MATHEMATICS OF OPERATIONS RESEARCH
Vol. 6, No. 3, August 1981
*Printed in U.S.A.*

# PROXIMAL SUBGRADIENTS, MARGINAL VALUES, AND AUGMENTED LAGRANGIANS IN NONCONVEX OPTIMIZATION

R. T. ROCKAFELLAR*†

*University of Washington*

The Clarke subgradients of a nonconvex function $p$ on $R^n$ are characterized in terms of limits of "proximal subgradients." In the case where $p$ is the optimal value function in a nonlinear programming problem depending on parameters, proximal subgradients correspond to saddlepoints of the augmented Lagrangian. When the constraint and objective functions are sufficiently smooth, this leads to a characterization of marginal values for a given problem in terms of limits of Lagrange multipliers in "neighboring" problems for which the standard second-order sufficient conditions for optimality are satisfied at a unique point.

**1. Introduction.** For a closed set $D$ in $R^m$ and point $u \in D$, a vector $\bar{y} \in R^m$ is said to be a *proximal normal* to $D$ at $\bar{u}$ if for $t > 0$ sufficiently small, $u$ is the unique nearest point of $D$ to $\bar{u} + ty$. (Here $\bar{y} = 0$ is degenerately always a proximal normal.) The *normal cone* to $D$ at $\bar{u}$ in Clarke's sense [1] is the set

$$N_D(\bar{u}) = \mathrm{cl}\,\mathrm{co}\{\bar{y} \in R^m \mid \exists u_k \in D,\ y_k \text{ proximal normal to } D \text{ at } u_k, \text{ with } u_k \to \bar{u},\ y_k \to \bar{y}\}, \tag{1}$$

where "cl" denotes closure and "co" convex hull. This closed convex cone contains a $\bar{y} \neq 0$ if and only if $\bar{u}$ is a boundary point of $D$ (cf. Rockafellar [2]).

Let $p$ be any lower semicontinuous function on $R^m$ with values in $[-\infty, +\infty]$, and let $E$ be its epigraph,

$$E = \{(u, \alpha) \in R^{m+1} \mid \alpha \geq p(u)\}.$$

The lower semicontinuity of $p$ is equivalent to the closedness of $E$. Clarke [1] has defined a generalized set of subgradients of $p$ at a point $\bar{u}$ where $p(\bar{u})$ is finite by

$$\partial p(\bar{u}) = \{\bar{y} \in R^m \mid (\bar{y}, -1) \in N_E(\bar{u}, p(\bar{u}))\}. \tag{2}$$

This set is closed and convex but possibly empty. It reduces to the subgradient set of convex analysis if $p$ is convex and to the single gradient $\nabla p(\bar{u})$ if and only if $p$ is strictly differentiable at $u$ (see [1]). It is nonempty and bounded if and only if $p$ is Lipschitz continuous in a neighborhood of $\bar{u}$ (see [2]). An alternative definition of $\partial p(\bar{u})$ can be given in terms of generalized directional derivatives of a sort for $p$, but this need not concern us here (cf. Rockafellar [3]–[5]).

Let us call $\bar{y}$ a *proximal subgradient* to $p$ at $\bar{u}$ if $(\bar{y}, -1)$ is actually a proximal normal to the epigraph $E$ at $(\bar{u}, p(\bar{u}))$. One can easily verify that this holds if and only if there is a function $g$ of class $C^2$ such that $\nabla g(\bar{u}) = \bar{y}$, $g(\bar{u}) = p(\bar{u})$ and $g(u) \leq p(u)$ in some neighborhood of $\bar{u}$. Our first objective is to provide a formula for $\partial p(\bar{u})$ as the closed convex hull of certain limits of proximal subgradients. This is done in Theorem 1. The

*Received July 22, 1980.
*AMS 1970 subject classification.* Primary 90C31. Secondary 90C30.
*OR/MS Index 1978 subject classification.* Primary: 650 Programming/nonlinear/theory.
*Key words.* Marginal values, subgradients, augmented Lagrangians.
†Research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under grant number 77-3204 at the University of Washington.



0364-765X/81/0603/0424$01.25

result is obviously based on the epigraph form of (1), but the argument is nontrivial. Trouble arises because proximal normals to $E$ can be of the degenerate form $(\bar{y}, 0)$ and can occur also at points $(\bar{u}, \alpha)$ of $E$ where $\alpha > p(\bar{u})$. It must be demonstrated that limits involving such normals can be replaced by limits of normals corresponding to proximal subgradients.

We proceed then to apply the result to the case where $p$ is the optimal value function for a parameterized nonlinear programming problem. Under a mild assumption furnished in §3, the problem

$$(P_u)\begin{cases}\text{minimize } f_0(x) \text{ over all } x \in C \text{ satisfying}\\ f_i(x) + u_i\begin{cases}\leqslant 0 & \text{for } i = 1, \dots, s,\\ = 0 & \text{for } i = s+1, \dots, m,\end{cases}\end{cases}$$

will have an optimal solution whenever it has a feasible solution, and the function $p(u) = \min(P_u)$ will be lower semicontinuous on $R^m$. Generalized derivatives of $p$ can be viewed in an economic framework as marginal values for the resources represented by the quantities $u_i$, and a description of $\partial p(\bar{u})$ takes on interest from this as well as from the insights it can provide in the study of various computational procedures.

Outer estimates for $\partial p(\bar{u})$ in such a setting have already been obtained by Gauvin [6] (see also [7]) in terms of first-order Lagrange multiplier vectors. These estimates assumed that the constraint and objective functions are of class $\mathcal{C}^1$, the abstract constraint set $C$ is all of $R^n$, the set of feasible solutions to $(P_u)$ is uniformly bounded in a neighborhood of $u = \bar{u}$, and the Mangasarian-Fromowitz constraint qualification is satisfied at all optimal solutions to $(P_{\bar{u}})$. Here we dispense with such restrictions and nevertheless derive an exact expression for $\partial p(\bar{u})$ in terms of limits of certain "augmented" multiplier vectors.

The (quadratic) *augmented Lagrangian* for $(P_u)$ is

$$L_u(x, y, r) = f_0(x) + \sum_{i=1}^{m} \varphi_i(y_i, f_i(x) + u_i, r) \quad \text{for } x \in C, r > 0, \tag{3}$$

where for equality constraints ($i = s+1, \dots, m$)

$$\varphi_i(y_i, f_i(x) + u_i, r) = y_i[f_i(x) + u_i] + (r/2)[f_i(x) + u_i]^2$$

and for inequality constraints ($i = 1, \dots, s$)

$$\varphi_i(y_i, f_i(x) + u_i, r) = \begin{cases} y_i[f_i(x) + u_i] + (r/2)[f_i(x) + u_i]^2 \\ \qquad\qquad \text{if } y_i + r[f_i(x) + u_i] \geqslant 0,\\ -y_i^2/2r \qquad \text{if } y_i + r[f_i(x) + u_i] \leqslant 0.\end{cases}$$

We have shown in [8, Theorem 5] that if the quadratic growth condition

$$\liminf_{|u| \to \infty} \frac{p(u)}{|u|^2} > -\infty \tag{4}$$

is satisfied (and this is equivalent to $L_u(x, 0, r)$ having a finite infimum in $x$ for some $u$ when $r$ is sufficiently large [8]), then the proximal normals to $p$ at any $\bar{u}$ are exactly the vectors $\bar{u}$ such that for $\bar{r}$ sufficiently large, $(\bar{x}, \bar{y}, \bar{r})$ is a global saddle point of $L_{\bar{u}}$ with respect to minimization in $x$ and maximization in $(y, r)$; here $\bar{x}$ denotes any optimal solution to $(P_{\bar{u}})$.

Clearly then, our general characterization of $\partial p(\bar{u})$ in terms of limits of proximal subgradients can be made to yield in this context a characterization of $\partial p(\bar{u})$ in terms of limits of multiplier vectors $y^k$ corresponding to saddle points of the augmented

Lagrangian in "nearby" problems $(P_{u^k})$ (Theorem 2). We go beyond this immediate consequence by demonstrating in Theorem 3 that one can get away with a more special case of multiplier vectors $y^k$ corresponding to the standard second-order sufficient conditions for optimality, provided the notion of a "nearby" problem is enlarged. This result ties in with work of Spingarn and Rockafellar [9] on the generic nature of the second-order conditions in problems with sufficiently smooth data (Theorem 4).

Incidentally, there is no real loss of generality in having the parameters on which $(P_u)$ depends appear only as additive constants in the constraint equations and inequalities as above. Other parameterizations can readily be placed in this mold, as will be explained and utilized in §§3, 4.

Our results lead in particular to estimates for $\partial p(\bar{u})$ in terms of multiplier vectors $\bar{y}$ that merely satisfy generalized first-order optimality conditions at the optimal solutions $\bar{x}$ to $(P_{\bar{u}})$. This application, which directly extends Gauvin's work [6], [7], has been written up separately [10].

**2. Subgradient formula.** One of the difficulties in characterizing $\partial p(\bar{u})$ in terms of limits of proximal subgradients is that such limits may fail to exist in the ordinary sense due to unboundedness. They have to be viewed in an extended (compactified) space consisting not only of the points in $R^m$ but also the *directions* in $R^m$, which represent "points at infinity." These directions correspond one-to-one with the *rays* in $R^m$ (half-lines emanating from the origin). It is possible to speak formally of convex hulls of mixed sets consisting of points and directions (cf. [8, §8 and §17]), and this idea underlies the theorem in this section, but we shall express such convex hulls here in a conventional manner.

For any nonempty set $Y$, the set

$$0^+ Y = \limsup_{\lambda \downarrow 0} \lambda Y = \{ y \in R^m \mid \exists y^k \in Y, \lambda_k \downarrow 0 \quad \text{with } \lambda_k y^k \to y \} \tag{5}$$

is the *recession cone* of $Y$. It contains a nonzero vector if and only if $Y$ is bounded. Then it is a union of rays which can be regarded as representing the "boundary points of $Y$ at infinity."

THEOREM 1. *Let $p: R^m \to [-\infty, \infty]$ be lower semicontinuous, and let $\bar{u}$ be any point with $p(\bar{u})$ finite. Define*

$$Y = \{ \bar{y} \in R^m \mid \exists y^k \text{ proximal subgradient of } p \text{ at } u^k \text{ with } u^k \to \bar{u},\ p(u^k) \to p(\bar{u}) \text{ and } y^k \to \bar{y} \},$$

$$Y_0 = \{ \bar{y} \in R^m \mid \exists \lambda_k \downarrow 0 \text{ and } y^k \text{ proximal subgradient of } p \text{ at } u^k \text{ with } u^k \to \bar{u},\ p(u^k) \to p(\bar{u}) \text{ and } \lambda_k y^k \to \bar{y} \}.$$

*Then $Y$ and $Y_0$ are closed sets satisfying $0 \in Y_0 \supset 0^+ Y$, and it is impossible to have both $Y = \emptyset$ and $Y_0 = \{0\}$. The formula*

$$\partial p(\bar{u}) = \operatorname{cl} \operatorname{co}[Y + Y_0] \tag{6}$$

*holds, where*

$$Y + Y_0 = \{ \bar{y} + \bar{y}^0 \mid \bar{y} \in Y, \bar{y}^0 \in Y_0 \}$$

*(a set which includes $Y$ and is empty if and only if $Y$ is empty).*

COROLLARY. *For $\partial p(\bar{u})$ to be nonempty and bounded ($p$ Lipschitz continuous on a neighborhood of $\bar{u}$), it is necessary and sufficient that $Y_0 = \{0\}$. Then*

$$\partial p(\bar{u}) = \operatorname{co} Y.$$

PROOF OF THEOREM 1. The initial assertions in Theorem 1 about $Y$ and $Y_0$ follow at once from the definitions and the fact, to be established below, that there do exist sequences $u^k \to \bar{u}$ such that $p$ has a proximal subgradient at $u^k$ and $p(u^k) \to p(\bar{u})$. (Any element of $0^+ Y$ can be represented as an element of $Y_0$ by a process of diagonalizing sequences.) To get this fact and the formula for $\partial p(\bar{u})$, we begin with an observation about the nature of proximal normals to the epigraph $E$ of $p$.

If $(\bar{y}, \bar{\eta})$ is a proximal normal to $E$ at $(\bar{u}, \alpha)$, then $\bar{\eta} \leqslant 0$ and $(\bar{y}, \bar{\eta})$ is also a proximal normal to $E$ at $(\bar{u}, p(\bar{u}))$. Indeed, $(\bar{y}, \bar{\eta})$ is a proximal normal at $(\bar{u}, \alpha)$ if and only if for some $t > 0$, there is a Euclidean ball $B$ in $R^{m+1}$ centered at $(\bar{u} + t\bar{y}, \alpha + t\bar{\eta})$ which contains $(\bar{u}, \alpha)$ but no other point of $E$. Since $E$, being an epigraph, includes with each of its points the half-line extending upward from that point, the half-line extending upward from $(\bar{u}, p(\bar{u}))$ meets $B$ only at $(\bar{u}, \alpha)$. Thus $(\bar{y}, \bar{\eta})$ cannot make an acute angle with the vertical vector $(0, 1)$: one must have $\bar{\eta} \leqslant 0$, in fact $\bar{\eta} = 0$ if $\alpha > p(\bar{u})$. In the latter case, if $B$ is shifted downward to have center at $(\bar{u} + t\bar{y}, p(\bar{u}))$ instead of $(\bar{u} + ty, \alpha)$, it will still meet $E$ in only one point, namely $(\bar{u}, p(\bar{u}))$ instead of $(\bar{u}, \alpha)$. Hence $(\bar{y}, 0)$ will be a proximal normal to $E$ at $(\bar{u}, p(\bar{u}))$ as well as at $(\bar{u}, \alpha)$.

We turn now to the construction of the normal cone $N_E(\bar{u}, p(\bar{u}))$ that appears in the definition (2) of $\partial p(\bar{u})$. This cone is by direct extension of (1) the closed convex hull of all limits $(\bar{y}, \bar{\eta})$ of sequences of elements $(y^k, \eta_k)$, where $(y^k, \eta_k)$ is a proximal normal to $E$ at $(u^k, \alpha_k)$, $\alpha_k \geqslant p(u^k)$ and $(u^k, \alpha_k) \to (\bar{u}, p(\bar{u}))$. Since $(\bar{u}, p(\bar{u}))$ is a boundary point of $E$, $N_E(\bar{u}, p(\bar{u}))$ cannot consist of just $(0, 0)$ [2]; hence at least one of the limits $(\bar{y}, \bar{\eta})$ is not $(0, 0)$.

From the observation above, therefore, it suffices in determining $N_E(\bar{u}, p(\bar{u}))$ to take $\alpha_k = p(u^k)$ and restrict attention to limits of sequences of proximal normals that are all of one of the following three types:

(a) $\lambda_k(y^k, -1)$, where $\lambda_k \to \lambda > 0$,

(b) $\lambda_k(y^k, -1)$, where $\lambda_k \downarrow 0$,

(c) $(y^k, 0)$ where $y^k \neq 0$.

In (a) and (b), $y^k$ is by definition a proximal subgradient at $u^k$. The limits in these cases are respectively of the forms $\lambda(\bar{y}, -1)$ with $\bar{y} \in Y$ and $(\bar{y}, 0)$ with $\bar{y} \in Y_0$. The crux of the matter is to show that any limit of type (c) is obtainable also as one of type (b). This will prove that $N_E(\bar{u}, p(\bar{u}))$ is the closed convex hull of the nonempty cone

$$\{\lambda(\bar{y}, -1) \mid \bar{y} \in Y, \lambda > 0\} \cup \{(\bar{y}, 0) \mid \bar{y} \in Y_0\}, \tag{7}$$

from which fact the desired formula for $\partial p(\bar{u})$ follows via (2). It will demonstrate at the same time that $u$ can be approached by at least one sequence of points $u^k$ such that $p(u^k) \to p(\bar{u})$ and a proximal subgradient exists at each $u^k$.

To verify that sequences of type (c) are superfluous, we demonstrate that a nonzero proximal normal $(y^k, 0)$ to $E$ at point $(u^k, p(u^k))$ can be approximated by a proximal normal $\lambda_k(\tilde{y}^k, -1)$ at an arbitrarily close point $(\tilde{u}^k, p(\tilde{u}^k))$. Notation is simplified by supposing (as is possible without loss of generality) that $(u^k, p(u^k)) = (0, 0)$ and denoting the given proximal normal at this point by $(y, 0)$, with $|y| = 1$. We shall construct a sequence of vectors $u_s$ and corresponding proximal normals $y_s$ such that $(u_s, p(u_s)) \to (0, 0)$ and $(y_s, -1)/|(y_s, -1)| \to (y, 0)$. This will complete the proof.

Fix $t > 0$ such that the ball of radius $t$ around $(ty, 0)$ meets $E$ only in the point $(0, 0)$ (which can be done since $(y, 0)$ is a nonzero proximal normal to $E$ at $(0, 0)$), and let $g$ be the function whose graph is the upper surface of this ball:

$$g(u) = (t^2 - |u - ty|^2)^{1/2}.$$

Clearly $g(u) < p(u)$ for all $u$ satisfying $|u - ty| \leqslant t$ except $u = 0$, where $g(0) = 0 = p(0)$ (because $|y| = 1$). For

$$h(u) = t - |u - ty|$$

and $s \geqslant 1$, one has

$$sh(u) \geqslant g(u) \quad \text{when } |u - ty| \leqslant t(s^2-1)/(s^2+1),$$

$$sh(u) \leqslant g(u) \quad \text{when } t(s^2-1)/(s^2+1) \leqslant |u - ty| \leqslant t,$$

so that by defining

$$f_s(u) = \begin{cases} g(u) & \text{when } |u - ty| \leqslant t(s^2-1)/(s^2+1), \\ sh(u) & \text{when } |u - ty| \geqslant t(s^2-1)/(s^2+1), \end{cases}$$

we get a continuous function on all of $R^m$ which is of class $\mathcal{C}^2$ where $|u - ty| \neq t(s^2-1)/(s^2+1)$ and has $f_s(u) \leqslant g(u)$ for all $u$ satisfying $|u - ty| \leqslant t$. Also, $f_s(0) = 0$. It follows that $f_s(u) < p(u)$ for all $u$ satisfying $|u - ty| \leqslant t$, except $u = 0$: $f_s(0) = 0 = p(0)$.

Recalling that $p$ is lower semicontinuous, and that a lower semicontinuous function attains a minimum over any compact set, we define

$$\alpha = \min_U p, \qquad \beta_s = \min_U \{ p - f_s \} \quad \text{for } s \geqslant 1,$$

where $U$ is the set of all $u$ satisfying $|u - ty| \leqslant t + 1$. For each $s$ we denote by $u_s$ one of the points where the minimum defining $\beta_s$ attained. From what we have noted about the relationship between $f_s$ and $p$, it is clear that

$$0 \geqslant \beta_s = p(u_s) - f_s(u_s) \geqslant p(u_s) \quad \text{and } |u_s - ty| \geqslant t. \tag{8}$$

Then $f_s(u) = sh(u)$ on a neighborhood of $u_s$, and in particular

$$\beta_0 \leqslant p(u_s) \leqslant f_s(u_s) = sh(u_s) = s(t - |u_s - ty|),$$

so that

$$|u_s - ty| \leqslant t - (\beta_0/s) < t + 1 \quad \text{when } s > -\beta_0. \tag{9}$$

Two conclusions may be drawn from (8) and (9):

for $s > -\beta_0$, one has $p(u) \geqslant \beta_s + sh(u)$ for all $u$ in some neighborhood of $u_s$, with equality when $u = u_s$; (10)

the sequence $\{u_s\}$ is bounded, and any cluster point $u_\infty$ satisfies $|u_\infty - ty| = t$ and $p(u_\infty) \leqslant 0$. (11)

Since $p(u) > g(u) \geqslant 0$ for all $u \neq 0$ satisfying $|u - ty| \leqslant t$, (11) says that $(u_s, p(u_s)) \to (0,0)$. On the other hand, (10) tells us, since $\beta_s + sh$ is a function of class $\mathcal{C}^2$ on a neighborhood of $u_s$, that the vector

$$y_s = s\nabla h(u_s) = -s(u_s - ty)/|u_s - ty|$$

is a proximal subgradient of $p$ at $u_s$. Moreover $|(y_s, -1)| = (1 + s^2)^{1/2}$, so that

$$(y_s, -1)/|(y_s, -1)| \to (ty, 0)/|ty| = (y, 0).$$

The sequences $\{u_s\}$ and $\{y_s\}$ thus meet all requirements.

REMARK. Parallel to (2), let us define

$$\partial^0 p(\bar{u}) = \{ \bar{y} \in R^m \,|\, (\bar{y}, 0) \in N_E(\bar{u}, p(\bar{u})) \}.$$

Since above proof shows that $N_E(\bar{u}, p(\bar{u}))$ is the closed convex hull of the cone (7), we may conclude that

$$\partial^0 p(u) \supset \operatorname{cl} \operatorname{co} Y_0.$$

**3. Saddle points and marginal values.** Next we investigate the parameterized problem $(P_u)$ in §1 under the assumption that $C$ is a topological space and

$$\text{for each compact set } U \subset R^m \text{ and number } \alpha \in R, \text{ the set } \{(u,x) \in U \times C \mid x \text{ feasible for } (P_u), f_0(x) \leq \alpha\} \text{ is compact.} \tag{12}$$

This is true, for instance, if $f_i$ is lower semicontinuous for $i = 0, 1, \ldots, s$, continuous for $i = s+1, \ldots, m$, and all level sets of the form $\{x \in C \mid f_0(x) \leq \alpha\}$ are compact. The assumption guarantees that $(P_u)$ has an optimal solution for each $u$ such that it has a feasible solution, and that the optimal value function $p(u) = \inf(P_u)$ is lower semicontinuous.

By an *augmented multiplier vector* for $(P_{\bar{u}})$, we shall mean a $\bar{y} \in R^m$ such that, for some $\bar{x} \in C$ and $\bar{r} > 0$, $(\bar{x}, \bar{y}, \bar{r})$ is a (global) saddle point of the augmented Lagrangian $L(\bar{u}, x, y, r)$ in (3) with respect to minimization over all $x \in C$ and maximization over all $y \in R^m$ and $r > 0$. As explained in [9, §4], when such a $\bar{y}$ exists at all, then for any $\bar{r}$ sufficiently large one has $(\bar{x}, \bar{y}, \bar{r})$ a saddlepoint if and only if $\bar{x}$ is an optimal solution to $(P_{\bar{u}})$. If the functions $f_i$ are all of class $\mathcal{C}^2$ and $\bar{x}$ is an interior point of $C$, the saddle point condition on $\bar{x}$ and $\bar{y}$ implies the standard second-order necessary conditions for optimality, and it is implied in turn by the corresponding sufficient conditions [9, §5]. Thus it is a sufficient condition for optimality that is close to the classical one but applicable in a much wider setting.

THEOREM 2. *Letting $p(u) = \inf(P_u)$, suppose the quadratic growth condition* (4) *and compactness condition* (12) *are satisfied. Then for any $\bar{u}$ such that $(P_{\bar{u}})$ has a feasible solution, the conclusions of Theorem* 1 *and its corollary hold with*

$$Y = \{\bar{y} \mid \exists y^k \text{ augmented multiplier for } (P_{u^k}) \text{ with } u^k \to \bar{u},\ p(u^k) \to p(\bar{u}) \text{ and } y^k \to \bar{y}\},$$

$$Y_0 = \{\bar{y} \mid \exists \lambda_k \downarrow 0 \text{ and } y^k \text{ augmented multiplier for } (P_{u^k}) \text{ with } u^k \to \bar{u},\ p(u^k) \to p(\bar{u}) \text{ and } \lambda_k y^k \to \bar{y}\}.$$

PROOF. This is immediate from the fact cited in §1 from [9, Theorem 5], that $y^k$ is a proximal subgradient of $p$ at $u^k$ if and only if it is, in the terminology above, an augmented multiplier vector for $(P_{u^k})$.

Although Theorem 2 ostensibly deals only with a rather special type of parameterization, it actually covers much more. To illustrate this, consider

$$(Q_\omega) \begin{cases} \text{minimize } F_0(\omega, x) \text{ over all } x \in C \text{ such that } F_i(\omega, x) \leq 0 \quad \text{for} \\ i = 1, \ldots, s \quad \text{and } F_i(\omega, x) = 0 \quad \text{for } i = s+1, \ldots, m. \end{cases}$$

Here $\omega$ is a parameter vector ranging over a set $\Omega$ in $R^d$, say, and we want to study subdifferential properties of $q(\omega) = \inf(Q_\omega)$. The same situation can be represented in the earlier format as

$$(Q'_\omega) \begin{cases} \text{minimize } F_0(\omega', x) \text{ over all } (\omega', x) \in D \text{ satisfying} \\ -\omega'_l + \omega_l = 0 \quad \text{for } l = 1, \ldots, d, \end{cases}$$

where $D$ is the set of $(\omega', x) \in \Omega \times C$ satisfying $F_i(\omega', x) \leq 0$ for $i = 1, \ldots, s$ and $F_i(\omega', x) = 0$ for $i = s+1, \ldots, m$. The augmented Lagrangian is then

$$\Lambda(\omega, \omega', x, \eta, r) = F_0(\omega', x) + \eta \cdot (\omega - \omega') + \frac{r}{2}|\omega - \omega'|^2 \quad \text{for } (\omega', x) \in D, \eta \in R^d, r > 0.$$

An augmented multiplier vector corresponding to $\overline{\omega}$ is a vector $\overline{\eta}$ which with any (globally) optimal solution $(\overline{\omega}', \overline{x})$ to $(Q'_{\overline{\omega}})$ and $\overline{r}$ sufficiently large forms a saddle point of $\Lambda$ with respect to minimization in $(\overline{\omega}', x) \in D$ and maximization in $\eta$. Since the optimal solutions to $(Q'_{\overline{\omega}})$ are of the form $(\overline{\omega}', \overline{x})$ where $\overline{x}$ is optimal for $(Q_{\overline{\omega}})$ and $\overline{\omega}' = \overline{\omega}$, the condition reduces to the following: $\overline{\eta}$ is an augmented multiplier vector corresponding to the parameter vector $\overline{\omega}$ if and only if for some (every) optimal solution $\overline{x}$ to $(Q_{\overline{\omega}})$, the minimum of

$$F_0(\omega', x) + \overline{\eta}(\omega' - \overline{\omega}) + \frac{r}{2}|\omega' - \overline{\omega}|^2$$

over all $\omega' \in \Omega$ and $x \in C$ satisfying

$$F_i(\omega', x)\begin{cases} \leqslant 0 & \text{for } i = 1, \dots, s, \\ = 0 & \text{for } i = s+1, \dots, m. \end{cases}$$

is attained at $(\overline{\omega}, \overline{x})$.

Obviously then, as a corollary to Theorem 2 one gets a parallel result for $q(\omega) = \inf(Q_\omega)$ in terms of augmented multiplier vectors of the latter type.

**4. Extended parameterization and the classical optimality conditions.** Our results about limits of augmented multiplier vectors take on special significance when translated by the device just described into the context of the problem

$$(Q_{w,u,v})\begin{cases} \text{minimize } f_0(v,x) + w \cdot x \text{ over all } x \in R^n, \\ \text{satisfying } f_i(v,x) + u_i \begin{cases} \leqslant 0 & \text{for } i = 1, \dots, s; \\ = 0 & \text{for } i = s+1, \dots, m; \end{cases} \end{cases}$$

where $v$ and $w$ are parameter vectors in $R^d$ and $R^n$, respectively, and the functions $f_i$ are all of class $\mathcal{C}^2$ on $R^d \times R^n$. Here we are interested in $q(w,u,v) = \inf(Q_{w,u,v})$, but properties of this obviously have bearing on more restricted parameterizations corresponding to $q(0,u,v), q(0,0,v), q(0,u,0)$, and so forth.

In $(Q_{w,u,v})$ there is no provision for a nontrivial abstract constraint $x \in C$, so there is actually somewhat less generality than in $(P_u)$, despite appearances. The inclusion of the "tilt" term $w \cdot x$, however, will allow us to characterize the subgradients of $q$ in terms of more classical kinds of Lagrange multiplier vectors.

We shall need to assume for $(Q_{w,u,v})$ the analog of the boundedness condition (12) for $(P_u)$. Since the functions $f_i$ are continuous and

$$\min\{w \cdot x \mid |w|\rho\} = -\rho|x|$$

for any $\rho > 0$, all that we need is the following:

for every bounded set $U \subset R^m$ and $V \subset R^d$ and every real number $\rho > 0$ and $\alpha$, the set of all $(u, v, x) \in U \times V \times R^n$ satisfying

$$f_i(v,x) + u_i \begin{cases} \leqslant 0 & \text{for } i = 1, \dots, s, \\ = 0 & \text{for } i = 1+1, \dots, n \end{cases} \tag{13}$$

$$f_0(v,x) - \rho|x| \leqslant \alpha$$

is bounded.This ensures that $(Q_{w,u,v})$ has an optimal solution when it has a feasible solution, and that the function $q$ is lower semicontinuous. (Actually, we are mainly interested in $w$ in a neighborhood of 0, and for this purpose it would be enough to consider in (13) a fixed, small $\rho > 0$.)

Our analysis will center on the standard second-order sufficient conditions [12, Chapter 1] for $\overline{x}$ to be a locally optimal solution to $(Q_{\overline{w},\overline{u},\overline{v}})$. These involve the function

$$l(v,x,y) = f_0(v,x) + \sum_{i=1}^{m} y_i f_i(v,x),$$

its gradient $\nabla_x l$ and Hessian matrix $\nabla_x^2 l$. They call for the existence of $\bar{y} \in R^m$ such that

(a) $\bar{y}_i \geq 0$, $f_i(\bar{v},\bar{x}) + \bar{u}_i \leq 0$, $\bar{y}_i[f_i(\bar{v},\bar{x}) + \bar{u}_i] = 0$ for $i = 1, \ldots, s$,
$f_i(\bar{v},\bar{x}) + \bar{u}_i = 0$ for $i = s+1, \ldots, m$,

(b) $\nabla_x l(\bar{v},\bar{x},\bar{y}) + \bar{w} = 0$,

(c) $h \cdot \nabla_x^2 l(\bar{v},\bar{x},\bar{y})h > 0$ for every nonzero $h \in R^n$ satisfying

$$h \cdot \nabla_x f_i(\bar{v},\bar{x}) \begin{cases} \leq 0 & \text{for } i \in I_0(\bar{u},\bar{v},\bar{x},\bar{y}), \\ = 0 & \text{for } i = I_1(\bar{u},\bar{v},\bar{x},\bar{y}), \end{cases}$$

where $I_0(\bar{u},\bar{v},\bar{x},\bar{y})$ consists of the indices $i \in \{1, \ldots, s\}$ such that $f_i(\bar{v},\bar{x}) + \bar{u}_i = 0$ and $\bar{y}_i = 0$, and $I_1(\bar{u},\bar{v},\bar{x},\bar{y})$ consists of the indices $i \in \{1, \ldots, s\}$ such that $f_i(\bar{v},\bar{x}) + \bar{u}_i = 0$ and $\bar{y}_i > 0$, together with all the indices $i \in \{s+1, \ldots, m\}$.

In the case where $\bar{x}$ and $\bar{y}$ satisfy these conditions and $\bar{x}$ is not only a locally optimal solution but the *unique globally* optimal solution to $(Q_{\bar{w},\bar{u},\bar{v}})$, we shall write

$$(\bar{x},\bar{y}) \in S^*(\bar{w},\bar{u},\bar{v}).$$

Two sets of limits will concern us:

$$\begin{aligned} S(\bar{w},\bar{u},\bar{v}) = \{(\bar{x},\bar{y}) \mid \exists (w^k,u^k,v^k) \to (\bar{w},\bar{u},\bar{v}) \quad &\text{and} \\ (x^k,y^k) \in S^*(w^k,u^k,v^k) \quad &\text{with } x^k \to \bar{x} \quad \text{and } y^k \to \bar{y}\}, \\ S_0(\bar{w},\bar{u},\bar{v}) = \{(\bar{x},\bar{y}) \mid \exists (w^k,u^k,v^k) \to (\bar{w},\bar{u},\bar{v}), \lambda_k \downarrow 0, \quad &\text{and} \\ (x^k,y^k) \in S^*(w^k,u^k,v^k) \quad &\text{with } x^k \to \bar{x} \quad \text{and } \lambda_k y^k \to \bar{y}\}. \end{aligned} \tag{14}$$

Note that for $(\bar{x},\bar{y})$ in $S(\bar{w},\bar{u},\bar{v})$, conditions (a) and (b) still hold by continuity. The same is true for $(\bar{x},\bar{y})$ in $S_0(\bar{w},\bar{v},\bar{u})$, except that $l$ is replaced in (b) by the degenerate Lagrangian

$$l_0(v,x,y) = \sum_{i=1}^m y_i f_i(v,x), \tag{15}$$

so as to get

$$\text{(b}_0\text{)} \qquad \nabla_x l_0(\bar{v},\bar{x},\bar{y}) + \bar{w} = 0.$$

THEOREM 3. *Let $q(w,u,v) = \inf(Q_{w,u,v})$ with every $f_i$ of class $\mathcal{C}^2$, and suppose the boundedness condition* (13) *holds. For any $(\bar{w},\bar{u},\bar{v})$ such that $(Q_{\bar{w},\bar{u},\bar{v}})$ has a feasible solution, let*

$$M = \{(\bar{x},\bar{y},\bar{z}) \mid \bar{x} \textit{ is (globally) optimal for } (Q_{\bar{w},\bar{u},\bar{v}}) \textit{ with } (\bar{x},\bar{y}) \in S(\bar{w},\bar{v},\bar{u}) \textit{ and } \bar{z} = \nabla_v l(\bar{v},\bar{x},\bar{y})\},$$

$$M_0 = \{(0,\bar{y},\bar{z}) \mid \exists \bar{x} \textit{ (globally) optimal for } (Q_{\bar{w},\bar{u},\bar{v}}) \textit{ with } (\bar{x},\bar{y}) \in S_0(\bar{w},\bar{v},\bar{u}) \textit{ and } \bar{z} = \nabla_v l_0(\bar{v},\bar{x},\bar{y})\}.$$

*Then $M$ and $M_0$ are closed sets satisfying $(0,0,0) \in M_0 \supset 0^+ M$, and it is impossible that both $M = \emptyset$ and $M_0 = \{(0,0,0)\}$. Moreover*

$$\partial q(\bar{w},\bar{u},\bar{v}) = \operatorname{cl}\operatorname{co}[M + M_0].$$

COROLLARY. *For $\partial q(\bar{w},\bar{u},\bar{v})$ to be nonempty and bounded ($q$ Lipschitz continuous on a neighborhood of $(\bar{w},\bar{u},\bar{v})$), it is necessary and sufficient that $M_0 = \{(0,0,0)\}$. Then*

$$\partial q(\bar{w},\bar{u},\bar{v}) = \operatorname{co} M.$$

*This holds in particular if no optimal solution $\bar{x}$ to $(Q_{\bar{w},\bar{u},\bar{v}})$ has a $\bar{y} \neq 0$ satisfying (a) and $(b_0)$.*

Notice that Theorem 3 does *not* require $(Q_{w,u,v})$ to satisfy the quadratic growth condition corresponding to the one in Theorem 2, namely that

$$\liminf_{|(w,u,v)| \to \infty} \frac{q(w,u,v)}{|(w,u,v)|^2} > -\infty. \tag{16}$$

This condition nevertheless plays a role temporarily in the proof.

PROOF OF THEOREM 3. Assume for the time being that (16) does hold. We begin by expressing $(Q_{w,u,v})$ equivalently in the form

$$(Q'_{w,u,v}) \begin{cases} \text{minimize } f_0(v',x) + w' \cdot x \text{ over all } (w',v',x) \text{ satisfying} \\ \quad f_i(v',x) + u_i \begin{cases} \leq 0 & \text{for } i = 1, \ldots, s, \\ = 0 & \text{for } i = s+1, \ldots, m, \end{cases} \\ \quad -v'_t + v_t = 0 \quad \text{for } t = 1, \ldots, d, \\ \quad -w'_j + w_j = 0 \quad \text{for } j = 1, \ldots, n, \end{cases}$$

which we identify as a specially structured case of the parameterization model $(P_u)$ treated earlier. In this framework our boundedness condition (13) corresponds to (12) and the quadratic growth condition (16) corresponds to (4). Theorem 2 is therefore applicable and gives a characterization of $\partial q(\bar{w},\bar{u},\bar{v})$ in terms of limits of augmented multiplier vectors $(\xi^k, y^k, z^k)$ for problems $(Q'_{w^k,u^k,v^k})$. Our result will be obtained by a closer analysis of such vectors.

To simplify notation, we focus on $(\bar{w},\bar{u},\bar{v})$ instead of $(w^k,u^k,v^k)$ as the parameter vector and look at the augmented Lagrangian for $(Q'_{\bar{w},\bar{u},\bar{v}})$. The latter turns out to be

$$\begin{aligned} \mathcal{L}_{\bar{u},\bar{v},\bar{w}}(w',v',x,\xi,y,z,r) = {} & L_{\bar{u}}(v',x,y,r) + w' \cdot x \\ & + z \cdot (\bar{v} - v') + \frac{r}{2}|\bar{v} - v'|^2 + \xi \cdot (\bar{w} - w') + \frac{r}{2}|\bar{w} - w'|^2, \end{aligned}$$

where $L_u(v',x,y,r)$ is the same as the function in (2), except that $f_i(v',x)$ appears in place of the earlier $f_i(x)$. We claim that $(\bar{w}',\bar{v}',\bar{x},\bar{\xi},\bar{y},\bar{z},\bar{r})$ is a (global) saddle point in (17) with respect to minimization in $(w',v',x)$ and maximization in $(\xi,y,z,r), r > 0$, if and only if $\bar{r}$ is sufficiently large and

$$(\bar{x},\bar{y}) \in S^*(\bar{w},\bar{u},\bar{v}), \quad \bar{z} = \nabla_v l(\bar{v},\bar{x},\bar{y}), \quad \bar{\xi} = \bar{x}, \quad \bar{v}' = \bar{v} \quad \text{and } \bar{w}' = \bar{w}. \tag{18}$$

In proving this, we investigate the optimality conditions for $(Q'_{\bar{w},\bar{u},\bar{v}})$ that correspond to (a), (b), (c) for $(Q_{\bar{w},\bar{u},\bar{v}})$. In these the role of $l$ is taken by

$$l'(w',v',x',\xi,y,z) = l(v',x,y) + w' \cdot x - \xi \cdot w' - z \cdot v'$$

and its gradient and Hessian with respect to $(w',v',x)$. For the local optimality of $(\bar{w}',\bar{v}',\bar{x})$, the conditions ask for the existence of $(\bar{\xi},\bar{y},\bar{z}) \in R^n \times R^m \times R^d$ such that

(a′) $\bar{y}_i \geq 0,\ f_i(\bar{v}',\bar{x}) + \bar{u}_i \leq 0,\ \bar{y}_i[f_i(\bar{v}',\bar{x}) + \bar{u}_i] = 0$ for $i = 1, \ldots, s$,

$$\begin{aligned} f_i(\bar{v}',\bar{x}) + \bar{u}_i &= 0 && \text{for } i = s+1, \ldots, m, \\ -\bar{v}'_t + \bar{v}_t &= 0 && \text{for } t = 1, \ldots, d, \\ -\bar{w}'_j + \bar{w}_j &= 0 && \text{for } j = 1, \ldots, n, \end{aligned}$$

(b′) $\nabla_x l(\bar{v}', \bar{x}, \bar{y}) = 0, \nabla_v l(\bar{v}', \bar{x}, \bar{y}) - \bar{z} = 0, \bar{x} - \bar{\xi} = 0,$

(c′) $h \cdot \nabla_x^2 l(\bar{v}', \bar{x}, \bar{y})h > 0$ for every nonzero $(h'', h', h) \in R^n \times R^d \times R^n$ satisfying

$$h \cdot \nabla_x f_i(\bar{v}', \bar{x}) \begin{cases} \leqslant 0 & \text{for } i \in I_0(\bar{u}, \bar{v}', \bar{x}, \bar{y}), \\ = 0 & \text{for } i \in I_1(\bar{u}, \bar{v}', \bar{x}, \bar{y}), \end{cases}$$

$$h'_t = 0 \quad \text{for } t = 1, \ldots, d,$$

$$h''_j = 0 \quad \text{for } j = 1, \ldots, n.$$

Clearly then, (18) is equivalent to these conditions along with the stipulation that $(\bar{w}', \bar{v}', \bar{x})$ be the *unique globally* optimal solution to $(Q'_{\bar{w},\bar{v},\bar{u}})$ (cf. the definition of $S^*$ above). According to a result we obtained in [8, Theorem 6], these properties imply that $(\bar{w}', \bar{v}', \bar{x}, \bar{\xi}, \bar{u}, \bar{z}, \bar{r})$ is a saddle point of the augmented Lagrangian for $(Q'_{\bar{w},\bar{v},\bar{u}})$ when $\bar{r}$ is sufficiently large. (The cited result speaks of the globally optimal solution being unique in the stronger sense that also every asymptotically minimizing sequence converges to it, but there is no difference here between this and simple uniqueness, by virtue of assumption (13) and the continuity of the functions $f_i$.)

We have just verified the *sufficiency* of (18) for a saddle point of the augmented Lagrangian (17). As for the necessity, we already know from general theory [8, §5] that if $(\bar{w}', \bar{v}', \bar{x}', \bar{\xi}, \bar{y}, \bar{z}, \bar{r})$ is a saddle point, then $(\bar{w}', \bar{v}', \bar{x})$ is globally optimal and the first and second-order *necessary* conditions for local optimality in $(Q'_{\bar{w},\bar{v},\bar{u}})$ must hold. These conditions coincide with (a′), (b′), (c′), except that only $h \cdot \nabla_x^2 l(\bar{v}', \bar{x}, \bar{y})h \geqslant 0$ is asserted. In particular they require $(\bar{w}', \bar{v}', \bar{x}) = (\bar{w}, \bar{v}, \bar{\xi})$ in (a′) and (b′). The saddle point property therefore precludes the existence of any globally optimal solution $(\tilde{w}', \tilde{v}', \tilde{x})$ different from $(\bar{w}', \bar{v}', \bar{x})$ (since $(\tilde{w}', \tilde{v}', \tilde{x}, \bar{\xi}, \bar{y}, \bar{z}, \bar{r})$ would have to be another saddle point in that case and again satisfy the necessary conditions in question). This saddle point property certainly implies all of (18) except conceivably for the strict inequality in (c). The latter can be deduced, however, from the special form of the augmented Lagrangian. If $(\bar{w}', \bar{v}', \bar{x}, \bar{\xi}, \bar{y}, \bar{z}, \bar{r})$ is a saddle point of $\mathfrak{L}$ in (17) with $\bar{\xi} = \bar{x}$, $\bar{w}' = \bar{w}$ and $\bar{v}' = \bar{v}$, then in particular $(\bar{x}, \bar{y})$ is a saddle point of the function

$$\begin{aligned} \lambda(x, y) &= \min_{w'} \mathfrak{L}(w', \bar{v}, x, \bar{\xi}, y, \bar{z}, \bar{r}) \\ &= L_{\bar{u}}(\bar{v}, x, y, \bar{r}) + \bar{w} \cdot x - (1/2\bar{r})|x - \bar{x}|^2. \end{aligned}$$

We showed in [8, §5] that if $(\bar{x}, \bar{y})$ is a saddle point of $L_{\bar{u}}(\bar{v}, x, y, \bar{r})$, then, among other things, condition (c) must hold at least with weak inequality. Now the same argument can be transferred to the case of $\lambda(x, y)$ simply by regarding the initial term $f_0(\bar{v}, x)$ in $L_{\bar{u}}(\bar{v}, x, y, \bar{r})$ as having been replaced by

$$f_0(\bar{v}, x) + \bar{w} \cdot x - (1/2\bar{r})|x - \bar{x}|^2.$$

Then in the Hessian condition, $h \cdot \nabla_x^2 l(\bar{v}, \bar{x}, \bar{y})h$ turns into

$$h \cdot \nabla_x^2 l(\bar{v}, \bar{x}, y)h - (1/2\bar{r})|h|^2.$$

The argument tells us that this expression is nonnegative for all vectors $h \neq 0$ satisfying the constraints in (c). But then indeed $h \cdot \nabla_x^2 l(\bar{v}, \bar{x}, y)h > 0$ for all such $h$. Therefore the full condition (18) is also *necessary* for the saddle point property of $\mathfrak{L}$ in (17), as claimed. It follows that (18) characterizes the augmented multiplier vectors which appear when Theorem 2 as applied to $(Q'_{w,v,u})$.

Summarizing what has been accomplished so far in the proof of Theorem 3, we have shown that under the additional assumption of the quadratic growth condition (16), we have via Theorem 2 that the assertions in Theorem 3 would at least be valid if $M$ and

$M_0$ were replaced by

$$\tilde{M} = \{(\bar{\xi}, \bar{y}, \bar{z}) \mid \exists (w^k, u^k, v^k) \to (\bar{w}, \bar{v}, \bar{u}) \text{ and } (\xi^k, y^k, z^k) \to (\bar{\xi}, \bar{y}, \bar{z})$$

$$\text{such that } q(w^k, u^k, v^k) \to q(\bar{w}, \bar{u}, \bar{v}) \text{ and for some } x^k,$$

$$\text{one has } (x^k, y^k) \in S^*(w^k, u^k, v^k), z^k = \nabla_v l(v^k, x^k, y^k) \text{ and } \xi^k = x^k\}$$

and $\tilde{M}_0$, which is the same except that $\lambda_k(\xi^k, y^k, z^k) \to (\bar{\xi}, \bar{y}, \bar{z})$ with $\lambda_k \downarrow 0$. It takes little effort to recognize that $\tilde{M}$ and $\tilde{M}_0$ are in fact identical to $M$ and $M_0$.

The remainder of the proof consists simply of the verification that the added hypothesis (16) is superfluous to the result that has been obtained. Fixing $(\bar{w}, \bar{v}, \bar{u})$ and any $\beta < q(\bar{w}, \bar{v}, \bar{u})$, choose a neighborhood $N$ of $(\bar{w}, \bar{v}, \bar{u})$ such that

$$(w, v, u) \in N \quad \text{implies } q(w, u, v) > \beta. \tag{19}$$

Such a neighborhood exists, because $q$ is lower semicontinuous from condition (13). Let $\theta$ be a function of class $\mathcal{C}^2$ on $R$ such that $\theta$ is bounded below, and $\theta(t) = t$ when $t > \beta$. Taking $\tilde{f}_0(v, x) = \theta(f_0(v, x))$ in place of $f_0(v, x)$ in $(Q_{w,u,v})$, we get a parameterized family of problems $(\tilde{Q}_{w,u,v})$ whose optimal value function $\tilde{q}$ is given by

$$\tilde{q}(w, u, v) = \theta(q(w, u, v)).$$

Since $\theta$ is bounded below, so is $\tilde{q}$, and the quadratic growth condition is therefore satisfied trivially by $\tilde{q}$. The assertions of Theorem 3 are valid then for $\tilde{q}$ and $\tilde{f}_0$. But from (19) and the specification about $\theta$ we know that $\tilde{q}$ agrees with $q$ on $N$, and indeed for every $(w, v, u) \in N$ and every feasible solution $x$ to $(\tilde{Q}_{w,v,u})$, the function $\tilde{f}_0(v, \cdot)$ agrees with $f_0(v, \cdot)$ in some neighborhood of $x$. Inasmuch as $N$ is a neighborhood of $(\bar{w}, \bar{v}, \bar{u})$, the formula obtained for $\partial\tilde{q}(\bar{w}, \bar{u}, \bar{v})$ reduces then to the one claimed for $\partial q(\bar{w}, \bar{u}, \bar{v})$.

**5. A generic differentiability result.** It will now be shown that the full force of the formula in Theorem 3 actually is needed only for relatively few choices of $(\bar{w}, \bar{u}, \bar{v})$, provided the constraint functions in the problem are smooth enough. In this we shall be dealing with the *strong* form of the optimality conditions for $(Q_{\bar{w},\bar{u},\bar{v}})$, which consist of (a), (b), (c) together with

(d) $\quad I_0(\bar{w}, \bar{u}, \bar{v}, \bar{x}) = \emptyset$, and the gradients $\nabla_x f_i(\bar{v}, \bar{x})$ for $i \in I_1(\bar{w}, \bar{u}, \bar{v}, \bar{x})$ are linearly independent.

It is well known (and follows readily from the implicit function theorem) that when (a), (b), (c) and (d) hold for some $(\bar{w}, \bar{u}, \bar{v})$ and $(\bar{x}, \bar{y})$, then for every $(\bar{w}', \bar{u}', \bar{v}')$ in some neighborhood of $(\bar{w}, \bar{u}, \bar{v})$ there is a *unique* pair $(\bar{x}', \bar{y}')$ satisfying (a), (b), (c), (d). Moreover the dependence of $(\bar{x}', \bar{y}')$ on $(\bar{w}', \bar{u}', \bar{v}')$ is of class $\mathcal{C}^1$, while that of the objective value $f_0(\bar{v}', \bar{x}')$ is of class $\mathcal{C}^2$. See Robinson [13].

THEOREM 4. *Again let $q(w, u, v) = \inf(Q_{w,u,v})$ and suppose the boundedness condition* (13) *holds and $f_0$ is of class $\mathcal{C}^2$, but for $i = 1, \ldots, m$ require $f_i$ to be of class $\mathcal{C}^r$ on $R^d \times R^n$, where $r = \max\{d + n, 2\}$. Let $\Omega$ be the set of all $(\bar{w}, \bar{u}, \bar{v})$ such that $(Q_{\bar{w},\bar{u},\bar{v}})$ has a unique globally optimal solution $\bar{x}$, and there is a unique $\bar{y}$ satisfying the strong optimality conditions* (a), (b), (c), (d) *along with $\bar{x}$. Let $\Omega_0$ be the set of all $(\bar{w}, \bar{u}, \bar{v})$ such that $(Q_{\bar{w},\bar{u},\bar{v}})$ fails to have this property, although it does have at least one feasible solution.*

*Then $\Omega$ is an open set on which $q$ is of class $\mathcal{C}^2$ with*

$$\nabla q(\bar{w}, \bar{u}, \bar{v}) = (\bar{x}, \bar{y}, \bar{z}) \quad \text{for } \bar{z} = \nabla_x l(\bar{v}, \bar{x}, \bar{y}),$$

*while $\Omega_0$ is a negligible set (i.e., of Lebesgue measure zero).*

PROOF. The assertions about $\Omega$ follow from a slight extension of the fact cited just prior to the statement of the theorem, namely that if $\bar{x}$ also happens to be the *unique* globally optimal solution to $(Q_{\bar{w},\bar{u},\bar{v}})$, then for $(\bar{w}',\bar{u}',\bar{v}')$ near enough to $(\bar{w},\bar{u},\bar{v})$, the corresponding $\bar{x}'$ is likewise the *unique* globally optimal solution to $(Q_{\bar{w}',\bar{u}',\bar{v}'})$ (and hence in particular $q(\bar{w}',\bar{u}',\bar{v}')$ equals $f_0(\bar{v}',\bar{x}')$ and exhibits $\mathcal{C}^2$ dependence on $(\bar{w}',\bar{u}',\bar{v}')$). The validity of this stems from our boundedness assumption (13), as we now demonstrate.

Conditions (a), (b), (c), (d) are known to guarantee an *isolated* local minimum, but what is more, the standard argument [12, Chapter 1] shows that this neighborhood can be taken uniformly in the parameters. In other words, there is an $\epsilon > 0$ such that for all $(\bar{w}',\bar{u}',\bar{v}')$ sufficiently near $(\bar{w},\bar{u},\bar{v})$, the corresponding $\bar{x}'$ is isolated by at least a distance of $\epsilon$ from any other locally optimal solution to $(Q_{\bar{w}',\bar{u}',\bar{v}'})$. Consider in light of this what would be the situation if $(\bar{w},\bar{u},\bar{v})$ could be approached by a sequence of vectors $(w^k,u^k,v^k)$ with corresponding $(x^k,y^k)$ satisfying (a), (b), (c), (d) (and therefore approaching $(\bar{x},\bar{y})$), but $x^k$ not *globally* optimal. For each $k$ there would exist an $\bar{x}^k$ globally optimal in $(Q_{w^k,u^k,v^k})$, $\bar{x}^k \neq x^k$, with

$$f_0(v^k,\bar{x}^k) < f_0(v^k,x^k) \to f_0(\bar{v},\bar{x}) = q(\bar{w},\bar{u},\bar{v}). \tag{20}$$

The boundedness condition (13) implies then that the sequence $\{\bar{x}^k\}$ is bounded; we can assume by passing to subsequences if necessary that $\bar{x}^k$ converges to some $\bar{x}^\infty$. Then from (20) and the continuity of every $f_i$ we may conclude that $\bar{x}^\infty$ is a feasible solution to $(Q_{\bar{w},\bar{u},\bar{v}})$ with $f_0(\bar{v},\bar{x}^\infty) \leqslant q(\bar{w},\bar{u},\bar{v})$. Thus $\bar{x}^\infty$ is a globally optimal solution to $(Q_{\bar{w},\bar{u},\bar{v}})$. If now it is true that $\bar{x}$ is the only globally optimal solution, we have $\bar{x}^\infty = \bar{x}$, so that the sequences $\{\bar{x}^k\}$ and $\{x^k\}$ converge to the same point. In this case $\bar{x}^k$ would eventually be within distance $\epsilon$ of $x^k$ in contradiction to there being no other locally optimal solution to $(Q_{w^k,u^k,v^k})$ within that distance. This verifies what we claimed at the start of the proof. The formula for $\nabla q$ on $\Omega$ can be viewed, of course, as a specialization of Theorem 3 to this case, as well as simply a consequence of the implicit function theorem being applied to the optimality conditions.

The proof of the assertions about $\Omega_0$ is two-pronged. First we apply a result of Spingarn and Rockafellar [9] to conclude that under our differentiability assumptions, the set of $(\bar{w},\bar{u},\bar{v})$ for which $(Q_{\bar{w},\bar{u},\bar{v}})$ has a locally optimal solution $\bar{x}$ *not* satisfying (a), (b), (c), (d) for some $\bar{y}$ is a set of Lebesgue measure zero. (The result in [9] is stated for inequality constrained problems only, but the generalization to equality constraints is obvious. It is also stated without a parameter vector $\bar{v}$, but this can be handled by introducing a decision vector $v'$ constrained by $v' = \bar{v}$, much as in the representation $(Q'_{\bar{w},\bar{u},\bar{v}})$ of $(Q_{\bar{w},\bar{u},\bar{v}})$ utilized in the preceding section. Alternatively, one can invoke more powerful theorems of Spingarn [14].)

The second part of the argument concerning $\Omega_0$ consists in showing that the set of $(\bar{w},\bar{u},\bar{v})$ for which $(Q_{\bar{w},\bar{u},\bar{v}})$ has a feasible solution (and hence a globally optimal solution), but not a *unique* globally optimal solution, is a set of Lebesgue measure zero. With the fact just cited, this will prove $\Omega_0$ is itself of measure zero.

We shall use convex analysis based on the observation that $q(w,u,v)$ is concave in $w$:

$$q(w,u,v) = \min_{x \in R^n} \{ w \cdot x - g_{u,v}(x) \} = g^*_{u,v}(w) \tag{21}$$

(concave conjugate function [11]), where

$$g_{u,v}(x) = \begin{cases} -f_0(v,x) & \text{if } x \text{ is feasible for } (Q_{w,u,v}), \\ -\infty & \text{otherwise.} \end{cases}$$

The boundedness condition (13) ensures that $q(w,u,v)$ is never $-\infty$, and that the minimum in (21) is indeed attained. Notice then from concavity that for each $(\bar{u}, \bar{v})$, either $q(w, \bar{u}, \bar{v})$ is finite for all $w$ or $+\infty$ for all $w$ [11, Theorem 7.2]. If we can show that for each fixed $(\bar{u}, \bar{v})$ with $q(\cdot, \bar{u}, \bar{v})$ finite, the set of all $\bar{w}$ for which $(Q_{\bar{w},\bar{u},\bar{v}})$ has more than one globally optimal solution is a set of measure zero, the desired conclusion can be obtained by integration with respect to $(\bar{u}, \bar{v})$. (A set in a product space is of measure zero if and only if all its sections with respect to one of the spaces are of measure zero; Fubini's theorem.)

Consider then the case where there are two distinct globally optimal solutions $\bar{x}$ and $\bar{x}'$ to $(Q_{\bar{w},\bar{u},\bar{v}})$. Then for $(\bar{w}, \bar{u}, \bar{v})$ the minimum in (21) is attained by both $\bar{x}$ and $\bar{x}'$. This tells us that

$$q(w, \bar{u}, \bar{v}) \leqslant w \cdot \bar{x} + g_{u,v}(\bar{x}) \quad \text{for all } w, \text{ with equality for } \bar{w},$$

or equivalently

$$q(w, \bar{u}, \bar{v}) \leqslant q(\bar{w}, \bar{u}, \bar{v}) + (w - \bar{w})\bar{x} \quad \text{for all } w,$$

and the same for $\bar{x}'$. Therefore both $\bar{x}$ and $\bar{x}'$ are subgradients of $q(\cdot, \bar{u}, \bar{v})$ at $\bar{w}$. It follows that $q(\cdot, \bar{u}, \bar{v})$ is not differentiable at $\bar{w}$ [11, Theorem 25.1]. But a finite concave function is differentiable except on a set of measure zero [11, Theorem 25.5]. Hence the multiplicity of globally optimal solutions can occur only for $\bar{w}$ in a set of measure zero.

## References


[1] Clarke, F. H. (1975). Generalized Gradients and Applications. *Trans. Amer. Math. Soc.* **205** 247–262.

[2] Rockafellar, R. T. (1979). Clarke's Tangent Cones and the Boundaries of Closed Sets in $R^n$. *Nonlin. Anal. Th. Math. Appl.* **3** 145–154.

[3] ———. (1980). Generalized Directional Derivatives and Subgradients of Nonconvex Functions. *Canad. J. Math.* **32** 257–280.

[4] ———. (1979). Directionally Lipschitzian Functions and Subdifferential Calculus. *Proc. London Math. Soc.* **39** 331–355.

[5] ———. (1979). *La Théorie des Sous-gradients et ses Applications à l'Optimization*, Collection Chaire Aisenstadt. Les Presses de l'Université de Montreal (C.P. 6129, succ. "A", Montréal, Québec, Canada H3C 3J7).

[6] Gauvin, J. (1979). The Generalized Gradient of a Marginal Function in Mathematical Programming. *Math. Oper. Res.* **4** 458–463.

[7] ——— and Dubeau, F. (to appear). Differential Properties of the Marginal Function in Mathematical Programming. *Math. Programming Studies.*

[8] Rockafellar, R. T. (1974). Augmented Lagrange Multiplier Functions and Duality in Nonconvex Programming. *SIAM J. Control Optimization* **12** 268–285.

[9] Spingarn, J. E. and Rockafellar, R. T. (1979). The Generic Nature of Optimality Conditions in Nonlinear Programming. *Math. Oper. Res.* **4** 425–430.

[10] Rockafellar, R. T. (to appear). Lagrange Multipliers and Subderivatives of Optimal Value Functions in Nonlinear Programming. *Math. Programming Studies.*

[11] ———. (1970). *Convex Analysis.* Princeton University Press, Princeton, New Jersey.

[12] Hestenes, M. R. (1966). *Calculus of Variations and Optimal Control Theory.* Wiley.

[13] Robinson, S. M. (1974). Perturbed Kuhn-Tucker Points and Rates of Convergence for a Class of Nonlinear-Programming Algorithms. *Math. Programming* **7** 1–16.

[14] Spingarn, J. E. (1977). Generic Conditions for Optimality in Constrained Minimization Problems. Dissertation, University of Washington, Seattle.


DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WASHINGTON, SEATTLE, WASHINGTON 98195