ZOR – Methods and Models of Operations Research (1990) 34:255–277

On the Numerical Solution of a Class of Stackelberg Problems

By J. V. Outrata[1]

Abstract: This study tries to develop two new approaches to the numerical solution of Stackelberg problems. In both of them the tools of nonsmooth analysis are extensively exploited; in particular we utilize some results concerning the differentiability of marginal functions and some stability results concerning the solutions of convex programs. The approaches are illustrated by simple examples and an optimum design problem with an elliptic variational inequality.

Zusammenfassung: Diese Arbeit zielt auf eine Entwicklung von neuen Verfahren für die numerische Lösung der Stackelbergproblemen. In beiden vorgeschlagenen Verfahren nützt man die Mittel der nichtglatten Analysis aus. Besonders handelt es sich um eine Charakterisierung der verallgemeinerten Gradienten von marginalen Funktionen und einige Stabilitätsergebnisse, die die Lösungen von konvexen Programmen betreffen. Die Verfahren sind durch einfache Beispiele und ein Optimum Design Problem mit einer elliptischen Variationsungleichung illustriert.

Key words: Nondifferentiable optimization, set-valued maps, generalized Jacobians.

1 Introduction

Stackelberg problems play an important role in economic modelling, optimum design and further areas of applied mathematics. They have been studied first from the point of view of the existence and the interpretation of their solutions and the construction of suitable optimality conditions, cf. e.g. Von Stackelberg (1952), Basar and Olsder (1982), Aubin and Ekeland (1984). Later the attention has been devoted also to the questions connected with their numerical solution, e.g. Shimizu and Aiyoshi (1981), Loridan and Morgan (1988). The aim of this work is to apply some results of nonsmooth analysis which, together with suitable optimization routines, would enlarge the

[1] J. V. Outrata, Institute of Information Theory and Automation, Prague. – Prepared while the author was visiting the Department of Mathematics, University of Bayreuth as a guest of the FSP „Anwendungsbezogene Optimierung und Steuerung".

0340–9422/90/4/255–277 $2.50

number of available numerical methods. We confine ourselves to the simplest situation with merely two players: The Leader (L) and the Follower (F). However, an increase of the number of followers would not cause principal difficulties as it is explained in the conclusion. Our problem attains the form

$$\begin{aligned} & f_L(x, y) \to \inf \\ & \text{subject to} \\ & y \in \arg\min_{s \in \Omega(x)} f_F(x, s) \\ & x \in \omega, \end{aligned} \tag{1.1}$$

where

$f_L[\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}]$ is the objective of L,

$f_F[\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}]$ is the objective of F,

$\omega \subset \mathbb{R}^n$ is the set of admissible strategies of L and

$\Omega[\mathbb{R}^n \to 2^{\mathbb{R}^m}]$ specifies the set of admissible strategies of F.

Problems of the type (1.1) can be found very often e.g. in optimum design. Consider e.g. the optimization problem

$$\begin{aligned} & J(x, y) \to \inf \\ & \text{subject to} \\ & Bx \in Ay + N_K(y) \\ & x \in \omega, \end{aligned} \tag{1.2}$$

where $x \in X$ is the *design* (*control*) variable, $y \in Y$ is the *state* variable, $J[X \times Y \to \mathbb{R}]$ is the *optimality criterion*, $N_K(y)$ is the normal cone to a convex set K at y, $A \in L[Y, Y^*]$ is a selfadjoint linear elliptic operator, $B[X \to Y^*]$ is the natural embedding and ω is the *set of feasible controls.* This problem possesses exactly the form (1.1) if one sets $f_L = J$, $f_F = \frac{1}{2}\langle y, Ay\rangle - \langle Bx, y\rangle$ and $\Omega(x) = K$. Indeed, the variational inequality in (1.2) may easily be rewritten into the form

$$\begin{aligned} &\frac{1}{2}\langle y, Ay\rangle - \langle Bx, y\rangle \to \inf \\ &\text{subject to} \\ &y \in K. \end{aligned} \tag{1.3}$$

Problems of the type (1.2) are investigated in Haslinger, Neittaanmäki (1988) from both the theoretical and numerical point of view. We have chosen a concrete problem of this type, having a direct practical interpretation, as the illustration for the both numerical approaches, proposed in the sequel.

Besides design problems of the above type a number of other important application of the Stackelberg model (1.1) may be found e.g. in the market theory, cf. Von Stackelberg (1952), Aubin and Ekeland (1984).

Let us return back to its general formulation. Assume that at any fixed $\bar{x} \in \omega$ the Follower's problem (F-problem)

$$\begin{aligned} &f_F(\bar{x}, y) \to \inf \\ &\text{subject to} \\ &y \in \Omega(\bar{x}) \end{aligned} \tag{1.4}$$

is a convex programming problem which possesses a solution. To be precise, we will henceforth suppose that

(i) f_F is continuous on $\mathbb{R}^n \times \mathbb{R}^m$ and $f_F(x, .)$ is convex over $\mathbb{R}^m$ for all $x \in \omega$;

(ii) Ω is a convex-valued closed-valued map over $\mathbb{R}^n$ and $\arg\min_{s\in\Omega(x)} f_F(x, s) \neq \emptyset$ for x from some open set $\omega' \supset \omega$.

We will also assume that the (generally nonconvex) problem (1.1) possesses a solution which, however, may be in some cases rather problematic, cf. Loridan, Morgan (1988).

Due to assumptions (i), (ii) problem (1.1) may be rewritten into the form

$$
\begin{aligned}
&f_L(x, y) \to \inf \\
&\text{subject to} \\
&0 \in \partial_y f_F(x, y) + N_{\Omega(x)}(y) \\
&x \in \omega,
\end{aligned}
\tag{1.5}
$$

where $\partial_y f_F$ is the partial subdifferential of f_F with respect to y. In this way, the hierarchical structure of (1.1) has been removed. The classical approach to the derivation of optimality conditions relies on this transcription. However, for the numerical solution of (1.1) this transcription is not especially suitable because the constraint structure of (1.5) is hardly tractable even in very simple situations, cf. Basar, Olsder (1982). Generally one has namely to introduce even the multipliers, corresponding to the constraint

$$y \in \Omega(x) \tag{1.6}$$

as unknown variables.

The appearance of multipliers as unknown variables may be in some cases avoided if we apply the method proposed in Shimizu, Aiyoshi (1981). They recommend to augment the constraint $y \in \Omega(x)$ into the F-objective by means of a smooth interior penalty which enables to replace the complicated constraint in (1.5) by the equality $\nabla_y P(x, y) = 0$, where P is the augmented F-objective.

In this paper we intend to study two different approaches which may be considered as alternative to the above mentioned treatments. In the next section we collect some fundamental results concerning marginal functions in mathematical programming which are then utilized in Sect. 4. Sect. 3 is devoted to a special class of problems (1.1) in which the map $x \mapsto \underset{y \in \Omega(x)}{\arg\min}\, f_F(x, y)$ is an operator and its generalized Jacobian may be computed. In this way, we obtain a nonsmooth optimization problem defined only over $x \in \omega$ which represents a substantial decrease of dimensionality. Unfortunately, the assumptions are rather severe so that in most cases we have to apply a more general approach of Sect. 4. This second approach enables us to solve fairly complicated problems, but we have to optimize over $\omega \times \mathbb{R}^m$ as in the metod of Shimizu, Aiyoshi (1981).

For the understanding of the paper a certain basic knowledge of nonsmooth analysis is necessary. We refer the reader to Clarke (1983) or to Aubin, Ekeland (1984), where all the required background is collected. The following notation is employed: $\partial f(x)$ is the generalized gradient or the generalized Jacobian of a function f at x, $\partial_x f(x, y)$ is the partial generalized gradient with respect to x, $\nabla^s f(x)$ is the strict derivative, $N_K(x)$ is the normal cone (in the sense of convex analysis) to a set K at $x \in K$, $\mathrm{cl}\, A$, $\mathrm{conv}\, A$ are the closure, the convex hull of a set A, respectively, Y^* is the topological dual to a linear normed space Y, $L[X, Y]$ is the space of continuous linear operators mapping X into Y, $\mathbb{R}^n_+$ is the nonnegative orthant or $\mathbb{R}^n$, E is the unit matrix, $|\,.\,|_n$ is a norm in $\mathbb{R}^n$, for an $\alpha \in \mathbb{R}$ $(\alpha)^+ = \max\{0, \alpha\}$ and x^j is the j-th coordinate of a vector $x \in \mathbb{R}^n$.

2 Some Preliminary Results

Consider the map Ω given by

$$\Omega(x) = \{y \in \mathbb{R}^m \mid \phi^i(x, y) \leqslant 0,\ i = 1, 2, \ldots, l\}, \tag{2.1}$$

where functions ϕ^i are twice continuously differentiable and functions $\phi^i(x, .)$ are convex for all $x \in \mathbb{R}^n$, $i = 1, 2, \ldots, l$. We may introduce the *marginal function* (*value function*) $h[\omega \to \mathbb{R}]$ in the standard way be the relation

$$h(x) = \inf_{y \in \Omega(x)} f_F(x, y), \tag{2.2}$$

and examine its differentiability properties. If f_F as well as all function ϕ^i are convex, it has been proved in Pschenichnyi (1980) that h is then also convex and for $x \in \omega$ a formula characterizing the subdifferential $\partial h(x)$ is also provided. However, one can only rarely meet Stackelberg problems satisfying such stringent convexity requirements. Therefore, we make use of a recent result of Outrata (1990) which directly implies the assertion of Prop. 2.1 below. To simplify the notation we introduce the map S which assigns to $\bar{x} \in \mathbb{R}^n$ the set of solutions of the F-problem

$$\begin{aligned} & f_F(\bar{x}, y) \to \inf \\ & \text{subject to} \\ & \phi^i(\bar{x}, y) \leqslant 0, \quad i = 1, 2, \ldots, l. \end{aligned} \tag{2.3}$$

In accordance with Rockafellar (1984), under the *inf-boundedness condition* we shall understand the requirement that for each $\bar{x}$ from ω' and each $\alpha \in \mathbb{R}$ there is an $\epsilon > 0$ such that the set of $(x, y) \in \omega' \times \mathbb{R}^m$ satisfying

$$|x - \bar{x}|_n \leqslant \epsilon, \quad f_F(x, y) \leqslant \alpha, \quad \phi^i(x, y) \leqslant \epsilon, \quad i = 1, 2, \ldots, l,$$

is bounded.

Proposition 2.1: Let the inf-boundedness condition hold, $\bar{x} \in \omega, \bar{y} \in S(\bar{x})$ and $U(\bar{x}, \bar{y}) \in \mathbb{R}^l_+$ be the set of Kuhn-Tucker vectors of (2.3) at $\bar{y}$. Assume furthermore that f_F is twice continuously differentiable and the following hypotheses hold:

(H_1) all couples $(\bar{y}, \lambda), \lambda \in U(\bar{x}, \bar{y})$ satisfy the 2-nd order sufficient optimality conditions of Robinson (1980) with a positive modulus;

(H_2) there exists a direction $\xi \in \mathbb{R}^m$ such that $\langle \nabla_y \phi^i(\bar{x}, \bar{y}), \xi \rangle < 0$ for all $i \in I(\bar{x}, \bar{y}) = \{j \in \{1, 2, \ldots, l\} | \phi^j(\bar{x}, \bar{y}) = 0\}$;

(H_3) the gradients $\nabla \phi^i(\bar{x}, \bar{y}), i \in I(\bar{x}, \bar{y})$, are positively linearly independent.

Then h is Lipschitz near $\bar{x}$, regular (in the sense of Clarke) at $\bar{x}$ and

$$\partial h(\bar{x}) = \{\nabla_x f_F(\bar{x}, \bar{y}) + \sum_{i=1}^{l} \lambda^i \nabla_x \phi^i(\bar{x}, \bar{y}) | \lambda \in U(\bar{x}, \bar{y})\}. \tag{2.4}$$

For the proof we refer to Outrata (1990). Note that due to (H_1) the set $S(\bar{x})$ shrinks to a singleton (namely $\bar{y}$). Hypothesis (H_2) is the well-known Mangasarian-Fromowitz constraint qualification. If we replace it by a more stringent requirement that the gradients $\nabla_y \phi^i(\bar{x}, \bar{y})$, $i \in I(\bar{x}, \bar{y})$, are linearly independent, then simultaneously hypothesis (H_3) is satisfied and $U(\bar{x}, \bar{y})$ also shrinks to a singleton – the unique Kuhn-Tucker vector λ. It implies that in this case h is strictly differentiable at $\bar{x}$ $(\nabla^s h(\bar{x}) = \nabla_x f_F(\bar{x}, \bar{y}) + \sum_{i=1}^{l} \lambda^i \nabla_x \phi^i(\bar{x}, \bar{y}))$ which is a well-known result proved in different contexts in various papers.

In many important cases the reather restrictive hypothesis (H_1) does not hold. In such a case we can sometimes utilize a result of Gauvin, Dubeau (1982) stated below.

Proposition 2.2: Let $\bar{x} \in \omega$ and f_F be continuously differentiable. Assume furthermore that the following hypotheses are fulfilled:

(H_4) Ω is uniformly compact near $\bar{x}$ (i.e. there is neighbourhood $\mathcal{O}$ of $\bar{x}$ such that $\text{cl} \bigcup_{x \in \mathcal{O}} \Omega(x)$ is compact);

(H_5) at all $\bar{y} \in S(\bar{x})$ the gradients $\nabla_y \phi^i(\bar{x}, \bar{y})$, $i \in I(\bar{x}, \bar{y})$, are linearly independent.

Then h is Lipschitz near $\bar{x}$, $-h$ is regular at $\bar{x}$ and

$$\partial h(\bar{x}) = \text{conv} \left\{ \nabla_x f_F(\bar{x}, \bar{y}) + \sum_{i=1}^{l} \lambda_{\bar{y}}^i \nabla_x \phi^i(\bar{x}, \bar{y}) \,|\, \bar{y} \in S(\bar{x}) \right\}, \tag{2.5}$$

where $\lambda_{\bar{y}} \in \mathbb{R}_+^l$ is the (unique) Kuhn-Tucker vector of (2.3) at its solution $\bar{y}$.

The proof may be found in Gauvin and Dubeau (1982).

Remark: As hypothesis (H_5) implies that the constraints in (2.3) satisfy the Slater constraint qualification, by using of Thms. 3.1.5 and 4.3.3 of Bank at al. (1982) we can conclude that S is upper semicontinuous at $\bar{x}$ and h is continuous at $\bar{x}$. It implies that the assertion of Prop. 2.2 remains true if we replace (H_4) by the inf-boundedness condition defined above.

Thus, dependantly on the nature of the F-problem (2.3), vectors from the generalized gradient of the marginal function h may be evaluated according to (2.4) or (2.5) provided the appropriate hypotheses hold. We utilize this possibility in Sect. 4 in the quasi-indirect numerical treatment of problem (1.1).

Remark: The assertions of Props. 2.1, 2.2 may also be used to the decomposition of the optimization problem

$$f_F(x, y) \to \inf$$

subject to (2.6)

$$\phi^i(x, y) \leqslant 0, \quad i = 1, 2, \ldots, l$$

$$x \in \omega$$

with respect to variables x and y. Indeed, instead of solving (2.6) over $\omega \times \mathbb{R}^m$ we may minimize the (generally nondifferentiable) marginal function h over ω by means of some nondifferentiable optimization (NDO) method. At each evaluation of $h(x)$ we solve then a convex program of the type (2.3) and the needed "subgradient information" is obtained by using of Props. 2.1, 2.2.

3 Direct Approach

The most elegant way of the numerical treatment of problem (1.1) would be to take it as a nondifferentiable optimal control problem with x being the control variable, y being the state variable and the system map $x \mapsto y$ being given by the set-valued map $S : x \mapsto \arg\min_{y \in \Omega(x)} f_F(x, y)$. To simplify the analysis, let us assume that $f_F(x, .)$ is strictly convex on $\mathbb{R}^m$ for all $x \in \mathbb{R}^n$ which implies that S is an operator. We impose also some further assumptions, namely that f_L is continuously differentiable over $\mathbb{R}^m \times \mathbb{R}^m$ and S is locally Lipschitz and directionally differentiable over ω. (The assumptions guaranteeing the satisfaction of the latter requirement for a concrete form of Ω are imposed below.)

Problem (1.1) may now be rewritten into the form

$$\Theta(x) = f_L(x, S(x)) \to \inf$$

$$\text{subject to} \qquad (3.1)$$

$$x \in \omega.$$

Under the above requirements Θ is locally Lipschitz over ω. It is also directionally differentiable and therefore the chance for a successful implementation of an NDO routine for the minimization of Θ over ω is satisfactory. However, we must be able to compute at any $x \in \omega$ at least one vector $\xi \in \partial\Theta(x)$. Let $P(x)$ be an $[m \times n]$ matrix from $\partial S(x)$, the generalized Jacobian of S at x. Then we may use the following assertion of Hiriart-Urruty (1978).

Proposition 3.1: Let $x \in \omega$ and $y = S(x)$. Then

$$\xi = \nabla_x f_L(x, y) + (P(x))^T \nabla_y f_L(x, y) \in \partial\Theta(x). \qquad (3.2)$$

However, the computation of matrices $P(x)$ is generally highly complicated and available results concern as to our knowledge only special cases, cf. Haslinger, Roubíček (1986), Outrata (1988). The approach we intend to apply here hinges upon a result of Malanowski (1985) (Prop. 3.2 below), and therefore we have to confine ourselves to the case when

$$\Omega(x) = \Omega = \{y \in \mathbb{R}^m \mid \phi^i(y) \leqslant 0,\ i = 1, 2, \ldots, l\}.$$

Recalling from the introduction that $\omega' \supset \omega$ is some open subset of $\mathbb{R}^n$, we impose the following further suppositions:

(i) for each $x \in \omega'$ the function $f_F(x, .)$ is twice continuously differentiable on $\mathbb{R}^m$. Moreover, there exists a constant $\beta > 0$ such that

$$\langle v, \nabla^2_{yy} f_F(x, y) v\rangle \geqslant \beta |v|^2_m \quad \text{for all } x \in \omega',\ y, v \in \mathbb{R}^m;$$

(ii) the functions f_F and $\nabla_y f_F$ are continuously differentiable on $\omega' \times \mathbb{R}^m$;

(iii) for each $x \in \omega'$ the functions ϕ^i, $i = 1, 2, \ldots, l$, are convex and twice continuously differentiable on $\mathbb{R}^m$;

(iv) the gradients $\nabla\phi^i(y)$, $i \in I(y) = \{j \in \{1, 2, \ldots, l\} \mid \phi^i(y) = 0\}$, are linearly independent at any $y = S(x), x \in \omega'$.

Under these assumptions S is indeed locally Lipschitz and directionally differentiable over ω due to the results of Hager (1979) and Jitorntrum (1984). At a fixed $\bar{x} \in \omega$ the F-problem attains the form

$$f_F(\bar{x}, y) \to \inf$$

$$\text{subject to} \tag{3.3}$$

$$\phi^i(y) \leqslant 0, \quad i = 1, 2, \ldots, l.$$

In what follows we employ for the sake of simplicity the following notation:

$$Q(\bar{x}) = \nabla^2_{yy} f_F(\bar{x}, S(\bar{x})) + \sum_{i=1}^{l} \lambda^i \nabla^2 \phi^i(S(\bar{x})),$$

where $\lambda = (\lambda^1, \lambda^2, \ldots, \lambda^l) \in \mathbb{R}^l_+$ is the (unique) Kuhn-Tucker vector of (3.3) at $\bar{y} = S(\bar{x})$,

$$q(\bar{x}, g) = \nabla^2_{yx} f_F(\bar{x}, S(\bar{x}))g,$$

where the vector $g \in \mathbb{R}^n$ will be specified later,

$$\mathbb{I}(\bar{x}) = \{i \in \{1, 2, \ldots, l\} \mid \phi^i(S(\bar{x})) = 0\},$$

$$\mathbb{J}(\bar{x}) = \{i \in \mathbb{I}(\bar{x}) \mid \lambda^i > 0\}$$

and for $\mathbb{L} \subset \{i = 1, 2, \ldots, l\}$

$$L_{\mathbb{L}}(\bar{x}) = \{v \in \mathbb{R}^m \mid \langle \nabla\phi^i(S(\bar{x})), v\rangle = 0, i \in \mathbb{L}\}.$$

Proposition 3.2: Let $\bar{x} \in \omega$ and the operator $\nabla^2_{yx} f_F(\bar{x}, S(\bar{x}))[\mathbb{R}^n \to \mathbb{R}^m]$ be surjective. Then a necessary and sufficient condition for S to be Fréchet differentiable at $\bar{x}$ is the strict complementarity

$$\mathbb{I}(\bar{x}) = \mathbb{J}(\bar{x}).$$

For any direction $g \in \mathbb{R}^n$ the Fréchet differential $\nabla S(\bar{x})g$ is given as the (unique) solution of the quadratic programming problem

$$\frac{1}{2}\langle v, Q(\bar{x})v\rangle + \langle q(\bar{x}, g), v\rangle \to \inf$$

subject to (3.4)

$$v \in L_{\mathbb{I}(\bar{x})}(\bar{x}).$$

If $\mathbb{J}(\bar{x})$ is a proper subset of $\mathbb{I}(\bar{x})$, then for any set of indices $\tilde{J}(\bar{x})$ satisfying the inclusions

$$\mathbb{J}(\bar{x}) \subset \tilde{J}(\bar{x}) \subset \mathbb{I}(\bar{x}) \qquad (3.5)$$

there exists a direction $g \in \mathbb{R}^n$ and a sequence $\{x_i\}$ converging to $\bar{x}$ along $-g$ such that

$$\mathbb{J}(x_i) = \mathbb{I}(x_i) = \tilde{J}(\bar{x}).$$

For the proof see Malanowski (1985). This assertion may directly be used for the construction of matrices from ∂S.

Proposition 3.3: Let $\bar{x} \in \omega$ and the operator $\nabla^2_{yx} f_F(\bar{x}, S(\bar{x}))$ be surjective. Assume that $\tilde{J}(\bar{x})$ satisfies inclusions (3.5) and $P(\bar{x})$ is the operator which assigns to $g \in \mathbb{R}^n$ the (unique) solution of the quadratic programming problem

$$\frac{1}{2}\langle v, Q(\bar{x})v\rangle + \langle q(\bar{x}, g), v\rangle \to \inf$$

$$\text{subject to} \qquad (3.6)$$

$$v \in L_{\tilde{J}(\bar{x})}(\bar{x}).$$

Then $P(\bar{x}) \in \partial S(\bar{x})$.

Proof: As S is Lipschitz near $\bar{x}$, we may approach $\bar{x}$ by points at which S is differentiable. By Prob. 3.2 we know that to any $\tilde{J}(\bar{x})$ satisfying the inclusions (3.5) there exists a direction g and sequence x_i converging to $\bar{x}$ along $-g$ such that $\nabla S(x_i)$ is the operator which assigns to g the solution of the quadratic programming problem

$$\frac{1}{2}\langle v, Q(x_i)v\rangle + \langle q(x_i, g), v\rangle \to \inf$$

$$\text{subject to} \qquad (3.7)$$

$$v \in L_{\tilde{J}(\bar{x})}(x_i).$$

We show that $\nabla S(x_i) \to P(\bar{x})$ with $P(\bar{x})$ being given by (3.6). Indeed, $\nabla S(x_i) = \Xi(x_i) \circ (-\nabla^2_{yx} f_F(x_i, S(x_i)))$, where $\Xi(x_i)$ is the projection operator which projects $(Q(x_i))^{-1}u$, $u \in \mathbb{R}^m$, onto $L_{\tilde{J}(\bar{x})}(x_i)$ in the $Q(x_i)$-metric. Due to the continuity assumptions being imposed, Ξ as well as $\nabla^2_{yx} f_F(., S(.))$ depend continuously on x over ω' so that

$$\nabla S(x_i) \to P(\bar{x}) \in \partial S(\bar{x}) \tag{3.8}$$

by definition. □

Of course, when computing points from $\partial\Theta$, it is not necessary to evaluate the operator Ξ explicitly, but we may proceed directly according to Prop. 3.1.

Proposition 3.4: Let $\bar{x} \in \omega, \tilde{J}(\bar{x})$ satisfy inclusions (3.5) and $\bar{v}$ be the (unique) solution of the quadratic programming problem

$$\frac{1}{2}\langle v, Q(\bar{x})v\rangle - \langle \nabla_y f_L(\bar{x}, S(\bar{x})), v\rangle \to \inf$$

subject to (3.9)

$$v \in L_{\tilde{J}(\bar{x})}(\bar{x}).$$

Then

$$\nabla_x f_L(\bar{x}, S(\bar{x})) - (\nabla^2_{yx} f_F(\bar{x}, S(\bar{x})))^T \bar{v} \in \partial\Theta(\bar{x}). \tag{3.10}$$

Proof: As Ξ is symmetric, $(P(\bar{x}))^T = -(\nabla^2_{yx} f_F(\bar{x}, S(\bar{x}))^T \Xi(\bar{x})$. It remains to apply Prop. 3.1. □

Let us consider a particular problem in which $n \geqslant m$ and

$$f_F(x, y) = \frac{1}{2}\langle y, Hy\rangle - \langle b(x), y\rangle, \tag{3.11}$$

where H is a symmetric positive definite $(m \times m)$ matrix and $b[\mathbb{R}^n \to \mathbb{R}^m]$ is a continuously differentiable operator with ∇b being surjective on ω'. Then $\nabla^2_{yx} f_F(x, y) = -\nabla b(x)$ so that the assumptions are met. For $\bar{x} \in \omega$

$$\nabla_x f_L(\bar{x}, S(\bar{x})) + (\nabla b(\bar{x}))^T \bar{v} \in \partial\Theta(\bar{x}),$$

provided $\bar{v}$ is the solution of the quadratic programming problem (3.9) with $Q(\bar{x}) = H + \sum_{i=1}^{l} \lambda^i \nabla^2 \phi^i(S(\bar{x}))$ and an index set $\tilde{J}(\bar{x})$ being appropriately chosen.

As we have pointed out in the introduction, Stackelberg problems with f_F given by (3.11) result from a discretization of optimum design problems with elliptic variational inequalities. To test the proposed numerical approach we have solved firstly five simple academic examples of this form and then a larger optimum design problem taken from Haslinger, Neittaanmäki (1988). In the first five problems we have $n = m = 2$, $\omega = \mathbb{R}^2$,

$$f_L = \frac{1}{2}(y^1)^2 + \frac{1}{2}(y^2)^2 - 3y^1 - 4y^2 + r[(x^1)^2 + (x^2)^2],$$

where $r > 0$ varies, and

$$\Omega = \{(y^1, y^2) \in \mathbb{R}^2_+ \mid -0.333y^1 + y^2 \leqslant 2, y^1 - 0.333y^2 \leqslant 2\}.$$

Example 1: $r = 0.1, H = \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix}, b(x) = x.$

Starting point: $x^1 = 0, x^2 = 0.$

Solution: $x^1 = 0.97, x^2 = 3.14, y^1 = 2.6, y^2 = 1.8, f_L = 3.58.$

Example 2: $r = 1, H$ and b like in Ex. 1.

Starting point: $x^1 = 0, x^2 = 0.$

Solution: $x^1 = 0.28, x^2 = 0.48, y^1 = 2.34, y^2 = 1.03, f_L = 4.92.$

Starting point: $x^1 = 10, x^2 = 10.$

Solution: $x^1 = 0, x^2 = 0, y^1 = 0, y^2 = 0, f_L = 12.5.$

Example 3: $r = 0, H = \begin{bmatrix} 1 & 3 \\ 3 & 10 \end{bmatrix}, b(x) = x.$

Starting point: $x^1 = 0, x^2 = 0.$

Solution: $x^1 = 3, x^2 = -0.14, y^1 = 2, y^2 = 0, f_L = 8.5.$

Starting point: $x^1 = 10, x^2 = 10.$

Solution: $x^1 = 20.26, x^2 = 42.81, y^1 = 3, y^2 = 3, f_L = 0.5.$

Example 4: $r = 0.1, H$ and b like in Ex. 3.

Starting point: $x^1 = -3, x^2 = 6$.

Solution: $x^1 = 2, x^2 = 0.06, y^1 = 2, y^2 = 0, f_L = 8.9$.

Example 5: r and H like in Ex. 4, $b(x) = \begin{bmatrix} -1 & 2 \\ 3 & -3 \end{bmatrix} x$.

Starting point: $x^1 = -3, x^2 = 6$.

Solution: $x^1 = 2.42, x^2 = -3.65, y^1 = 0, y^2 = 1.58, f_L = 9.35$.

In all examples we have to solve two quadratic programming problems at each computation of a vector from $\partial\Theta$; the first corresponds to the solution of the F-problem (3.3) at the current x, the second is the appropriate problem (3.9). To this purpose the code SOL/QPSOL of Gill and al. has been applied, for the unconstrained minimization of Θ the NDO code M1FC1 of Cl. Lemaréchal has been used. In all examples we needed less that 10 iterations of the bundle algorithm.

Consider now the problem (1.2) in which the design (control) space $X = L_2(0, 1)$, the state space $Y = \mathring{H}^2(0, 1)$,

$$J(x, y) = \int_0^1 (y(s) - g(s))^2 ds, \tag{3.12}$$

where the function $g \in H^2(0, 1)$ is given, $K = \{y \in Y | y(s) \geqslant g(s) \text{ on } [0, 1]\}, Ay = y''''$, B is the natural embedding of $L_2(0, 1)$ into $H^{-2}(0, 1)$ and

$$\omega = \{x \in L_\infty(0, 1) | \alpha \leqslant x(s) \leqslant \beta \text{ for a.a. } s \in [0, 1], \int_0^1 x(s) ds = M\}, \tag{3.13}$$

where the nonnegative scalars α, β, M are given.

The elliptic variational inequality $Bx \in Ay + N_K(y)$ assigns in this case to a given load density x the deflection y of a clamped beam limited from below by a rigid obstacle described by the function g. The aim of the optimization is to find such a load density x that maximizes the contact area between the beam and the obstacle.

The discretization has been performed by dividing of [0, 1] into ten equidistant subintervals of the length $d = 1/10$, and representing X by means of functions, constant on each subinterval and Y by means of third order polynomials on each subinterval. In this way we have obtained a Stackelberg problem of the form (1.1) in which $n = 10$, $m = 18, x = (x_1, x_2, \ldots, x_{10})$ gives us the load density and the odd, even coordinates of y approximate the beam deflection and its derivative at the grid points, respectively. f_L corresponds to a chosen rule for the quadrature of (3.12) (we have chosen the trapezoidal rule), f_F is given by (3.11) with the *rigidity* matrix H and the map b being given by the used discretization,

$$\omega = \left\{ x \in \mathbb{R}^{10} \,|\, \alpha \leqslant x^i \leqslant \beta,\ 1/10 \sum_{i=1}^{10} x^i = M \right\}$$

and

$$\Omega = \{ y \in \mathbb{R}^{18} \,|\, y^{2i-1} \geqslant g(i/10),\ i = 1, 2, \ldots, 9 \}.$$

We have set $\alpha = 0, \beta = 10, M = 5, g \equiv -0.001$, the product of the Young modul and the cross-sectional moment of innertia to be equal one and applied the proposed direct approach with the numerical code M2FC1 of Cl. Lemaréchal. We had, however, to augment the constraint $1/10 \sum_{i=1}^{10} x^i = M$ to the objective by means of an exact penalty because this code can handle only the box constraints $\alpha \leqslant x^i \leqslant \beta$ directly within the optimization procedure. With the initial guess $x^i = \beta$ for all i and the final accuracy being specified by $\epsilon = 0.001$ we have obtained the solution after 7 iterations (17 function evaluations). The results are depicled on Fig. 1.

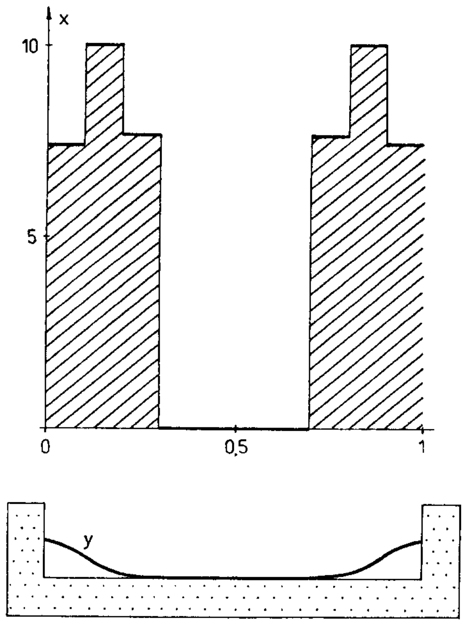

Fig. 1

Remark: Due to the discretization being chosen b is not surjective so that relation (3.10) need not be true. This obstacle is of a purely technical nature and may easily be overcome by using a different discretization. However, as we wanted to compare our results with those of Haslinger, Neittaanmäki (1988), we have used their discretization and the proposed direct approach proved itself to be sufficiently robust.

4 Quasi-Indirect Approach

In this section we propose a numerical approach to the problem (1.1) without any substantial restrictions, needed in the previous section. It uses a special form of optimality conditions which justifies the adjective quasi-indirect. Problem (1.1) can namely be rewritten into the form

$$\begin{aligned} & f_L(x, y) \to \inf \\ & \text{subject to} \\ & f_F(x, y) - \inf_{s \in \Omega(x)} f_F(x, s) \leq 0 \\ & y \in \Omega(x) \\ & x \in \omega, \end{aligned} \tag{4.1}$$

where we have a generally nondifferentiable function appearing in an inequality constraint. We will assume that f_L is regular locally Lipschitz and distinguish between the following two situations:

(α) $\Omega(x) = \Omega$ is a nonempty compact subset of $\mathbb{R}^m$ not depending on x, $f_F(.,y)$ is continuously differentiable on some open set $\omega' \supset \omega$ for all $y \in \mathbb{R}^m$ and $\nabla_x f_F$ is continuous as a function of (x, y).

Then it is a consequence of the calculus of generalized gradients that the function

$$g(x, y) = f_F(x, y) - \inf_{s \in \Omega} f_F(x, s) = f_F(x, y) + \sup_{s \in \Omega} (-f_F(x, s))$$

is regular locally Lipschitz over $\mathbb{R}^n \times \mathbb{R}^m$ and

$$\begin{bmatrix} \nabla_x f_F(x, y) - \nabla_x f_F(x, z) \\ \\ \mu \end{bmatrix} \in \partial g(x, y), \tag{4.2}$$

provided $z \in S(x) = \arg\min_{s \in \Omega} f_F(x, s)$ and $\mu \in \partial_y f_F(x, y)$. If f_F is continuously differentiable on $\omega' \times \mathbb{R}^m$ and attains its infinum with respect to y on Ω at a single point for all $x \in \omega'$, then g is in fact continuously differentiable on $\omega' \times \mathbb{R}^m$.

Let

$$\Omega = \{y \in \mathbb{R}^n \mid \phi^i(y) \leqslant 0, i = 1, 2, \ldots, l\}, \tag{4.3}$$

where functions $\phi^i, i = 1, 2, \ldots, l$, are convex and continuous over $\mathbb{R}^m$.

If f_L, g are continuously differentiable on $\omega' \times \mathbb{R}^m$ and all functions ϕ^i are continuously differentiable on $\mathbb{R}^m$, then some standard (smooth) method can be applied to the nonlinear programming problem

$$\begin{aligned} & f_L(x, y) \to \inf \\ & \text{subject to} \\ & g(x, y) \leqslant 0 \\ & \phi^i(y) \leqslant 0, \quad i = 1, 2, \ldots, l \\ & x \in \omega \end{aligned} \tag{4.4}$$

with ∇g computed according to (4.2).

Provided either f_L or g or at least one of the functions ϕ^i is nondifferentiable and problem (4.4) is calm with respect to vertical perturbations of constraints $g(x, y) \leqslant 0$, $\phi^i(y) \leqslant 0$, $i = 1, 2, \ldots, l$ (cf. Clarke 1983), we recommend to augment all inequality constraints to the objective by means of exact nondifferentiable penalties. We arrive then at the problem

$$f_L(x, y) + r_0(g(x, y))^+ + \sum_{i=1}^{l} r_i(\phi^i(y))^+ \to \inf$$

subject to (4.5)

$$x \in \omega$$

which may be solved by some NDO method provided the structure of ω is sufficiently simple and $r_i > 0, i = 0, 1, \ldots, l$, are suitably chosen penalty parameters.

(β) $\Omega(x)$ is given by (2.1) and either the assumptions of Prop. 2.1 or the assumptions of Prop. 2.2 hold for all $\bar{x} \in \omega$.

Then we may use the respective assertion and conclude that in both cases the function

$$G(x, y) = f_F(x, y) - \inf_{s \in \Omega(x)} f_F(x, s) \tag{4.6}$$

is locally Lipschitz directionally differentiable on $\omega \times \mathbb{R}^m$ and

$$\xi = \begin{bmatrix} \nabla_x f_F(x, y) - \nabla_x f_F(x, z) - \sum_{i=1}^{l} \lambda^i \nabla_x \phi^i(x, z) \\ \nabla_y f_F(x, y) \end{bmatrix} \in \partial G(x, y) \tag{4.7}$$

provided $z \in S(x)$ and $\lambda \in U(x, z)$. Note that $S(x)$ is a singleton in the case of Prop. 2.1 and $U(x, z)$ is a singleton in the case of Prop. 2.2.

From the available NDO software only some methods from Kiwiel (1985) can directly be applied to problems with nondifferentiable inequality constraints. For the application of most NDO codes one has to assume that the mathematical program

$$f_L(x, y) \to \inf$$

subject to (4.8)

$$G(x, y) \leqslant 0$$

$$\phi^i(x, y) \leqslant 0, \quad i = 1, 2, \ldots, l$$

$$x \in \omega$$

is calm with respect to vertical perturbations of constraints $G(x, y) \leqslant 0$, $\phi^i(x, y) \leqslant 0$, $i = 1, 2, \ldots, l$. We can then augment these inequalities to the objective by means of exact nondifferentible penalties and arrive at the problem

$$f_L(x, y) + r_0(G(x, y))^+ + \sum_{i=1}^{l} r_i(\phi^i(x, y))^+ \to \inf$$

subject to (4.9)

$$x \in \omega,$$

where $r_i > 0, i = 0, 1, \ldots, l$, are suitably chosen penalty parameters. Provided ω is given by lower and upper bounds for single coordinates, e.g. the code M2FC1 of Cl. Lemaréchal, mentioned already in the previous section, may be applied. Vectors from the generalized gradient of the augmented objective in (4.9) may be computed by using of (4.7) and the standard calculus rules of Clarke (1983) except of the following situation: G is positive and nondifferentiable at (x, y) and there exists an index $i \in \{1, 2, \ldots, l\}$ for which $\phi^i(x, y) = 0$. Then, namely, the sum of points from the generalized gradients $\partial G(x, y)$ and $\partial\phi^i(x, y)$ need not belong to $\partial(G + \phi^i)(x, y)$ whenever G is nonregular. Such a "subgradient information" could theoretically mislead the algorithm, but if G is differentiable at the solution of (4.1), we do not expect any difficulties caused by this obstacle.

If for all x from some open set $\omega' \supset \omega$ the assumptions of Prop. 2.2 hold and the F-problems possess unique solutions, then G is in fact continuously differentiable on ω' (Clarke 1983) and any standard (smooth) method can be applied to the nonlinear programming problem (4.8) whenever f_L is continuously differentiable. ∇G may be be computed by means of (4.7).

Remark: In some concrete problems of the type (β) it may be difficult to check, whether for all $x \in \omega$ indeed $\Omega(x) \neq \emptyset$. Constraints arising from this requirement are termed *induced constraints* and may cause substantial difficulties.

The approach was used to solve the following simple test examples with increasing complexity. In all of them $x \in \mathbb{R}_+$, $y \in \mathbb{R}^2$, $f_L = \frac{1}{2}[(y^1 - 3)^2 + (y^2 - 4)^2]$ and

$$f_F = \frac{1}{2}\langle y, H(x)y\rangle - \langle b(x), y\rangle, \quad b(x) = \begin{bmatrix} 3 + 1.333x \\ x \end{bmatrix},$$

where the $(m \times m)$ matrix of continuously differentiable functions $H(x)$ varies. Also the starting point was in all of them but Ex. 2 the same: $x = y^1 = y^2 = 0$.

Example 1: $H(x) = E$, $\Omega(x) = \Omega =$
$\{(y^1, y^2) \in \mathbb{R}^2_+ \mid -0.333y^1 + y^2 \leqslant 2, y^1 - 0.333y^2 \leqslant 2\}$.

Solution: $x = 2.07, y^1 = 3, y^2 = 3, f_L = 0.5$.

Example 2: $H(x) = \begin{bmatrix} 1+x & 0 \\ 0 & 0 \end{bmatrix}$, $\Omega(x)$ like in Ex. 1.

Starting point: $x = 5, y^1 = 0, y^2 = 0$.

Solution: $x = 0, y^1 = 3, y^2 = 3, f_L = 0.5$.

Example 3: $H(x) = \begin{bmatrix} 1+x & 0 \\ 0 & 1+0.1x \end{bmatrix}$, $\Omega(x)$ like in Ex. 1.

Solution: $x = 3.456, y^1 = 1.707, y^2 = 2.569, f_L = 1.859$.

Example 4: $H(x) = E$, $\Omega(x) =$
$\{(y^1, y^2) \in \mathbb{R}^2_+ \mid (-0.333 + 0.1x)y^1 + y^2 \leqslant x, y^1 + (-0.333 - 0.1x)y^2 \leqslant 2\}$.

Solution: $x = 2.498, y^1 = 3.632, y^2 = 2.8, f_L = 0.919$.

Example 5: $H(x) = \begin{bmatrix} 1+x & 0 \\ 0 & 1 \end{bmatrix}$, $\Omega(x)$ like in Ex. 4.

Solution: $x = 3.999, y^1 = 1.665, y^2 = 3.887, f_L = 0.897$.

Example 6: $H(x) = \begin{bmatrix} 1+0.2x & 0 \\ 0 & 1+0.1x \end{bmatrix}$, $\Omega(x) = \{(y^1, y^2) \in \mathbb{R}^2_+ \mid (-0.333$
$+ 0.1x)y^1 + y^2 \leqslant 2 - 0.1x, y^1 + (-0.333 - 0.1x)y^2 \leqslant 2 - 0.1x\}$.

Solution: $x = 1.909, y^1 = 2.979, y^2 = 2.232, f_L = 1.562$.

All examples have been solved first by the NDO code M1FC1 of Cl. Lemaréchal with penalties. As the functions g in Ex. 1, 2, 3 and functions G in Ex. 4, 5, 6 are continuously differentiable, we could apply also the sequential quadratic programming code NLPQL (Schittkowski 1985) to the appropriate problems (4.4), (4.8). The inner quadratic programming problems have again been solved by the SOL/QPSOL code of Gill and al. M1FC1 needed approximately 2–3 times more function evaluations than NLPQL. Maximal number of iterations of M1FC1 was 22.

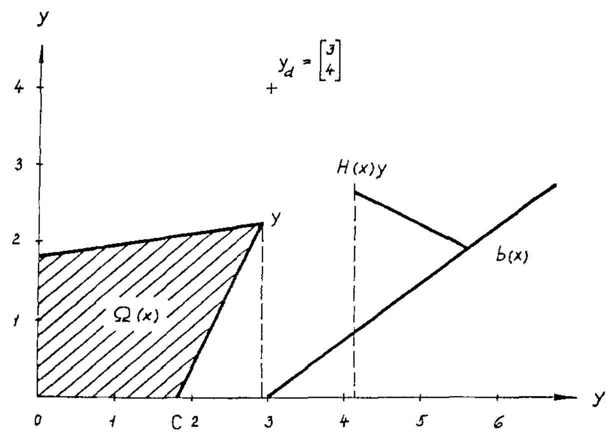

Fig. 2

The geometric situation of Ex. 6 at its solution is depicted on Fig. 2. Observe that the line segment $[H(x)y, b(x)]$ is perpendicular to the edge $[C, y]$ of $\Omega(x)$ so that

$$H(x)y - b(x) \in -N_{\Omega(x)}(y)$$

and hence y is indeed the solution of the F-problem at x.

Besides the above academic examples we have applied the quasi-indirect approach also to the design problem with the elliptic variational inequality, described in the previous section. We have treated the appropriate inequality $g(x, y) \leqslant 0$ as well as the equality $1/10 \sum_{i=1}^{10} x^i = M$ by means of exact penalties, arriving thus at the augmented objective

$$J_d(y) + r_1(g(x, y))^+ + r_2 \left| 1/10 \sum_{i=1}^{10} x^i - M \right|,$$

where J_d is the discretized original objective and r_1, r_2 are positive penalty parameters. The constraints $\alpha \leqslant x^i \leqslant \beta$, $i = 1, 2, \ldots, 10$, and $y^{2j-1} \geqslant -0.001$, $j = 1, 2, \ldots, 9$, have been treated directly within the used NDO code M2FC1. However, we have needed 106 iterations to obtain results comparable with those computed by the direct approach and the final accuracy specified by $\epsilon = 0.001$ has not been achieved. The reason lies probably in a very bad scaling of this problem (cf. Haslinger, Neittaanmäki 1988) which makes a proper choice of the penalty parameters r_1, r_2 extremely difficult.

5 Conclusion

In both approaches being proposed we have a certain outer minimization procedure for which the information about the minimized function is provided by solving of another inner optimization problem(s). The solution procedure of the inner problem(s) cannot be too time-consuming and must be sufficiently accurate. Therefore, we have in fact to confine ourselves to those inner problems that are either linear or quadratic. In the simple academic examples of this paper as well as in the solved optimum design problem these inner optimizations are convex quadratic programming problems and there have been difficulties neither with the time-consumption nor with the accuracy. On the other hand, the outer problems are generally nonconvex and especially function Θ from Sect. 3 may possess a lot of stationary points even in the simplest examples. This phenomenon is well illustrated by Examples 1 and 3 (Sect. 3).

Let us conclude with a remark concerning a possible extension of the quasi-indirect approach to Stackelberg problems with more, let us say, two followers F_1, F_2 playing according to the Nash solution concept between themselves.

Under the appropriate assumptions the quasi-indirect approach combined with an exact penalization technique leads to the optimization problem

$$f_L(x, y_1, y_2) + R_1 |y_1 - v_1|_{m_1} + R_2 |y_2 - v_2|_{m_2} \to \inf$$

subject to (5.1)

$$G_1(x, y_1, y_2, v_2) = f_{F_1}(x, y_1, y_2) - \inf_{\{s \mid (s, v_2) \in \Omega(x)\}} f_{F_1}(x, s, v_2) \leqslant 0$$

$$G_2(x, y_1, y_2, v_1) = f_{F_2}(x, y_1, y_2) - \inf_{\{s \mid (v_1, s) \in \Omega(x)\}} f_{F_2}(x, v_1, s) \leqslant 0$$

$$x \in \omega,$$

where R_1, R_2 are some positive penalty parameters, v_1, v_2 are auxiliary variables, enabling us to decouple the followers' problems and the other symbols have an analogous meaning as in the previous text. To the solution of (5.1) a suitable NDO method can be applied with functions G_1, G_2 treated as shown in Sect. 4. The dimensionality has, unfortunately, further increased due to the introduction of variables v_1, v_2.

Acknowledgement: I would like to express my deep thanks to Prof. Cl. Lemaréchal, Prof. J. Zowe, Prof. K. Schittkowski and Prof. J. Haslinger for their important help during my work on this subject.

References

Aubin JP, Ekeland I (1984) Applied nonlinear analysis. Wiley, New York

Bank B, Guddat J, Klatte D, Kummer B, Tammer K (1982) Non-linear parametric optimization. Akademie-Verlag, Berlin

Basar T, Olsder GJ (1982) Dynamic noncooperative game theory. Academic Press, London

Clarke FH (1983) Optimization and nonsmooth analysis. Wiley, New York

Gauvin J, Dubeau F (1982) Differential properties of the marginal function in mathematical programming. Math Prog Study 19:101–119

Hager WW (1979) Lipschitz continuity for constrained processes. SIAM J Contr Optim 17:321–338

Haslinger J, Neittaanmäki P (1988) Finite element approximation for optimal shape design. Theory and applications. Wiley, Chichester

Haslinger J, Roubíček T (1986) Optimal control of variational inequalities. Approximation and numerical realization. Appl Math Optim 14:187–201

Hiriart-Urruty JB (1978) Gradients generalises de fonctions marginales. SIAM J Contr Optim 16: 301–316

Jittorntrum K (1984) Solution point differentiability without strict complementarity in nonlinear programming. Math Prog Study 21:127–138

Kiwiel KC (1985) Methods of descent for nondifferentiable optimization. Lecture Notes Math, vol 1133. Springer-Verlag, Berlin

Lemaréchal C, Strodiot JJ, Bihain A (1980) On a bundle algorithm for nonsmooth optimization. NPS 4, Madison

Loridan P, Morgan J (1988) Approximate solutions for two-level optimization problems. International Series of Numerical Mathematics, vol 84. Birkhäuser, Basel, pp 181–196

Malanowski K (1985) Differentiability with respect to parameters of solutions to convex programming problems. Math Prog 33:352–361

Outrata JV (1988) On the usage of bundle methods in optimal control of nondifferentiable systems. In: Hoffman KH, Hiriart Urruty JB, Lemaréchal C, Zowe J (eds) Trends in mathematical optimization. Birkhäuser, Basel, pp 233–246

Outrata JV (1990) On generalized gradients in optimization problems with set-valued constraints. To appear in Math Oper Res

Pschenichnyi BN (1980) Convex analysis and extremal problems. Nauka, Moscow (in Russian)

Robinson SM (1980) Generalized equations and their solutions. Part II: Application to nonlinear programming. Univ. Wisconsin-Madison, Technical Summary Rp. #2048

Rockafellar RT (1984) Directional differentiability of the optimal value function in a nonlinear programming problem. Math Prog Study 21:213–226

Schittkowski K (1985) NLPQL: A Fortran subroutine solving constrained nonlinear programming problems. Annals Oper Res: 485–500

Shimizu K, Aiyoshi E (1981) A new computational method for Stackelberg and min-max problems by use of a penalty method. IEEE Trans Autom Contr AC-26:460–466

Von Stackelberg H (1952) The theory of market economy. Oxford University Press, Oxford

Received August 1988
Revised version received May 1989