\bar{\partial}_\mu f|_{x_o}$ can be choosen in such a way, that they are compact sets, then f is called (ε,μ)-quasidifferentiable in x_o.

With the introduction of these two sets, we now give a modified steepest descent algorithm to find an ε-inf-stationary point x^* of f.

Let us assume that $f\colon \mathbb{R}^n \to \mathbb{R}$ is quasidifferentiable and moreover that, for given $\varepsilon,\ \mu > 0$, it is (ε,μ)-quasidifferentiable. Then choose an arbitrary $x_o \in \mathbb{R}^n$. Suppose that x_k has already been defined.

If $-\bar{\partial} f|_{x_k} \subset \underline{\partial}_\varepsilon f|_{x_k}$ then x_k is an ε-inf-stationary point and the algorithm stops.

Otherwise, if $-\bar{\partial} f|_{x_k} \not\subset \underline{\partial}_\varepsilon f|_{x_k}$, then compute

$$G(x_k) := \Big\{ g := -\frac{v_o+w_o}{\|v_o+w_o\|_2} \in \mathbb{R}^n \;\Big|\; \max_{w \in \bar{\partial}_\mu f|_{x_k}} \Big(\min_{v \in \underline{\partial}_\varepsilon f|_{x_k}} \|v+w\|_2 \Big) = \|v_o+w_o\|_2 \Big\} .$$

For $g \in G(x_k)$ let us denote

$$\bar{\alpha}(g) := \sup\{\alpha \,|\, f(x_k+\beta g) \leq f(x_k) \text{ for all } 0 \leq \beta \leq \alpha \} ,$$

and let

$$g(x_k) := \operatorname*{argmin}_{g \in G(x_k)} f(x_k + \bar{\alpha}(g)\cdot g) \qquad \alpha_k := \bar{\alpha}(g(x_k)) .$$

Now, we define

$$x_{k+1} := x_k + \alpha_k g(x_k)$$

In this paper we want to apply this modification for finding an ε-inf stationary point for a class of quasidifferentiable functions.

2. A MOTIVATING EXAMPLE

Let $F,G:\mathbb{R}^n \longrightarrow \mathbb{R}$ be two arbitrary functions with $F,G\in C_1(\mathbb{R})$. Then define the following, quasidifferentiable function $f:\mathbb{R}^n \longrightarrow \mathbb{R}$ by

$$f; = \max(|G|,-F-|G|) - ||G| -2|F|| .$$

This type of function is considered in [1] and obviously does not belong to the class $\mathcal{F}$ defined in the introduction. For illustration, Figure 1 shows the graph of a function f of this type for

$$F:\mathbb{R}^2 \longrightarrow \mathbb{R}, \quad F(x_1,x_2) = x_1^2-x_2$$

$$G:\mathbb{R}^2 \longrightarrow \mathbb{R}, \quad G(x_1,x_2) = -x_1^2-x_2^2+1.2$$

in the set $\Omega=[-1,1.4] \times [-2,1.25]$.

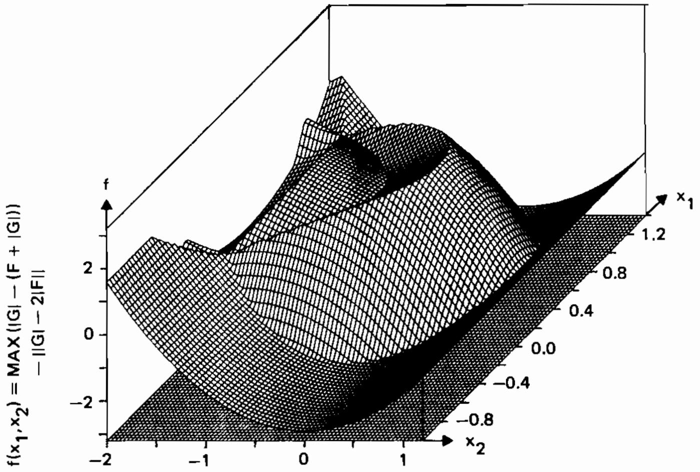


Figure 1

For functions of that type, as well as for the class $\mathcal{F}$, the following properties are valid, as observed in [3].

I. If for all $x\in\mathbb{R}^n$, the convex, compact sets $\underline{\partial} f\big|_x$ and $\overline{\partial} f\big|_x$ are computed as in [3] the two mappings

$$x \longmapsto \underline{\partial} f\big|_x \quad \text{and} \quad x \longmapsto \overline{\partial} f\big|_x$$

are upper-semi-continuous. Moreover for suitable $\varepsilon,\mu > 0$ the functions $\underline{\partial}_\varepsilon f$, $\overline{\partial}_\mu f$ are also upper-semi-continous.

II. If $x \in \mathbb{R}^n$ is not a stationary point, then there exist a real number $M > 0$ and a neighbourhood U_o of $0 \in \mathbb{R}^n$, such that for all $y \in U_o$

$$\left| \frac{df}{dg}\bigg|_x - \frac{df}{d(g+y)}\bigg|_x \right| \le M \cdot \|y\|_2 .$$

3. A CONVERGENCE THEOREM

Theorem:

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a quasidifferentiable function with the following properties:

(i) There exist real numbers $\varepsilon > 0$, $\mu > 0$ such that for all $x \in \mathbb{R}^n$ f is (ε, μ)-quasidifferentiable and the mappings

$$x \longmapsto \underline{\partial}_{\varepsilon} f\big|_x \quad , \quad x \longmapsto \overline{\partial}_{\mu} f\big|_x$$

and

$$x \longmapsto \underline{\partial} f\big|_x \quad , \quad x \longmapsto \overline{\partial} f\big|_x$$

are upper semi-continuous (u.s.c.)

(ii) If $x \in \mathbb{R}^n$ is not an ε-inf stationary point, then there exist an $M > 0$ and a neighbourhood U_o of $0 \in \mathbb{R}^n$ such that for all $y \in U$, $g \in \mathbb{R}^n$

$$\left| \frac{df}{dg}\bigg|_x - \frac{df}{d(g+y)}\bigg|_x \right| \le M \cdot \|y\|_2 \quad .$$

Then: Every limit point of the sequence $(x_n)_{n \in \mathbb{N}}$, constructed by the modified steepest descent algorithm, is an ε-inf stationary point of f.

Proof:

Let x^* be a limit point of $(x_n)_{n \in \mathbb{N}}$ and let us assume that x^* is not ε-inf stationary.

Hence there exist a $v_o \in \underline{\partial}_{\varepsilon} f|_{x^*}$ and a $w_o \in \overline{\partial} f|_{x^*}$ such that

$$\| v_o + w_o \|_2 = \sup_{w \in \overline{\partial} f|_{x^*}} \left(\inf_{v \in \underline{\partial}_{\varepsilon} f|_{x^*}} \| v + w \|_2 \right) = a > 0.$$

Thus $g := -\frac{v_o+w_o}{\|v_o+w_o\|_2}$ is a normalized descent direction in x^*.

Observe that $w_o \in \bar{\partial}_\mu f\big|_x$.

Since $x \longmapsto \underline{\partial}_\varepsilon f\big|_x$ is u.s.c.,there exist a neighbourhood $\tilde{U}$ of $\underline{\partial}_\varepsilon f\big|_{x^*}$ and a neighbourhood U of x^* such that for all $x \in U$

$$\underline{\partial}_\varepsilon f\big|_x \subset \tilde{U} .$$

Moreover,to $\bar{\partial}_\mu f\big|_{x^*}$ there exist a neighbourhood $\tilde{V}$ of $\bar{\partial}_\mu f\big|_{x^*}$ and a neighbourhood V of x^* such that for all $x \in V$

$$\bar{\partial}_\mu f\big|_x \subset \tilde{V} .$$

Choose U_o according to assumption (ii) of the theorem. To $W := U \cap V \cap (U_o+x^*)$ there exists a $k_o \in \mathbb{N}$ such that for all $k \geq k_o$, $x_k \in W$. (Here k isthe index of the convergent subsequence .)
Let us denote by $w_k^* \in \bar{\partial}_\mu f\big|_{x_k}$ the point which is nearest to w_o.

From the upper semicontinuity of $\bar{\partial}_\mu f$ we have

$$\lim w_k^* = w_o \quad .$$

Now, let $v_k \in \underline{\partial}_\varepsilon f\big|_{x_k}$ be a point of minimal distance to $-w_k^*$.

Then $\lim_k (\mathrm{dist}(v_k, \underline{\partial}_\varepsilon f\big|_{x^*}) = 0$.

The neighbourhoods of $\underline{\partial}_\varepsilon f\big|_{x^*}$ can be assumed to be bounded, since $\underline{\partial}_\varepsilon f\big|_{x^*}$ is compact.

Hence, there exists a subsequence $(v_k)_{k \in \mathbb{N}}$, also indexed by k, which converges to $\tilde{v} \in \underline{\partial}_\varepsilon f\big|_{x^*}$.

Thus, for a suitable subsequence and an index K we have:

$$\lim_{k\to\infty} \|w_k+v_k\|_2 = \|w_o+\tilde{v}\|_2 \geq \mathrm{dist}(w_o, \underline{\partial}_\varepsilon f\big|_{x^*}) = a \quad .$$

We see that $\tilde{v}=v_o$ since the Euclidian norm is strict. Therefore, for all $k \geq K$

$$\|w_k + v_k\|_2 \geq \frac{a}{2} \quad .$$

Now, we want to show that for k large enough,

$$\hat{g}_k = - \frac{v_k+w_k}{\|v_k+w_k\|_2}$$

is a descent direction in x^*.
For this, let $\alpha > 0$. Then:

$$f(x_k+\alpha\hat{g}_k)=f(x^*) + \left.\frac{df}{d[(x_k-x^*)+\alpha\hat{g}_k]}\right|_{x^*} +o(\|x_k-x^*+\alpha\hat{g}_k\|) \quad .$$

From assumption (ii) follows

$$\left.\frac{df}{d[(x_k-x^*)+\alpha\hat{g}_k]}\right|_{x^*} = \alpha \left.\frac{df}{d\hat{g}_k}\right|_{x^*} + O\,(\|x_k-x^*\|_2)$$

and therefore

$$f(x_k+\alpha\hat{g}_k)=f(x^*)+\alpha \left.\frac{df}{d\hat{g}_k}\right|_{x^*} +o(\|x_k-x^*+\alpha\hat{g}_k\|_2)+O(\|x_k-x^*\|_2)$$

$$=f(x^*)+\alpha \left.\frac{df}{d\hat{g}_k}\right|_{x^*} +O(\|x_k-x^*\|_2)+o(\alpha) \quad .$$

From the definition of quasidifferentiability we have:

$$\left.\frac{\partial f}{\partial\hat{g}_k}\right|_{x_k} = \min_{w\in\bar{\partial}_\mu f|_{x_k}} \Big(\max_{v\in\underline{\partial}_\varepsilon f|_{x_k}} \langle w+v,\hat{g}_k\rangle\Big)$$

and therefore, from the definition of v_k:

$$\left.\frac{\partial f}{\partial\hat{g}_k}\right|_{x_k} \leq \max_{v\in\underline{\partial}_\varepsilon f|_{x_k}} \langle w_k+v,\hat{g}_k\rangle$$

$$\leq \max_{v\in\underline{\partial}_\varepsilon f|_{x_k}} \Big(- \langle w_k+v,w_k+v_k\rangle\cdot\|w_k+v_k\|_2^{-1}\Big)$$

$$= - \| w_k + v_k \|_2^2 \cdot \| w_k + v_k \|_2^{-1} = - \| v_k + w_k \|_2 \leq - \frac{a}{2} .$$

Since $\left.\frac{df}{d\hat{g}_k}\right|_{x^*} \leq \max\limits_{\substack{v \in \tilde{U} \\ \underline{\partial}_\varepsilon f|_{x^*} \subset \tilde{U}}} \langle v, \hat{g}_k \rangle + \min\limits_{w \in \bar{\partial} f|_{x^*}} \langle w, \hat{g}_k \rangle$,

$w_o \in \bar{\partial} f|_{x^*}$ and $\lim\limits_{k \to \infty} w_k = w_o$

we find for a given $\delta > o$ an index K_1 such that for all $k \geq K_1$

$$\left.\frac{df}{d\hat{g}_k}\right|_{x^*} \leq \max_{\substack{v \in \tilde{U} \\ \underline{\partial}_\varepsilon f|_{x^*} \subset \tilde{U}}} \langle v, \hat{g}_k \rangle + \min_{w \in \bar{\partial} f|_{x^*}} \langle w, \hat{g}_k \rangle$$

$$\leq \Big(\max_{v \in \underline{\partial}_\varepsilon f|_{x_k}} \langle v, \hat{g}_k \rangle + \delta \Big) + \langle w_o, \hat{g}_k \rangle$$

$$\leq \Big(\max_{v \in \underline{\partial}_\varepsilon f|_{x_k}} \langle v, \hat{g}_k \rangle + \delta \Big) + \langle w_k, \hat{g}_k \rangle + \| w_k - w_o \|_2$$

$$\leq \left.\frac{df}{d\hat{g}_k}\right|_{x_k} + 2\delta \leq - \frac{a}{2} + 2\delta .$$

Thus, for all $k \geq K_1$, we see that $\hat{g}_k$ is a descent direction in x^*.

Hence there is $\tau_o > o$ such that for all $\tau \leq \tau_o$

$$f(x_k + \tau \hat{g}_k) < f(x^*) .$$

Now by the definition of the sequence $(x_k)_{k \in \mathbb{N}}$ via the modified steepest descent algorithm and by condition ii.) of the theorem we have:

$$f(x_{k+1}) = f(x_k + \alpha_k \cdot g(x_k))$$

$$\leq \min_{o \leq \alpha \leq \alpha_k} f(x_k + \alpha g_k) = f(x_k + \hat{\alpha}_k \hat{g}_k)$$

$$\leq f(x_k + \tau \hat{g}_k) < f(x^*)$$

for a suitable $\tau \leq \tau_o$.

This contradicts the facts that $(f(x_k))_{k\in\mathbb{N}}$ is monotonously decreasing and $\lim f(x_k)=f(x^*)$.

QED.

Remark: The proof also remains valid for $\varepsilon=o$, i.e. replacing "ε-inf-stationary" by "inf-stationary".

4. NUMERICAL EXPERIENCES

The above mentioned modification of the steepest descent method was implemented on the Siemens 7780 at the Computer Center of the University of Karlsruhe.

Applying this procedure to the motivating example of Section 2, ε-inf stationary points could easily be found (this is also true for problems under constraints, see [2]).

Let us now discuss a further example.

Example

let $f:\mathbb{R}^3 \longrightarrow \mathbb{R}$ be given by

$$f_1(x_1,x_2,x_3) = ((x_1+x_2) + \sqrt{(x_1-x_2)^2+4x_3^2}\ )\ /\ 2$$

and

$$f_2(x_1,x_2,x_3) = ((x_1+x_2) - \sqrt{(x_1-x_2)^2+4x_3^2}\ )\ /\ 2$$

with:

$$f(x_1,x_2,x_3) = |f_1(x_1,x_2,x_3)| - |f_2(x_1,x_2,x_3)|$$

Obviously $f_1, f_2 \notin C_1(\mathbb{R}^3)$.

This function occurs naturally in the investigation of the condition of matrices, i.e., if we assign to any symmetric $(n\times n)$-matrix $A=(a_{ij})_{1\leq i,j\leq n}$ the difference of moduli of the maximal and minimal eigenvalue $|\lambda_{\max}|$ and $|\lambda_{\min}|$ respectively, i.e.

$$\varphi : L(\mathbb{R}^n, \mathbb{R}^n) \longrightarrow \mathbb{R}$$

$$\varphi_n(A) := |\lambda_{\max}| - |\lambda_{\min}| \quad .$$

This function is quasidifferentiable, since $\lambda_{\max} = \sup\limits_{\|x\|=1} \langle Ax,x\rangle$ is a convex function and $\lambda_{\min} = \inf\limits_{\|x\|=1} \langle Ax,x\rangle$ is a concave function.

For n = 2, ϕ_n coincides with the above defined function $f:\mathbb{R}^3 \longrightarrow \mathbb{R}$. Morover, the properties i) and ii) of the theorem are valid for the sets $\underline{\partial}_{\varepsilon} f$ and $\overline{\partial}_{\mu} f$ for suitable ε and μ. Figure 2 below gives an illustration of the graph of the function f for 4 different values of x_3, i.e. x_3 = 0.3; x_3 = 0.2; x_3 = 0.1; x_3 = 0.0.

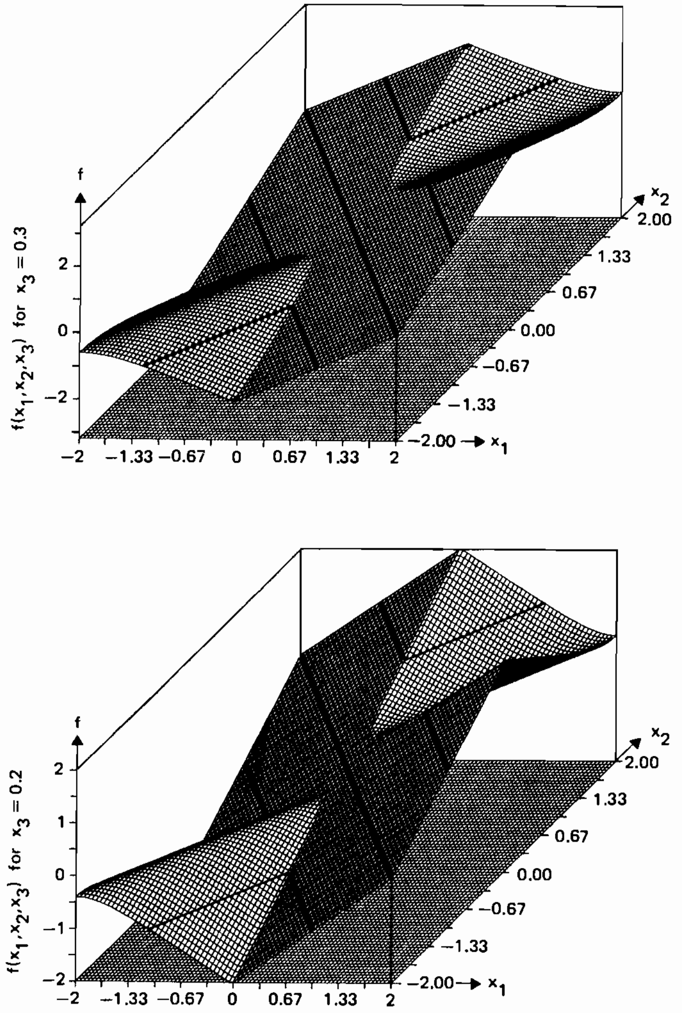


Figure 2

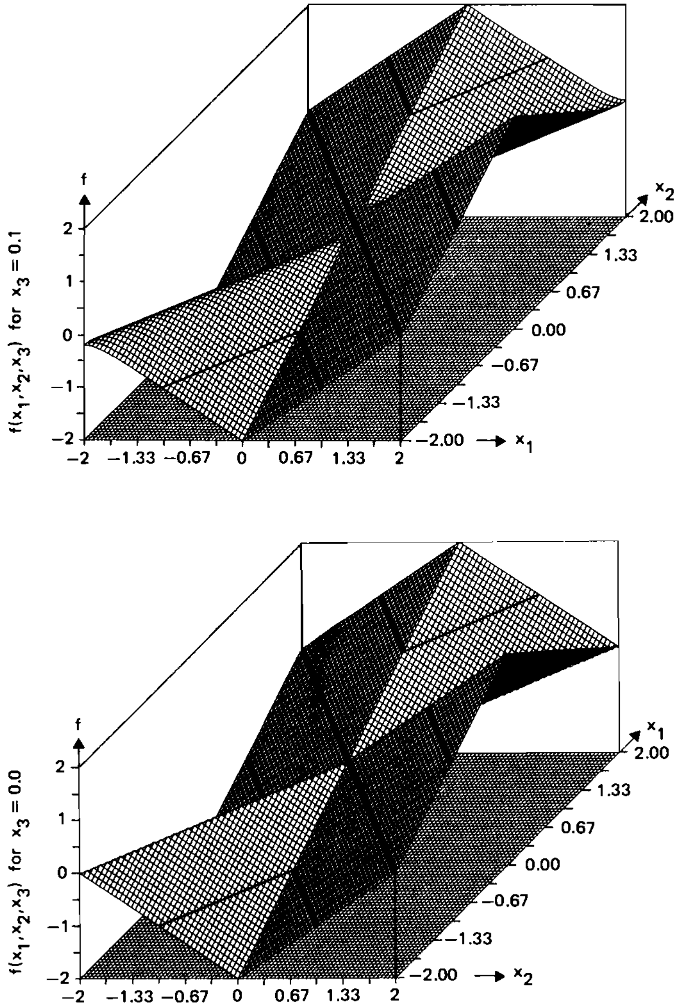


Figure 2

The behaviour of this function $x_3 = 0$ is similar to that given in example 2.1 of [4].

In Clarke's sense, the point (0,0,0) is stationary, but is neither minimum or maximum, nor a saddle-point. It is a monkey-saddle point. Moreover, $0 \in \text{int } (\partial_{cl} f|_0)$, i.e., 0 is an inner point of the Clarke subdifferential. Of course, using quasidifferentials, the algorithm could find a descent direction (0,0,0).

The "cumulative character" of Clarke's subdifferential can be clearly observed in Figure 2.

REFERENCES

[1] V.F.Demyanov, A.M.Rubinov
On Quasidifferentiable Mappings.
Math.Operationsforschung und Statistik, Ser. Optimization
14 (1983) pp. 3-21.

[2] V.F.Demyanov
Quasidifferentiable functions: Necessary Conditions and
Descent Directions.
IIASA-Working Papers, WP-83-64 (June 1983).

[3] V.F.Demyanov, S.Gamidov and T.I.Sivelina
An Algorithm for Minimizing a Certain Class of Quasidifferentiable Functions.
IIASA-Working Papers, WP-83-122 (Dec. 1983).

[4] V.F.Demyanov, L.N.Polykova, A.M.Rubinov
Nonsmoothness and Quasidifferentiability.
IIASA-Working Papers, WP-84-22 (March 1984).

A MODIFIED ELLIPSOID METHOD FOR THE MINIMIZATION OF CONVEX FUNCTIONS WITH SUPERLINEAR CONVERGENCE (OR FINITE TERMINATION) FOR WELL-CONDITIONED C^3 SMOOTH (OR PIECEWISE LINEAR) FUNCTIONS

G. Sonnevend
Department of Numerical Analysis, Eötvös University,
Muzeum körut 6–8, 1088 Budapest, Hungary

INTRODUCTION

The motivations for constructing algorithms with the properties specified in the title of this paper come from two sources. The first is that the ellipsoid method (see e.g. Shor (1982) and Sonnevend (1983)) has a slow (asymptotic) convergence for functions of the above two classes. The second arises since the popular idea (practice) that the globalization of convergence for the asymptotically fast quasi-Newton methods should be achieved by the application of line search strategies (these are described in Stoer (1980); bundle methods are described in Lemarechal et al. (1981)) becomes rather questionable if function and subgradient evaluations are costly and if the function is "stiff", i.e. has badly conditioned or strongly varying second derivatives (Hesse matrixes).
Indeed, line search uses - intuitively speaking - the local information about the function only for local prediction, while in the ellipsoid method the same information is used to obtain a global prediction (based on a more decisive use of the convexity). In the bundle (ε-subgradient) methods the generation of a "useable" descent direction (not speaking about the corresponding line search) may require - for a nonsmooth f (in the "zero-th" steps) - a lot of function (subgradient evaluations).
The important feature of the ellipsoid method, which will be used here to obtain a method with finite termination (i.e. exact computation of f*) for piecewise linear functions (which is very important for the solution of general linear programming problems), is that it provides us with (asymptotically exact) lower bounds for the value of f*.

Of course, for nonconvex functions or when n, the dimension of the independent variable x, is very large and we have some special (sparsity) structure, the "optimal" choice of a globalization method may fall on another method (using line searches or homotopy), especially if sensitivity (stability) aspects (with respect to rounding or measurement errors) are important. Concerning the sensitivity of a much more stable ellipsoid method we refer to Sonnevend (1983).

The two sources mentioned above are, in fact not very different: it is very important to understand that for C^{∞}, but "stiff" convex functions the "initial" behaviour of any algorithm is the same as for the class of general convex functions: any convex functions can be arbitrarily closely (uniformly) approximated (say, over a simplex) by C^{∞} convex functions for which the Hesse matrixes are nonsingular at their (unique) minimum points. Concerning test results supporting the competitiveness of "ellipsoid" methods we can refer e.g. to those cited in Ech-Cherif, Ecker (1984).

Of course, when we wish to prove - for the proposed method - the two (asymptotic) convergence properties mentioned above it is natural (in fact, almost necessary) to assume that the (function, near to its) minimum is "well conditioned" in respective sense, see below.

The interest (coming from different fields of applications) in constructing methods for the computation of the minimal value f* of a general (nonsmooth) convex function f (over R^n) should not be stressed here, see e.g. Zowe (1984); neither is a detailed, formal description of the allowed algorithms necessary. It will be enough to recall that an algorithm consists in the sequential choice of points $x_j \in R^n$, $j=1,2,...,$ where the values $f(x_j)$ and $g(x_j) \in \partial f(x_j)$, i.e. one subgradient of f at x_j, are evaluated. A positive and important feature of the algorithm presented below is that it provides - at each step s - an easily computed and good (asymptotically exact) upper bound $\delta(s,f)$ for the unknown value

$$\varepsilon(s,f) := \min_{j \le s} f(x_j) - f^* , \qquad (1.1)$$

i.e. a lower bound ℓ_s for the value of f^*. The global error of an N-step algorithm A - over a class of functions F - is defined by

$$\varepsilon(N,F):=\sup\{\varepsilon(N,f)\mid f\in F\}, \tag{1.2}$$

where it is understood that in (1.1) $x_j=x_j(A,f(x_k),g(x_k), k\le j, F)$, for $j=1,\dots,N$.

The function f will be assumed (in Section 2) to belong - for some, finite, known values m,M - only to the class

$$F=F(m,M,L_o)=\{h\mid h \text{ convex on } R^n,\ X^*(h)\cap L_o\neq\emptyset,\ m\le h(x)\le M, \text{ for } x\in L_o\}, \tag{1.3}$$

where L_o is a ball of radius R around the origin in R^n, and $X^*(h)=\{z\mid h(z)=\inf\{h(x)\mid x\in R^n\}\}$. It is well known that a general (finitely constrained) convex programming problem can be reduced - via exact penalty functions - to an unconstrained problem.

The proposed method is a nontrivial, stepwise combination of a modified, graph ellipsoid method (GEM) - presented in section 2 - of a simple quasi - Newton method and of (a proximal point) cutting plane method: roughly speaking one chooses - at each step - that method of the three which leads to an ellipsoid of smallest volume. All three "next" ellipsoids (possible followers of the present one) are constructed to contain all "minimumpairs" (z^*,h^*) - with $z^*\in L_o$ - of functions h compatible with (i.e. indistinguishable from) f based on the information collected up to that step. It will be, in fact, enough to update (resp. apply) the quasi-Newton (resp. cutting plane) method only after each (consecutive) n steps. The global (linear) rate of convergence of the method is the same as that of the ellipsoid method (per one function and subgradient evaluation, i.e. "step", which requires $0(n^2)$ arithmetical operations: in the average, over periods of n steps). We emphasize that the proposed method is a "stationary iteration" method which "automatically" tunes itself to the required, asymptotic behaviour.

2. A MODIFIED (GRAPH) ELLIPSOID METHOD

As a result of search for a (global) acceleration of the method of centers of gravity (CGM) we proposed in Sonnevend (1984) a graph method of centers of gravity (GCGM), whose (global) convergence rate is $\exp(-n^{-1})$ and - as an easy implementable approximation for the latter - we also proposed a graph ellipsoid method (GEM) described below in a more detailed manner.

Let us begin with a definition: we say that - for a convex function h - the vector $(u,v)\in R^{n+1}$ is a minimumpair (of h) if $h(u)=v=h^*=\inf\{h(z)\mid z\in R^n\}$. The underlying idea of GEM is to localize the set of minimumpairs of f (which is supposed - see (1.3) and (2.1) - to have a nonempty intersection with an initial ellipsoid E_0) into a sequence of recursively (i.e. stepwise) updated ellipsoid E_s, $s=0,1,\dots,$ of regularly decreasing volumes. In GCGM these sets of localizations (polyhedrons in R^{n+1}, if L_0 is assumed to be a polyhedron, e.g. a simplex) are computed exactly and the x-projections of their, recursively computed centers of gravity are taken as the places of the next function evaluations. It can be proved - at least for n=1 - that GCTM has a better (global) convergence rate than CGM, and that the same holds for arbitrary n is indicated by the following observation: for piecewise linear functions the asymptotic rate of convergence of GCGM is - in the worst case - n/(n+2), while for CGM this number is n/(n+1).

We describe the construction of E_s inductively with respect to the value of s. Let E_0 be the ellipsoid of smallest volume containing the set

$$\{(u,v)\mid u\in L_0,\ m\le v\le M\}\,. \tag{2.1}$$

It is easy to prove that the vertical width of $E_0=E_0(m,M,L_0)$ is equal to $(M-m)\sqrt{n+1}$ and

$$\mathrm{vol}\ E_0=\frac{M-m}{2R}\sqrt{n+1}\ \left(1-\frac{1}{n+1}\right)^{\frac{n}{2}}\ \mathrm{vol}\ L_0. \tag{2.2}$$

Suppose now that $s\ge1$ and an ellipsoid E_{s-1} is known (i.e. constructed in the previous step) to contain all minimumpairs of functions $h\in F\ (m,M,L_0)$, for which

$$h(x_j)=f(x_j),\ \partial h(x_j)\ni g(x_j),\ 1\le j\le s-1.$$

In order to define E_s we first define x_s, $s=1,2,\ldots$, to be the projection of the centre of E_{s-1} to the X space:

$$x_s:=x(c(E_{s-1})). \tag{2.3}$$

Having computed (measured) the values of $f(x_s)$ and $g(x_s)$ we define the sets

$$H_2^s:=\{(u,v)\mid v\ge f(x_s)+\langle u-x_s,g(x_s)\rangle\}=H_2(x_s,f(x_s),g(x_s))$$

$$H_1^s:=\{(u,v)\mid v\le f(x_s)\},\ T_s:=H_1^s\cap H_2^s\cap E_{s-1}. \tag{2.4}$$

We shall present a simple (suboptimal) method - which will suffice for our purposes - for the construction of an ellipsoid E_s of small (i.e. not necessarily minimal) volume containing T_s for the somewhat more general case, when T_s is replaced by $T=E\cap H_1\cap H_2$, where H_1 is an arbitrary "horizontal, lower" half-space,

$$H_1=\{(u,v)\mid v\le h\},\ H_2=\{(u,v)=t\mid \langle t,p\rangle\ge c\}, \tag{2.5}$$

where $p=(-g,1)$, $g\in R^n$ and the ellipsoid E is arbitrary, but non-degenerate. The computation of a minimal volume ellipsoid, $E^*(T_s)$ containing T_s - by some rank two update formula - would not be very difficult, for special cases this was done already, see e.g. Eh-Cherif, Ecker (1984) or Shor (1982).

It is important to note that in the special case $T=T_s$, either $H_1^s\cap E_{s-1}$ or $H_2^s\cap E_{s-1}$ is contained in a "half ellipsoid" $E_{s-1}\cap H$, where the boundary of H contains the centre of E_{s-1}. Indeed, if $f(x_s)\ge v(c(E_{s-1}))$, i.e. the last coordinate of the centre of E_{s-1}, then we can choose H be parallel to H_2^s and if $f(x_s)<(c(E_{s-1}))$ we can choose H be parallel to H_1^s, which amounts to moving H_2^s, resp. H_1^s downward, resp. upward.

We shall need first to compute the minimal "horizontal", or parallel to H_2 layer(depending on the alternative defined just above) $S(E,H_1,H_2)$ containing T. This clearly amounts - say in the case of the horizontal layer - to computing the value $m(E,H_2):=\min\{v\mid \text{there is an } u \text{ such that } (u,v)\in E\cap H_2\}$, (2.6).

Further we have to compute the minimal volume ellipsoid $E^*(E,S)$ containing the intersection of an ellipsoid E and a (horizontal) layer S.

Finally E_s will be defined as the ellipsoid

$$E_s := E^*(E_{s-1},\ S(E_{s-1},\ H_1^s,\ H_2^s)) =: \bar{E}(E_{s-1},\ H_1^s,\ H_2^s), \tag{2.7}$$

(and x_{s+1} is chosen according to (2.3) for $s \to s+1$).

The ellipsoids $E=E(w,A)$ will be represented by their centers $w \in R^{n+1}$ and symmetric, positive definite matrixes $A=A^T \in R^{(n+1)x(n+1)}$:

$$E(w,A) := \{t \mid \langle t-w,\ A^{-1}(t-w) \rangle \le 1\}. \tag{2.8}$$

The value $m(E,H_2)$, thus the "width" of $S(E,H_1,H_2)$, for the data (E,H_1,H_2), see (2.5), (2.8), can be computed as follows (for simplicity - but, of course, without loss of generality - again for the case of the horizontal layer)

$$m(E,H_2) = v(w) - q\langle Ap,e_o\rangle ||Ap||^{-1} - \sqrt{1-q^2}\langle A^p e_o^p, e_o^p\rangle^{1/2},$$

where $e_o^p := e_o - \langle e_o,p\rangle ||p||^{-1})(1-\langle e_o,p\rangle^2 ||p||^{-2})^{-1/2}$, $e_o=(0,0,\ldots,0,1)$

$$A^p := A - A\ (Ap)^*\langle Ap,p\rangle^{-1},\quad q := (\langle p,w\rangle - c_2)\langle Ap,p\rangle^{-1/2}.$$

The parameters of the ellipsoid $E^*(E,S)$, for E in (2.8) and $S=\{t \mid \xi \le \langle \gamma, t-w\rangle \langle A\gamma,\gamma\rangle^{-1/2} \le \eta\}$, $0 \le \xi < \eta \le 1$ are given by the following formulae (note that the alternative stated above assures that the chosen layers always do not contain the centre of E in their interior, thus $\xi \geq 0$ can be assumed);

$$A^* := \rho^2(A-(1-(\Psi/\rho)^2)A\gamma(A\gamma)^*\langle A\gamma,\gamma\rangle^{-1}, \tag{2.9}$$

$$w^* := w - \varkappa A\gamma\langle A\gamma,\gamma\rangle^{-1/2},\quad \varkappa := \xi + \Psi(1-(1-\xi^2)\rho^{-2})^{1/2},$$

where

$$\rho^2 := \frac{(n+1)^2}{n^2+2n}\left[1 - \frac{\eta^2+\xi^2}{2} + \left[(\frac{\eta^2-\xi^2}{2})^2 + \frac{(1-\eta^2)(1-\xi^2)}{(n+1)^2}\right]^{1/2}\right]$$

$$\Psi := (\eta-\xi)(\sqrt{1-(1-\eta^2)\rho^{-2}} + \sqrt{1-(1-\xi^2)\rho^{-2}}), \tag{2.10}$$

For the cases when $\eta=1$; $E^*(E,S)=E^*(E \cap H_1)$ or $E^*(E,S)=E^*(E \cap H_2)$.

Moreover the volume of $E^*(E,S)$ is equal to $\Psi\rho^n$. For a proof of these formulae see e.g. König, Pallaschke (1981) or the references in Shor (1982). Note that in the special cases, where $E\cap S=E\cap H_1$ or $E\cap S=E\cap H_2$, (i.e. when $\eta=1$) these formulae are more simple, morever they would be enough for assuring the next fundamental inequality:

$$\mathrm{vol}\ E^*(E,S(E,H_1,H_2))\leq \mathrm{vol}\ E^*(E,H_i)\leq\lambda_{n+1}(1-\xi_i)(1-\xi_i^2)^{\frac{n-1}{2}}\ \mathrm{vol}\ E_{s-1}\ ,$$

for $i=1$, or $i=2$, where

$$\lambda_n = \frac{n^n}{n+1}(n^2-1)^{-\frac{n-1}{2}} < e^{-\frac{1}{2(n+1)}}.$$

Consequently by our construction we shall have

$$\mathrm{vol}\ E_s\leq\exp(-(2(n+2))^{-1})\ \mathrm{vol}\ E_{s-1},\ \text{for all } s. \qquad (2.11)$$

We have now almost everything needed for the proof of the next theorem.

Theorem 1. The algorithm GEM, described above by (2.9) assures the existence of a constant $k_{2,n}$ (such that $\lim k_{2,n}=1$ for $n=\infty$) for which

$$\varepsilon(N,f)\leq k_{2,n}(M-m)\exp(-N(2(n+1)(n+2))^{-1}).$$

holds for all $f\in F(m,M,L_0)$.

Proof. As for the original ellipsoid method, here also the following Lemma will be useful. It is well known in the theory of ellipsoid method; for the simple proof see e.g. Sonnevend (1984).

Lemma 1. The information that a convex set $L\subset R^n$ contains all points z (of a convex set L_0), where a convex function is less than a constant c, implies that

$$c-\inf_{L_0} f \leq \left(\frac{\mathrm{Vol}\ L}{\mathrm{Vol}\ L_0}\right)^{1/n} (\sup_{L_0} f - \inf_{L_0} f). \qquad (2.12)$$

We apply this Lemma with $c=\min\{f(x_j)\mid j\leq N\}=:f_N$, $L=\{u\mid(u,c)\in E_N\}$ and L_0 as defined earlier. First we note that $\mathrm{vol}\ L \leq \mathrm{vol}\ K_N$, where K_N is the horizontal, central section of E_N, and

$$\mathrm{vol}\ E_N = \mathcal{H}_{n+1}\,\mathcal{H}_n^{-1}\delta(E_N)\ \mathrm{vol}\ K_N\ , \qquad (2.13)$$

where $\delta(E_N):=\langle A_N e_0,e_0\rangle^{1/2} \geq \frac{1}{2}\, e(N,f)$, $\mathcal{H}_n = \dfrac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}$ is the volume of the unit ball in R^n. From (2.2), (2.13), and (2.11) we obtain - for $\lambda=\exp(-N(2(n+1))^{-1})$ -

$$\lambda \geq \frac{\text{vol } E_N}{\text{vol } E_o} = \frac{\text{vol } K_N}{\text{vol } L_o} \frac{\delta(E_N)}{M-m} k_{1,n} \geq \frac{\text{vol } L}{\text{vol } L_o} \frac{e(N,f)}{M-m} k_{1,n} ,$$

where $k_1 n^{-3/2} \leq k_{1,n} \leq k_2 n^{-3/2}$, for some finite, positive constants and from this follows that we have

$$\text{either } \lambda^{1/n+1} \geq k_{2,n}^{-1} \left(\frac{\text{volL}}{\text{volL}_o}\right)^{1/n} \text{ or } \lambda^{1/n+1} \geq k_{2,n}^{-1} \frac{e(N,f)}{M-m} ,$$

where $k_{2,n}$ tends to 1 for $n\to\infty$. This finishes the proof by the definition of λ and by Lemma 1.

Remark 1. Notice that we could replace - in the definition of H_1^s - the value $f(x_s)$ by f_s i.e. the minimum of the f values computed up to step s. Since as a by-product of the update formulae (2.10) we can compute the volume of E_s, s=1,2,..., a lower bound ℓ_s for the value f* can be updated: $\ell_s:=f_s-k_{2,m}(M-m)\exp(-s(2(n+1)(n+2))^{-1})$. The values f_s and ℓ_s can be used for narrowing a horizontal layer $S(E_{s-1},H_1^s,H_2^s)$.

Remark 2. Even if the volume of E_s decreases regularly (if $g(x_s)\neq 0$), the diameter of E_s may tend to infinity for $s\to\infty$, which then leads to amplified rounding errors in the update formulae. It has been noted by several researchers, see e.g. Gill et al. ((1981) that - for reasons of stability - the update formulae should be written for the matrixes $I_S=B_S Q_s$, where Q_s is an (arbitrarily chosen) orthogonal matrix and $B_k B_k^*=A_k$. In Sonnevend (1983) it is shown that these diameters can be kept bounded by introducing "stabilization steps" in which the intersection $E_{s_k} \cap E_o$ is included in an ellipsoid of (uniformly in (n,k)) bounded diameter and small volume (i.e. proportional to vol E_{s_k}). It is shown there that by a suitable stopping rule one obtains thus an algorithm in which - in order to compute f* within accuracy ε - it is enough to have rounding and measurement errors not greater than ε^7 const (if - for f - the existence of a finite Lipschitz constant is assumed), moreover the sequence of stabilizing steps (and some other safeguards) can be chosen so that the essential complexity, convergence features of the original ellipsoid method are maintained.

3. MODIFICATIONS YIELDING THE REQUIRED ASYMPTOTIC BEHAVIOUR

We shall give the most simple modifications by which the required (asymptotic) properties can be ascertained for functions with well conditioned minimum. Here the (usual) notion of a well conditioned minimum for "smooth" functions is given below; for piecewise linear functions we define this notion by requiring the following assumption to be fulfilled for f: there exist finite and positive numbers d_1 and D_1 such that - for the unique minimumpair of f -

$$f^*+d_1||x-x^*||\le f(x)\le D_1||x-x^*||+f^*, \tag{3.1}$$

holds in a (convex) neighbourhood of x^*, V_0, where there is no point z - other than x^* - for which $(z,f(z))$ is a vertex of the graph of f. (Let us note that for GCGM the analogous, in fact more simple modifications allow us to obtain finite termination for arbitrary, piecewise linear functions).

The existence of a well-conditioned minimum for a "smooth" function f will be ensured by requiring that $f\in C^2$, and for its unique minimumpoint x^*,

$$g(x^*)=0,\ \frac{\partial g(x)}{\partial x} =: B(x),\ \text{ is nonsingular at } x=x^*, \tag{3.2}$$

$$||B(x)-B(x^*)||\le L||x-x^*||, \text{ for some, finite } L, \tag{3.3}$$

and for x in some convex neighbourhood V_1 of x^*.
Without loss of generality we can assume that

$$d_2||z||^2\le\langle B(x)z,z\rangle\le D_2||z||^2, \text{ for all } z \text{ in } V_1,$$

and some positive finite constants d_2, D_2.

Let $e_1,\dots,e_n$ be the orthonormed system of coordinate vectors in the X space. We define a matrix function $\tilde{B}(x)$ as the unique solution of

$$\tilde{B}(x)e_j=(g(\zeta_j)-g(x))||g(x))||^{-1}\qquad, j=1,\dots,n, \tag{3.4}$$

where $\zeta_j=x+||g(x)||e_j$, $j=1,\dots,n$, for all x such that $||g(x)||\ge\varepsilon_1$ is a prefixed, small number. In order to simplify the phrasing of the proofs below, we shall set $\varepsilon_1=0$.

Now we define (the construction of) the ellipsoids $\hat{E}_s$, $s=0,1,\dots,$ in the modified GEM by induction with respect to s. Let $\tilde{z}_{-1}:=x_0$, $\hat{E}_0:=E_0$. Suppose that - at step s - we have already computed an ellipsoid $\hat{E}_{s-1}$ and a vector $\tilde{z}_{s-1}$ (the latter is that point among those where f and g has been evaluated,which yields the smallest value for f). We define $x_s:=x(c(\hat{E}_{s-1}))$, (see (2.3)) and compute

$$\hat{z}_s:= \tilde{z}_{s-1}- B_{s-s \bmod n}\, g(\tilde{z}_{s-1}), \text{where } B_{nq}:=\tilde{B}(\tilde{z}_{nq-1}), q=0,1,\dots, \tag{3.5}$$

if in the computation of $\tilde{B}^{-1}(z)$ for $z=\tilde{z}_{nq-1}$ - say by a QR factorization - we obtain an inverse whose maximal element (or Frobenius norm) is not larger than a prefixed (large) number Ω, otherwise we define $\hat{z}_s:=x_s$.

Next we evaluate the functions f and g at x_s and at $\hat{z}_s$ and compute the ellipsoids, see (2.7)

$$E_s:=\bar{E}(\hat{E}_{s-1},H_1^s,H_2^s) \text{ and } \tilde{E}_s:=\bar{E}(\hat{E}_{s-1},\tilde{H}_1^s,\tilde{H}_2^s), \tag{3.6}$$

where $\tilde{H}_i^s$, $i=1,2$ are defined as in (2.4) but replacing x_s by $\hat{z}_s$.

In order to define the (proximal point) cutting plane step, which will be fulfilled only once after each n,consecutive iterations, i.e. for $s=nq+r$, $q=0,1,\dots$, r fixed (arbitrarily: say $r:=0$) - we need the values of the (asymptotically exact) lower bounds, ℓ_{s-1}, (see Remark 1 above, of course ℓ_j can now be computed from the volume of $\hat{E}_j$). We fix a number $\lambda>1$ and solve the problem (if $||g(\tilde{z}_{s-1})||\geq\varepsilon_1$)

$$\inf\{||\tilde{z}_{s-1}-y||\,|\,f(\tilde{z}_{s-1})+<y-\tilde{z}_{s-1},g(\tilde{z}_{s-1})>=\ell_{s-1}\}, \tag{3.7}$$

(we set $\ell_0:=m$), and evaluate f and g at its unique solution point, $x_{s,1}$, if it is defined and belongs to λL_0. Suppose now - by induction with respect to the value of k - that $x_s=x_{s,0}$, $x_{s,1},\dots,x_{s,k}$, $k< n$ are already defined and denote the linear functions, corresponding to these points by $L_j(y):=f(x_{s,y}) + <y-x_{x,j}, g(x_{s,j})>$, $j=0,\dots,k$. We define $x_{s,k+1}$ as the unique solution point (if it exists and belongs to λL_0) of

$$\inf\{||\tilde{z}_{s-\bar{1}}\ y||\,|\min \max_{j\leq k} L_j(y))=:h_{s,k}=\ell_{s-1}\}, \tag{3.8}$$

where for $k=n$ the equality sign before ℓ_{s-1} should be replaced by the inequality sign $\geq$. Finally we evaluate the function f at $x_{s,n}$ and compute the minimal volume ellipsoid, see (2.8)-(2.10)

$$\check{E}_s:=E^*(\hat{E}_{s-1},\hat{S}_s),\ \hat{S}_s:=\{(u,v)\mid h_{s,n}\le v\le f(x_{s,n})\}, \tag{3.9}$$

in the cases when either (3.8) or (3.9) has no solutions (inside λL_0) we set $\check{E}_s:=\hat{E}_{s-1}$.

Now the description of the proposed algorithm is finished by defining $\hat{E}_s$ to be that one among E_s, $\tilde{E}_s$, $\check{E}_s$ which has the smallest volume.

Remark 3. Note that one could use - instead of the values $g(\zeta_j)$ and $f(x_{s,j})$, $g(x_{s,j})$, $j=,1,\ldots,n$ - the values of f and g at points computed earlier, say at x_{s-j}, $j=1,\ldots,n$ in order to define recursively updated quadratic (resp. piecewise linear) approximations. We did not do so both for simplicity and for reasons of stability. What is important is that the number of arithmetical operations per function evaluation remains in the modified method $O(n^2)$ (in the average: over periods $[s, s+n]$), while·for the volumes of $\hat{E}_s$ we have - as a consequence of (3.6) and (2.11)

$$\text{vol } \hat{E}_s\le\exp(-(2(n+1))^{-1})\text{vol } \hat{E}_{s-1},\ s=1,2,\ldots\ . \tag{3.10}$$

Thus we have proved the first part of the following theorem.

Theorem 2. The modification of GEM described above has the required global and asymptotic convergence properties.

Proof. From (3.10) and Lemma 1 and the assumption (3.1) follows that - unless the algorithm is stopped: a trivial alternative, which we shall neglect in what follows - there exists a finite value for q_0, such that $\tilde{z}_{nq_0}-1\in V_0$, which is estimable in terms of the constants n,m,M,L_0,d_1,D_1 and V_0. Since $f(\tilde{z}_{s-1})$ is monotonically decreasing in s, and the lower bounds ℓ_s are asymptotically exact (with a predictable convergence rate for $(f^*-\ell_s)$), there exists a q_0 so large that for $s=nq_0$

$$(f^*-\ell_{s-1})\,\frac{D_1}{d_1}\le H-f^*\ , \tag{3.11}$$

where H is the maximum of the values such that - except (x^*,f^*) - no vertex $(x,f(x))$ of the graph of f exists for which $f(x)<H$. Now (3.11) implies that the ellipsoid $\hat{E}_s$ has zero volume: i.e. finite termination occurs.

The equality $x_{s,n}=x^*$ is established by showing (inductively) that the linear functions L_j, $j=0,1,\ldots,n$ are then all different (and defined) as a consequence of the definitions of $x_{s,j}$ and of the assumptions (3.1), (3.11).

In order to study the asymptotic behaviour of the proposed method for functions f satisfying (3.2), (3.3) we show first that

$$||\tilde{B}(x)-B^*||\leq K||x-x^*||, \text{ if } x\in V_2 , \qquad (3.12)$$

where V_2 is another neighbourhood of x^*, whose size - as well as the corresponding value of K - can be estimated from below in terms of (d_2,D_2,L,V_1).

Indeed, from the identity

$$\frac{\partial f(b)}{\partial x^i} = \frac{\partial f(a)}{\partial x^i} + \sum_{j=1}^{n} \int_o^1 \frac{\partial f(a+s(b-a))}{\partial x^i \partial x^j} (b^j-a^j)ds, \; i=1,\dots,n ,$$

one obtains that, for $z\in V_1$

$$||g(z)-B^*(z-x^*)||\leq L||z-x^*||^2 , \qquad (3.13)$$

Therefore, if z is such that $\zeta_j\in V_1$ for $j=1,\dots,n$,

$$\|(g(\zeta_j)-g(z))\|\|g(z)\|^{-1}-B^*e_j\|\leq L(||\zeta_j-x^*||^2+||z-x^*||^2)||g(z)||^{-1}.$$

Now observe that - again from (3.13) -

$$||z-x^*||(d_2-L||z-x^*||)\leq||g(z)||\leq(D_2+L||z-x^*||)||z-x^*||, \qquad (3.14)$$

By construction we have the inequalities

$$||\zeta_j-x^*||\leq||g(z)||+||z-x^*|| \qquad (3.15)$$

From all these the existence of V_2 and K with the property (3.12) follows by simple calculations.

We now need a well known fact from the theory of quasi-Newton methods, see e.g. Ortega, Rheinbolt (1970): suppose that for the iteration

$$z_{i+1}:=z_i-B_i^{-1}g(z_i), \; i=0,1,\dots,$$

where g satisfies the conditions (3.2), (3.3) one has an estimation

$$||B_i-B^*||\leq K||z-x^*||, \text{ for all } i=0,1,\dots,.$$

Then there exists a neighborhood V_3 of x^* and a finite number c, whose size (resp. value) can be estimated from below (resp. above) in terms of (d_2, D_2, V_2, L, K) only, such that if $x_0\in V_3$ then

$$||x_{i+1}-x^*||\leq c||x_i-x^*||^2, \text{ for all } i=0,1,\dots . \qquad (3.16)$$

From the fact that the sequence $f(\tilde{z}_{s-1})$, $s=1,2,\ldots$ is monotonically nonincreasing and tending - for $s\to\infty$ - to f^* by (3.10), we obtain, in view of the conditions (3.2) and (3.3), that $\tilde{z}_{s-1} \to x^*$, for $s\to\infty$. Therefore we shall have

$$\tilde{z}_{nq-1} \in V_3, \text{ if } q\geq q_0 \qquad (3.17)$$

and - if Ω (and q_0) is chosen to be large enough - the matrixes $\tilde{B}_{nq}$ will be defined for $q>q_0$ so that the iteration (3.5)-(3.9) assures that $\tilde{z}_{s-1}\in V_3$ for all $s\geq nq_0$, (notice that the maps $(I-\tilde{B}_{qn})$ are contractive for all q large enough, in fact their norms tend to zero).

It remains only to prove the next Lemma.

Lemma 2. Suppose that an ellipsoid E is contained in a ball of radius bR, H_1 and H_2 are halfspaces as specified in (2.5), with $p=(-g,1)$, such that the X projection of the intersection of their boundaries has a common point with the the X projection of E, then

$$\text{vol } \bar{E}(E,H_1,H_2) \leq \|g\| b^{n+1} R^n \Psi_n \qquad (3.18)$$

where $\Psi_n \to 0$ for $n\to\infty$.

Proof. By the assumptions made, the minimal horizontal layer containing the intersection $E\wedge H_1\wedge H_2$ has a width not greater than $2b\|g\|R$. Therefore the minimal valume ellipsoid $E^*(E,S)$ has - see (2.2) - also a volume not greater than

$\|g\| b\sqrt{n+1}\,\varkappa_n (1-(n+1)^{-1})^{\frac{n}{2}} R^n b^n$.

Now we shall apply this Lemma for $E=\hat{E}_{s-1}$ for $s=nq$, $q\geq q_0$, in order to estimate the volume of $\bar{E}(\hat{E}_{s-1},\tilde{H}_1^s,\tilde{H}_2^s)$, see (3.6). Note that if one is not making the stabilization mentioned in Remark 2 and guaranteeing the existence of a finite constant b (for all n uniformly) then everything remains true with b=1 if in the definition (3.6) we set

$$\tilde{E}_s := \bar{E}(E_0,\tilde{H}_1^s,\tilde{H}_2^s). \qquad (3.19)$$

We obtain from (3.16) - (3.18) that

$$\text{vol}\|\hat{E}_{qn}\| \leq c_4 \text{ vol}^2(\hat{E}_{qn-1}), \text{ for } q\geq q_0 \qquad (3.20)$$

where the constant c_4 is independent of q. From this by Lemma 1 we obtain the superlinear convergence of the values of $(f(\tilde{z}_{s-1})-f^*)$, which implies by the conditions (3.1)-(3.2) the superlinear convergence of $||\tilde{z}_{s-1}-x^*||$, for $s\to\infty$.

REFERENCES

Ech-Cherif, A., Ecker K.G. (1984), "A class of rank two ellipsoid algorithms for convex programming", Math. Programming 29(2):187-202.

Gill, P.E, Murray, W., Saunders, M.A. and Wright, M.H., (1982) A numerical investigation of ellipsoid algorithms for large scale linear programming, in G.B. Dantzig, M.A.H. Dempster and M.J. Kallio (eds.) Large-Scale Linear Programming, IIASA Proceedings Series, CP-81-S1: 487-511 (vol.1.)

König, H., Pallaschke, D. (1981), "On Khachian's Algorithm and Minimal Ellipsoids", Numerische Mathematik (36),2: 197-211

Lemarechal, C., Strodiot, J.J., and Bihain, A. (1981), "On a Bundle Algorithm for Nonsmooth Optimization", in O.L. Mangasarian, R.R. Meyer and S.M. Robinson (eds.), Nonlinear Programming, Academic Press, New York.

Ortega, J.M, Rheinbolt, W.C. (1970), Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York

Shor, N.Z. (1982), "Generalized Gradient Methods of Nondifferentiable Optimization Employing Space Dilatation Operations", in A. Bachem, M. Grötschel, and B. Korte (eds.), Mathematical Programming, The State of Art, Springer, Berlin.

Sonnevend, Gy. (1983), "Stabilization of the ellipsoid method for computing the minimal value of a convex function" to appear in Applied Mathematics and Optimization,

Sonnevend, Gy. (1984), "Acceleration and Implementations of the method of centers of gravity, IIASA Collaborative Paper, System and Decision Sciences Program, IIASA, Laxenburg, Austria, 15 p.

Stoer, J. (1980) Introduction to Numerical Analysis, Springer, Berlin

Zowe, J. (1984) Nondifferentiable Optimization, to appear in ASI Proceedings on Computational Mathematical Programming (Conf. at Bad Windsheim, July 1984)

NUMERICAL METHODS FOR MULTIEXTREMAL NONLINEAR PROGRAMMING PROBLEMS WITH NONCONVEX CONSTRAINTS

Roman G. Strongin
Gorky State University, Gorky, USSR

1. INTRODUCTION

Existing approaches to multiextremal optimization (see Evtushenko, 1971; Ivanov, 1972; Mockus, 1977; Strongin, 1978; Zilinskas, 1978) mostly focus on numerical methods for unconstrained problems. Constraints are usually handled by introducing penalty functions since other techniques (see, for example, Demyanov and Vasiliev, 1981) require the minimizing function and the constraints to be convex, unimodal, or to have other properties. Below we present a new algorithm for multiextremal problems with nonconvex constraints which does not make use of penalties.

2. ONE-DIMENSIONAL CASE

Let us consider the problem

$$\min\{h(x) : x \in [a,b],\ g_i(x) \leqslant 0,\ 1 \leqslant i \leqslant m\} \quad , \tag{1}$$

where the function $h(x)$ to be minimized (denoted below by $g_{m+1}(x)$) and the left-hand sides $g_i(x)$, $1 \leqslant i \leqslant m$, of the constraints are all Lipschitz functions. We also assume that the functions g_i, $1 \leqslant i \leqslant m+1$, are defined and computable only in the corresponding domains Q_i, where

$$Q_1 = [a,b] \ , \quad Q_{i+1} = \{x \in Q_i : g_i(x) \leqslant 0\} \ , \quad 1 \leqslant i \leqslant m \quad ,$$

and the following inclusions obviously hold

$$[a,b] = Q_1 \supset Q_2 \supset \ldots \supset Q_{m+1} \supset Q_{m+2} \quad ,$$

where $Q_{m+2} \neq \emptyset$. With each point $x \in (a,b)$ we associate an index

$$s = s(x) \quad , \qquad 1 \leqslant s \leqslant m+1 \quad , \tag{2}$$

defined by the requirements $x \in Q_s$ and $x \notin Q_{s+1}$. The maximum value of the index (2) over the domain $[a,b]$ will be denoted by N, i.e.,

$$1 \leqslant N \overset{df}{=} \max \{s(x) : x \in [a,b]\} \quad . \tag{3}$$

Now we introduce the optimization problem

$$g_N^* \overset{df}{=} \min \{g_N(x) : x \in Q_N\} \quad , \tag{4}$$

which is defined for any value N from (3). If $N = m+1$ the solution of this problem is simultaneously the solution of the source problem (1). If, on the other hand, $N < m+1$ (i.e., the constraints in (1) are incompatible) we obtain the inequality $g_N^* > 0$, which provides a test for this case. The function

$$H(x) = g_s(x) - \begin{cases} 0 & , \text{ if } s = s(x) < N \\ g_N^* & , \text{ if } s = s(x) = N \end{cases} \quad , \tag{5}$$

is associated with the problem (4) in the following way; the point x^* representing the absolute minimum of $H(x)$ over $[a,b]$ is such that $g_N(x) = g_N^*$, $x^* \in Q_N$ and $H(x^*) = 0$, i.e., unconstrained minimization of the function $H(x)$, $x \in [a,b]$, yields the solution of problem (4).

Since the value denoted in (4) and (5) by g_N^* is not known *a priori*, the method described below employs an adaptive estimate of this value.

3. ALGORITHM FOR ONE-DIMENSIONAL MULTIEXTREMAL PROGRAMS

Each iteration of the proposed method at any arbitrary point $x \in [a,b]$ involves the determination of some corresponding value $f(x) = g_s(x)$ (where $s = s(x)$ is the index from (2)), obtained by successive calculation of the value of the functions $g_i(x)$, $1 \leqslant i \leqslant s$. It is a condition that $g_{i+1}(x)$ can be calculated only if $g_i(x) \leqslant 0$. The calculations are terminated when either the inequality $g_s(x) > 0$ or the equality $s = m+1$ is satisfied. The above process therefore results in the evaluation of both $f(x)$ and $s(x)$ for any given point x.

The first iteration is carried out at an arbitrary point $x^1 \in (a,b)$. The choice of any subsequent point x^{k+1}, $k \geqslant 1$, is determined by the following rules:

(a) points $x^1,\dots,x^k$ from previous iterations are renumbered using subscripts in the following way:

$$a = x_0 < x_1 < \dots < x_i < \dots < x_k < x_{k+1} = b$$

and associated with values $z_i = f(x_i)$, $1 \leqslant i \leqslant k$, computed at these points (values z_0 and z_{k+1} are undefined);

(b) the following sets of indices are constructed:

$$I_0 = \{0,k+1\} \quad , \qquad I_s = \{i : 1 \leqslant i \leqslant k \,, \, s = s(x_i)\} \quad ,$$

$$S_s = I_0 \cup \dots \cup I_{s-1} \quad , \quad T_s = I_{s+1} \cup \dots \cup I_{m+1} \quad , \quad 1 \leqslant s \leqslant m+1$$

and the following values calculated:

$$M_s = \max \{|z_i - z_j|(x_i - x_j)^{-1} : i,j \in I_s \,, \, i < j\} \quad , \tag{6}$$

$1 \leqslant s \leqslant m+1$. If $|I_s| < 2$ or M_s from (6) is equal to zero, it is assumed that $M_s = 1$;

(c) for all nonempty sets I_s, $1 \leqslant s \leqslant m+1$, the following values are determined:

$$z_s^* = \begin{cases} 0 \quad , \quad \text{if } T_s \neq \emptyset \quad , \\ \min \{z_i : i \in I_s\} \quad , \quad \text{if } T_s = \emptyset \quad ; \end{cases}$$

(d) for each interval (x_{i-1},x_i), $1 \leqslant i \leqslant k+1$, the value $R(i)$ (called the *characteristic*) is computed, where

$$R(i) = (x_i - x_{i-1}) + (z_i - z_{i-1})^2/M_s^2(x_i - x_{i-1}) - 2(z_i + z_{i-1} - 2z_s^*)/rM_s \;,$$
$$i-1, i \in I_s \quad ;$$

$$R(i) = 2(x_i - x_{i-1}) - 4(z_i - z_s^*)/rM_s \quad , \quad i \in I_s \quad , \quad i-1 \in S_s \quad ;$$

$$R(i) = 2(x_i - x_{i-1}) - 4(z_{i-1} - z_s^*)/rM_s, \quad i-1 \in I_s, \quad i \in S_s$$

(here r is a parameter, with a value greater than 1). The interval (x_{t-1},x_t) with maximal characteristic $R(t) = \max \{R(i) : 1 \leqslant i \leqslant k+1\}$ is then determined.

If $s = s(x_{t-1}) = s(x_t)$ then the next iteration is carried out at a point

$$x^{k+1} = [(x_t+x_{t-1})/2] - [(z_t-z_{t-1})/2rM_s] \quad ;$$

otherwise, i.e., if $s(x_{t-1}) \neq s(x_t)$, the second term in the above formula is omitted.

4. SUFFICIENT CONVERGENCE CONDITIONS

THEOREM 1. *Assume that for* N *from* (3) *the following conditions are satisfied:*

(a) *domains* Q_i, $1 \leqslant i \leqslant N$, *are the finite unions of intervals of positive length in* [a,b];

(b) *functions* $g_i(x)$, $1 \leqslant i \leqslant N$, $x \in Q_i$, *admit Lipschitz extensions (with corresponding constants* K_i*) over* [a,b];

(c) *point* x^* *is a solution to problem* (4);

(d) *the inequality* $rM_s > 2K_s$, $1 \leqslant s \leqslant N$, *for* N *from* (3) *and for* M_s *from* (6), *is satisfied for some step in the search process.*

Then:

(1) x^* *is an accumulation point of the sequence* $\{x^k\}$ *generated by the algorithm described above and convergence to* x^* *is bilateral if* $x^* \neq a$ *and* $x^* \neq b$;

(2) *any other accumulation point* x' *of the sequence* $\{x^k\}$ *is also a solution to problem* (4).

Computer simulations of the search process for a given one-dimensional problem with two constraints yield the results presented in Figure 1. The plotted curves represent functions g_i, $1 \leqslant i \leqslant 3$, the labels corresponding to the values of subscript i. Vertical bars indicate the iteration points $x_1,\ldots,x_{57}$ and are arranged in three rows according to the values of indices $\nu(x_k)$, $1 \leqslant k \leqslant 57$.

The points marked on the broken line in the lower part of the figure represent pairs (x^k,k), where k is the step number and x^k is the coordinate of the corresponding iteration. This simulation terminated at the 58th step when the condition $x_t-x_{y-1} \leqslant 0.001$ (the stopping rule) was satisfied. (The right-hand side of this condition is of course the required accuracy.)

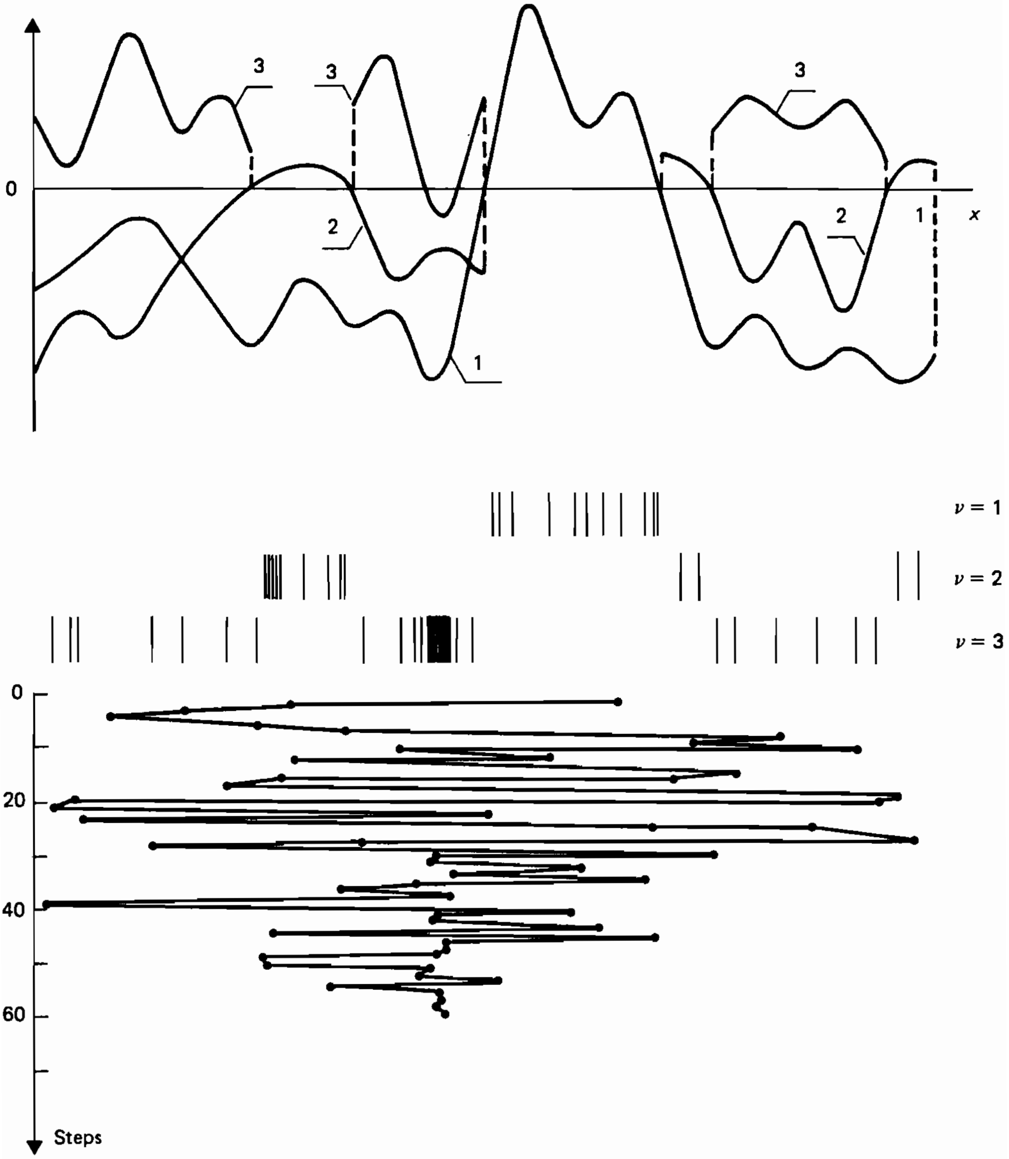


FIGURE 1 Computer simulation of the one-dimensional search process

5. MULTIDIMENSIONAL MULTIEXTREMAL NONLINEAR PROGRAMS

Program (1) could be generalized to the multidimensional problem

$$\min \{h(y) : y \in D , g_i(y) \leqslant 0 , 1 \leqslant i \leqslant m\} \quad , \tag{7}$$

where

$$D = \{y \in R^n : a_j \leqslant y_j \leqslant b_j , 1 \leqslant j \leqslant n\} \quad . \tag{8}$$

This problem can be reduced to one dimension by employing a Peano-type space-filling curve mapping a unit interval [0,1] on the x axis onto the n-dimensional domain (8). Thus it is possible to find the minimum in (7) by solving the one-dimensional problem

$$\min \{h(y(x)) : x \in [0,1] , g_i(y(x)) \leqslant 0 , 1 \leqslant i \leqslant m\} \quad . \tag{9}$$

As shown in Strongin (1978), the Peano transformation y(x) provides a function $g_i(y(x))$ that satisfies Hölder's condition if the source function $g_i(y)$ satisfies the Lipschitz condition. Thus problem (9) could be solved by a generalized version of the above algorithm. The difference between these two algorithms is that all distances of the type $(x_i - x_{i-1})$ in the original algorithm must be replaced by values $(x_i - x_{i-1})^{1/n}$ in the new algorithm, for which some analog of Theorem 1 will hold.

REFERENCES

Evtushenko, Yu.G. (1971). Numerical method for seeking the global extremum of a function (nonuniform grid search). Journal of Computational Mathematics and Mathematical Physics, 11: 1390-1403 (in Russian).

Ivanov, V.V. (1972). On optimal minimization algorithms for functions of some classes. Kibernetika, (4): 81-94 (in Russian).

Mockus, J. (1977). On Bayesian methods of seeking the extremum and their applications. In Gilchrist (Ed.), Information Processing 77. North-Holland, Amsterdam, pp. 195-200.

Demyanov, V.F. and Vasiliev, L.V. (1981). Nondifferentiable Optimization. Nauka, Moscow (in Russian).

Strongin, R.G. (1978). Numerical Methods in Multiextremal Problems. Nauka, Moscow (in Russian).

Zilinskas, A. (1978). On one-dimensional multimodal minimization. In Transactions of the Eighth Prague Conference on Information Theory, Stat. Dec. Functions, and Random Processes, pp. 393-402.

A MODIFICATION OF THE CUTTING-PLANE METHOD WITH ACCELERATED CONVERGENCE

V.N. Tarasov and N.K. Popova
Department of Physics and Mathematics, Syktyvkar State University, Syktyvkar, USSR

1. INTRODUCTION

The cutting-plane method of J.E. Kelley [3] is widely used in convex programming. There are some modifications of this method (see, e.g. [4]), which in some cases accelerate its convergence. In this paper we discuss another modification of the Kelley method based on the idea described in [2] for solving equation f(x) = 0 with multiple roots by the Newton method. It is well-known that if an initial approximation is close enough to the root (and some additional conditions are satisfied) then the Newton method is of quadratic rate of convergence. But it is not the case if, for example, $f(x) = x^2$ where $x \in E_1$. Then the multiplicity of the root $x^* = 0$ is $m = 2$. The Newton method implies

$$x_{k+1} = x_k - f(x_k)/f'(x_k) = x_k - x_k^2/2x_k = \frac{1}{2}x_k$$

i.e. the rate of convergence is geometric (and its coefficient is $\frac{1}{2}$). But if we take

$$x_{k+1} = x_k - m\,f(x_k)/f'(x_k)$$

(where m is the multiplicity of the root of the equation f(x) = 0); then in our example we get

$$x_{k+1} = x_k - 2x_k^2/2x_k = 0$$

Generally speaking, such a modification has quadratic rate of convergence. This idea is behind the approach we are going to present (a short description can be found in [6,7]).

2. AN ALGORITHM

Let f be a convex function defined and finite on the n-dimensional Enclidean space E_n, $\Omega \subset E_n$ be a compact convex set. It is necessary to find a point $x^* \in \Omega$ such that

$$f(x^*) = \min_{x \in \Omega} f(x)$$

By $\partial f(u)$ we denote the subdifferential of f at u, i.e.,

$$\partial f(u) = \{v \in E_n \mid f(x) \geq f(u) + (v, x-u) \quad \forall x \in E_n\} \tag{1}$$

Choose $v(u) \in \partial f(u)$ and let us introduce the function

$$F(x,u,\varepsilon) = f(u) + (\tfrac{1}{2} + \varepsilon)\ (v(u),\ x-u) \tag{2}$$

Take an arbitrary point $x_0 \in \Omega$ and put $\sigma_0 = \{x_0\}$. Let $\sigma_K = \{x_0, x_1, \ldots, x_k\}$ have been found. Let us choose $\varepsilon_k = \varepsilon(x_k) \geq 0$ such that

$$F(x^*, x_k, \varepsilon_k) \leq f(x^*) \tag{3}$$

Such an ε_k exists for any k since for $\varepsilon = \frac{1}{2}$ it follows from (1) that

$$f(x^*) \geq f(x_k) + (v(x_k),\ x^* - x_k)$$

Therefore we can assume that $0 \leq \varepsilon_k \leq \frac{1}{2}$. Now let us introduce the function

$$\phi_k(x) = \max_{i \in 0:k} F(x, x_i, \varepsilon_i) \tag{4}$$

and find

$$x_{k+1} = \arg\min \{\phi_k(x) \mid x \in \Omega\} \tag{5}$$

Now take $\sigma_{k+1} = \sigma_k \cup \{x_{k+1}\}$ and continue in the same manner

Theorem 1. If for some k

$$\phi_\kappa(x_{k+1}) = f(x_{k+1})$$

then x_{k+1} is a minimum point of f on Ω. Otherwise any limit point of the sequence $\{x_k\}$ is a minimum point of the function f on Ω.

Proof. The first part of the theorem is obvious:

$$f(x_{k+1}) \geq f(x^*) \geq \phi_k(x^*) \geq \phi_k(x_{k+1}) = f(x_{k+1})$$

which implies

$$f(x^*) = f(x_{k+1})$$

(inequality $f(x^*) \geq \phi_k(x^*)$ above follows from (3) and (4)).

To prove the rest of the theorem assume the opposite: then there exists a subsequence $\{x_{k_s}\}$ such that $x_{k_s} \to \bar{x}$, $k_s \to \infty$ and

$$f(\bar{x}) > f(x^*) \tag{6}$$

By construction

$$\phi_{k_s}(x_{k_{s+1}}) = \max_{i \in 0:x} F(x_{k_{s+1}}, x_i, \varepsilon_i) \geq F(x_{k_{s+1}}, x_{k_s}, \varepsilon_{k_s}) =$$

$$= f(x_{k_s}) + (\frac{1}{2} + \varepsilon_{k_s}) (v(x_{k_s}), (x_{k_{s+1}} - x_{k_s}))$$

$$\xrightarrow[k_s \to \infty]{} f(\bar{x}) .$$

On the other hand, since $\phi_k(x) \leq \phi_{k+\ell}(x) \; \forall \ell > 0, \forall x$ then

$$\phi_{k_s} (x_{k_{s+1}}) \leq \phi_{k_{s+1}-1} (x_{k_{s+1}}) \leq \phi_{k_{s+1}-1} (x^*) \leq f(x^*) ,$$

i.e. $f(\bar{x}) \leq f(x^*)$ which contradicts (6).

Remark 1. If $\varepsilon_k = \frac{1}{2}$ $\forall k \in 0 : \infty$ the method becomes the Kelley method.

Remark 2. Let f be a quadratic function

$$f(x) = (Ax,x) + (b,x)$$

where A is an $n \times n$ positive definite matrix, $b \in E_n$; $\Omega \subset E_n$ is a convex set. If $x^* = \arg\min \{f(x) | x \in \Omega\} \in \text{int}\,\Omega$ then by the necessary condition for a minimum

$$f'(x^*) = 0 \quad . \tag{7}$$

Therefore

$$f(x) + \frac{1}{2}(f'(x), x^* - x) - f(x^*) = (Ax,x) + (b,x) + \frac{1}{2}(2Ax + b, x^* - x) -$$

$$- (Ax^*, x^*) - (b,x^*) = (Ax^*, x-x^*) + \frac{1}{2}(b, x-x^*) = (Ax^* + \frac{1}{2}b, x - x^*) \quad .$$

Since $f'(x^*) = 2Ax^* + b$, then from (7)

$$f(x) + \frac{1}{2}(f'(x), x^*-x) - f(x^*) = (f'(x^*), x-x^*) = 0 \quad \forall x \in \Omega$$

i.e.

$$F(x^*, x, 0) \leq f(x^*) \qquad \forall x \in \Omega \tag{8}$$

and in (3) we can choose $\varepsilon_k = 0$.

Thus, for a quadratic convex function we can always take $\varepsilon_k = 0$ $\forall k$.

Theorem 2. If f is a strongly convex twice continuously differentiable function then there exists a sequence $\{\varepsilon_k\}$ satisfying condition (3) such that $\varepsilon_k \to 0$ as $k \to \infty$.

Proof. Since f is twice continuously differentiable then the matrix of the second derivatives is strictly positive definite. Let $x^* = \arg\min \{f(x) | x \in \Omega\}$. Assume that $x^* \in \text{int}\ \Omega$. Since f is strongly convex then there exists $\mu > 0$ such that

$$f(x^*) \geq f(x) + (f'(x),\ x^* - x) + \mu\|x^*-x\|^2 \quad \forall\, x \in E_n \quad .$$

It implies

$$(f'(x),\ x^* - x \leq - \mu\|x^*-u\|^2 \tag{9}$$

Let us introduce the function

$$f_1(x) = f(x^*) + (f'(x^*),\ x-x^*) + \frac{1}{2}(f''(x^*)(x-x^*),\ (x-x^*)) \ .$$

Clearly

$$f(x) = f(x^* + (x-x^*)) = f_1(x) + o(\|x-x^*\|^2) \tag{10}$$

where

$$o(\| x-x^*\|^2) = \frac{1}{2}(f''(x^*+\theta(x)(x-x^*)) -$$

$$- f''(x^*)(x-x^*),(x-x^*)),\theta = \theta(x) \in (0,1) \quad .$$

and

$$f_1(x^*) = f(x^*) \quad . \tag{11}$$

Since f_1 is a quadratic function then it follows from (8) that

$$f_1(x) + \frac{1}{2}(f_1'(x),\ x^*-x) \leq f_1(x^*) \quad \forall\, x \in \Omega \quad . \tag{12}$$

From (10)

$$f_1(x) = f(x) + o(\|x-x^*\|^2) \quad .$$

Therefore

$$f_1'(x) = f'(x) + o(\|x-x^*\|)$$

and (11) and (12) imply

$$f(x) + o(\|x-x^*\|^2) + \frac{1}{2}(f'(x),x^*-x) + (o(\|x-x^*\|),x^*-x) \leq f_1(x^*) = f(x^*)$$

or

$$f(x) + \frac{1}{2}(f'(x), x^*-x) + o(\|x-x^*\|^2) \leq f(x^*) \quad . \qquad (13)$$

Since

$$o(\|x-x^*\|^2) = \alpha(x)\|x-x^*\|^2 \text{ where } \alpha(x) \xrightarrow[x\to x^*]{} 0$$

then from (9) it follows that

$$\frac{|\alpha(x)|}{\mu}(f'(x), x^*-x) \leq -|\alpha(x)|\|x-x^*\|^2 \quad .$$

Moreover,

$$\frac{|\alpha(x)|}{\mu}(f'(x), x^*-x) \leq \alpha(x)\|x-x^*\|^2 \quad .$$

Hence (13) implies

$$f(x) + (\frac{1}{2} + \varepsilon(x))(f'(x), x^*-x)) \leq f(x^*) \qquad (14)$$

where

$$\varepsilon(x) = \frac{|\alpha(x)|}{\mu} \xrightarrow[x\to x^*]{} 0 \quad . \qquad (15)$$

Thus, if in the method described above (see (5)) we choose $\varepsilon_k = \varepsilon(x_k)$ then

1) $x_k \to x^*$ (since (14) implies (3))

2) $\varepsilon_k \to 0$ (due to (15)). Q.E.D.

Remark 3. Computational experiments have shown that the method described is very efficient (and for a quadratic function under some additional conditions it is even finite).

REFERENCES

1. Demyanov, V.F. and V.N. Malozemov, "Introduction to minimax", J. Wiley, N.Y. 1974.
2. Krylov, V.I., V.V. Bobkov, and P.N. Monastyrnyi, "Computational methods in mathematics", vol. 1, Moscow, Nauka, 1976 (in Russian)

3. Kelley, J.E. "The cutting-plane method for solving convex programs". SIAM J. Applied Math. vol. 8, N. 4, pp. 703-712.
4. More, J.J., B.S. Garbow, and K.E. Hillstrom, "Testing unconstrained optimization", AMC Trans. Math. Software, vol. 7, N. 1, 1981, pp. 17-41.
5. Wolfe, Ph. "Accelerating the cutting-plane method for nonlinear programming", SIAM J. Applied Math. vol. 9, N. 3, 1961, pp. 481-488.
6. Popova, N.K. and V.N. Tarasov, "A modification of the Kelley method with accelerated convergence". Deposits of VINITI, N. 3648-82 DEP, 1982, (in Russian).
7. Popova, N.K. and V.N. Tarasov, "On convergence of one modification of the Kelley method", (to be published in Vestnik of Leningrad University).

A FINITE ALGORITHM FOR SOLVING LINEAR PROGRAMS WITH AN ADDITIONAL REVERSE CONVEX CONSTRAINT

Nguyen Van Thuong and Hoang Tuy
Institute of Mathematics, Vien Toan Hoc, P.O. Box 631, Hanoi, Vietnam

1. INTRODUCTION

This paper presents an algorithm for solving the following problem :

$$\text{(P)} \quad \text{Minimize} \quad cx \ , \quad \text{s.t.}$$

$$x \in D \tag{1}$$

$$g(x) \leq 0 \tag{2}$$

where $D \subset R^n$ is a polytope and g is a finite concave function on R^n . Problems of this kind occur in certain economic and engineering applications.

Clearly, without the additional constraint (2) the problem would reduce merely to the ordinary linear program

$$\text{Minimize} \quad cx \ , \quad \text{s.t.} \quad x \in D \ . \tag{3}$$

Therefore, all the difficulties of the problem arise from the presence of the constraint (2) which is called a reverse convex constraint, meaning that it is the reverse of a convex constraint.

Linear programs with an additional reverse convex constraint like (P) have been first studied by Bansal and Jacobsen [3,4], Hillestad [6] and also Hillestad and Jacobsen [7]. In [3,4] the special problem of optimizing a network flow capacity under economies-of-scale was discussed. In [6] a branch and bound edge search procedure was developed for the problem (P) under the assumption that the concave function g is differentiable. In [7] , it was shown that an optimal solution for (P) lies on an edge of the polytope D . From this basic property, a characterization of the set of edge of D that can contain such an optimal solution was given and a pivot type algorithm for solving (P) was derived.

Problems more general than (P) have been treated by Rosen [11], Avriel and Williams [1,2], Meyer [9], Hillestad and Jacobsen [8], and also Hoang Tuy [13]. In the latter paper, a finite method was developed for globally minimizing a concave function under the constraints (1) (2). As specialized to problem (P), it provides an algorithm different from that of Hillestad and Jacobsen [7] and having the advantage of being still valid when D is an unbounded polyhedral convex set.

It should be noted that the method in [13] is based on an extension of a method of concave minimization under linear constraints due to Vu Thien Ban. On the other hand Hillestad and Jacobsen [7,8] have shown that cuts originally devised for concave programming could be as well used for reverse convex programming. Thus, the problem (P) and, more generally, the reverse convex programming problem, is closely related to the concave minimization problem.

The purpose of the present paper is to develop a finite procedure for solving (P) which exploits this relationship in a more systematic way than has been done in the previously cited references. It turns out that a linear program with an additional reverse convex constraint can be decomposed into an alternating sequence of linear programs (minimizing cx under constraints (1)) and concave programs (minimizing g(x) under constraints (1) and one additional constraint of the form $cx \le \alpha$) . Roughly speaking, the proposed algorithm switches between steps of two types: in the open region $g(x) < 0$, we use simplex pivots to improve the current feasible solutions, while at a feasible point on the boundary $g(x) = 0$, we must solve a concave program in order to decide whether or not the current feasible solution is optimal and if not to move to a better feasible solution in the region $g(x) < 0$. To solve these concave programs we can use any available finite algorithm, for instance the algorithm of Thieu-Tam-Ban [12] or that of Falk and Hoffman [5].

2. THE ALGORITHM

For the sake of convenience we shall make the following assumptions in this section:

(i) $\text{Min}\,\{ cx : x \in D \} < \text{Min}\,\{ cx : x \quad D, g(x)=0 \}$;

(ii) The function $g(x)$ is strictly concave and does not vanish at any vertex of D .

Assumption (i) simply means that the constraint (2) is essential : if (i) does not hold, then (P) is equivalent to the linear program (3). In the sequel we shall use this assumption in the following form :

For any feasible vertex u of D , there is a neighbouring vertex v such that

$$cv < cu \quad .$$

Assumption (ii) is not a too stringent one. Later we shall see that any concave function g can be made to satisfy this assumption by a slight " perturbation " . For our purpose, this assumption is convenient in that it will allow a significant simplification of the algorithm.

Let us first explain the basic ideas of the method to be proposed.

Suppose that a vertex x^o of the polytope D is available such that

$$g(x^o) < 0 \quad . \tag{4}$$

(We shall discuss later the case where such a vertex is not readily available : see Remark 1)

Since the constraint (2) is not binding for x^o it is natural to first improve x^o by moving only on the polytope D . We do this by applying the simplex procedure to the linear program (3) : if a neighbouring vertex x^1 to x^o exists such that $cx^1 < cx^o$ and $g(x^1) < 0$ we perform a simplex pivot to move to x^1 . This procedure can be continued until we find a pair of vertices u , v of D such that $g(u) < 0$, $g(v) \geq 0$. (This must occur in view of assumption (i)). Then we can move along the line segment $[u,v]$ to the point $\overline{x}$ where this segment meets the boundary $g(x) = 0$ (since g is strictly concave, and $g(u) < 0$, $g(v) \geq 0$, there is on the line segment $[u,v]$ just one point $\overline{x}$ satisfying $g(\overline{x}) = 0$) . Clearly $\overline{x}$ is the best feasible solution obtained so far. Therefore, it only remains to consider the polytope

$$D(\overline{x}) = \{ x \in D : cx \leq c\overline{x} \} \quad . \tag{5}$$

The question to be examined now is whether $D(\overline{x})$ has a vertex z such that $g(z) < 0$. For if we can find such a vertex, then the same procedure as before can obviously be repeated, with $D(\overline{x})$ and that vertex replacing D and x^o .

The best way to check whether $D(\overline{x})$ has a vertex z such that $g(z) < 0$, and to find such a vertex if it exists, is to solve the concave programming problem

$$\text{Min } \{ g(x) : x \in D(\overline{x}) \} \quad . \tag{6}$$

It turns out that, under Assumption (ii) if the optimal value of g in this program is zero, i.e. if there is no z in $D(\overline{x})$ such that $g(z) < 0$, then $\overline{x}$

is necessarily optimal to the original problem (Theorem 1 below). Otherwise, we shall find an optimal solution z of (6), i.e. a vertex of $D(\bar{x})$ such that $g(z) < 0$. Using then z in place of x^o, and $D(\bar{x})$ in place of D, we can restart the whole process in a new round.

In a formal way, the algorithm can be described as follows.

Initialization. Take a vertex x^o of D such that $g(x^o) < 0$. Set $D_o = D$.

Step 1. Starting from x^o, pivot via the simplex algorithm for solving the linear program

$$\text{Min } \{ cx : x \in D_o \} \tag{7}$$

until a pair of vertices u, v of D_o is found so that $g(u) < 0$, $g(v) \geq 0$, and $cv < cu \leq cx^o$. Let $\bar{x}$ be the (unique) point of the line segment $[u,v]$ such that $g(\bar{x})=0$. Go to Step 2.

Step 2. Solve the concave program

$$\text{Min } \{ g(x) : x \in D(\bar{x}) \} \tag{8}$$

where $D(\bar{x}) = \{ x \in D : cx \leq c\bar{x} \}$.

a) If the optimal value in this concave program is zero, stop : $\bar{x}$ is an optimal solution to (P).

b) Otherwise, obtain an optimal solution z to (8), which is a vertex of $D(\bar{x})$ satisfying $g(z) < 0$.

Set $x^o \leftarrow z$, $D_o \leftarrow D(\bar{x})$ and go back to Step 1.

Remark 1. Unless the problem has no feasible solution, a vertex x^o of D satisfying $g(x^o) \leq 0$ always exists (for otherwise $g(v) > 0$ for every vertex v of D, hence $g(x) > 0$ for every $x \in D$). If such a vertex is not readily available, it can be found, in any case, by solving the concave program $\min\{ g(x) : x \in D\}$.

If $g(x^o) = 0$, one can set $\bar{x} = x^o$ and go directly to Step 2.

Remark 2. Since $g(x^o) < 0$, by virtue of Assumption (i) x^o can not be an optimal solution of the linear program (7). Therefore, a pair u,v satisfying the conditions mentioned in Step 1 can always be found.

3. JUSTIFICATION

To justify the above algorithm we first establish the following optimality criterion which includes Theorem 2 in [7] as a special case.

Theorem 1 (Optimality criterion). Under Assumptions (i) and (ii) a feasible solution $\bar{x}$ to (P) is optimal if and only if the optimal value in the concave program (8) is zero.

Proof. Suppose that $\bar{x}$ is an optimal solution to (P), while the optimal value in (8) is not zero. Since $g(\bar{x}) \leq 0$, this optimal value must be < 0 . Then there is an $\hat{x} \in D$ such that $g(\hat{x}) < 0$, $c\hat{x} = c\bar{x}$. In view of the continuity of g , one must still have $g(x) \leq 0$ for all x in some ball V around $\hat{x}$. On the other hand $\bar{x}$ being optimal to (P), one must have $cx \geq c\bar{x} = c\hat{x}$ for all $x \in D \cap V$. The latter implies that x is an optimal solution to the linear program (7). Since $c\hat{x} = c\bar{x}$, this conflicts with Assumption (i). Therefore, if $\bar{x}$ is optimal to (P) , then

$$0 = \min \{ g(x) : x \in D , cx \leq c\bar{x} \}. \qquad (9)$$

Conversely, suppose that (9) holds and consider any $\hat{x} \in D$ satisfying $g(\hat{x}) \leq 0$, $c\hat{x} \leq c\bar{x}$. Then (9) implies

$g(\hat{x}) = 0$, so that $\hat{x}$ is an optimal solution to the concave program (9). But, the function g being strictly concave (Assumption (ii)) its minimum over the polytope $D(\overline{x}) = \{ x \in D , cx \leq c\overline{x} \}$ can be achieved only at a vertex of $D(\overline{x})$. Therefore $\hat{x}$ is a vertex of $D(\overline{x})$. Since by Assumption (ii) the function g does not vanish at any vertex of D , since $g(\hat{x}) = 0$, it follows that $\hat{x}$ is not a vertex of D , and hence, $c\hat{x} = c\overline{x}$. Thus for any $x \in D$ such that $g(x) \leq 0$, $cx \leq c\overline{x}$, one must have $cx = c\overline{x}$. This proves the optimality of $\overline{x}$. □

We can now prove :

<u>Theorem 2.</u> <u>Under Assumptions (i) and (ii), the algorithm described in the previous section is finite</u>.

<u>Proof</u>. The algorithm consists of a sequence of consecutive loops of execution of Steps 1 and 2. Denote by u^k, v^k, $\overline{x}^k$ the points u, v, $\overline{x}$ obtained at the end of Step 1 of round k . Since $c\overline{x}^{i+1} < c\overline{x}^i$ the set D_o at round k is clearly

$$D_o = D(\overline{x}^{k-1}) = \{ x \in D : cx \leq c\overline{x}^{k-1} \} \quad .$$

We now show that $[u^k, v^k]$ is contained in some edge of D . Indeed by construction $[u^k, v^k]$ is an edge of $D(\overline{x}^{k-1})$, and since $cv^k < cu^k \leq c\overline{x}^{k-1}$ it cannot be contained in the face $cx = c\overline{x}^{k-1}$ of $D(\overline{x}^{k-1})$. Hence it must be contained in some edge of D .

Now let M denote the set of all $x \in D$ such that $g(x) = 0$ and x is contained in some edge of D . By the above, $\overline{x}^k \in M$ for every $k = 1,2,\ldots$. But the number of edges of D is finite and by the strict concavity of the function g there can be on each edge of D at most two points where $g(x) = 0$. Therefore, the set M is finite. The finiteness of the algorithm follows then from the finiteness of the set M and the fact that each round generates a point $\bar{x}^k \in M$ and $c\bar{x}^{k+1} < c\bar{x}^k$ $(k = 1,2,\ldots)$ (indeed $c\bar{x}^{k+1} < cx^{o,k} = c\bar{x}^k$ where $x^{o,k}$ is the point x^o at round k) . □

4. DISCUSSION

1. In Step 2 of round k , we have to solve the concave program

$$(Q_k) \qquad \text{Min } \{ g(x) : x \in D(\bar{x}^k) \} \quad .$$

But, since $c\bar{x}^k < c\bar{x}^{k-1}$, it is clear that

$$D(\bar{x}^k) = \{ x \in D(\bar{x}^{k-1}) : cx \leq c\bar{x}^k \} \quad .$$

Thus (Q_k) can be obtained by adding to (Q_{k-1}) the constraint

$$cx \leq c\bar{x}^k$$

(which, by the way, makes the previous constraints $cx \leq c\bar{x}^i$, $i = 1,\ldots,k-1$, redundant). In view of this fact, to economize the computational effort, one should use for solving (Q_k) an algorithm which could take advantage of the information obtained in solving (Q_{k-1}). For example, the algorithm given by Thieu-Tam-Ban in [12] satisfies this requirement (see e.g. [14] for details).

2. The point $\bar{x}$ obtained at the completion of Step 1 is always a vertex of $D(\bar{x})$ (since $\bar{x}$ lies on an edge $[u, v]$ of D). Therefore, it can be used to start the process of solving the concave program (8).

Also note that it is not always necessary to solve (Q_k) to the end. In fact, we can take as z any vertex of $D(\bar{x}^k)$ such that $g(z) < 0$, and not necessarily an optimal solution of $D(\bar{x}^k)$. It is easily seen that with this modification the algorithm will still be finite.

3. Let D be defined by the system of linear

inequalities :

$$h_i(x) \leq 0 \quad (i = 1,\dots,m)$$

An alternative variant of the algorithm is the following:

Pick a polytope $S_1 \supset D$ whose vertices are known.

Stage k = 1,2,... . Apply the basic algorithm to the problem

$$\text{Min } \{ cx : x \in S_k , g(x) \leq 0 \} ,$$

obtaining an optimal solution $\tilde{x}^k$.

If $\tilde{x}^k \in D$, stop : $\tilde{x}^k$ is an optimal solution to (P). Otherwise, $h_{i_k}(\tilde{x}^k) = \max_i h_i(\tilde{x}^k) > 0$. Let

$$S_{k+1} = S_k \cap \{ x : h_{i_k}(x) \leq 0 \}$$

and go to stage k+1 .

It seems that for large problems this variant should work more efficiently than the basic algorithm.

4. So far we assumed that condition (ii) is fulfilled. To deal with the general case where this condition may not hold, we use the following propositions.

Proposition 1. Let

$$g_\varepsilon(x) = g(x) - \varepsilon(|x|^2 + 1)$$

There is $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0,\varepsilon_0)$ the function g_ε is strictly concave and does not vanish at any vertex of D .

Proof. Denote by V_0 the set of vertices x of D such that $g(x) = 0$, and by V_1 the set of remaining vertices of D . Let $\delta = \min \{ |g(x)| : x \in V_1 \} > 0$, and pick ε_0 so small that $\varepsilon_0(|x|^2 + 1) < \delta$ for all $x \in V_1$.

Then for every $\varepsilon \in (0,\varepsilon_0)$ we have

$$g(x) - \varepsilon(|x|^2 + 1) \leq -\varepsilon < 0 \qquad \forall x \in V_0$$

$$|g(x) - \varepsilon(|x|^2 + 1)| \geq \delta - \varepsilon_0(|x|^2 + 1) > 0$$
$$\forall x \in V_1 \quad .$$

Thus the function g_ε does not vanish at any vertex of D. Since the strict concavity of g_ε is obvious, the Proposition is proved.

Proposition 2. *Consider the problem*

$$(P_\varepsilon) \qquad \text{Min}\{cx : x \in D , g_\varepsilon(x) \leq 0\} \quad (0 < \varepsilon < \varepsilon_0).$$

If x_ε *is an optimal solution to* (P_ε) *and* x *is an accumulation point of* $\tilde{x}$ as $\varepsilon \to 0+$ *then* $\tilde{x}$ *is an optimal solution to* (P).

Proof. For all $x \in D$ satisfying $g(x) \leq 0$ we have $g_\varepsilon(x) \leq g(x) \leq 0$, hence $cx_\varepsilon \leq cx$, hence $c\tilde{x} \leq cx$. But clearly $\tilde{x} \in D$, $g(\tilde{x}) \leq 0$, hence $\tilde{x}$ is optimal to (P) . □

On the basis of these Propositions, if condition (ii) fails to hold, we can solve (P_ε) with $\varepsilon > 0$ arbitrarily small and then make $\varepsilon = 0$ in the result.

5. The algorithm given by Hillestad and Jacobsen in [7] can also be described as consisting of consecutive rounds requiring each two steps. The first step of that algorithm is exactly the same as Step 1 of the algorithm presented above, so the main difference between the above algorithm and that of Hillestad and Jacobsen is in the second step.

5. ILLUSTRATIVE EXAMPLE

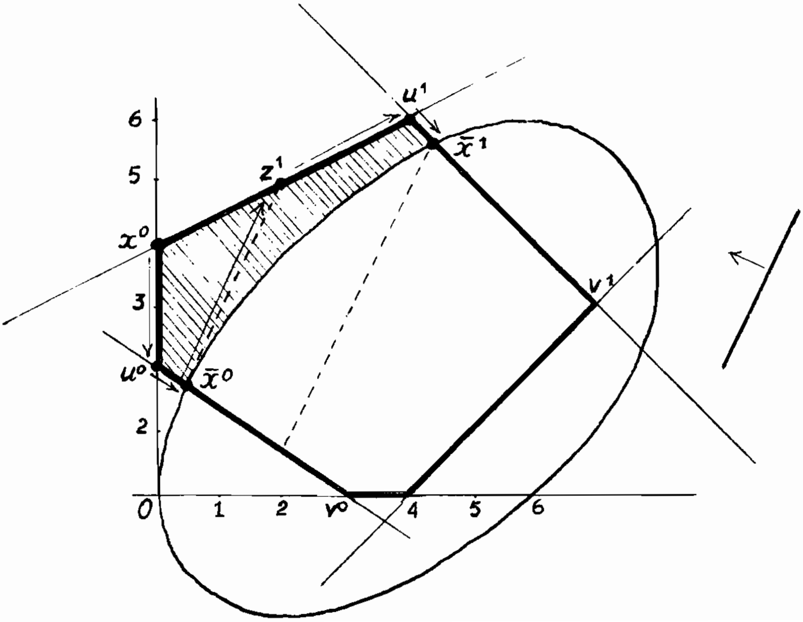
$$\text{Minimize} \quad -2x_1 + x_2 \quad , \quad \text{subject to}$$
$$x_1 + x_2 - 10 \leq 0$$

$$
\begin{aligned}
-x_1 + 2x_2 - 8 &\leq 0\\
x_1 - x_2 - 4 &\leq 0\\
-2x_1 - 3x_2 + 6 &\leq 0\\
x_1 \geq 0 \ , \quad x_2 &\geq 0\\
-x_1^2 + x_1x_2 - x_2^2 + 6x_1 &\leq 0
\end{aligned}
$$

The algorithm begins at the vertex $x^o = (0;4)$.

Iteration 1. Step 1 finds the vertices $u^o = (0;2)$, $v^o = (3;0)$ and the point $\bar{x}^o = (0.4079356 \ ; \ 1,728049)$. Step 2 solves the concave program $\max\{g(x) : x \in D(\bar{x}^o)\}$, and finds the point $z^1 = (2;5)$.

Iteration 2. Here $x^1 = (2;5)$. Step 1 finds the vertices $u^1 = (4;6)$, $v^1 = (7;3)$ and the point $\bar{x}^1 = (4.3670068 \ ; \ 5.6329334)$. Since the optimal value of the concave program $\max\{g(x) : x \in D(\bar{x}^1)\}$ is 0 , Step 2 concludes that $\bar{x}^1$ is the optimal solution of the problem.



REFERENCES

[1] M. Awriel and A.C. Williams. Complementary geometric programming, SIAM, J. Appl. Math., 19(1974), 125-141.

[2] M. Awriel and A.C. Williams. An extension of geometric programming with applications in engineering optimization, J. Eng. Math., 5(1971), 187-194.

[3] P.P. Bansal and S.E. Jacobsen. Characterization of basic solutions for a class of nonconvex program, J. Optim. Theory and Applications, 15(1975), 549-564.

[4] P.P. Bansal and S.E. Jacobsen. An algorithm for optimizing network flow capacity under economies-of-scale. J. Optim. Theory and Applications, 15(1975), 565-586.

[5] J. Falk and K.L. Hoffman. A successive underestimation method for concave minimization problem, Math. of Oper. Research, N$^{\underline{o}}$1, 1976, 251-259.

[6] R.J. Hillestad. Optimization problems subject to a budget constraint with economies of scale, Operations Research 23, No.6, Nov-Dec 1975.

[7] R.J. Hillestad and S.E. Jacobsen. Linear program with an additional reverse-convex constraint, Appl. Math. Optim., 6(1980), 257-269.

[8] R.J. Hillestad and S.E. Jacobsen. Reverse-convex programming, Appl. Math. Optim., 6(1980), 63-78.

[9] R. Meyer. The validity of a family of optimization methods, SIAM, J. Control, 8(1970), 41-54.

[10] R.T. Rockafellar. Convex Analysis, Princeton Univ. Press, 1970.

[11] J.B. Rosen. Iterative solution of non-linear optimal control problems, SIAM, J. Control, 4(1966), 223-244.

[12] T.V. Thieu, B.T. Tam and V.T. Ban. An outer approximation method for globally minimizing a concave function over a compact convex set, IFIP Working Conference on Recent Advances on System Modelling and Optimization, Hanoi, January 1983.

[13] H. Tuy. Global minimization of a concave function subject to mixed linear and reverse-convex constraints, IFIP Working Conference on Recent Advances on System Modelling and Optimization, Hanoi, January 1983.

[14] H. Tuy. On outer approximation methods for solving concave minimization problems, Report N$^{\underline{o}}$108, Forschungwerkpunkt Dynamische System, Univ. Bremen, 1983, Acta Mathematica Vietnamica (to appear).

IV. STOCHASTIC PROGRAMMING AND APPLICATIONS

SOME REMARKS ON QUASI-RANDOM OPTIMIZATION

Walter Bayrhamer
Institute of Mathematics, University of Salzburg, Austria

1. INTRODUCTION

In the theory and practice of optimization it often happens that the objective function has a very low degree of regularity or that it is defined only empirically. Another critical point in optimization is that many algorithms deliver only local convergence. So for these two reasons it is adviseable to analyze methods of direct search like random and quasi-random search techniques. In this paper we consider error estimates for deterministic analogues of random search.

2. ERROR ESTIMATES FOR THE FUNCTION VALUES

Let (K,d) be a compact, metric space and let f be a continous function from K into the real numbers. Then we are interested in the maximum of f and in one point where this maximum is attained. Such a point exists by the compactness of K and the continuity of f. As the exact computation of these values is in most of the cases very complicated or impossible we try to approximate them. Define $M:=\max\{f(x)/x\in K\}$ and $\bar{x}=\mathrm{argmax}\{f(x)/x\in K\}$ for the requested values. For the approximation take a sequence of finite subsets of K: $A_1, A_2, \ldots$ where A_k has k elements and let them have the property that then $\lim h(A_N,K)=0$ for $N\to\infty$, where h is the Hausdorff-metric in the space of compact subsets of K. For example take a sequence $x_1, x_2, \ldots$ which is dense in K and take for A_N the first N elements of this sequence. In $K=[0,1]^s$ for such a sequence we can take $x_n=(\{n\theta_1\},\ldots,\{n\theta_s\})$ n=1,2,... where $\{x\}$ is the fractional part of x and $\theta_1,\ldots,\theta_s$ are real numbers so that 1, θ_1, ..., θ_s form a basis of a real algebraic number field of degree s+1 over the rationals. (See Niederreiter 1983b)

By the above mentioned limit-property of A_N we can interprete it as an approximation to the space K. So we conclude that the extrema of f on A_N will approximate the extrema of f on K. For the approximation of M we take $M_N := \max\{f(x_i)/\ 1 \leq i \leq N\}$ and for $\bar{x}$ we take $x_{kN} = \operatorname{argmax}\{f(x_i)/1 \leq i \leq N\}$. From the definition of h we receive $h(K,A_N) = \sup_{x \in K} \min_{1 \leq i \leq N} d(x,x_i)$ and similarily we have $h(f(K),f(A_N)) = \sup_{x \in K} \min_{1 \leq i \leq N} |f(x)-f(x_i)|$. Before establishing an error estimate we must define the modulus of continuity of f by $w(t) := \sup_{d(x,y) \leq t} |f(x)-f(y)|$.

Now the following error estimate can be proved:

$$O \leq M - M_N \leq h(f(K),f(A_N)) \leq w(h(K,A_N)) \qquad (1)$$

<u>Proof</u>: $O \leq M-M_N = f(\bar{x}) - f(x_{kN}) = \min_{1 \leq i \leq N} |f(x) - f(x_i)| \leq h(f(K),f(A_N)) \leq \sup_{x \in K} \min_{1 \leq i \leq N} w(d(x,x_i)) = w(\sup_{x \in K} \min_{1 \leq i \leq N} d(x,x_i)) = w(h(K,A_N))$.

(For this result see also Niederreiter 1983 a,b and Sobol 1982)

The quantity $h(K,A_N)$ is often denoted by $d_N(x_1,\ldots,x_N)$ and is called the dispersion of the points.

By using terms of levelsets we can establish another plain error estimate. The set $L(\alpha) := \{x \in K/\ f(x) \geq \alpha\}$is called the <u>levelset of f at level α</u>,where α is a real number. If $\alpha \leq \inf f(x)$, then $L(\alpha)=K$ and if$\alpha > \sup f(x)$ then $L(\alpha)=\emptyset$. Another important term that we need is the <u>diameter-function $v(\alpha)$</u>$:=\sup\{d(x,y)$ $x,y \in L(\alpha)\}$ and it is the diameter of the levelset $L(\alpha)$. Then we can show the following error estimate:

1. $d(\bar{x},x_{kN}) \leq v(M_N)$ (2a)
2. $O \leq M - M_N \leq w(v(M_N))$ (2b)

<u>Proof</u>: As $f(x_{kN})=M_N$ and $f(\bar{x})=M \geq M_N$ it follows that $x_{kN} \in L(M_N)$ and $\bar{x} \in L(M_N)$ which implies (2a) and this implies (2b) by the definition of w.

For the special case where $K = [o,1]^s$, where s is a natural number, we can derive another error estimate from the paper of Hellekalek (1979) by the use of Lebesgue-measure. Let λ be the Lebesgue-measure and assume $f \geq O$. Then he defines the function $r(\varepsilon) := \inf\{\delta > O\colon \lambda(L(M-\delta)) \geq \varepsilon\}$. He proved that $r(\varepsilon) \leq$ $\leq w(\varepsilon^{1/s})$. If we set $\varepsilon = \lambda(L(M_N))$ we obtain:

$r(\lambda(L(M_N))) = \inf\{\delta > O\colon \lambda(L(M-\delta)) \geq \lambda(L(M_N))\}$ and $\lambda(L(M-\delta)) \geq \lambda(L(M_N)) \iff$ $L(M-\delta) \supseteq L(M_N) \iff M_N \geq M-\delta \iff M-M_N \leq \delta \Rightarrow r(\lambda(L(M_N)) \geq M - M_N \Rightarrow$ $M - M_N \leq w((\lambda(L(M_N)))^{1/s})$.

3. <u>SOME PROPERTIES OF LEVELSETS</u>

So far we described error estimates for the function values. For a refined procedure and for a more accurate analysis of the problem we will need some error estimate for the argument. But this question is closely connected with the theory of levelsets. So we like to consider some theorems resp.:

Theorem 1: Let (α_n) be a monotonically increasing sequence of real numbers converging to α. Then for a continous function f follows:

$h(L(\alpha_n),L(\alpha))\to 0$ for $n\to\infty$ (3)

Proof: As it is easily seen, the sets $L(\alpha)$, $L(\alpha_n)$ are compact for all n and $L(\alpha) = \bigcap_{n=1}^{\infty} L(\alpha_n)$ and $L(\alpha_1) \supset L(\alpha_2) \supset \dots$. It is known from topology that in a compact, metrical space the topological convergence of nonempty pointsets is equivalent to the convergence in the metrical sense and so statement (3) follows. (See for example: Alexandroff-Hopf 1935)

Theorem 2: $\alpha\le\alpha'$ implies

$$0\le v(\alpha) - v(\alpha') \le 2\, h(L(\alpha),L(\alpha')) \qquad (4)$$

for $L(\alpha)$, $L(\alpha')$ are nonempty.

Proof: Take $x,y \in L(\alpha)$ arbritrarily and choose z and z' so that $d(x,z) = \inf_{u\in L(\alpha')} d(x,u)$, $d(y,z') = \inf_{u'\in L(\alpha')} d(y,u')$ and $z, z' \in L(\alpha')$.

Then by the triangle-inequality we obtain:

$$d(x,y)\le d(x,z)+d(z,z')+d(z',y)\le \sup_{z,z'\in L(\alpha')} d(z,z')+\inf_{u\in L(\alpha')} d(x,u)+\inf_{u\in L(\alpha')} d(y,u)$$

and this implies $\sup_{x,y\in L(\alpha)} d(x,y)=v(\alpha)\le v(\alpha')+ 2 \sup_{x\in L(\alpha)} \inf_{u\in L(\alpha')} d(x,u)\le v(\alpha') + 2\, h(L(\alpha),L(\alpha'))$.

So we have proved the right-handpart of (4) and the left-handpart is obvious.

Theorem 3: $L(\alpha) \subset B_N^*(\alpha)$, where $B_N^*(\alpha)= \bigcup B(x_i,d_N)$ for i with $f(x_i)\ge\alpha-w(d_N)$ and $B(x,t)= \{y/d(x,y)\le t\}$

Proof: Take $x \in L(\alpha)$, so by the definition of d_N there exists x_i, so that $d(x,x_i)\le d_N$, and so $f(x) - f(x_i) \le w(d_N)$ and this implies $f(x)-w(d_N)\le f(x_i)$ and so $f(x_i)\ge \alpha -w(d_N)$ and so our statement follows.

4. ERROR ESTIMATES FOR THE ARGUMENT

By the definition of d_N it is clear that the balls $B(x_1,d_N),\dots,B(x_N,d_N)$ cover the whole space K. So we consider the following idea: Take $x\in K$, and let $x \in B(x_i,d_N)$ and by using the modulus of continuity of f we obtain $f(x) - f(x_i) \le w(d_N)$ and so $f(x) \le f(x_i) + w(d_N)$. so if $f(x_i)+w(d_N)<M_N$ then it follows that the extremal point $\bar{x}$ cannot lie in such a ball and therefore $\bar{x}$ lies in $B_N^*(M_N)$. For theorem 4 let us assume that there is exactly one index KN with $f(x_{kN})=M_N$ and $L(M)=\{\bar{x}\}$. If we define $p_N(A_N,f)= \max\{d(x_{kN},x_k)/ k \in I_N \}$, where $I_N= \{i/f(x_i)\ge M_N - w(d_N)\}$ then the following theorem results:

Theorem 4:

1. $d(\bar{x}, x_{kN}) \leq d_N + p_N$ (5a)
2. $\lim p_N = 0$ for $N \to \infty$ (5b)

Proof: 1. For $\bar{x} \in B^*_N(M_N)$ there exists an index $i \in I_N$ with $d(\bar{x},x_i) \leq d_N$ and so $d(\bar{x},x_{kN}) \leq d(\bar{x},x_i) + d(x_i, x_{kN}) \leq d_N + p_N$.

2. Consider now the increasing sequence $\alpha_N := M_N - w(d_N)$, converging to M and that for all $i \in I_N$ values $f(x_i) \geq \alpha_N$, so $xi \in L(\alpha_N)$ for all $i \in I_N$ and that implies $p_N \leq v(\alpha_N)$. But the last term converges to $v(M)$ and $v(M)$ equals 0 by the assumptions, and theorems 1 and 2. Thus statement (5b) holds.

Theorem 5:

$$1/2\ v(M_N) \leq d_N + p_N \leq v(M_N) + 2\ h(L(\alpha_N), L(M_N)) + d_N \qquad (6)$$

Proof: From $p_N \leq v(\alpha_N)$ follows by theorem 2 that $p_N \leq v(M_N) + 2h(L(\alpha_N), L(M_N))$ and so the right-hand side of (6) is proved. From theorem 3 we have $L(M_N)$
$B^*_N(M_N) \Rightarrow \forall\, x,y \in L(M_N) \Rightarrow \exists\, i,j \in I_N$: $d(x,x_i) \leq d_N$, $d(y,x_j) \leq d_N \Rightarrow$
$d(x,y) \leq d(x,x_i) + d(x_i,x_j) + d(x_j,y) \leq 2d_N + d(x_i,x_{kN}) + d(x_{kN},x_j) \Rightarrow$
$d(x,y) \leq 2(d_N + p_N) \Rightarrow v(M_N) \leq 2(d_N + p_N) \Rightarrow 1/2\ v(M_N) \leq d_N + p_N$ and so the statement (6) is proved.

Remark: The error estimate (5a) and its behaviour partially depends on the behaviour of the function $\varphi(\alpha,\varepsilon) := h(L(\alpha-\varepsilon), L(\alpha))$, which can be interpreted as an index of flatness of the objective function f. It indicates the behaviour of the function f with respect to flat regions and local extrema which are both bad for global optimization. The study of φ is closely connected with the theory of parametric optimization.

5. SOME REMARKS ON ADAPTIVE PROCESSES

The rate of convergence of error estimates depends partially on the magnitude of the dispersion and so can be rather slow. Therefore it is adviseable to study algorithms which deliver a better convergence rate. This can be reached by adapting the search-area to the function. From estimate (5a) we know that it suffices to search in the ball with center x_{kN} and radius $d_N + p_N$. If this radius is acceptable small then we restrict our search to this ball and we can repeat the preceding error estimates. But if we think that there are more than one point in $L(M)$ we should prefer another adaptive algorithm. Take each of the balls $B(x_i,d_N)$ with $i \in I_N$ to perform the global search there. So if I_N does not contain a large number of indices then the number of additional function-evaluations will still be acceptable. You can also use the global search for determining a starting point for a gradient method or another local optimization technique to search a local maximum point.

6. REFERENCES

Alexandroff P./Hopf H.(1935): Topologie I. Springer, Berlin.

Hellekalek (1979): Über das Wachstum von L_p-Normen. Arbeitsbericht 3/79 des Institutes für Mathematik der Universität Salzburg.

Niederreiter H. (1983a): A quasi Monte-Carlo Method for the approximate computation of the extreme values of a function. Studies in Pure Mathematics(To the memory of Paul Turàn)pp523-529, Akadêmiai Kiadô, Budapest(Submitted in 1977)

Niederreiter H. (1983b): Quasi Monte-Carlo methods for global optimization (Manuscript)

Sobol I.M. (1982): On an estimate of the accuracy of a simple multidimensional search. Soviet. Math. Dokl. Vol.26,No. 2, pp 398-401

OPTIMAL SATELLITE TRAJECTORIES: A SOURCE OF DIFFICULT NONSMOOTH OPTIMIZATION PROBLEMS

L.C.W. Dixon, S.E. Hersom and Z. Maany
Numerical Optimization Centre, Hatfield Polytechnic, College Lane, Hatfield, UK

1. INTRODUCTION

In this paper we will show that optimal satellite trajectory problems can be posed as difficult nonsmooth optimisation problems. The aim is not to advocate solving satellite trajectory problems by using nonsmooth optimisation algorithms; they can be solved more simply by other means. The aim is simply to challenge the designers of nonsmooth optimisation codes to test them on these problems; which we believe will prove to be very difficult. We look forward to hearing the results of such tests.

2. THE SATELLITE TRAJECTORY PROBLEM

In this paper our intention is to define a set of N.S.O. problems by reformulating a particular satellite trajectory problem.

The problem we will consider is a rendezvous with the asteroid VESTA; the details of Vesta's orbit are given in Appendix 1. In the problem we will asume that the satellite is launched from earth on a particular day and that the trajectory to be optimised commences at a point sufficiently removed from the earth for earth's gravity to be ignored. The time, position and velocity of the satellite at that starting point are also given in Appendix 1; starting from these values of t_o, $\underline{r}_o$, $\underline{v}_o$ the satellite's trajectory is then integrated by fourth order Runge Kutta with a standard step size of 24 days.

The satellite's motion is governed by gravity and controlled by a low thrust motor, so that

$$\dot{\underline{r}} = \underline{v} \qquad ; \quad \underline{r}(t_o) = r_o \tag{1}$$

$$\dot{\underline{v}} = -\frac{\mu\underline{r}}{r^3} + \underline{T}/M; \quad \underline{v}(t_o) = \underline{v}_o \tag{2}$$

where $\underline{T}$ is the thrust and at any point is constrained by

$$0 \leq ||\underline{T}|| \leq T_m . \tag{3}$$

The mass flow equation for the fuel used is then

$$\dot{m} = -\,||\,\underline{T}\,||/gI \qquad (m(t_o) = m_o \text{ given}) \tag{4}$$

and the thrust level is restricted by power considerations

$$T_m = 2\eta P_o/(gIr^k). \tag{5}$$

Two values of k are of interest, k = 0 corresponds to conventional RTG motors but k = 1.7 is more appropriate for solar powered motors.

The problem is then to determine that trajectory r(t) that rendezvous with Vesta's trajectory while using least fuel. Assuming the rendezvous takes place at time t_R then Vesta's trajectory specifies the values of $\underline{r}_v(t_R)$ and $\underline{v}_v(t_R)$.

The optimal control problem is therefore:-

$$\text{Maximise } m(t_r) \tag{6}$$

$$\text{s.t.} \quad \underline{r}(t_R) = \underline{r}_v(t_R) \tag{7}$$

$$\underline{v}(t_R) = \underline{v}_v(t_R) \tag{8}$$

and equations 1 - 5 by varying t_R, $\underline{r}(t)$, $\underline{T}(t)$.

There are a number of specialised codes for solving optimal satellite trajectory problems but it is not our intention to discuss them in this paper. Instead we wish to show that this problem can be posed in different ways that lead to N.S.O. problems. The solution to the problem has been obtained by other means and is also given in Appendix 1.

3. THE INDIRECT PONTRYAGIN FORMULATION

Pontryagin [3] showed that the optimal control problem could be converted to an optimisation problem by the introduction of adjoint variables and a Hamiltonian function.

We will denote the adjoint variables for equations (1), (2) and (4) by $\underline{M}$, $\underline{L}$ and p respectively and will let the Lagrange multiplier for (5) be represented by λ. Also we will denote the thrust $\underline{T}$ by $||\underline{T}||\hat{\underline{T}}$ where $\hat{T}$ is a unit vector, then the Hamiltonian is given by

$$H = \underline{M}.\underline{v} + \underline{L}.(-\mu\frac{\underline{r}}{r^3} + ||\frac{T}{m}||\hat{\underline{T}}) - p||T||/gI$$

$$+ \lambda(T_m - 2\eta P_o/(gIr^k)) \tag{9}$$

$$\text{where } \dot{\underline{M}} = \frac{\mu\underline{L}}{r^3} - 3\frac{\underline{L}.\underline{r}}{r^5}\underline{r} + \frac{2k\lambda\eta P_o}{gI}\frac{\underline{r}}{r^{k+2}} \tag{10}$$

$$\dot{\underline{L}} = -\underline{M} \tag{11}$$

$$\dot{p} = +\frac{||T||}{m^2}\underline{L}.\hat{\underline{T}}. \tag{12}$$

Again the maximum principle implies that on an optimal trajectory

$$\hat{\underline{T}} = \hat{\underline{L}} \tag{13}$$

and as H is linear in $||\underline{T}||$, then the optimal values of $||\underline{T}||$ is either 0 or T_m for all t. A period during which $||\underline{T}|| = 0$ will be termed a coast arc; if $\|\underline{T}\| = T_m$ it will be termed a thrust arc, then $\lambda = p/gI - \|L\|/m$ on a thrust arc and 0 on a coast arc.

The optimal trajectory consists of a thrust arc, followed by a coast arc, followed by a final thrust arc which we can express as

$$\begin{aligned} \|T\| &= T_m & t_o < t < t_1 \\ \|T\| &= 0 & t_1 < t < t_2 \\ \|T\| &= T_m & t_2 < t < t_R \end{aligned} \qquad (14)$$

with implied constraints $t_o < t_1 < t_2 < t_R$. (15)

Given the values of $L_o = L(t_o)$; $M_o = M(t_o)$; $p_o = p(t_o)$, t_1, t_2 and t_R then equations (1), (2), (4), (10), (11), (12) can be integrated forward in time sufficiently accurately using RK4 with a step of 24 days (with suitable modifications at t_1, t_2 and t_k). In integrating these equations the constraints (5), (13) and (14) are automatically applied, so at t_R the values of $m(t_R)$, $\underline{r}(t_R)$ and $\underline{v}(t_R)$ that correspond to these variables can be computed.

We may then pose the NLP problem

FORMULATION 1

$$\left.\begin{aligned} &\text{Max} \quad m(t_R) \\ &\text{s.t.} \quad \underline{r}(t_R) = \underline{r}_v(t_R) \\ &\qquad\;\; \underline{v}(t_R) = \underline{v}_v(t_R) \\ &\qquad\;\; 0 < t_1 < t_2 < t_R \end{aligned}\right) \qquad (16)$$

where the optimisation variables are

$\underline{L}_o$, $\underline{M}_o$, p_o, t_1, t_2 and t_R. (17)

This is of course a standard NLP problem but our experience reported in [1] is that it is too difficult for most codes, even when the adjoint-variable transformation [2] is applied. For completeness the transformation used is given:-

$$L_o = \begin{bmatrix} \cos\alpha_o \cos\beta_o \\ \sin\alpha_o \cos\beta_o \\ \sin\beta_o \end{bmatrix}$$

$$M_o = \dot{L}_o = \begin{bmatrix} \dot{S}_o \cos\alpha_o \cos\beta_o - \dot{\alpha}_o \sin\alpha_o \cos\beta_o - \dot{\beta}_o \cos\alpha_o \sin\beta_o \\ \dot{S}_o \sin\alpha_o \cos\beta_o + \dot{\alpha}_o \cos\alpha_o \cos\beta_o - \dot{\beta}_o \sin\alpha_o \sin\beta_o \\ \dot{S}_o \sin\beta_o + \dot{\beta}_o \cos\beta_o \end{bmatrix} \qquad (18)$$

FORMULATION 1b

is therefore to solve (16) using the optimisation variables

$$\dot{S}_o,\ \alpha_o,\ \dot{\alpha}_o,\ \beta_o,\ \dot{\beta}_o,\ t_1,\ t_2 \text{ and } t_R. \qquad (19)$$

The NLP problem (16) could be solved by minimizing the exact nonsmooth penalty function, so our first NSO problem consists of

FORMULATION 2

$$\text{Min} \quad -m(t_R) + c_R \sum_i |r_i(t_R) - r_{vi}(t_R)| + c_v \sum_i |v_i(t_R) - v_{v_i}(t_R)| \qquad (20)$$

$$0 < t_1 < t_2 < t_R$$

with respect to variables (17).

In FORMULATION 2b variables (19) would be used.

As Formulation 1 is an NLP and Formulation 2 its EPF; then Formulation 2 has a rather special structure as an NSO and codes have been written for NSO problems with this structure. It is therefore interesting to find that we can pose the problem as an NSO without this structure.

4. THE DOCKING FORMULATION

As Pontryagin's path is optimal if we were to replace part of the path by an alternative feasible stategy the solution must be worse. In particular if we were to stop the second thrust arc at $t_3 > t_2$ and were to replace the thrust strategy in $t_3 < t < t_R$ by a nonoptimal feasible strategy that ensures

$$\underline{r}(t_R) = \underline{r}_v(t_R)$$

and $\underline{v}(t_R) = \underline{v}_v(t_R)$

and then maximise $m(t_3) - m_D$ (21)

where m_D is the mass used in this manoeuvre, then the optimum must occur with $t_3 = t_R$ and $m_D = 0$. But we have converted the NLP (16) into the simpler problem

FORMULATION 3

Maximise $m(t_3) - m_D$

s.t. $0 < t_1 < t_2 < t_3 \leq t_R.$

w.r.t. either $\underline{L}_o$, $\underline{M}_o$, p_o, t_1, t_2 and t_3

or $\dot{S}_o$, α_o, $\dot{\alpha}_o$, β_o, $\dot{\beta}_o$, p_o, t_1, t_2 and t_3.

The method proposed for the final manoeuvre is described in Appendix 2. The function m_D is nonsmooth, so Formulation 3 is an unstructured NSO, which will we believe prove difficult if not impossible for most NSO codes.

5. THE POSITION SPACE FORMULATION

A very different approach to the same problem also leads to a difficult NSO problem. Let us consider the following approximate problem. Let us divide the range $0 < t < t_R$ into a number of intervals by grid points $t_1, \dots, t_i$, ... (for convenience let $t_R = t_{10}$). Let us take the position $\underline{r}_i = \underline{r}(t_i)$ and velocity $\underline{v}_i = \underline{v}(t_i)$ as optimisation variables; then in the interval $t_i < t < t_{i+1}$ we may approximate the trajectory $\underline{r}(t)$ by a cubic variation in each component, for instance, if we represent $\underline{r} = (x, y, z)^T$ then each of x, y and z can be matched by a cubic to the values at t_i, t_{i+1}.

As $\underline{r}$ is cubic in t, $\underline{v}$ is quadratic and $\ddot{\underline{r}}$ linear, we have an implied thrust from equation (2) of

$$\underline{T} = m(\ddot{\underline{r}} + \frac{\mu\underline{r}}{r^3}) \tag{22}$$

So
$$\|T\| = m\,\|\ddot{\underline{r}} + \frac{\mu\underline{r}}{r^3}\| \tag{23}$$

and constraint (3) becomes

$$0 \leq m^2 r^{2k} \|\ddot{\underline{r}} + \frac{\mu\underline{r}}{r^3}\|^2 \leq (\frac{2\eta P_o}{gI})^2. \tag{24}$$

Due to the smooth nature of the function it is probably sufficient to apply these constraints only at the endpoints of the intervals t_i, t_{i+1}.

For any value of t_R we can ensure that the initial and final positions and velocities are correct so we now need only consider the objective function which is governed by (4)

$$\dot{m} = -\frac{\|T\|}{gI} = -\frac{m\|\ddot{\underline{r}} + \frac{\mu\underline{r}}{r^3}\|}{gI} \tag{25}$$

$$\frac{\dot{m}}{m} = -\|\ddot{\underline{r}} + \frac{\mu\underline{r}}{r^3}\|/gI$$

$$[\log m] = -\int \|\ddot{\underline{r}} - \frac{\mu\underline{r}}{r^3}\| \frac{dt}{gI}$$

$$\frac{m_{i+1}}{m_i} = \exp\{-\int \|\ddot{\underline{r}} - \frac{\underline{r}}{r}\| dt/gI\}. \tag{26}$$

For the given values of r_i, $\dot{r}_i$ we can therefore compute the values of m_i, m_{i+1} given m_o and therefore both the objective function and the m_i to be used in (24). The problem is nonsmooth due to the square roots in (26). For simplicity we will standardise the formulation by approximating the integral in (26) by

$$\int_{t_i}^{t_{i+1}} \|\ddot{\underline{r}} - \frac{\mu\underline{r}}{r^3}\| dt = (\frac{t_{i+1} - t_i}{2})(\|\ddot{\underline{r}}_i - \frac{\mu\underline{r}_i}{r_i^3}\| + \|\ddot{r}_{i+1} - \frac{\mu\underline{r}_{i+1}}{r_{i+1}^3}\|) \tag{27}$$

FORMULATION 4

Maximise m_R calculated via (26) and (27) subject to the constraints (24) using the variables t_R, $\underline{r}_i$, $\underline{v}_i$ $\quad i = 1, \ldots , 9$.

In this paper we have posed 7 formulations of a satellite trajectory problem. The purpose of the paper is unusual, namely to challenge the designers of NSO codes to apply their codes to Formulations 2-4. We will be interested to hear the results.

REFERENCES

Dixon, L.C.W., Hersom, S.E., and Maany, Z.A. Low Thrust Satellite Trajectory Optimisation. (to appear in McKee, S and Elliot, C, Industrial Numerical Analysis, Academic Press, 1985).

Dixon, L.C.W. and Bartholomew-Biggs, M.C. (1981). Adjoint-Control Transformations for solving practical optimal control problems. In Optimal Control Applications and Methods, Vol. 2, pp.365-381.

Pontryagin, L.S., Boltyanski, V.G., and Gambrelidze, R.W. (1962). The Theory of Optimal Processes. Interscience Press.

APPENDIX 1. TRAJECTORY DETAILS

Characteristics of Target Orbit

Semi major axis	2.361680 Au	
Aphelion	2.573452 Au	
Perihelion	2.149908 Au	
Eccentricity	0.089670	
Inclination	7.144 Deg	
Right Ascension	103.489 Deg	
Arg of Perihelion	150.618	
True Anomaly at launch	16.377	(Launch Feb 1st 1993)

Constants Used

$P_o = 20$ KW, $\eta = 68\%$, $I = 3900$ secs

$g = 9.81$ m/sec², $\mu = 1.32715 \times 10^{11}$ Km³/sec².

(These should be converted to AU/DAY/Kg units).

Trajectory Details. The trajectory commences on February 1st 1993 at

$r_o = (-.661201, .730588, 0.000)$Au

$v_o = (-24.04952, -21.395403, .371891)$Km/sec

$m_o = 2000$ Kg

where the optimal values of the optimisation variables for Formulation 1b are

$\dot{S}_o = -.01990306$; $\alpha_o = -64.774°$; $\dot{\alpha}_o = 4.553824$ Deg/Day

$\beta_o = 86.88414°$; $\dot{\beta}_o = -1.109517$ Deg/Day; $p_o = 1.388167$ secs

$t_1 = 474.5221$ Days; $t_2 - t_1 = 237.4531$ Days

$t_R - t_2 = 202.3659$ Days.

The final mass at rendezvous is 1537.414 Kg. From this point a 10 day coast arc is prescribed before optimisation commences, the values after the coast arc are

$$r_o = (-.789214,\ .596415,\ 0.002137)$$

$$v_o = (-20.174716,\ -24.941806,\ 0.366192).$$

The starting point we used for our optimisation run Formulation 1b was

$$\dot{S}_o = 0,\ \alpha_o = -137.854°,\ \dot{\alpha}_o = 1 \text{ Deg/Day},\ \beta_o = 0,\ \dot{\beta}_o = 0,$$

$$p_o = 10{,}000 \text{ secs},\ t_1 = 20 \text{ days},\ t_2 - t_1 = 50 \text{ days},\ t_R - t_2 = 40 \text{ days}.$$

APPENDIX 2. DOCKING

We take $\underline{x}(t)$, $\underline{v}(t)$ to be the relative distance and velocity vectors of the S/C with respect to the target where t is the time after the end of normal thrusting. "Docking" is defined as attaining, after a time T, $\underline{x}(T) = 0$ and $\underline{v}(T) = 0$. The manoeuvre is to apply an acceleration of constant amplitude in each co-ordinate but, in each, the direction is reversed at some time $t_i (i = 1,2,3$ and $0 \leq t_i \leq T)$.

It is assumed that the magnitude of the acceleration is equal to the ratio of the maximum thrust/mass at the end of normal thrusting, i.e. change in thrust due to change in the power available and change in mass due to the loss of propellant are ignored. Further, it is assumed that the S/C and target are in a uniform gravitational field. The motion in each co-ordinate direction can therefore be considered independently.

If a is the acceleration up to the switching time, t, and -a is the acceleration from t to T, then if x and v are the values in one co-ordinate of $\underline{x}(0)$ and $\underline{v}(0)$ respectively, the final values are

$$x(T) = x + vt + at^2/2 + (v + at)(T - t) - a(T - t)^2/2$$

and $v(T) = v + a(2t - T)$.

Since both must always be zero, these can be written as:

$$x + vT + a(2tT - t^2 - T^2/2) = 0 \qquad (1)$$

$$2at = - v + aT. \qquad (2)$$

Eliminating t between (1) and (2) we obtain

$$a^2T^2 + 2a[2x + vT] - v^2 = 0 \qquad (3)$$

or $aT^2 = - D \pm \surd[D^2 + v^2T^2]$

where $D = 2x + vT$.

Since $0 \leq t \leq T$, it is readily shown from (2) that $T \geq |v/a|$

or $a^2T^2 \geq v^2$.

From (3), therefore, we obtain $aD \leq 0$

i.e. SIGN(a) = - SIGN(D) = S, say.

Hence $aT^2 = - D + S\surd[D^2 + v^2T^2]. \qquad (4)$

This expression, for a given x, v and T, gives the value of the acceleration required. If this is a_i for the ith co-ordinate, then docking is achieved when a value of T is found such that $\sqrt{\Sigma(a_i^2)}$ is equal to the acceleration available. In the program this is achieved by an iterative procedure. The propellant used is calculated as the flow-rate at the end of normal thrusting multiplied by the docking time, T.

A REDUCED SUBGRADIENT ALGORITHM FOR NETWORK FLOW PROBLEMS WITH CONVEX NONDIFFERENTIABLE COSTS

M.A. Hanscom[1], V.H. Nguyen[2] and J.J. Strodiot[2]
[1] *IREQ, Varennes, Canada*
[2] *FNDP, Rempart de la Vièrge 8, 5000 Namur, Belgium*

1. PROBLEM FORMULATION

Consider a single-commodity directed network with m nodes and n arcs. The general nonlinear network flow problem (Dembo et al. 1981) consists in finding a vector flows $x = (x_1,\dots,x_n)$ solution of

$$(P) \quad \begin{cases} \text{Minimize} \quad f(x) \\ \text{s.t.} \quad A\,x = b \\ \qquad\quad x \leq x \leq \overline{x} \end{cases}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$, A is the $m \mathrm{x} n$ node-arc incidence matrix of the network, $A\,x = b$ expresses the flow conservation constraints and x and $\overline{x}$ denote the lower and upper bound on the flow x.

An important class of problems of this type is the hydrogeneration scheduling problem. This problem consists in the maximization of the profit obtained by producing hydroenergy along a time horizon (one year) in a multi-reservoir power system as, for example, that of Hydro-Québec (Hanscom et al. 1980). The decision variables are the amount of water to be released from and stored in each reservoir and in each time period (one week). Let K be the number of periods and L the number of reservoirs. The associated network is a temporally expanded arborescence (Kennington et al. 1980). Each node corresponds to a time period-reservoir pair and has two outgoing arcs : the storage s_{kl} of reservoir l at the end of period k and the release r_{kl} of reservoir l at period k.

Several types of **differentiable** objective functions have been used for this problem. In this paper we consider a **nondifferentiable** function **f** (to be minimized) of the following type

$$\sum_{k=1}^{K} C_k \, (W_k(x^k)) - R(x^K)$$

where $x^k = (r_{k1},\dots,r_{kL},s_{k1},\dots,s_{kL})$, $1 \leq k \leq K$, $W_k(\cdot)$ is the energy deficit at week k , $C_k(\cdot)$ is the cost of generating energy deficit W_k at period k and $R(\cdot)$ is the economic value associated with the final storage of the reservoirs. The functions W_k and $-R$ are convex and differentiable. Here C_k is modeled as a nondecreasing piecewise-linear function from $\mathbb{R}$ to $\mathbb{R}$ to take into account an energy market structure. Under these assumptions, f is a convex **nondifferentiable** function. If we denote by $\bar{g}_k$, $1 \leq k \leq K$, the gradient of $W_k(\cdot)$ at x^k , by $\bar{h}_K$ the gradient of $-R(\cdot)$ at x^K and by u_k and $\bar{u}_k$ respectively the left-hand side and the right-hand side derivative of $C_k(\cdot)$ at $W_k(x^k)$, then the subdifferential of f at x can be expressed as follows :

$$\partial f(x) = \{ \sum_{k=1}^{K} u_k \, g_k + h_K \mid u_k \leq u_k \leq \bar{u}_k \, , \quad k=1,\dots,K \} \qquad (1)$$

where $g_k^T = (0,\dots,0,\bar{g}_k^T,0,\dots,0)$ and $h_K^T = (0,\dots,0,\bar{h}_K^T)$.

2. A REDUCED SUBGRADIENT ALGORITHM

For solving this special scheduling problem a reduced subgradient strategy is adopted (Bihain et al. 1984). As usual the matrix A is partitioned into two submatrices B and H so that B is of full rank m. Let (x_B,x_H) be the corresponding partitioning of x in **basic** and **out-of-basis arcs.** We recall (Kennington et al. 1980) that the basic arcs form a spanning tree in the network and that each out-of-basis arc forms a unique cycle with basic arcs. Once the classical reduction is performed, the reduced problem becomes :

$$(RP) \quad \begin{cases} \text{Minimize} \quad \tilde{f}(x_H) \\ \text{s.t.} \quad x_H \leq x_H \leq \bar{x}_H \, , \end{cases}$$

where $\tilde{f}(x_H) = f(B^{-1}b - B^{-1}Hx_H \, , \, x_H)$.

If we denote by Z_H the $nx(n-m)$ matrix $\begin{pmatrix} -B^{-1}H \\ I \end{pmatrix}$ then the subdifferential of the convex function $\tilde{f}$ is given by : $\partial\tilde{f}(x_H) = \{Z_H^T g \mid g \in \partial f(x)\}$ and is called the **reduced subdifferential** of f at x . A feasible descent direction d_H (if it exists) in the space of out-of-basis arcs can be obtained by checking the optimality conditions of problem (RP). More precisely, using (1) we have to solve the following linear least-squares problem :

$$(Q_H) \quad \begin{cases} \text{Minimize} \quad \frac{1}{2} \| Z_H^T \left(\sum_{k=1}^{K} u_k\, g_k + h_K \right) - \lambda_H + \mu_H \|^2 \\ \\ \text{s.t.} \quad u_k \leq u_k \leq \overline{u}_k \ , \quad k=1,\dots,K \ , \\ \\ \qquad \lambda_H \geq 0 \ , \quad \mu_H \geq 0 \ , \end{cases}$$

$$\lambda_H^T\,(x_H - x_H) = 0 \ , \quad \mu_H^T\,(\overline{x}_H - x_H) = 0 \ .$$

Let u_k^* , λ_H^* and μ_H^* be a solution to (Q_H) . Then set

$$d_H = \lambda_H^* - \mu_H^* - Z_H^T\,(\Sigma\, u_k^*\, g_k + h_K) \ .$$

If $d_H = 0$, then x_H is a solution to (RP) and (x_B, x_H) is a solution to (P) . If $d_H \neq 0$, then set $d_B = -B^{-1}\, H\, d_H$ and $d^T = (d_B^T, d_H^T)$. It is easy to see that d is a descent direction which is feasible with respect to the bounds if : $(x_B)_i < (x_B)_i < (\overline{x}_B)_i$ is satisfied for each basic arc. If it is not satisfied, d_B need not to be feasible with respect to the bounds on the basic arcs. This is known as the **degeneracy** problem.

As d_B depends on d_H , the degeneracy problem can be solved by partitioning the matrix H into two submatrices (Murtagh et al. 1978) : S and N so that if we set $d_N = 0$ then d_B is feasible. Let $\overline{H}$, $\overline{S}$ and $\overline{N}$ be the arc index sets corresponding to matrices H , S and N . The arcs corresponding to S and N are called the superbasic and nonbasic arcs respectively. The problem is to decide for each $i \in \overline{H}$ if we put i in $\overline{S}$ or not. Two cases are possible : the variable corresponding to arc i is free or it is at its bound. If $i \in \overline{H}$ is free and if each arc of the cycle associated with arc i is also free, then we put i in S . As we want to have the set $\overline{S}$ as large as possible, we try to obtain a basis B containing the maximum number of free arcs. Such a basis B is called a maximal basis (Dembo et al. 1981) and has the property that there can only be free basic arcs in the cycle associated with a free out-of-basis arc. If $i \in \overline{H}$ is at its bound, we have to examine the cycle associated to arc i , arc by arc in order to see if the flow can be changed on arc i without violating the bounds on the basic arcs of the cycle. The arc $i \in \overline{H}$ will be called blocked and put in $\overline{N}$ if a basic arc of the cycle is in the same orientation as arc i but at the opposite bound or if a basic arc of the cycle is in the opposite orientation with respect to arc i but at the same bound.

Now we have $H = (S\ N)$ and we want to compute d_S (we know already that $d_N = 0$) by solving :

$$(Q_S) \quad \begin{cases} \text{Minimize} \quad \frac{1}{2} \| Z_S^T \left(\sum_{k=1}^{K} u_k\, g_k + h_K \right) - \lambda_S + \mu_S \|^2 \\ \text{s.t.} \quad u_k \le u_k \le \overline{u}_k \ , \quad k=1,\dots,K \ , \\ \qquad \lambda_S \ge 0 \ , \quad \mu_S \ge 0 \ , \\ \qquad \lambda_S^T (x_S - x_S) = 0 \ , \quad \mu_S^T (\overline{x}_S - x_S) = 0 \ . \end{cases}$$

If u_k^* , λ_S^* , μ_S^* denote a solution of (Q_S) then

$$d_S = \lambda_S^* - \mu_S^* - Z_S^T \left(\sum_{k=1}^{K} u_k^*\, g_k + h_K \right) .$$

Observe that $\Sigma_{k=1}^{K}$ can be replaced by $\Sigma_{k \in J_S}$ where J_S is the set of time periods covered by the cycles associated with S and that $(\lambda_S)_i = 0$ if $(x_S)_i$ is free or at its upper bound and $(\mu_S)_i = 0$ if $(x_S)_i$ is free or at its lower bound. The number of variables of (Q_S) is then the number of time periods $k \in J_S$ such that $u_k < \overline{u}_k$ plus the number of superbasic variables at their bound.

If $d_S = 0$, we check

$$\begin{cases} Z_t^T \left(\Sigma\, u_k^*\, g_k + h_K \right) \ge 0 \ , \quad t \in N \ , \quad (x)_t = (x)_t \ , \\ Z_t^T \left(\Sigma\, u_k^*\, g_k + h_K \right) \le 0 \ , \quad t \in N \ , \quad (x)_t = (\overline{x})_t \ . \end{cases}$$

If these conditions are satisfied, then $d_H = 0$ and x is optimal. Otherwise we have to solve (Q_H) to obtain d_H . If $d_H = 0$ then x is optimal; otherwise we have to check the feasibility of d_B . If it is the case we perform a line search along d ; in the other case we find a feasible descent direction d by solving :

$$(\mathrm{PL})\quad \begin{cases} \text{Minimize} \quad f'(x;d) \\ \text{s.t.} \quad A\,d = 0\ , \\ \qquad 0 \leq d_j \leq 1 \quad \text{if} \quad (x)_j = (x)_j\ , \\ \qquad -1 \leq d_j \leq 0 \quad \text{if} \quad (x_j) = (\bar{x})_j\ , \\ \qquad -1 \leq d_j \leq 1 \quad \text{if} \quad (x)_j < (x)_j < (\bar{x})_j\ , \end{cases}$$

where

$$f'(x;d) = h_K^T d + \sum_{k=1}^{K} \max\,\{u_K\, g_K^T d\ ,\ \bar{u}_K\, g_K^T d\}\ .$$

An experimental FORTRAN code implementing this algorithm has been written and tested on two scheduling problems related to the medium term energy generation planning problem for the Hydro-Québec multireservoir system.

The first test problem is a small-scale problem : it involves 8 reservoirs and 10 time periods and represents a network of 80 nodes and 168 arcs. The second test problem is a medium-scale problem : it also involves 8 reservoirs but 52 time periods. Here the network has 416 nodes and 840 arcs. The numerical results will appear in a forthcoming paper.

REFERENCES

Bihain, A., Nguyen, V.H., and Strodiot, J.J. (1984). A reduced subgradient algorithm. Internal report 84/3. Department of Mathematics. Facultés Universitaires N.-D. de la Paix, Namur, Belgium.

Dembo, R.S., and Klincewicz, J.G. (1981). A scaled reduced gradient algorithm for network flow problems with convex separable costs. Mathematical Programming Study, 15 : 125-147.

Hanscom, M., Lafond, L., Lasdon, L., and Pronovost, G. (1980). Modeling and resolution of the medium term energy generation planning problem for a large hydro-electric system. Management Science, 26 : 659-668.

Kennington, J., and Helgason, R. (1980). Algorithms for network programming. Wiley, New York.

Murtagh, B., and Saunders, M. (1978). Large scale linearly constrained optimization. Mathematical Programming, 14 : 41-72.

AN ALGORITHM FOR SOLVING A WATER-PRESSURE-CONTROL PLANNING PROBLEM WITH A NONDIFFERENTIABLE OBJECTIVE FUNCTION

Yoshikazu Nishikawa and Akihiko Udo
Department of Electrical Engineering, Kyoto University, Kyoto 606, Japan

1. INTRODUCTION

In this paper we develop an algorithm for a nondifferentiable optimization problem arising in pressure-control planning of water distribution networks (WDN).

Although the problem is of the nonlinear programming type, it is solved by iterating solutions of linear programs and descents along V-shaped ravines caused by the nondifferentiability of the objective function. The equations of the V-shaped ravines are derived from the physical law governing the steady-state flow of WDN. The resulting solution procedure is then widely applicable to large-scale networks.

Our early work on this problem has already been reported (Nishikawa and Udo, 1982). In this paper, the problem is reformulated in a mathematically more refined manner, the character of the V-shaped ravine is clarified, and a revised algorithm is constructed.

2. FORMULATION

The problem is to minimize the total energy, or equivalently the cost, expended in pumping while keeping the water heads (pressures) at all nodes in an allowable range. Every pipe link where a pump or a valve is introduced is considered.

This enables us to find desirable locations for pumps and/or valves as well as their scheme of operation, which is especially useful in the planning stage.

Let z denote a pressure gap due to a pump or a valve. Then the characteristic equation of pipe link i equipped with a pump or a valve is written as

$$h_i = r_i q_i |q_i|^{0.85} - z_i \qquad (1)$$

where h_i is the head differential (the difference of the heads at both ends of a link), r_i is the resistance factor determined by the diameter, length and smoothness of the pipe, and q_i is the flow rate, all of link i.

If $q_i z_i > 0$, z_i denotes the pressure gap given by the pump, while if $q_i z_i < 0$, that by the valve.

Then our problem is formulated as follows:

$$\text{(P1)} \quad \text{minimize} \quad f = \Sigma_i (q_i z_i + |q_i z_i|)/2 \qquad \text{(Pumping cost)} \quad (2)$$

subject to

$$L(Z,\ Q_c) = (l_\phi) \triangleq \text{(Linear function of } Z\text{)}$$
$$+ \text{(Nonlinear function of } Q_c\text{)}$$
$$\text{(Head-differential loop law: HDLL)} \quad (3)$$

$$\underline{P}_j \leq P_j(Z,\ Q_c) \triangleq \text{(Linear function of } Z\text{)}$$
$$+ \text{(Nonlinear function of } Q_c\text{) and}$$
$$P_j \leq \overline{P}_j \qquad \text{(Node-head condition)} \quad (4)$$

Here $Z \triangleq (z_i)$, Q_c is the vector of the flow rates of cotree links, i.e., a set of necessary and sufficient variables to describe all q_i's. The HDLL is equivalent to Kirchhoff's voltage law and implies that the total head differential around any loop is zero. $\underline{P}_j = (\underline{p}_\mu)$ and $\overline{P}_j = (\overline{p}_\mu)$ are the vectors denoting the lowest and the highest allowable values of $P_j = (p_\mu)$, the heads at the consumption nodes.

By way of example, the problem (P1) is formulated as follows for Network-1 of Fig. 1.

minimize $q_1z_1+|q_1z_1|+q_2z_2+|q_2z_2|+(2-q_1-q_2)z_3$

$+|(2-q_1-q_2)z_3|+(1-q_2)z_4+|(1-q_2)z_4|$

subject to $r_1q_1|q_1|^{0.85}-z_1-r_3(2-q_1-q_2)|2-q_1-q_2|^{0.85}+z_3=0$

$r_2q_2|q_2|^{0.85}-z_2-r_4(1-q_2)|1-q_2|^{0.85}+z_4$

$-r_3(2-q_1-q_2)|2-q_1-q_2|^{0.85}+z_3=0$

$\underline{p}_1\leq p_1=p_0-r_1q_1|q_1|^{0.85}+z_1\leq\overline{p}_1$

$\underline{p}_2\leq p_2=p_0-r_2q_2|q_2|^{0.85}+z_2\leq\overline{p}_2$

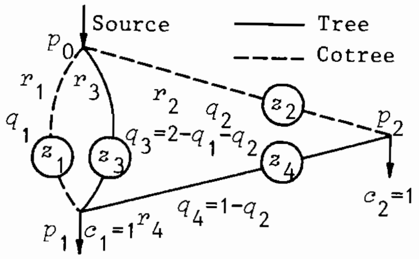


Fig. 1. Diagram of Network-1.

3. BASIC ALGORITHM

(P1) is obviously a nonlinear programming problem. However, if Q_c is fixed at some value, and if z_i is written as $z_i=x_i-y_i$ $(x_i,\ y_i\geq 0)$, (P1) is reduced to a linear programming problem (LPP) whose unknowns are $X=(x_i)$ and $Y=(y_i)$. Let us denote the optimal value of the LPP with Q_c fixed by $f^*(Q_c)$ and the optimal point by $(X^*,\ Y^*)$. Then the gradient of $f^*(Q_c)$ with respect to Q_c can be calculated using the shadow prices of the LPP as follows;

$$\nabla f^*(Q_c)=\frac{\partial f^*}{\partial Q_c}+\sum_{\mu\in S_L}\frac{\partial f^*\partial p_\mu}{\partial \underline{p}_\mu\partial Q_c}+\sum_{\mu\in S_U}\frac{\partial f^*\partial p_\mu}{\partial \overline{p}_\mu\partial Q_c}+\sum_{\phi}\frac{\partial f^*}{\partial l_\phi}\frac{\partial l_\phi}{\partial Q_c}$$

$$=\frac{\partial f^*(X^*,Y^*,Q_c)}{\partial Q_c}+\begin{bmatrix}\partial P_j(X^*,Y^*,Q_c)/\partial Q_c{}^T\\ \partial P_j(X^*,Y^*,Q_c)/\partial Q_c{}^T\\ \partial L(X^*,Y^*,Q_c)/\partial Q_c{}^T\end{bmatrix}^T\Lambda \qquad (5)$$

T: Transposition of a vector/matrix

S_L and S_U denote the set of active lower node-head constraints and that of active higher node-head constraints at $(X^*,\ Y^*)$, respectively. Λ is the vector of the shadow prices at $(X^*,\ Y^*)$ and its size is equal to the number of constraints (3) and (4).

Suppose that $\nabla f^*(Q_c)$ is always defined. If $\nabla f^*(Q_c)\neq 0$, then there is a positive number δ (step size) which satisfies

$$f^*(Q_c - \delta \nabla f^*(Q_c)) < f^*(Q_c) \tag{6}$$

Hence, the optimal solution of (P1) can be found by iterating solutions of the LPP and computing the gradient, Eq. (5).

Since a pump or a valve can be located on every link, the head at every node can be set arbitrarily for any Q_c. Hence, the LPP is always solvable.

It is difficult to know how best to determine the step size δ. One-dimensional search is far from efficient because of the time needed to compute $f^*(Q_c)$. In fact, this involves the solution of an LPP. We therefore use the following algorithm:

Basic Algorithm (Algorithm 1).

Suppose that Q_c^0 and $\delta^0 > 0$ are given.

```
(Step 1)  Set $k=0$.
(Step 2)  Set $Q_c^k = Q_c^0$ and set $\delta^k = \delta^0$.
(Step 3)  Compute $Q_c^{k'} = Q_c^k - \delta^k \nabla f^*(Q_c^k) / |\nabla f^*(Q_c^k)|$.
            $|\cdot|$: Euclid norm of a vector
(Step 4)  If $f^*(Q_c^{k'}) < f^*(Q_c^k)$, go to Step 5;
              otherwise, go to Step 7.
(Step 5)  Set $Q_c^{k+1} = Q_c^{k'}$, and set $\delta^{k+1} = \begin{cases} 1.5\delta^k & \text{(if } \delta^k \geqq \delta^{k-1} \geqq \delta^{k-2}) \\ \delta^k & \text{(except the above)} \end{cases}$
(Step 6)  Set $k=k+1$, and go to Step 3.
(Step 7)  Set $\delta^k = \delta^k/2$.
(Step 8)  If $\delta < \varepsilon$ stop; otherwise, return to Step 3.
            $\varepsilon$: a reference small positive quantity for stopping
               the algorithm.
```

4. V-SHAPED RAVINE

The basic algorithm stops on the subspace of Q_c-space where $\nabla f^*(Q_c)$ is not defined. Let us call such a subspace a *V-shaped ravine*.

The V-shaped ravine is caused by the nondifferentiability of the objective function f: f is nondifferentiable with respect to q_i and z_i at $q_i=0$ and $z_i=0$, respectively. In fact, if the sign of $q_i z_i$ switches, the objective function of the LPP changes, and consequently the V-shaped ravine is formed. It should be noted that the sign of $z_i(z_i^*)$ cannot be known until the LPP is solved.

The subspace of q_i=0 is a hyperplane in Q_c-space, because q_i is a linear function of the components of Q_c. z_i=0 is the subspace where the basis (the set of basic variables) of the LPP changes, i.e., at least one of x_i^* and y_i^* switches between zero and positive.

The V-shaped ravine can also be explained through Eq. (5). The first term of Eq. (5) is discontinuous at the subspace of q_i=0 and z_i^*=0. The second term is also discontinuous at the subspace of z_i^*=0, because some components of Λ change discontinuously there due to the change in the basis of the LPP. (Note: a V-shaped ravine can thus emerge even if the objective function is smooth.)

Now let us consider the subspace of z_i^*=0 in detail. It must be noted that we use the linear graph Ξ where inflows from sources and outflows from consumption nodes are represented by the flow rates of the reference-node connected links. Figure 2 shows Ξ of Network-1. Consider the neighbourhood of $\overline{Q}_c$, a point on a V-shaped ravine. First, if neither the upper constraint nor the lower constraint on p_μ is active at $\overline{Q}_c$, and the constraints are also not active in the proper neighbourhood of $\overline{Q}_c$, then the consumption link μ is not involved in the change in basis of the LPP. Second, if x_i^*>0 holds for pipe link i, x_i^* remains positive for a small change in q_i and also for the small change in h_i caused by small changes in the flow rates in other links, as far as such changes are in the proper neighbourhood of $\overline{Q}_c$. The same is valid for the case when y_i^*>0 holds for pipe link i. Thus, a change in basis is possible only on the subgraph Ξ_b obtained by deleting the above mentioned consumption links and pipe links from Ξ.

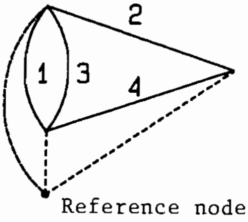


Fig. 2. Ξ of Network-1.

Consider the consumption link μ in Ξ_b. Since its node head is constrained to the lowest or the highest allowable value, the node-head condition is equivalent to the HDLL. By way of example, in Network-1, the HDLL of the loop of links 1 and 3, and the node-head condition at node 1 (assumed to be active) are written as follows:

$$x_1^*-y_1^*-(x_3^*-y_3^*)=r_1q_1|q_1|^{0.85}-r_3(2-q_1-q_2)|2-q_1-q_2|^{0.85}$$
$$x_1^*-y_1^*=\underline{p}_1-p_0+r_1q_1|q_1|^{0.85}$$

where the components of X and Y are collected on the left sides of the equations, and the components of Q_c and constants on the left sides.

Find a full set of independent loops in Ξ_b and write down the HDLL of those loops. Let us denote the set of the right sides of these equations by $G(Q_c)=(g_i(Q_c))$, which are called loop head-loss terms.

Now, since the left sides of the equations are all zero,

$$G(Q_c)=0 \tag{7}$$

is satisfied at $\overline{Q}_c$. If some $g_i(Q_c)$ becomes positive or negative, at least one of $z_i^*=x_i^*-y_i^*$ switches its sign. That is to say, Eq. (7) describes a V-shaped ravine.

If Q_c, i.e., the flow pattern, is changed along the V-shaped ravine, a new descent of $f^*(Q_c)$ becomes possible. The descent along the ravine does not put a pump or valve in any pipe link in Ξ_b.

5. REVISED ALGORITHM

Based on the foregoing discussion, an algorithm which descends along the bottom of a ravine is constructed in this section.

5.1 Algorithm for the Search of Ravine Equations (Algorithm 2)

Consider the k-th iteration of the basic algorithm. Let LP(Q_c) denote the LPP at Q_c, and let δ_r be a small positive value for judging an encounter with a V-shaped ravine.

(Step 0) If $f^*(Q_c^{k'})>f^*(Q_c^k)$, for $Q_c^{k'}=Q_c^k-\delta^k\nabla f^*(Q_c)/|\nabla f^*(Q_c)|$ with $\delta^k<\delta_r$, that is, if the cost cannot be improved even if the step size δ^k is small, go to (Step 1); otherwise, iterate the basic algorithm.

(Step 1) Find the subgraph Ξ_b based upon the solutions $X^k=(x_i)$ and $Y^k=(y_i)$ of LP(Q_c^k), and the solutions $X^{k'}=(x_i')$ and $Y^{k'}=(x_i')$ of LP$(Q_c^{k'})$.

a) Let all the source links be included in Ξ_b.

b) Consider pipe link i. If $(x_i-y_i)(x_i'-y_i')>0$, since link i is not involved in the change of basis, let link $i\notin\Xi_b$; otherwise, let link $i\in\Xi_b$.

c) Consider consumption link μ. If $\underline{p}_\mu<p_\mu^k<\overline{p}_\mu$ and $\underline{p}_\mu<p_\mu^{k'}<\overline{p}_\mu$, since the constrainits are not active both at p_μ^k and at $p_\mu^{k'}$, let link $\mu\notin\Xi_b$; otherwise, let link $\mu\in\Xi_b$. Here, p_μ^k and $p_\mu^{k'}$ denote the heads at node μ in the solutions of LP(Q_c^k) and LP$(Q_c^{k'})$, respectively.

(Step 2) In Ξ_b, find a full set of independent loops by spanning a tree, and construct their loop head-loss terms $g_i(Q_c)$ $(i=1, 2, \ldots, \tau_0)$.

(Step 3) If the sign of q_i at $Q_c^{k'}$ differs from that at Q_c^k, the equation of $q_i=0$ is added to the ravine equations as $g_i(Q_c)$ $(i=\tau_0+1, \ldots, \tau)$.

5.2 Algorithm for Descent along a V-shaped Ravine (Algorithm 3)

Suppose that the flow pattern is now Q_c^k and is close to a V-shaped ravine. Further, suppose that, by Algorithm 2, the ravine equations turn out to be

$$G(Q_c)\triangleq\{g_i(Q_c)=0 \quad (i=1,\ 2,\ \ldots,\ \tau)\} \tag{8}$$

(Step 1) Let v be the projected vector of $-\nabla f^*(Q_c)$ on the tangential hyperplane of the V-shaped ravine of Eq. (8). Change the flow pattern to $Q_c^{k'}$ which is in the direction of v by a step size δ :

$$Q_c^{k'} = Q_c^k - \delta^k v/|v| \tag{9}$$

where

$$v = (I - D^T(DD^T)^{-1}D)(-\nabla f^*(Q_c^k)), \qquad D \triangleq (\partial G^T(Q_c)/\partial Q_c)^T |_{Q_c = Q_c^k}$$

(Step 2) By use of the Newton-Raphson method, change the flow pattern from $Q_c^{k'}$ to $Q_c^{k''}$ which satisfies Eq. (8) (see Fig. 3).

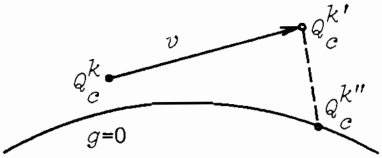


Fig. 3. Descent along the V-shaped ravine.

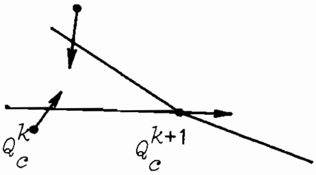


Fig. 4. V-shaped ravines.

It must be noted that more than one ravine at a time may be found by Algorithm 2 (see Fig. 4). In such a case, Eq. (8) denotes the intersection of those ravines. In general, the ravines terminate at an intersection and a new ravine starts there. (Note: If the intersection is a point, it may be the optimal point.) Then the nearest point on the intersection from Q_c^k is chosen as Q_c^{k+1} and descent is restarted from Q_c^{k+1} by using Algorithm 1. The following step is appended for this purpose.

(Step 3) If $f^*(Q_c^k) < f^*(Q_c^{k''})$, then execute Step 1 and Step 2 for a smaller step size, $0.1\delta^k$. If f^* does not decrease even for this step size, find the nearest point on the subspace of Eq. (8) from Q_c^k and let the point be Q_c^{k+1}.

5.3 Revised Algorithm

Algorithms 1 through 3 are combined as follows.

(1) Start minimization by using Algorithm 1, the basic algorithm.

(2) When the cost cannot be improved even if step size δ^k is made smaller than δ_r as in the Step 0 of Algorithm 2, switch into Algorithm 2 and find the ravine equations.

(3) The descent along the ravine bottom may be stopped at some

point by an encounter with another V-shaped ravine. If it is stopped, restart minimization by using Algorithm 1 from that point. This is because the equations of the present ravine are not necessarily included in the set of equations of the new ravine.

Since the step size δ^{k^r} becomes smaller and smaller as Q_c approaches the optimal point, we halve δ_r on the application of Algorithm 2, bearing in mind the balance between δ^k and δ_r.

Finally, the procedure is stopped when both $\delta^k < \varepsilon$ and $\delta_r < \varepsilon$ hold.

6. EXAMPLE

The revised algorithm was applied to some networks of practical size. For each network, computations were started from some different initial values of Q_c, and the unique minimum-cost solution was obtained. It is practical to use the steady-state flow without use of any pumps and valves as the initial value.

In the case of the network of 36 nodes, 42 links and 4 sources presented in our earlier paper, the solution is obtained by solving a chain of 80 linear programs which consist of 84 unknowns and 42 constraints (only the lower constraint for each node head), with some extra time for descent along the V-shaped ravines.

REFERENCES

Nishikawa, Y., Udo. A., (1982). Uniform Pressure Control in Water Distribution Networks Including Location of Control Points, Lecture Notes in Control and Information Sciences, 44, 656-669, Springer-Verlag.

QUASI-DIFFERENTIABLE FUNCTIONS IN THE OPTIMAL CONSTRUCTION OF ELECTRICAL CIRCUITS

E.F. Voiton
Department of Applied Mathematics, Leningrad State University,
Universiteskaya Nab. 7/9, Leningrad 199164, USSR

When considering the optimization problems which arise in the design of technical devices, it is clear that a central role is played by minimax problems, i.e., the problem of finding

$$\min_{X\in\Omega} \phi(X) \tag{1}$$

where $\phi(X) = \max\limits_{y\in G} f(X,y)$ is a maximum function, $\Omega \subset E_n$, $G \subset E_m$.

The minimax formulation of the problem is in many cases preferable to other models and in some cases is crucial.

Problems of form (1) appear, in particular, in the design of electric circuits if it is necessary to find either the values of components of a circuit of given structure (parametric synthesis) or the values of parameters of a circuit function of given type (the approximate synthesis problem).

In what follows we consider some examples of problems based on the structural synthesis of electrical circuits which contain linear elements with constant parameters, linear elements with variable (so-called controlled) parameters and non-inertial nonlinear elements.

We consider the possibility of stating a wide class of optimization problems of form (1) and suggest a unified approach to solving these problems.

1. Analytical methods for solving (1) can be applied only to a limited number of one-dimensional approximation problems where the function f(X,y) is an algebraic or trigonometric polynomial or a rational function with a given polynomial in the denominator.

These functions describe characteristics of certain classes of electrical circuits, the most sophisticated of them being the so-called frequency filter, i.e., a device with different properties on two nonintersecting sets of some variable.

2. A wider class of devices is described by functions f(X,y) which are continuous with $\frac{\partial f(X,y)}{\partial X}$ jointly in both variables on a set $\Omega \times G$. It is necessary to use a more complicated criterion function since in many circuits there exist elements with fixed values of parameters, and additional constraints to be satisfied by the circuit.

If f is a linear function then one can apply the Remez polynomial algorithm [1,2]. But in general the function f(X,y) is nonlinear in X therefore we cannot use this algorithm.

Some effective minimization methods are based on the directional differentiability of a maximum function [2]. Since

$$\frac{\partial\phi(X)}{\partial g} \equiv \lim_{\alpha\to+0} \frac{\phi(X+\alpha g)-\phi(X)}{\alpha} = \max_{y\in R(x)} \left(\frac{\partial f(X,y)}{\partial X}\ ,\ g\right)$$

where $R(X) = \{X\in G \mid f(X,y) = \phi(X)\}$. Then the necessary condition of an unconstrained minimum

$$\frac{\partial\phi(X^*)}{\partial g} \geq 0 \qquad \forall g \tag{2}$$

is equivalent to the condition

$$0 \in \underline{\partial}\phi(X^*) \tag{3}$$

where $\underline{\partial}\phi(X) = \text{co}\left\{\frac{\partial f(X,y)}{\partial X}\middle|\ y\in R(X)\right\}$.

If at $X_0 \in E_n$ condition (3) is not satisfied then the direction

$$g(X_0) = -\frac{Z(X_0)}{\|Z(X_0)\|}$$

where

$$\|Z(X_0)\| = \min_{Z\in\underline{\partial}\phi(X_0)} \|Z\|$$

is the uirection of steepest descent (of f at X_0).

Problem (1) can be discretized (i.e., the set G can be replaceu by a finite number of points) and we shall have a discrete minimax problem which can be solved by well-known methods. There are different approaches to discretize G. The "direct" method (to replace G by a "thick" grid) is too "expensive" from the computational standpoint. Much more effective is "the extremal basis method" which uses only n + 2 points (where n is the dimensionality of the space) at each step, but these points ("a basis") are being adjusted at each step (see [5]). Computational experiments showed that the extremal basis method is highly effective, especially if the method of equalizing maxima (see [7]) is applied at the final stage of computations.

EXAMPLE 1. (The Mandelshtam problem).

Let

$$f(X,t) = \cos t + \sum_{k=1}^{15} \cos ((k+1)t+x_k)) \quad .$$

It is required to find $X^* = (x_1^*,\dots,x_{15}^*) \in E_n$ such that

$$\phi(X^*) = \max_{t\in[-\pi,\pi]} f(X^*,t) = \min_{X\in E_n} \max_{t\in[-\pi,\pi]} f(X,t) \quad . \quad (4)$$

This is the problem of finding the "phase" shifts in the circuits of 16 harmonic generators (it is assumed that $x_0 = 0$) which guarantee the resulting signal with the minimal value of the "maximal" level.

Note that the maximal possible value is 16 (it is achieved if $x_i = 0 \ \forall\, i \in 1:15$). After solving problem (4) (by any available minimax technique) we get the following optimal solution

$$\begin{array}{llll} X^{*} = (1.57625; & 1.99774, & -2,91176; & -2.00577 \\ \quad 0.35074, & 2.60184, & -1.13237; & 2.83581 \\ \quad 0.55282, & -1.34677, & 2.19116; & 2.26278 \\ \quad 0.75618, & 0.97112, & 0.06332). & \end{array}$$

This peak value of the total signal was reduced from 16 to $3.89755 = \phi(X^{*})$.

The function $h(t) \equiv f(X^{*},t)$ achieves its maximal value (with respect to t) at 16 points.

$$\begin{array}{llll} t_1 = -2.2295; & t_2 = -1.8311, & t_3 = -1,5243; & t_4 = -1,1755 \\ t_5 = -0.5162; & t_6 = \ \ 0.1445; & t_7 = \ \ 0.3655, & t_8 = \ \ 0.7931 \\ t_9 = 1.0087; & t_{10} = 1.2297; & t_{11} = 1.4687; & t_{12} = 1.8025 \\ t_{13} = 2.1754; & t_{14} = 2.3969; & t_{15} = 2.7671; & t_{16} = 3.1318 \end{array}$$

The signs of $f(X^{*},t)$ were as follows

$$+,-,-,+,+,-,+,+,-,+,-,+,-,+,-,+$$

This fact shows that there is no "alternance" property (as was the case in linear minimax problems).

3. In solving practical problems it is often necessary to minimize a function which is a composition of max-type functions.

Let a function $F(a_1,a_2,\ldots,a_p)$ be continuously differentiable on E_p, and $\phi_k(x), k \in 1:p$, be functions of the form $\phi_k = \max\limits_{y \in G_k} f_k(X,y)$ (or $\phi_k = \min f_k(X,y)$), where G_k are compact in E_m, and function $f_k(X,y)$ as before are continuous together with $\dfrac{\partial f_k(X,y)}{\partial X}$ on $E_n \times G_k$. Let $\phi(X) = F(\phi_1(X),\ldots,\phi_p)X))$ be a superposition of functions $\phi_1(X),\ldots,\phi_p(X)$. Without loss of generality we assume that $\phi_k(X)$ are max-type functions.

Since ϕ_k are directionally differentiable then ϕ is also directionally differentiable and

$$\frac{\partial \phi(X)}{\partial g} = \sum_{k=1}^{p} \frac{\partial F}{\partial \phi_k} \quad \frac{\partial \phi_k(X)}{\partial g} \quad .$$

Since $$\frac{\partial \phi_k(X)}{\partial g} = \max_{y \in R_k(X)} \left(\frac{\partial f_k(X,y)}{\partial X} , g \right) ,$$

where

$$R_k(X) = \{ y \in G_k \mid f_k(X,y) = \phi_k(X) \} \quad , \quad k \in 1:p \quad ,$$

then

$$\frac{\partial \phi(X)}{\partial g} = \sum_{k=1}^{p} \frac{\partial F}{\partial \phi_k} \max_{y \ R_k(X)} \left(\frac{\partial f_k(X,y)}{\partial X} , g \right) \quad . \tag{5}$$

Putting

$$\frac{\partial F(\phi_1(X),\ldots,\phi_p(X))}{\partial \phi_k} = \Psi_k(X) \quad ,$$

we can rewrite (5) in the form

$$\frac{\partial \phi(X)}{\partial g} = \sum_{k=1}^{p} \Psi_k(X) \max_{y \in R_k(X)} \left(\frac{\partial f_k(X,y)}{\partial X} , g \right) =$$

$$= \sum_{k \in J_+(X)} \Psi_k(X) \max_{y \in R_k(X)} \left(\frac{\partial f_k(X,y)}{\partial X} , g \right) +$$

$$+ \sum_{k \in J_-(X)} \Psi_k(X) \max_{y \in R_k(X)} \left(\frac{\partial f_k(X,y)}{\partial X} , g \right) \equiv A + B$$

where

$$J_+(X) = \{ k \in 1:p \mid \Psi_k(X) \geq 0 \}, \; J_-(X) = \{ k \in 1:p \mid \Psi_k(X) < 0 \} \quad ,$$

$$A = \sum_{k\in J_+(X)} \max_{y\in R_k(X)} \left(\Psi_k(X)\ \frac{\partial f_k(X,y)}{\partial X}\ ,\ g\right) =$$

$$= \max_{\substack{y_1\in R_1(X)\\ y_2\in R_2(X)\\ \cdots\cdots\cdots\\ y_p\in R_p(X)}} \left(\sum_{k\in J_+(X)} \Psi_k(X)\ \frac{\partial f_k(X,y_k)}{\partial X},\ g\right)$$

$$B = \sum_{k\in J_-(X)} \Psi_k(X) \max_{y\in R_k(X)} \left(\frac{\partial f_k(X,y)}{\partial X}\ ,\ g\right) =$$

$$= \sum_{k\in J_-(X)} \min_{y\in R_k(X)} \left(\Psi_k(X)\ \frac{\partial f_k(X,y)}{\partial X}\ ,\ g\right) =$$

$$= \min_{\substack{y_1\in R_1(X)\\ y_2\in R_2(X)\\ \cdots\cdots\cdots\\ y_p\in R_p(X)}} \left(\sum_{k\in J_-(X)} \Psi_k(X)\ \frac{\partial f_k(X,y_k)}{\partial X}\ ,\ g\right) \quad .$$

Thus,

$$\frac{\partial\phi(X)}{\partial g} = \max_{\substack{y_1\in R_1(X)\\ \cdots\cdots\cdots\\ y_p\in R_p(X)}} \left(\sum_{k\in J_+(X)} \Psi_k(X)\ \frac{\partial f_k(X,y_k)}{\partial X}\ ,\ g\right) +$$

$$+ \min_{\substack{y_1\in R_1(X)\\ \cdots\cdots\cdots\\ y_2\in R_p(X)}} \left(\sum_{k\in J_-(X)} \Psi_k(X)\ \frac{\partial f_k(X,y_k)}{\partial X}\ ,\ g\right) \quad . \quad (6)$$

Recall (see [5]) that a function f is called quasidifferentiable at a point $X \in E_n$ if it is differentiable at the point X in any direction $g \in E_n$ and if there are convex compacts $\underline{\partial} f(X) \subset E_n$ and $\overline{\partial} f(X) \subset E_n$ such that

$$\frac{\partial f(X)}{\partial g} \equiv \lim_{\alpha\to+0} \frac{f(X+\alpha g)-f(X)}{\alpha} = \max_{v\in\underline{\partial} f(X)} (v,g) + \min_{w\in\overline{\partial} f(X)} (w,g).$$

The pair of sets $Df(X) = [\underline{\partial} f(X), \overline{\partial} f(X)]$ is called a quasidifferential of the function f at the point X, and sets $\underline{\partial} f(X)$ and $\overline{\partial} f(X)$ are respectively called a subdifferential and a superdifferential of the function f at the point X.

Now it is easy to see that the function ϕ is quasidifferentiable.

We say that a quasidifferentiable function f has a vertex-type quasidifferential at a point X, if the subdifferential $\underline{\partial} f(X)$ and the superdifferential $\overline{\partial} f(X)$ may be represented as convex hulls of a finite number of points. In the case of a vertex-type quasidifferential the formulas of quasidifferential calculus are readily applicable in practice.

For example, let $f(X) = \max_{i\in I} f_i(X)$, where $X \in E_n$, and functions f_i are quasidifferentiable, and

$$\underline{\partial} f_i(X) = \text{co } A_i(X) = \text{co } \{a_1^i, \ldots, a_{m_i}^i\} ,$$

$$\overline{\partial} f_i(X) = \text{co } B_i(X) = \text{co } \{b_1^i, \ldots, b_{n_i}^i\} .$$

Here each of $A_i(X)$ and $B_i(X)$ consists of a finite number of points in E_n.

Let

$$R(X) = \{i \in I \mid f(X) = f_i(X)\}. \quad \text{Then}$$

$$\underline{\partial} f(X) = \text{co } A(X), \overline{\partial} f(X) = \text{co } B(X) ,$$

where

$$A(X) = \left\{ a = a^{i'}_{j(i')} - \sum_{\substack{i\in R(X)\\ i\neq i'}} b^i{}_{k(i)} \,\middle|\, i' \in R(X), j(i') \in 1:m_i, k(i) \in 1:n_i \right\},$$

$$B(X) = \left\{ b = \sum_{i\in R(X)} b^i_{k(i)} \,\middle|\, k(i) \in 1:n_i \right\} .$$

It is easy to see that the number of points in the set A(X) is equal to $\sum_{i'\in R(X)} (m_{i'} \times \prod_{\substack{i\in R(X)\\ i\neq i'}} n_i)$ and in the set B(X) is equal to $\prod_{i\in R(X)} n_i$.

A simple structure of vertex-type quasidifferentials enables one to apply well-known methods for finding the distance between sets $\underline{\partial} f(X)$ and $-\overline{\partial} f(X)$ and at the same time to check whether the necessary condition for an unconstrained minimum

$$-\overline{\partial} f(X) \subset \underline{\partial} f(X) \tag{7}$$

is satisfied and to determine a direction of steepest descent.

It is easy to see from (6) that ϕ has a vertex-type quasidifferential, if the sets $R_k(X)$, $k\in 1:p$, are finite.

4. Let $X\in\Omega\subset E_n$, $y\in G\subset E_m$, $Z\in\omega\subset E_s$. Here G and ω are compact sets in proper spaces, Ω is a convex compact set. Let

$$\phi(X,y) = \min_{Z\in\omega} f(X,y,Z) \quad ,$$

$$Q(X,y) = \{Z\in\omega \mid f(X,y,Z) = \phi(X,y)\} \quad .$$

Consider the function

$$\phi(X) = \max_{y\in G} \phi(X,y)$$

and the set

$$R(X) = \{y\in G \mid \phi(X,y) = \phi(X)\} \quad .$$

The problem of minimizing ϕ on $\Omega\subset E_n$ is reduced to that of finding parameters $X\in\Omega$ and determining the relation Z(y), which provide the minimal values of f(X,y,Z) on the set G i.e., to the problem:

$$\max_{y\in G}\min_{Z\in\omega} f(X,y,Z) \longrightarrow \min_{x\in\Omega} \quad .$$

By discretizing sets G and ω, we have the problem of minimizing the function

$$\phi(X) = \max_{i\in I} \min_{j\in J} f_{ij}(X) \quad .$$

It is easy to see that this function is quasidifferentiable. Consider the following example.

Let

$$f(X,y,z) = \frac{1}{4x_1yz} \{(z+x_1y)^2+[x_1z(1+y^2)- -(z^2+x_1^2y^2)] \cos^2(x_2\sqrt{x_1z})\} - 1$$

where

$$X = (x_1,x_2) \in E_2,\ y\in [\bar{y}_1,\bar{y}_2] \subset E_1,\ z\in [\bar{z}_1,\bar{z}_2] \subset E_1 \quad .$$

It is necessary to find $\min_{X\in E_2} \phi(X)$ where

$$\phi(X) = \max_{y\in[\bar{y}_1,\bar{y}_2]} \min_{z\in[\bar{Z}_1,\bar{Z}_2]} f(X,y,Z) \ .$$

This is the problem of optimizing the operational attenuation of a ferrite impedance transformer by choosing the proper values of parameters x_1,x_2 (the dielectrical permeability and the electrical length) and determining an optimal rule for controlling the magnetic permeability $z(y)$ if the transformer load changes on the interval $[\bar{y}_1,\bar{y}_2]$.

Fix numbers N_1,N_2 and put

$$y_i = \bar{y}_1 + \frac{\bar{y}_2-\bar{y}_1}{N_1} i;\ z_j = \bar{z}_1 + \frac{\bar{z}_2-\bar{z}_1}{N_2} j;\ i\in 0:N_1 \equiv I,\ j\in 0:N_2 \equiv J \quad .$$

The initial function ψ can be approximated by the function

$$\phi(X) = \max_{i\in I} \min_{j\in J} f_{ij}(X)$$

where $f_{ij}(X) = f(X,y_i,z_j)$.

The function ϕ is quasidifferentiable. Find its quasidifferential.

We have

$$\phi(X) = \max_{i\in I} f_i(X)$$

where

$$f_i(X) = \tilde{f}(X,y_i) = \min_{j\in J} f_{ij}(X) \quad .$$

Since the functions f_{ij} are continuously differentiable, we can take

$$\underline{\partial} f_{ij}(X) = \{0\}; \overline{\partial} f_{ij}(X) = \left\{ \frac{\partial f_{ij}(X)}{\partial X} \right\} \quad .$$

Using the rules of quasidifferential calculus [5] we obtain

$$\underline{\partial} f_i(X) = 0,\ \overline{\partial} f_i(X) = \mathrm{co}\left\{ \frac{\partial f_{ij}(X)}{\partial X} = \frac{\partial f(X,y_i,Z_j)}{\partial X} \;\middle|\; Z_j \in Q_i(X) \right\}$$

where

$$Q_i(X) = Q(X,y_i) = \{Z_j \mid j \in J, f_i(X) = f_{ij}(X)\} \quad .$$

Finally we get

$$\underline{\partial}\phi(X) = - \mathrm{co} \sum_{\substack{y_k \in R(X) \\ y_k \neq y_i}} \overline{\partial} f_k(X) \mid y_i \in R(X)\} \ ,$$

$$\overline{\partial}\phi(X) = \sum_{y_i \in R(X)} \overline{\partial} f_i(X)$$

where

$$R(X) = \{y_i \mid i \in I, \phi(X) = f_i(X)\} \quad .$$

It is clear that the function ϕ has a vertex-type quasidifferential.

5. Let us now consider the problem of designing smoothly tuning frequency filters. Mathematically this can be stated as the following problem:

$$F(X) = \max_{t\in[a,b]} \min_{z(t)\in w} \max_{y\in[t,t+\delta]} f(X,y,Z) \longrightarrow \min_{X\in\Omega} \qquad (8)$$

where a,b are constants, $\delta < \lessdot$ b-a, w is some class of functions defined on [a,b]. Replacing each of the intervals [a,b] and [t,t+δ] by a finite number of points and a function Z(t) by the corresponding vector we shall approximate the function F by the function

$$F_1(X) = \max_{i\in 0:N} \min_{j\in J} \max_{k\in i:(i+\ell)} f_{ij}(X)$$

where

$$f_{kj}(X) = f(X,y_k,Z_j) \quad .$$

Clearly, F_1 is a quasidifferentiable function and has a vertex-type quasidifferential.

The problem of designing discrete controllable frequency band filters is of particular interest. The returning of these filters within the workable frequency band, for example, by the passband is performed in steps by switching filter element groups. Capacities are often used as components of such groups.

The problem of optimal synthesis (in the Chebyshev sense) of the discretized controllable filter may be presented as follows:

$$\max_{i\in I} \min_{Z\in\omega_i} \max_{t\in S_i} f(X,Z,t) \longrightarrow \min_{x\in\Omega}$$

where S_i is the set of workable band frequencies, $S_i \subset E_1$;
I is an index set, $I = 1:p$;
p is a number of filter subbands
ω_i is a set of groups of discretized tunable elements, $\omega_i \subset E_m$;
Ω is a set of unvariable filter elements, $\Omega \subset E_n$.

6. Now let us discuss the problem of synthesising non-linear circuits. Mathematically this can be stated as the problem of minimizing the function

$$\phi(X) = \max_{t\in[0,T]} |f(X,t) - F(t)|$$

where

$$f(X,t) = \bar{f}(X,u(t)), \bar{f}(X,u) = \sum_{i=1}^{m} f_i(X,u) \quad ;$$

F(t) is a given function; $x \in E_n$, u(t) is a periodic function of a given period T; $f_i(X,u)$ are so-called module functions. The function f(X,t) is the result of transforming the function u(t) by a nonlinear element, the volt/ampere characteristics of which are given by the module function $\bar{f}(X,u)$.

Consider two examples of solving practical problems.

Let

$$f(X,u) = (u-x_0+|u-x_0|), \; u = x_1 \cos t, \; X = (x_1,x_2) \quad ,$$

$$F(t) = \alpha_0 + \alpha_1 \cos t + \alpha_2 \text{ cost } 2t + \alpha_3 \cos 3t \quad .$$

The problem is to find

$$\min_{x_0,x_1} \max_{t\in[0,\pi]} \left| x_1\cos t - x_0 + \left| x_1\cos t - x_0 \right| -F(t) \right|$$

or in the discrete form

$$\min_{X\in E_2} \phi(X)$$

where

$$\phi(x_0,x_1) = \max_{j\in J} | x_1\cos t_j - x_0 + | x_1\cos t_j - x_0 | - F(t_j) | \, ,$$

$$t_j = \frac{\pi}{N} j, \; J = 0:N,$$

N is a fixea natural number.

The problem is reduced to that of finding a cosinusoidal pulse x_1, and a level for the cut-off of cosinusoid x_0, which guarantee that the periodic pulse constructed is approximated in the best way by a polyharmonic oscillation with given amplitudes of the first, second and third harmonics and a constant component.

For the initial approximation let us choose the solution obtained via Fourier Series. Let $x_1 = 1$, $x_0 = 0.5$ (the cut-off angle $\theta = 60^0$) to which Berg coefficients $\alpha_0 = 0.218$, $\alpha_1 = 0.391$, $\alpha_2 = 0.276$, $\alpha_3 = 0.138$ correspond. Thus

$$F(t_j) = 0.218 + 0.391 \cos t_j + 0.276 \cos 2t_j + 0.138 \cos 3t_j \quad ,$$

$$\phi(x_0, x_1) = \max_{j \in J} \left| x_1 \cos t_1 - x_0 + \left| x_1 \cos t_j - x_0 \right| - F(t_j) \right| \quad .$$

The initial value $\phi(0,5;1) = 0.12074$.

For computational reasons we introduce an ε-subdifferential and an ε-superdifferential of the functions (they are approximations of a subdifferential and a superdifferential)

$$r(x_0, x_1, t) = x_1 \cos t - x_0 + |x_1 \cos t - x_0| - F(t) \quad ,$$

$$s(x_0, x_1, t) = |r(x_0, x_1, t)| \quad .$$

Here $\varepsilon > 0$. We obtain (see [5])

$$\underline{\partial}_\varepsilon r(x_0, x_1, t) = \begin{cases} (2\cos t, -2), \text{ if } x_1 \cos t - x_0 > \frac{\varepsilon}{2} & , \\ (0,0), \text{ if } x_1 \cos t - x_0 < -\frac{\varepsilon}{2} & , \\ \{(2\cos t, -2), (0,0)], \text{ if } -\frac{\varepsilon}{2} \leq x_1 \cos t - x_0 < \frac{\varepsilon}{2} & , \end{cases}$$

$$\overline{\partial}_\varepsilon r(x_0, x_1, t) = (0,0);$$

$$\underline{\partial}_\varepsilon s(x_0,x_1,t) = \begin{cases} \underline{\partial}_\varepsilon r(x_0,x_1,t),\ \text{if}\ r(x_0,x_1,t) > \frac{\varepsilon}{2} , \\ -\overline{\partial}_\varepsilon r(x_0,x_1,t),\ \text{if}\ r(x_0,x_1,t) < -\frac{\varepsilon}{2} , \\ \text{co}\{2\underline{\partial}_\varepsilon r(x_0,x_1,t), -2\overline{\partial}_\varepsilon r(x_0,x_1,t)\}, \\ \qquad\qquad \text{if}\ -\frac{\varepsilon}{2} \le r(x_0,x_1,t) \le \frac{\varepsilon}{2} , \end{cases}$$

$$\overline{\partial}_\varepsilon s(x_0,x_1,t) = \begin{cases} \overline{\partial}_\varepsilon r(x_0,x_1,t),\ \text{if}\ r(x_0,x_1,t) > \frac{\varepsilon}{2} , \\ -\underline{\partial}_\varepsilon r(x_0,x_1,t),\ \text{if}\ r(x_0,x_1,t) < -\frac{\varepsilon}{2} , \\ \overline{\partial}_\varepsilon r(x_0,x_1,t) - \underline{\partial}_\varepsilon r(x_0,x_1,t), \\ \qquad\qquad \text{if}\ -\frac{\varepsilon}{2} \le r(x_0,x_1,t) \le \frac{\varepsilon}{2} . \end{cases}$$

Then

$$\underline{\partial}_\varepsilon \phi(x_0,x_1) = \text{co}\{\underline{\partial}_\varepsilon s(x_0,x_1,t_k) - \sum_{\substack{t_i \in R_\varepsilon(x_0,x_1) \\ t_i \ne t_k}} \overline{\partial}_\varepsilon s(x_0,x_1,t_i) \mid t_k \in R_\varepsilon(x_0,x_1)\},$$

$$\overline{\partial}_\varepsilon \phi(x_0,x_1) = \sum_{t_k \in R_\varepsilon(x_0,x_1)} \overline{\partial}_\varepsilon s(x_0,x_1,t_k) ,$$

where

$$R_\varepsilon(x_0,x_1) = \{t = t_j = \tfrac{\pi}{N} j \mid \phi(x_0,x_1) - s(x_0,x_1,t_j) \le \varepsilon\} .$$

At the initial point $-\overline{\partial}_\varepsilon \phi(0.5,1) \not\subset \underline{\partial}_\varepsilon \phi(0.5,1)$, therefore $X_0 = (0.5;1)$ is not a stationary point. When using the method of ε-steepest descent after 13 steps on a grid having $N = 50$, we obtained point $X_N^* = (0,347541,\ 0.822896)$. At this point $\phi(x_0^*,x_1^*) = s(x_0^*,x_1^*,\ 1.134724) = 0.072292$. Assume $R_\varepsilon(x_0^*,x_1^*) = \{t = t_k \in [0,\pi] \mid \phi(x_0^*,x_1^*) - s(x_0^*,x_1^*,t_k) \le \varepsilon\}$, where t_k is a local maximum of function $s(x_0^*,x_1^*,t)$ with respect to t (between grid points) and take $\varepsilon = 0,0001$. Then $R_\varepsilon(x_0^*,x_1^*) = \{t_1=0, t_2=0.760555, t_3=1.134724\}$. Finally we get

$$\underline{\partial}_\varepsilon \phi(x_0^*,x_1^*) = \text{co}\{\underline{\partial}_\varepsilon s(x_0^*,x_1^*,t_1) - \overline{\partial}_\varepsilon s(x_0^*,x_1^*,t_2) - \overline{\partial}_\varepsilon s(x_0^*,x_1^*,t_3) ,$$

$$\underline{\partial}_{\varepsilon}s(x_0^*,x_1^*,t_2) - \overline{\partial}_{\varepsilon}s(x_0^*,x_1^*,t_1) - \overline{\partial}_{\varepsilon}s(x_0^*,x_1^*,t_3) \quad ,$$

$$\underline{\partial}_{\varepsilon}s(x_0^*,x_1^*,t_3) - \overline{\partial}_{\varepsilon}s(x_0^*,x_1^*,t_1) - \overline{\partial}_{\varepsilon}s(x_0^*,x_1^*,t_2)\} =$$

$$= \mathrm{co}\{[(0.844765,-2.0),(0.0)],\ (3.448907,-4.0) +$$

$$+[(0.844765,-2.0),(0.0)],(+2.0,-2.0)\} \quad ;$$

$$\overline{\partial}_{\varepsilon}\phi(x_0^*,x_1^*) = \overline{\partial}_{\varepsilon}s(x_0^*,x_1^*,t_1) + \overline{\partial}_{\varepsilon}s(x_0^*,x_1^*,t_2) + \overline{\partial}_{\varepsilon}s(x_0^*,x_1^*,t_3) =$$

$$= (-2.0,+2.0) + [(-0.844765,+2.0),(0,0)] \quad ;$$

$$-\overline{\partial}_{\varepsilon}\phi(x_0^*,x_1^*) = (2.0,-2.0) + [(0.844765,-2.0),(0,0)] \quad .$$

Thus at the point $x^* = (0.347541, 0.822896)$ we have

$$-\overline{\partial}_{\varepsilon}\phi(x_0^*,x_1^*) \subset \underline{\partial}_{\varepsilon}\phi(x_0^*,x_1^*) \quad ,$$

i.e. at this point the necessary condition for a minimum of the function $\phi(x_0,x_1)$ (condition (7)) is satisfied (up to ε-accuracy (Fig. 1)).

It is interesting to note that the solution of the inverse problem of finding amplitudes of the three harmonics and a constant component which provide the best approximation of the periodic cosinusoidal pulse of the same form ($x_1 = 1$, $x_0 = 0.5$) leads to the following values of coefficients: $\alpha_0 = 0.20918$, $\alpha_1 = 0.37849$, $\alpha_2 = 0.27542$, $\alpha_3 = 0.18398$; $\phi(\alpha)$ being 0.077876. The comparison shows that the solution given differs essentially from the coefficient determined by applying the Fourier Series and provides a better (in the Chebyshev sense) approximation of the initial function.

Another example relates to the problem of designing amplitude harmonic filters. Let some signal be of the form $u(t) = b_1\cos t + b_2\cos 2t$, $0 < b_1 \leq 1$, $0 < b_2 < 1$, i.e. the signal has the first and second harmonics. It is required to reduce the level of the second harmonic with respect to the first one in the output signal spectrum by choosing the proper transducer parameters. The transducer consists of n diode nonlinear elements and its output signal is

$$f(X,t) = \frac{1}{2} \sum_{i=1}^{n} \alpha_i (u(t) - a_i + | u(t) - a_i | + \Delta \quad .$$

The problem of synthesis is formulated as follows. Determine a vector $X = (a_1, \ldots, a_n, \alpha_1, \ldots, \alpha_n, \Delta)$ which minimizes the function

$$\phi(X) = \max_{t \in [0,\pi]} | f(X,t) - F(t) | \quad ,$$

where $F(t) = b_0 \cos t$, α_i is the characteristic curvature of the i-th diode, α_i is the current cut-off angle of the i-th diode, Δ is the constant component of the output signal. Let $n = 3$, $b_0 = -1$, $b_1 = 1$, $b_2 = 0.2$. The initial approximation was the following: $a_1^0 = -0.9$, $a_2^0 = -0.7$, $a_3^0 = 0$, $\alpha_1^0 = -2.5$, $\alpha_2^0 = 1.5$, $\alpha_3^0 = = 0.3$, $\Delta^0 = 1$. The maximum signal slope for the given case was $\varphi(X) = 0.25$.

By using the ε-steepest descent method the vector $X^* = (a_1^* = -0.822$, $a_2^* = -0.624$, $a_3^* = -0.054$, $\alpha_1^* = -2.608$, $\alpha_2^* = 1.416$, $\alpha_3^* = 0.505$, $\Delta^* = 1.031)$ was obtained and the max-type function was $\phi(X^*) = 0.027$. At this point the sets of sub- and superdifferentials $\underline{\partial}_\varepsilon \phi(X^*)$ and $\overline{\partial}_\varepsilon \phi(X^*)$ represent convex polyhedra having respectively 23 and 4 vertices in 7-dimensional space. Since the distance between the sets $\underline{\partial}_\varepsilon \phi(X^*)$ and $-\overline{\partial}_\varepsilon \phi(X^*)$ is small $(\rho_\varepsilon(X^*) = 0.002)$, the point X^* can be regarded as an ε-stationary one.

The resulting suppression of the second harmonic is easy to determine by representing the found signal F(t) as a Fourier Series. In this example the suppression value amounts to 24 dB. So the transducer considered is in fact a nonlinear harmonic filter. Within the interval where the frequency of the nonlinear element operates, the suppression level does not depend on a frequency.

however, it is necessary to underline that unlike the characteristics of frequency filters the characteristics of amplitude filters are sensitive to the input signal level.

Thus, the examples discussed show that Quasidifferential Calculus enables one to greatly extend the class of electrical circuit problems which can be successfully solved.

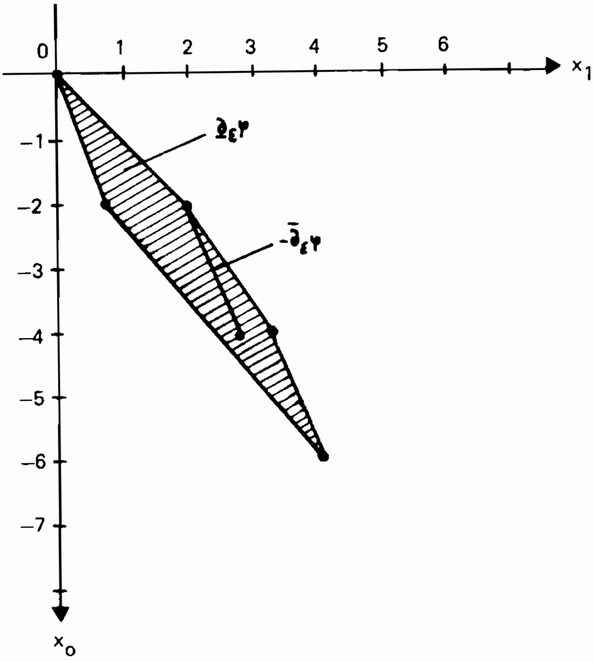

0
1
2
3
4
5
6
x_1
-1
-2
-3
-4
-5
-6
-7
x_o
$\underline{\partial}_\varepsilon \varphi$
$-\overline{\partial}_\varepsilon \varphi$

Fig. I

REFERENCES

1. Remes, E.Y. "The Principles of Chebyshev Approximation Numerical Methods". (In Russian), Kiev, Naukova dumka, 1969).
2. Demyanov, V.F., and B.N. Malozemov, "Introduction to Minimax". (Wiley, New York, 1974).
3. Voiton, E.F. "A Minimax Method of Electrical Circuits Optimization by Absence of Constraints on Variable Parameters". (In Russian), Izvestia VUZOV, radioelektronika, 15(2), (1972).
4. Voiton, E.F. "A Minimax Method of Electrical Circuits Optimization under the Limitation on Variable Parameters". (In Russian), Izvestia VUZOV, radioelektronika, 18(8), (1975).
5. Demyanov, V.F., and L.V. Vasiliev. "Nondifferentiable Optimization". (In Russian), Nauka, Moscow, 1981).

ь. Voiton, E.F. "On Methods for Solving some Extremum Problems of Electrical Circuits Synthesis". (In Russian), Issleaovanie operatziy, Vol. 5, (Computing Centre of the USSR Academy of Sciences, Moscow, 1976).

7. Demyanov, V.F., (ed.), "Nonsmooth Problems of Control Theory and Optimization". (In Russian), (Leningrad Univ. Press, Leningraa, 1982). (see Ch. 1 by Demyanov V.F. and Rubinov A.M.).

THE INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS

is a nongovernmental research institution, bringing together scientists from around the world to work on problems of common concern. Situated in Laxenburg, Austria, IIASA was founded in October 1972 by the academies of science and equivalent organizations of twelve countries. Its founders gave IIASA a unique position outside national, disciplinary, and institutional boundaries so that it might take the broadest possible view in pursuing its objectives:

To promote international cooperation in solving problems arising from social, economic, technological, and environmental change

To create a network of institutions in the national member organization countries and elsewhere for joint scientific research

To develop and formalize systems analysis and the sciences contributing to it, and promote the use of analytical techniques needed to evaluate and address complex problems

To inform policy advisors and decision makers about the potential application of the Institute's work to such problems

The Institute now has national member organizations in the following countries:

Austria
The Austrian Academy of Sciences

Bulgaria
The National Committee for Applied
Systems Analysis and Management

Canada
The Canadian Committee for IIASA

Czechoslovakia
The Committee for IIASA of the
Czechoslovak Socialist Republic

Finland
The Finnish Committee for IIASA

France
The French Association for the
Development of Systems Analysis

German Democratic Republic
The Academy of Sciences of the German
Democratic Republic

Federal Republic of Germany
Association for the Advancement
of IIASA

Hungary
The Hungarian Committee for Applied
Systems Analysis

Italy
The National Research Council

Japan
The Japan Committee for IIASA

Netherlands
The Foundation IIASA–Netherlands

Poland
The Polish Academy of Sciences

Sweden
The Swedish Council for Planning and
Coordination of Research

Union of Soviet Socialist Republics
The Academy of Sciences of the Union
of Soviet Socialist Republics

United States of America
The American Academy of Arts and
Sciences

This series reports new developments in mathematical economics, economic theory, econometrics, operations research, and mathematical systems, research and teaching – quickly, informally and at a high level. The type of material considered for publication includes:

1. Preliminary drafts of original papers and monographs
2. Lectures on a new field or presentations of a new angle in a classical field
3. Seminar work-outs
4. Reports of meetings, provided they are
   a) of exceptional interest and
   b) devoted to a single topic.

Texts which are out of print but still in demand may also be considered if they fall within these categories.

The timeliness of a manuscript is more important than its form, which may be unfinished or tentative. Thus, in some instances, proofs may be merely outlined and results presented which have been or will later be published elsewhere. If possible, a subject index should be included. Publication of Lecture Notes is intended as a service to the international scientific community, in that a commercial publisher, Springer-Verlag, can offer a wide distribution of documents which would otherwise have a restricted readership. Once published and copyrighted, they can be documented in the scientific literature.

Manuscripts

Manuscripts should be no less than 100 and preferably no more than 500 pages in length. On request, the publisher will supply special paper with the typing area outlined and essentials for the preparation of camera-ready manuscripts. Manuscripts should be sent directly to Springer-Verlag Heidelberg or Springer-Verlag New York.

Springer-Verlag, Heidelberger Platz 3, D-1000 Berlin 33
Springer-Verlag, Tiergartenstraße 17, D-6900 Heidelberg 1
Springer-Verlag, 175 Fifth Avenue, New York, NY 10010/USA
Springer-Verlag, 37-3, Hongo 3-chome, Bunkyo-ku, Tokyo 113, Japan

ISBN 3-540-15979-7
ISBN 0-387-15979-7