Mathematical Programming 9 (1975) 87–99.
North-Holland Publishing Company

NECESSARY AND SUFFICIENT CONDITIONS FOR A PENALTY METHOD TO BE EXACT*

Dimitri P. BERTSEKAS
University of Illinois, Urbana-Champaign, Ill., U.S.A.

Received 27 December 1973
Revised manuscript received 19 December 1974

This paper identifies necessary and sufficient conditions for a penalty method to yield an optimal solution or a Lagrange multiplier of a convex programming problem by means of a single unconstrained minimization. The conditions are given in terms of properties of the objective and constraint functions of the problem as well as the penalty function adopted. It is shown among other things that all linear programs with finite optimal value satisfy such conditions when the penalty function is quadratic.

1. Introduction

Considerable attention has been given recently to devising penalty function methods which solve a constrained minimization problem by means of a single unconstrained minimization rather than by means of a sequence of such minimizations [3,4]. The methods proposed require, however, minimization of a nondifferentiable cost functional and it is as yet an unsettled question whether they offer any advantage over sequential penalty methods [7]. One of the purposes of this paper is to show that, except for trivial cases, nondifferentiabilities are a necessary evil if the penalty method is to yield an optimal solution in a single minimization.

On the other hand, it is possible for a penalty method to yield a Lagrange multiplier of the problem in a single minimization even though it may require a second minimization to yield an optimal solution. This fact holds true for a wide class of differentiable penalty functions provided the problem has a special structure which is characterized in this

* This work was performed while the author was with the Engineering-Economic Systems Department, Stanford University, Stanford, California.

paper. It is shown further that linear programs with finite optimal value possess such a structure.

The analysis is conducted for the case of a convex programming problem. Extensive use of the tools of convex analysis is made and it is assumed that the reader is familiar with the notions and methodology of Rockafellar's text [13].

2. Problem formulation

Consider the following convex programming problem

$$\text{minimize} \quad f_0(x)$$
$$\text{subject to } x \in X \subset \mathbf{R}^n, \qquad f_1(x) \leqslant 0, \ldots, f_m(x) \leqslant 0. \tag{1}$$

The functions $f_i : \mathbf{R}^n \to \mathbf{R}$ are assumed real valued and convex, and the set X is assumed closed and convex.

We shall make the following assumptions:

A.1. Problem (1) has a nonempty and compact set of optimal solutions.

A.2. Problem (1) has at least one Lagrange multiplier (or Kuhn–Tucker vector as defined in [13, p. 274]).

One may verify that under assumption A.1 the ordinary perturbation function $q : \mathbf{R}^m \to (-\infty, +\infty]$

$$q(u) = \inf\{f_0(x) \colon x \in X,\ f_i(x) \leqslant u_i,\ i = 1, \ldots, m\} \tag{2}$$

is a closed proper convex function [13]. (This fact is obtained by direct application of [13, Theorem 9.2].) Furthermore the ordinary dual functional $g : \mathbf{R}^m \to [-\infty, \infty)$,

$$g(y) = \inf_u \left\{ q(u) + \sum_{i=1}^m y_i u_i \right\} \tag{3}$$

is a closed proper concave function. In addition,

$$q(0) = \min_{u \leqslant 0} q(u) = \sup_y g(y). \tag{4}$$

The dual functional g is of course maximized at points which are Lagrange multipliers of the problem and the existence of at least one such point is guaranteed by A.2.

We shall consider scalar penalty functions $p : \mathbf{R} \to \mathbf{R}$ which satisfy the following conditions:

C.1. p is convex.

C.2. $p(t) = 0$ for all $t \leqslant 0$ and $p(t) > 0$ for all $t > 0$.

Now consider the problem

$$\inf_{x \in X} \left\{ f_0(x) + \sum_{i=1}^{m} p_i[f_i(x)] \right\}, \tag{5}$$

where each of the functions $p_i : \mathbf{R} \to \mathbf{R}$, $i = 1, ..., m$ satisfies C.1. and C.2.

Let $\tilde{x} \in X$ be a solution of problem (5) (assuming one exists). We shall be interested in identifying the conditions under which $\tilde{x}$ is also a solution of the original problem (1). We shall also be interested in characterizing the class of problems for which a Lagrange multiplier can be directly obtained from $\tilde{x}$ even though $\tilde{x}$ may not be a solution of problem (1).

Denote by $p_i^* : \mathbf{R} \to (-\infty, +\infty]$ the convex conjugate function of p_i [13]

$$p_i^*(t^*) = \sup_t \{t \cdot t^* - p_i(t)\}.$$

The form of three typical pairs p_i, p_i^* is depicted in Fig. 1. One may easily verify the following relationship

$$\inf_{x \in X} \left\{ f_0(x) + \sum_{i=1}^{m} p_i[f_i(x)] \right\} = \inf_u \left\{ q(u) + \sum_{i=1}^{m} p_i(u_i) \right\}$$

$$= \max_y \left\{ g(y) - \sum_{i=1}^{m} p_i^*(y_i) \right\}, \tag{6}$$

where the last equality follows by direct application of the Fenchel duality theorem [13, Theorem 31.1]. Furthermore, by the nature of f_0, f_i, p_i, condition (a) of [13, Theorem 31.1] is satisfied and the maximum is attained in (6) by some vector $\tilde{y}$ (even though this maximum may equal $-\infty$).

The maximization on the right-hand side of (6) for various cases is depicted in Fig. 2. This figure, which has been used in [9] to interpret geometrically the method of multipliers, is the key to the understanding of the ideas of this paper. It shows clearly that in order that the "penalized" problem (6) has the same optimal value as problem (1) it is necessary that the conjugate p_i^* be "flat" at least along the interval $[0, \bar{y}_i]$,

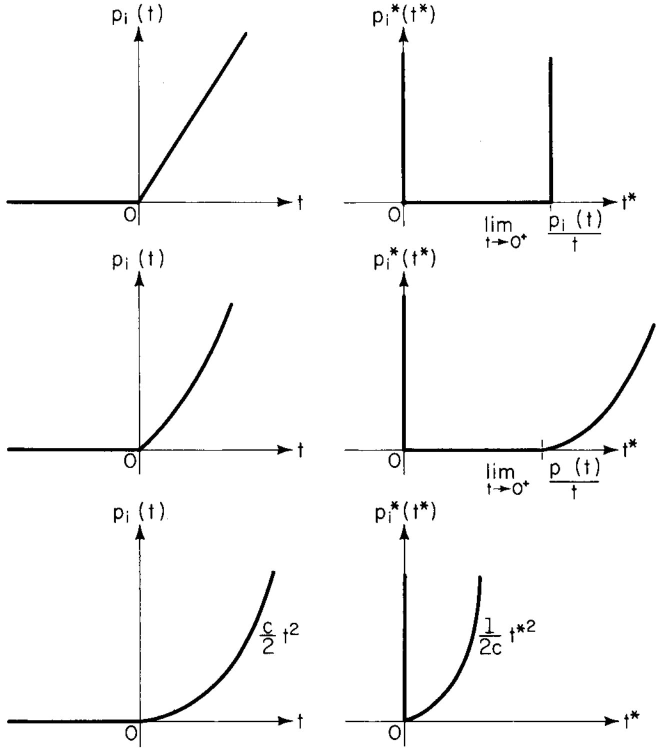

Fig. 1.

where $\bar{y} = (\bar{y}_1, ..., \bar{y}_m)$ is some Lagrange multiplier of the problem. At the same time $\bar{y}$ must be a maximizing point in (6). This in turn will imply the necessity for p_i to be nondifferentiable at 0 with right slope greater or equal to $\bar{y}_i$. The figure also shows that in order for a maximizing point $\tilde{y}$ in (6) to be a Lagrange multiplier it is necessary that either p_i^* be flat in a sufficiently large area or that the dual functional g be "pointed" or at least have a "corner" at a maximizing point $\tilde{y}$. The remainder of the paper is devoted to making these ideas precise.

3. Conditions for an exact penalty

Let $\tilde{x} \in X$ be an optimal solution of the penalized problem

$$\inf_{x \in X} \left\{ f_0(x) + \sum_{i=1}^{m} p_i[f_i(x)] \right\} \tag{7}$$

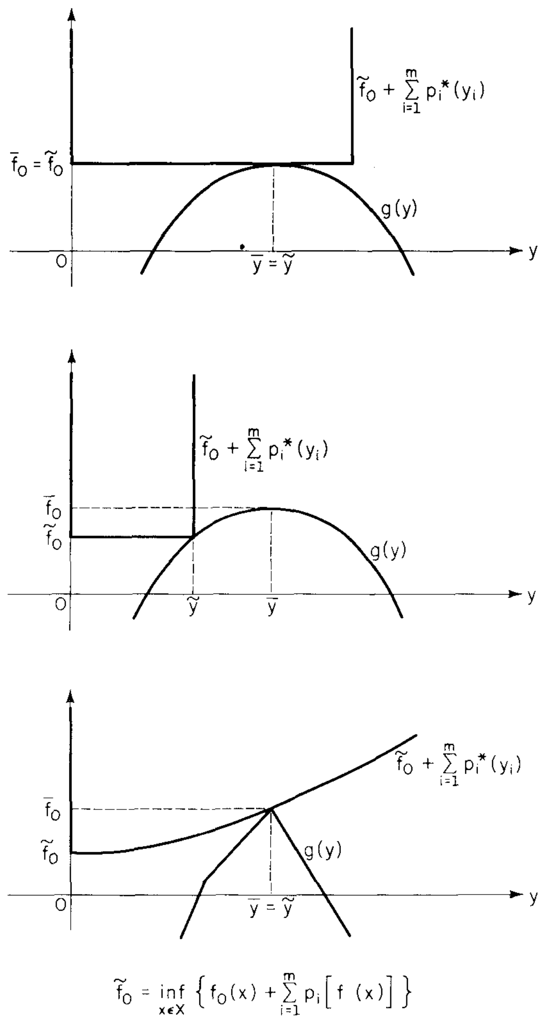

Fig. 2.

assuming that one such solution exists. Then we have the following proposition:

Proposition 1. (a) *In order that $\tilde{x}$ is also an optimal solution of the original problem* (1), *it is necessary that*

$$\lim_{t\to 0^+} \frac{p_i(t)}{t} \geqslant \bar{y}_i, \qquad i = 1, 2, ..., m \tag{8}$$

for some Lagrange multiplier $\bar{y} = (\bar{y}_1, ..., \bar{y}_m)$ *of problem* (1).

(b) *In order that problems* (1) *and* (7) *have exactly the same solutions, it is sufficient that*

$$\lim_{t\to 0^+} \frac{p_i(t)}{t} > \bar{y}_i, \qquad i = 1, 2, ..., m \tag{9}$$

for some Lagrange multiplier $\bar{y} = (\bar{y}_1, ..., \bar{y}_m)$ *of problem* (1).

Proof. (a) If $\tilde{x}$ is an optimal solution of problem (1), we have from (6),

$$\bar{f}_0 = f_0(\tilde{x}) + \sum_{i=1}^{m} p_i[f_i(\tilde{x})] = \max_y \left\{ g(y) - \sum_{i=1}^{m} p_i^*(y_i) \right\},$$

where $\bar{f}_0$ denotes the optimal value of problem (1). Let $\bar{y}$ be any vector attaining the maximum above. Then

$$\bar{f}_0 + \sum_{i=1}^{m} p_i^*(\bar{y}_i) = g(\bar{y}) \leqslant \bar{f}_0 .$$

Since $p_i^*(t^*) \geqslant 0$ for all t^*, we obtain

$$p_i^*(\bar{y}_i) = 0, \qquad i = 1, 2, ..., m, \qquad g(\bar{y}) = \bar{f}_0$$

showing that $\bar{y}$ must be a Lagrange multiplier of problem (1). Now the relation $p_i^*(\bar{y}_i) = 0$, $i = 1, 2, ..., m$ by the definition of p_i, p_i^* implies (8) which was to be proved.

(b) If $\bar{x}$ is an optimal solution of problem (1), then by (9) and the definition of a Lagrange multiplier we have

$$f_0(\bar{x}) + \sum_{i=1}^{m} p_i[f_i(\bar{x})] = f_0(\bar{x}) = f_0(\tilde{x}) + \sum_{i=1}^{m} \bar{y}_i f_i(\bar{x})$$

$$\leqslant f_0(x) + \sum_{i=1}^{m} \bar{y}_i f_i(x) \leqslant f_0(x) + \sum_{i=1}^{m} p_i[f_i(x)] \quad \text{for all } x \in X.$$

Hence $\bar{x}$ is also a solution of problem (7). Conversely, if $\tilde{x}$ is a solution of problem (7), then $\tilde{x}$ is either a feasible point in which case it is also a solution of problem (1), or it is infeasible in which case $f_i(\tilde{x}) > 0$ for some i. Then since $p_i(t) > 0$ for all $t > 0$ we have by (9) for any solution $\tilde{x}$ of problem (1),

$$f_0(\tilde{x}) + \sum_{i=1}^{m} p_i[f_i(\tilde{x})] > f_0(\tilde{x}) + \sum_{i=1}^{m} \bar{y}_i f_i(\tilde{x})$$

$$\geqslant \bar{f}_0 = f_0(\bar{x}) + \sum_{i=1}^{m} p_i[f_i(\bar{x})]$$

which is a contradiction. Hence problems (1) and (7) have exactly the same optimal solutions which was to be proved.

Proposition 1(b) generalizes a known result which has been proved for the particular penalty function $p_i(t) = c \max [0, t]$ with c being a penalty parameter. This result has been pointed out in various forms in [6,16,14, 12,8], and it has formed the basis for exact penalty methods described in [3,4]. With the exception of [12], these references do not observe that it is sufficient that c be greater than some Lagrange multiplier of the problem. A fact worth mentioning is that a useful upper bound to the maximum magnitude of the Lagrange multipliers can be obtained if an interior point to the constraints and a lower bound to the optimal value are known. The statement of this result may be found in [1, p. 647]. Part (b) also generalizes a result of Evans, Gould and Tolle [5], which was proved for the particular penalty function $p_i(t) = \max [0, e^{ct} - 1]$. The same reference proves a similar result for the penalty function

$$p_i(t) = \begin{cases} 0 & \text{if } t \leqslant 0, \\ \dfrac{ct}{\beta(\beta - t)} & \text{if } 0 < t < \beta, \\ +\infty & \text{if } \beta \leqslant t, \end{cases}$$

where $\beta > 0$ is some scalar. Proposition 1 may also be proved for this penalty function by verbatim repetition of its proof. It should be noted that some of the references cited above consider nonconvex programming problems while they require differentiability and other assumptions so that Proposition 1(b) is only a partial generalization of the results mentioned.

Proposition 1(a) shows that unless some Lagrange multiplier is zero (and the problem is essentially unconstrained) we cannot expect to obtain an optimal solution to problem (1) by solving problem (7) using a differentiable penalty function within the class considered. A similar result has not been pointed out in the literature to the author's knowledge.

Let us restrict now the class of penalty functions under consideration by requiring

C.3. p is differentiable everywhere, and $\lim_{t\to\infty} \mathrm{d}p(t)/\mathrm{d}t = \infty$.

It may be seen that C.3 implies that p^* is strictly convex on $[0,\infty)$, $p^*(t^*) = +\infty$ for all $t^* < 0$, and $p^*(t^*) < +\infty$ for all $t \geqslant 0$.

Assume that p_i in problem (7) satisfy C.1, C.2, C.3, and let $\tilde{x}$ be a solution of problem (7). Consider the vector $\tilde{u} = (\tilde{u}_1, ..., \tilde{u}_m)$ given by

$$\tilde{u}_i = f_i(\tilde{x}), \qquad i = 1, ..., m \tag{10}$$

and the vector $\tilde{y} = (\tilde{y}_1, ..., \tilde{y}_m)$ given by

$$\tilde{y}_i = \frac{\mathrm{d}}{\mathrm{d}t} p_i(\tilde{u}_i), \qquad i = 1, ..., m. \tag{11}$$

We have the following lemma which is similar to a corresponding result pointed out in [9].

Lemma 1. (a) *Problem* (7) *has a nonempty and compact optimal solution set denoted by* $\tilde{X}$.

(b) *Every* $\tilde{x} \in \tilde{X}$ *defines the same vector* $\tilde{y}$ *via* (10), (11), *which is the unique maximizing point of* $g(y) - \Sigma_{i=1}^m p_i^*(y_i)$.

(c) *For every* $\tilde{x} \in \tilde{X}$ *the vectors* $\tilde{u}$ *and* $\tilde{y}$ *defined by* (10), (11) *satisfy*

$$p_i(\tilde{u}_i) + p_i^*(\tilde{y}_i) = \tilde{u}_i \tilde{y}_i, \qquad i = 1, 2, ..., m, \tag{12}$$

$$g(\tilde{y}) = q(\tilde{u}) + \sum_{i=1}^m \tilde{y}_i \tilde{u}_i. \tag{13}$$

Proof. (a) is a special case of [9, Proposition 6.1]. The proof consists of showing that the objective function of problem (7) has no common directions of recession [13] with the set X. This fact is a direct consequence of assumption A.1 and the assumption $\lim_{t\to\infty} \mathrm{d}p_i(t)/\mathrm{d}t = \infty$.

(b) and (c) follow from [13, Theorem 31.3] except for uniqueness of $\tilde{y}$ which follows from the differentiability assumption on p_i which in turn implies strict convexity of p_i^*.

Lemma 1 implies the following proposition.

Proposition 2. *In order that every solution $\tilde{x}$ of problem* (7) *yields a Lagrange multiplier $\bar{y}$ of problem* (1) *by means of the equation*

$$\bar{y}_i = \frac{\mathrm{d}}{\mathrm{d}t} p_i[f_i(\tilde{x})], \qquad i = 1, ..., m, \tag{14}$$

it is necessary and sufficient that there exists a vector $\tilde{y} \in \mathbf{R}^m$ such that

$$0 \in \partial g(\tilde{y}), \qquad 0 \in \partial \left(g - \sum_{i=1}^{m} p_i^*\right)(\tilde{y}), \tag{15}$$

where ∂ denotes the subdifferential of the corresponding function [13]. *Furthermore, if a vector $\tilde{y}$ satisfying* (15) *exists, then it is unique and it is equal to $\bar{y}$ as defined by* (14).

Proof. Simply notice that the subdifferential conditions (15) are equivalent to the condition that $\tilde{y}$ maximizes both $g(y)$ and $g(y) - \Sigma_{i=1}^m p_i^*(y_i)$. Then the result follows from Lemma 1.

Now let $\tilde{u}$ be a vector where the infimum is attained in (6) and $\tilde{y}$ be the unique vector where the supremum is attained in (6). By the conjugacy relation between g and q, and equality (13), we have [13]

$$\tilde{u} \in \partial g(\tilde{y}), \qquad -\tilde{y} \in \partial q(\tilde{u}). \tag{16}$$

As a result, the conditions (15) imply both

$$0 \in \partial g(\tilde{y}), \qquad \tilde{u} \in \partial g(\tilde{y}), \tag{17}$$

$$-\tilde{y} \in \partial q(0), \qquad -\tilde{y} \in \partial q(\tilde{u}). \tag{18}$$

The condition (17) shows that in order for the solution of problem (7) to yield a Lagrange multiplier via (14), it is necessary that the dual functional g have a corner at $\tilde{y}$, unless $\tilde{y} = 0$. This fact is essentially revealed by examination of Fig. 2. The conditions (18) together with (13) imply that the points $[0, q(0)]$, $[\tilde{u}, q(\tilde{u})]$ lie on the same hyperplane of $\mathbf{R}^{m+1}$ which is defined by $\tilde{y}$. Hence the perturbation function q must be linear along the line segment joining 0 and $\tilde{u}$.

A conclusion that can be drawn from the above observations is that one must have a problem with very special structure in order to be able to use a differentiable penalty which is exact in the sense that it yields a Lagrange multiplier in a single minimization. Even if such structure is present, it is necessary to have a sufficiently "steep" function p_i (or equivalently a sufficiently "flat" function p_i^*) in order to obtain a Lagrange multiplier. This fact is best illustrated by considering a quadratic penalty function

$$p_i(t) = \tfrac{1}{2} c\,[\max(0, t)]^2, \tag{19}$$

where $c > 0$ denotes the penalty parameter. For this case we have

$$p_i^*(t^*) = \begin{cases} \dfrac{1}{2c}\, t^{*2} & \text{if } t^* \geqslant 0, \\ +\infty & \text{if } t^* < 0. \end{cases} \tag{20}$$

The conditions (15) can be written for this case as

$$0 \in \partial g(\tilde{y}), \qquad \frac{1}{c}\tilde{y} \in \partial g(\tilde{y}). \tag{21}$$

This fact may be seen either by directly comparing (15) and (21) or by observing that the point $\tilde{y}$ is a maximizing point in (6) with p_i^* given by (20) if and only if it is a maximizing point of $g(y) - \Sigma_{i=1}^m y_i^2/2c$. Then (21) follows immediately. One should also note that a point $\tilde{y}$ which satisfies (21) is also the vector of minimum Euclidean norm on the set of Lagrange multipliers Y^* of problem (1), i.e.,

$$0 \in \partial g(\tilde{y}), \qquad \frac{1}{c}\tilde{y} \in \partial g(\tilde{y}) \Rightarrow \|\tilde{y}\| = \min_{y \in Y^*} \|y\|, \tag{22}$$

where $\|y\| = (\Sigma_{i=1}^m y_i^2)^{\frac{1}{2}}$. Now the conditions (21), (22) together with Lemma 1 and Proposition 2 yield the following result.

Proposition 3. *Let $\bar{y} = (\bar{y}_1, ..., \bar{y}_m)$ be the vector of minimum norm on the set of Lagrange multipliers Y^* of problem* (1). *Consider also problem* (7) *with p_i being the quadratic penalty function* (19). *Then in order that for some scalar $c^* > 0$ and for all $c \geqslant c^*$, every solution $\tilde{x}$ of problem* (7) *satisfies*

$$\bar{y}_i = \frac{\mathrm{d}}{\mathrm{d}t} p_i[f_i(\tilde{x})] = \max[0, c f_i(\tilde{x})], \qquad i = 1, ..., m, \tag{23}$$

it is necessary and sufficient that the line segment joining 0 *and $\bar{y}/c^*$ is contained in $\partial g(\bar{y})$.*

It is easy to show that for every linear program (or, more generally, polyhedral convex program [13]) having a unique Lagrange multiplier $\bar{y}$, the origin is an interior point of $\partial g(\bar{y})$. Hence, when A.1, A.2 are satisfied, the Lagrange multiplier of such linear programs can be found by means of the quadratic penalty method in a single minimization, provided the penalty parameter is sufficiently high. On the other hand, in the presence of polyhedral convexity it is possible to prove a stronger result than the one suggested by Proposition 1. To this end, consider in place of assumptions A.1 and A.2 the following assumption.

A′. Problem (1) (viewed as an ordinary convex program [13, §28]) is a polyhedral convex program and has a finite optimal value.

Assumption A′ guarantees that problem (1) has a nonempty solution set and a nonempty Lagrange multiplier set [13, Theorem 29.2]. Furthermore the perturbation function q, the dual functional g and the set of Lagrange multipliers are polyhedral. Under A′, Lemma 1 holds with the exception of part (a), where now we can guarantee nonemptiness of $\tilde{X}$ but not compactness. Proposition 2 also holds under A′. We have the following proposition which strengthens the results obtained from Proposition 3 for polyhedral programs.

Proposition 4. *Assume A′ in place of* A.1, A.2. *Let* $\bar{y} = (\bar{y}_1, ..., \bar{y}_m)$ *be the vector of minimum norm in the set of Lagrange multipliers* Y^* *of problem* (1). *Consider also problem* (7) *with* p_i *being the quadratic penalty function* (19). *Then there exists a scalar* $c^* \geqslant 0$ *such that for all* $c \geqslant c^*$ *every solution* $\tilde{x}$ *of problem* (7) *satisfies*

$$\bar{y}_i = \frac{\mathrm{d}}{\mathrm{d}t} p_i[f_i(\tilde{x})] = \max[0, cf_i(\tilde{x})], \qquad i = 1, ..., m.$$

Proof. Since $g(y)$ is polyhedral and $\bar{y}$ maximizes g, we have the representation [13, p. 172]

$$g(y) = \min_{i \in I_1} \{\min[g(\bar{y}), \langle a_i, y\rangle - b_i]\} - \sum_{i \in I_2} \delta(y \colon \langle a_i, y\rangle \geqslant b_i)$$

for some finite index sets I_1, I_2, vectors $a_i \in \mathbf{R}^m$ and scalars b_i. In the above equation, $\delta(\cdot \colon \langle a_i, y\rangle \geqslant b_i)$ denotes the indicator function [13] of the set $\{y \colon \langle a_i, y\rangle \geqslant b_i\}$, and $\langle \cdot, \cdot \rangle$ denotes the usual inner product in $\mathbf{R}^m$. Denote

$$\bar{I}_1 = \{i \in I_1 : g(\bar{y}) = \langle a_i, \bar{y} \rangle - b_i\},$$

$$\bar{I}_2 = \{i \in I_2 : \langle a_i, \bar{y} \rangle = b_i\}.$$

Then the subdifferential $\partial g(\bar{y})$ is given by

$$\partial g(\bar{y}) = \text{conv}(\{0, a_i : i \in \bar{I}_1\}) + \left\{ \sum_{i \in \bar{I}_2} \lambda_i a_i : \lambda_i \geqslant 0, i \in \bar{I}_2 \right\} \tag{24}$$

where conv($\cdot$) denotes convex hull. The set of Lagrange multipliers of the problem is the set

$$Y^* = \{y : \langle a_i, y \rangle - b_i \geqslant g(\bar{y}), i \in I_1\} \cap \{y : \langle a_i, y \rangle \geqslant b_i, \ i \in I_2\}.$$

Since $\bar{y}$ is the vector of minimum norm on this set, we have by the Kuhn–Tucker conditions,

$$\bar{y} = \sum_{i \in \bar{I}_1} \mu_i a_i + \sum_{i \in \bar{I}_2} \lambda_i a_i, \tag{25}$$

where μ_i, $i \in \bar{I}_1$, λ_i, $i \in \bar{I}_2$ are some nonnegative scalars. Take $c^* = \Sigma_{i \in \bar{I}_1} \mu_i$. Then from (24), (25) we have that for every $c \geqslant c^*$, $\bar{y} \in c\,\partial g(\bar{y})$. Hence $\bar{y}$ satisfies the conditions (21) and the result follows from Proposition 2.

Another interesting result which will not be proved here but may be easily deduced from the analysis of [9,10] and the present paper, concerns the application of the method of multipliers (with a quadratic penalty) to problems satisfying assumption A′, or A.1 and A.2 with the origin being an interior point of $\partial g(\bar{y})$. For such problems it may be proved that the method of multipliers will yield a Lagrange multiplier $\bar{y}$ in a single or at most a finite number of unconstrained minimizations for any fixed positive value of the penalty parameter c.

While this paper was under review, the author became aware of some papers in the Soviet literature [2,11,15] which prove a result similar to the one of Proposition 4 for linear programs. The approach in these references is entirely different than the one of this paper and necessitates the assumption that the linear program has a unique solution as well as further assumptions which guarantee that problem (7) has a unique solution for every $c > 0$. The same references provide some interesting finite procedures for solving such linear programs by the quadratic penalty method.

References

[1] D.P. Bertsekas and S.K. Mitter, "A descent numerical method for optimization problems with nondifferentiable cost functionals", *SIAM Journal on Control* 11 (4) (1973) 637–652.

[2] S.P. Chebotarev, "Variation of the penalty coefficient in linear programming problems", *Automation and Remote Control* 34 (7) (1973) 102–107.

[3] A.R. Conn, "Constrained optimization using a nondifferentiable penalty function", *SIAM Journal on Numerical Analysis* 10 (1973) 760–784.

[4] A.R. Conn and T. Pietrzykowski, "A penalty function method converging directly to a constrained minimum", Research Rept. 73-11, Dept. of Combinatorics and Optimization, University of Waterloo (May 1973).

[5] J.P. Evans, F.J. Gould and J.W. Tolle, "Exact penalty functions in nonlinear programming", *Mathematical Programming* 4 (1) (1973) 72–97.

[6] Y.M. Ermol'ev and N.Z. Shor, "On the minimization of nondifferentiable functions", *Kibernetica* 3 (1) (1967) 101–102.

[7] A.V. Fiacco and G.P. McCormick, *Nonlinear programming: sequential unconstrained minimization techniques* (Wiley, New York, 1968).

[8] S. Howe, "New conditions for exactness of a simple penalty function", *SIAM Journal on Control* 11 (2) (1973) 378–381.

[9] B.W. Kort and D.P. Bertsekas, "Multiplier methods for convex programming", in: *Proceedings of 1973 IEEE decision and control conference,* San Diego, Calif., Dec. 1973, pp. 428–432.

[10] B.W. Kort and D.P. Bertsekas, "Combined primal dual and penalty methods for convex programming", *SIAM Journal on Control,* to appear.

[11] A.T. Kutanov, "Refining the solution of the linear programming problem in the method of penalty functions", *Automation and Remote Control* 39 (4) (1970) 127–132.

[12] D.G. Luenberger, "Control problems with kinks", *IEEE Transactions on Automatic Control* AC-15 (1970) 570–575.

[13] R.T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, N.J., 1970).

[14] T. Pietrzykowski, "An exact potential method for constrained maxima", *SIAM Journal on Numerical Analysis* 6 (2) (1969) 269–304.

[15] N.O. Vil'chevskii, "Choosing the penalty coefficient in linear programming problems", *Automation and Remote Control* 39 (4) (1970) 121–126.

[16] W.I. Zangwill, "Nonlinear programming via penalty functions", *Management Science* 13 (1967) 344–358.