Jiří V. Outrata

# A note on the usage of nondifferentiable exact penalties in some special optimization problems

*Kybernetika*, Vol. 24 (1988), No. 4, 251--258

Persistent URL: `http://dml.cz/dmlcz/124419`

KYBERNETIKA — VOLUME 24 (1988), NUMBER 4

# A NOTE ON THE USAGE OF NONDIFFERENTIABLE EXACT PENALTIES IN SOME SPECIAL OPTIMIZATION PROBLEMS

JIŘÍ V. OUTRATA

The usage of exact nondifferentable penalties for the numerical solution of optimization problems with a special constraint structure is recommended. Vectors from generalized gradients of appropriate objectives are computed so that effective nondifferentiable minimization methods can be applied.

## 1. INTRODUCTION

Based on the connection of an exact penalization technique with nondifferentiable optimization (NDO) methods we propose a numerical approach for the treatment of special inequality constraints involving min-terms.

In the next section we study constraints of the form

$$\psi(x, y) = f_2(x, y) - \min_{s \in \Omega} f_2(x, s) \leqq 0 , \tag{1.1}$$

where $f_2[\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}]$ is continuously differentiable with respect to $x$, convex continuous with respect to $y$ and $\nabla_x f_2$ is continuous on $\mathbb{R}^n \times \mathbb{R}^m$. $\Omega \subset \mathbb{R}^m$ is assumed to be nonempty, convex and compact. Such constraints arise if we solve optimization problems of the form

$$f_1(x, y) \to \inf \tag{1.2}$$

subj. to

$$y \in \operatorname*{argmin}_{s \in \Omega} f_2(x, s) \tag{1.3}$$

and replace relation (1.3) by the optimality condition

$$\psi(x, y) \leqq 0 , \quad y \in \Omega . \tag{1.4}$$

Problems (1.2)–(1.3) are termed Stackelberg problems and occur frequently in economic modelling or optimum design problems, cf. [2], [6].

Section 3 is devoted to constraints of the form

$$(1.5) \qquad \beta(x) = \min_{i=1,\dots,m} \{q^i(x)\} \leqq 0$$

or

$$(1.6) \qquad \tilde{\beta}(x) = \min_{s \in \varkappa} Q(x, s) \leqq 0\,,$$

where functions $q^i[\mathbb{R}^n \to \mathbb{R}]$, $i = 1, 2, \dots, m$, are continuously differentiable on $\mathbb{R}^n$, function $Q[\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}]$ is continuously differentiable on $\mathbb{R}^n \times \mathbb{R}^m$ and convex with respect to $s$ and $\varkappa \subset \mathbb{R}^m$ is nonempty, convex and compact. Constraints of the type (1.5) arise mostly due to a combinatorial structure in the problem in question. Semi-infinite constraint (1.6) may appear in some CAD problems or special control problems.

For the understanding of the paper a certain basic knowledge of nonsmooth analysis is required. We refer the reader to Chapter 2 of [1]. The following notation is employed:

$\partial f(x)$ is the generalized gradient of a function $f$ at $x$, $\partial_x f(x, y)$ is the partial generalized gradient with respect to $x$, for an $\alpha \in \mathbb{R}$ $(\alpha)^+ = \max\{0, \alpha\}$, $\mathbb{R}^n_+$ is the nonnegative orthant of $\mathbb{R}^n$, $x^j$ is the $j$th coordinate of a vector $x \in \mathbb{R}^n$ and $E$ is the unit matrix.

## 2. STACKELBERG PROBLEMS

We will assume that in problem (1.2)–(1.3) the function $f_1[\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}]$ is regular (in the sense of Clarke), locally Lipschitz on $\mathbb{R}^n \times \mathbb{R}^m$ and

$$(2.1) \qquad \Omega = \{y \in \mathbb{R}^m \mid \Phi^i(y) \leqq 0,\ i = 1, 2, \dots, k\}\,,$$

where the functions $\Phi^i[\mathbb{R}^m \to \mathbb{R}]$ are convex continuous. Under the assumptions being imposed (cf. [1]) $\psi$ is locally Lipschitz. The inner optimization problem $\min_{s \in \Omega} f_2(x, s)$ possesses a solution so that constraints (1.4) are consistent. Hence, the existence of a solution $(\hat{x}, \hat{y})$ of problem (1.2)–(1.3) may be guaranteed by some coercivity assumption on $f_1$ with respect to $x$ or by adding an additional constraint

$$(2.2) \qquad x \in \omega \subset \mathbb{R}^n\,,$$

where $\omega$ is nonempty and compact. Throughout this section it is assumed that a solution $(\hat{x}, \hat{y})$ exists.

Let us assume that the rewritten problem

$$(2.3) \qquad \begin{gathered} f_1(x, y) \to \inf \\ \text{subj. to} \\ \psi(x, y) \leqq 0\,, \quad y \in \Omega \end{gathered}$$

is calm at its solution $(\hat{x}, \hat{y})$ with respect to vertical perturbations of the constraint $\psi(x, y) \leqq 0$. Then it has been proved in [1] that there exists a positive scalar $r_0$

such that for $r \geqq r_0$ the function

$$\Theta = f_1 + r(\psi)^+ \tag{2.4}$$

attains its minimum over $\mathbb{R}^n \times \Omega$ at $(\hat{x}, \hat{y})$. Hence, we may solve instead of (1.2)−(1.3) the augmented problem

$$\begin{aligned} &\Theta(x, y) \to \inf \\ &\text{subj. to} \\ &y \in \Omega \end{aligned} \tag{2.5}$$

with a suitably chosen penalty parameter $r > 0$. $\Theta$ is nondifferentiable so that for its numerical solution an NDO method is needed. Then, under the appropriate calmness assumption with respect to vertical perturbations of constraints $\Phi^i(x) \leqq 0$, $i = 1, 2, \ldots, k$, we may handle also the constraint $y \in \Omega$ by the same technique, arriving thus at the unconstrained minimization problem

$$\begin{aligned} &\tilde{\Theta}(x, y) = f_1(x, y) + r(\psi(x, y))^+ + \sum_{i=1}^{k} r_i(\Phi^i(y))^+ \to \inf \\ &\text{subj. to} \\ &(x, y) \in \mathbb{R}^n \times \mathbb{R}^m, \end{aligned} \tag{2.6}$$

where $r_i$, $i = 1, 2, \ldots, k$, are positive penalty parameters. The objective $\tilde{\Theta}$ is locally Lipschitz and directionally derivable; hence the chance for a successful implementation of an NDO routine is satisfactory. However, if we want to use a bundle or subgradient algorithm, we must be able to compute at any pair $(x, y)$ one arbitrary vector from $\partial\tilde{\Theta}(x, y)$.

**Proposition 2.1.** $\Theta$ is regular on $\mathbb{R}^n \times \mathbb{R}^m$ and one has

$$\begin{bmatrix} \xi \\ \eta \end{bmatrix} + r \begin{bmatrix} \nabla_x f_2(x, y) & -\nabla_x f_2(x, z) \\ \mu & \end{bmatrix} \in \partial\Theta(x, y) \tag{2.7}$$

provided $(\xi, \eta) \in \partial f_1(x, y)$, $z \in \arg\min_{s \in \Omega} f_2(x, s)$, $\mu \in \partial_y f_2(x, y)$ and $\psi(x, y) \geqq 0$. If $\psi(x, y) \leqq 0$, then

$$\begin{bmatrix} \xi \\ \eta \end{bmatrix} \in \partial\Theta(x, y)\,. \tag{2.8}$$

Proof. $\Theta = f_1 + rg \circ \psi$, where $g = (\cdot)^+$. $f_1$ is regular by assumption and $\psi(x, y) = f_2(x, y) + \sup_{s \in \Omega}(-f_2(x, s))$ so that it is also regular due to the assumptions being imposed, cf. Th. 2.8.2 of [1]. For any $\alpha \in \mathbb{R}$ $\partial g(\alpha) \subset \mathbb{R}_+$ which implies that $g \circ \psi$ is regular and

$$\gamma \begin{bmatrix} \lambda \\ \gamma \end{bmatrix} \in \partial(g \circ \psi)(x, y)\,,$$

whenever $\gamma \in \partial g(\psi(x, y))$ and $(\lambda, \gamma) \in \partial\psi(x, y)$ because of Chain Rule I of Clarke. $\lambda$ may be computed according to Cor. 2 of the above mentioned Th. 2.8.2, meanwhile the computation of $v$ and $\gamma$ is trivial. The assertion has been proved. □

The same argumentation implies also the regularity of penalty terms $(\Phi^i(y))^+$, $i = 1, 2, \ldots, k$, and the validity of relations

$$\vartheta_i \in \partial((\Phi^i(y))^+) , \tag{2.9}$$

where

$$\vartheta_i \in \partial\Phi^i(y) \quad \text{if} \quad \Phi^i(y) > 0$$

$$\vartheta_i = 0 \qquad \text{if} \quad \Phi^i(y) \leqq 0 , \quad i = 1, 2, \ldots, k .$$

All terms of $\tilde{\Theta}$ are regular functions and hence the desired gradient information for a bundle or subgradient algorithm can be obtained by summing up a vector from $\partial\Theta$ computed according to Proposition 2.1 with a vector

$$\begin{bmatrix} 0 \\ \sum_{i=1}^{k} r_i \vartheta_i \end{bmatrix},$$

$\vartheta_i$ being given by (2.9). Of course, the solution $z$ of the inner optimization problem must be sufficiently precise, otherwise the NDO algorithm could fail.

This approach was used to solve the three following simple test examples. In all of them $x \in \mathbb{R}$, $y \in \mathbb{R}^2$, $f_1 = \frac{1}{2}[(y^1 - 3)^2 + (y^2 - 4)^2]$, $\Omega = \{(y^1, y^2) \in \mathbb{R}^2_+ \mid - 0{\cdot}333y^1 + y^2 \leqq 2,\ y^1 - 0{\cdot}333y^2 \leqq 2\}$ and

$$f_2 = \tfrac{1}{2}\langle y, H(x)\, y\rangle - \langle b(x), y\rangle,\ b(x) = \begin{bmatrix} 1 + 1{\cdot}333x \\ x \end{bmatrix},$$

where the $[2 \times 2]$ matrix $H(x)$ varies.

**Example 1.**

$$H(x) = E .$$

Starting point: $x = y^1 = y^2 = 0$ .
Solution: $x = 2{\cdot}07$ , $y^1 = 3$ , $y^2 = 3$ , $f_1 = 0{\cdot}5$ .

**Example 2.**

$$H(x) = \begin{bmatrix} 1 + x & 0 \\ 0 & 0 \end{bmatrix}$$

Starting point: $x = 5$ , $y^1 = y^2 = 0$ .
Solution: $x = 0$ , $y^1 = 3$ , $y^2 = 3$ , $f_1 = 0{\cdot}5$.

**Example 3.**

$$H(x) = \begin{bmatrix} 1 + x & 0 \\ 0 & 1 + 0{\cdot}1x \end{bmatrix}$$

Starting point: $x = y^1 = y^2 = 0$ .
Solution: $x = 3{\cdot}456$ , $y^1 = 1{\cdot}707$ , $y^2 = 2{\cdot}569$ , $f_1 = 1{\cdot}859$ .

All examples have been solved by means of the code M1FC1 written by Cl. Lemaréchal according to the bundle method [4]. The inner quadratic programming problems have been solved by the SOL/QPSOL code of Gill and al.

## 3. COMBINATORIAL INEQUALITY CONSTRAINTS

Let us investigate the optimization problem

$$
\begin{aligned}
& f_0(x) \to \inf \\
(3.1) \qquad & \text{subj. to} \\
& \beta(x) \leqq 0\,, \quad x \in \Omega\,,
\end{aligned}
$$

where $f_0[\mathbb{R}^n \to \mathbb{R}]$ is continuously differentiable on $\mathbb{R}^n$, $\beta$ is given by (1.5) and $\Omega \subset \mathbb{R}^n$ is nonempty, convex and compact. As function $q^i$, $i = 1, 2, \ldots, m$, are continuously differentiable, $\beta$ is locally Lipschitz and hence problem (3.1) possesses a solution $\hat{x}$ whenever

$$(3.2) \qquad \{x \in \Omega \mid \beta(x) \leqq 0\} \neq \emptyset\,.$$

We will assume that relation (3.2) holds and problem (3.1) is calm at $\hat{x}$ with respect to vertical perturbations of the constraint $\beta(x) \leqq 0$. Then, as in Section 2, we may conclude that $\hat{x}$ provides a minimum of the function

$$(3.3) \qquad \Xi = f_0 + r(\beta)^+$$

over $\Omega$, whenever the penalty parameter $r > 0$ is sufficiently large. The calmness property can be ensured e.g. by using the generalized Mangasarian-Fromowitz constraint qualification, cf. [1]. The augmented objective $\Xi$ is clearly locally Lipschitz and semismooth (cf. [5]) so that a bundle or subgradient algorithm may be applied to the problem

$$
\begin{aligned}
& \Xi(x) \to \inf \\
(3.4) \qquad & \text{subj. to} \\
& x \in \Omega\,,
\end{aligned}
$$

provided the constraint $x \in \Omega$ can be handled directly within the used minimization routine. The vectors from $\partial\Xi(x)$ may be computed according to the following assertion.

**Proposition 3.1.** Let $x \in \mathbb{R}^n$, $\beta(x) > 0$ and $i \in I(x) = \{i \in \{1, 2, \ldots, m\} \mid q^i(x) = \beta(x)\}$.

Then

$$(3.5) \qquad \nabla f_0(x) + r\, \nabla q^i(x) \in \partial\Xi(x)\,.$$

If $\beta(x) \leqq 0$, then

$$(3.6) \qquad \nabla f_0(x) \in \partial\Xi(x)\,.$$

Proof. $\Xi = f_0 + rg \circ \beta$, where $g = (\cdot)^+$. If $\beta(x) > 0$, then due to Chain Rule I of Clarke

$$\partial(g \circ \beta)(x) = \partial\beta(x) = -\partial(\max_{i=1,\ldots,m} \{-q^i(x)\})\,.$$

Hence, by Prop. 2.3.12 of [1] for $i \in I(x)$

$$\nabla q^i(x) \in \partial(g \circ \beta)(x)$$

so that relation (3.5) holds. If $\beta(x) \leqq 0$, then $x$ is a global minimizer of $g \circ \beta$ which implies relation (3.6). $\square$

Differently from function $(\psi)^+$ discussed in the previous section, function $(\beta)^+$ is nonregular (in the sense of Clarke). This is the reason why we require $f_0$ to be continuously differentiable; otherwise relations (3.5), (3.6) do not hold.

The structure of $\Omega$ is also important. If $\Omega$ consists merely of lower and upper bounds on single coordinates of $x$, e.g. the effective code M2FC1 of Cl. Lemaréchal written according to the bundle method [4] may be applied with the necessary gradient information being computed according to Proposition 3.1.

If, however, $\Omega$ is given by (2.1) and we use (under the appropriate calmness assumption) the same penalization technique to the constraints $\Phi^i(x) \leqq 0$, $i = 1, 2, \ldots, k$, we may have difficulties with the computation of a vector from $\partial\tilde{\Xi}$, where

$$\tilde{\Xi} = \Xi + \sum_{i=1}^{k} r_i(\Phi^i)^+ , \tag{3.7}$$

$r_i$, $i = 1, 2, \ldots, k$, being some suitably chosen positive penalty parameters. $\tilde{\Xi}$ is locally Lipschitz and semismooth, but we do not know any computationally acceptable way of evaluating a vector $\xi \in \partial\tilde{\Xi}(x)$ provided

$$\beta(x) > 0 ,$$

cardinality of $I(x)$ is greater than 1, and

$$\exists i \in \{1, 2, \ldots, k\} \quad \text{such that} \quad \Phi^i(x) = 0 .$$

In all other situations one has

$$\xi = \xi_1 + \sum_{i=1}^{k} r_i \vartheta_i \in \partial\tilde{\Xi}(x) ,$$

where $\xi_1 \in \partial\Xi(x)$ is computed according to (3.5), (3.6) and vectors $\vartheta_i$, $i = 1, 2, \ldots, k$, are computed according to (2.9).

This obstacle will certainly not cause any difficulties in a majority of problems. If, however, some line-search difficulties occur, it might be due to a bad gradient information and we have then either to augment the constraints $\Phi^i(x) \leqq 0$ by some smooth penalty or apply some algorithm of Kiwiel [3], capable of treating general inequality constraints within the nonsmooth minimization method.

If the constraint $\beta(x) \leqq 0$ is replaced in (3.1) by the semi-infinite constraint $\tilde{\beta}(x) \leqq 0$ with $\tilde{\beta}$ given by (1.6), then all the above considerations remain true, only Proposition 3.1 must be replaced by the following statement:

**Proposition 3.2.** Let $x \in \mathbb{R}^n$, $\tilde{\beta}(x) > 0$ and $R(x) = \{y \in \varkappa \mid Q(x, y) = \tilde{\beta}(x)\}$. Then, on denoting

$$\Lambda = f_0 + r(\tilde{\beta})^+ , \quad r > 0 , \tag{3.8}$$

$$\nabla f_0(x) + r \nabla_x Q(x, z) \in \partial\Lambda(x) \tag{3.9}$$

provided $z \in R(x)$. If $\tilde{\beta}(x) \leqq 0$, then

$$\nabla f_0(x) \in \partial\Lambda(x) . \tag{3.10}$$

The proof can be performed along the same lines as the proof of Prop. 3.1, but instead of Prop. 2.3.12 we have now to exploit Th. 2.8.2 of [1]. □

We conclude this section by an illustrative optimal control example. Let us consider the problem

$$F(x_m) + \sum_{i=0}^{m-1} \varphi_i(x_i, u_i) \to \inf$$

(3.11) $$\text{subj. to}$$

$$x_{i+1} = f_i(x_i, u_i), \quad i = 0, 1, \ldots, m-1, \quad x_0 = a, \quad u_i \in \omega \subset \mathbb{R}^k,$$

(3.12) $$\max_{i=1,\ldots,m} x_i^1 \geqq L,$$

where $x = (x_1, x_2, \ldots, x_m) \in \mathbb{R}^{n \times m}$ is the trajectory, $u = (u_0, u_1, \ldots, u_{m-1}) \in \mathbb{R}^{k \times m}$ is the control, $\omega$ is the set of admissible controls, $a \in \mathbb{R}^n$ is a given initial state and the functions $F[\mathbb{R}^n \to \mathbb{R}]$, $\varphi_i[\mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}]$, $f_i[\mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n]$, $i = 0, 1, \ldots, m-1$, are supposed to be continuously differentiable. The inequality (3.12) expresses the requirement that the first coordinate $x_i^1$ must be at some $i$ greater or equal to the given scalar $L$.

*Remark.* If $x$ is a discretized trajectory of a rocket, we may require that a certain prescribed altitude must be achieved. If $x^1$ represents the temperature measured at a given point of a heated object, condition (3.12) means that during the heating a certain prescribed temperature must be reached.

If we denote by $S_i$ the operator which assigns to each control $u \in \mathbb{R}^{k \times m}$ the state $x_{i+1} \in \mathbb{R}^n$, corresponding to $u$ with respect to the system equation, the problem (3.11)–(3.12) may be written in the form

$$J(u) = F \circ S_{m-1}(u) + \varphi_0(a, u_0) + \sum_{i=1}^{m-1} \varphi_i(S_{i-1}(u), (u_i)) \to \inf$$

(3.13) $$\text{subj. to} \qquad u_i \in \omega, \quad i = 0, 1, \ldots, m-1$$

$$\min_{i=1,\ldots,m} \{L - (S_{i-1}(u))^1\} \leqq 0.$$

Problem (3.13) is exactly of the type (3.1). Proposition 3.1, the assumptions of which are here clearly satisfied, implies the following assertion:

**Proposition 3.3.** Let $u$ be an admissible control, $x$ be the corresponding trajectory and $I(u) = \{i \in \{1, 2, \ldots, m\} \mid x_i^1 = \max_{j=1,\ldots,m} x_j^1\}$.

Assume that

$$\Xi(u) = J(u) + r\left(\min_{i=1,\ldots,m} \{L - x_i^1\}\right)^+$$

is the augmented objective with a suitably chosen penalty parameter $r > 0$. Finally, let $(p_1, p_2, \ldots, p_m) \in \mathbb{R}^{n \times m}$ be the solution of the adjoint equation

(3.14) $$p_i = [\nabla_{x_i} f_i(x_i, u_i)]^{\mathrm{T}} p_{i+1} - \nabla_{x_i} \varphi_i(x_i, u_i), \quad i = 1, 2, \ldots, m-1,$$

with the terminal condition

$$p_m = -\nabla F(x_m)\,.$$

Then $(v_0, v_1, \ldots, v_{m-1}) \in \partial \Xi(u)$ if

$$v_i = \nabla_{u_i} \varphi_i(x_i, u_i) - [\nabla_{u_i} f_i(x_i, u_i)]^T p_{i+1}\,, \quad i = 0, 1, \ldots, m-1 \tag{3.15}$$

and there exists a $j \in \{1, 2, \ldots, m\}$ such that $x_j^1 \geqq L$. If $x_i^1 < L$ for all $i = 1, 2, \ldots, m$, then formula (3.15) remains true provided we replace the adjoint equation (3.14) at some $i \in I(u)$ by the equation

$$p_i = [\nabla_{x_i} f_i(x_i, u_i)]^T p_{i+1} - \nabla_{x_i} \varphi_i(x_i, u_i) + \xi\,,$$

$\xi = (r, 0, 0, \ldots, 0)^T \in \mathbb{R}^n$.

In the proof we just need to combine Proposition 3.1 with the standard way of constructing the adjoint equations. Thus, combinatorial optimal control problems of the type (3.11)–(3.12) can easily be solved by the proposed approach provided the appropriate calmness assumption holds.

(Received September 14, 1987.)

REFERENCES

[1] F. H. Clarke: Optimization and Nonsmooth Analysis. Wiley, New York 1983.

[2] J. Haslinger and T. Roubíček: Optimal control of variational inequalities. Approximation theory and numerical realization. Appl. Math. Optimiz. *14* (1986), 187–201.

[3] K. Kiwiel: Methods of Descent for Nondifferentiable Optimization. (Lecture Notes Mathematics 1133.) Springer-Verlag, Berlin—Heidelberg—New York 1985.

[4] Cl. Lemaréchal, J. J. Strodiot and A. Bihain: On a bundle algorithm for nonsmooth optimization. NPS 4, Madison 1980.

[5] R. Mifflin: Semismooth and semiconvex functions in constrained optimization. SIAM J. Control Optimiz. *15* (1977), 959–972.

[6] J. V. Outrata: On the usage of bundle methods in optimal control of nondifferentable systems. In: Trends in Mathematical Optimization — 4th French-German Conference on Optimization (K.-H. Hoffmann, J.-B. Hiriart-Urruty, C. Lemarechal, J. Zowe, eds.). Birkhäuser Verlag, Basel—Boston 1988, 233–246.

*Ing. Jiří V. Outrata, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Prague 8. Czechoslovakia.*