Mathematical Programming 26 (1983) 245–286
North-Holland

MARGINAL VALUES AND SECOND-ORDER NECESSARY CONDITIONS FOR OPTIMALITY

R. Tyrrell ROCKAFELLAR*

Department of Mathematics, University of Washington, Seattle, WA, U.S.A.

Received 2 April 1981
Revised manuscript received 5 August 1982

Second-order necessary conditions in nonlinear programming are derived by a new method that does not require the usual sort of constraint qualification. In terms of the multiplier vectors appearing in such second-order conditions, an estimate is obtained for the generalized subgradients of the optimal value function associated with a parameterized nonlinear programming problem. This yields estimates for 'marginal values' with respect to the parameters. The main theoretical tools are the augmented Lagrangian and, despite the assumption of second-order smoothness of objective constraints, the subdifferential calculus that has recently been developed for nonsmooth, nonconvex functions.

Key words: Optimality Conditions, Marginal Values, Sensitivity Analysis, Lagrange Multipliers.

1. Introduction

For $i = 0, 1, \dots, m$, let f_i be a function of class $\mathscr{C}^2$ on $\boldsymbol{R}^n$ and consider the parameterized nonlinear programming problem:

$$(\mathrm{P}_u) \qquad \text{minimize} \quad f_0(x) \quad \text{over all } x \in \boldsymbol{R}^n$$
$$\text{satisfying} \quad f_i(x) + u_i \begin{cases} \le 0 & \text{for } i = 1, \dots, s, \\ = 0 & \text{for } i = s+1, \dots, m, \end{cases}$$

where $u = (u_1, \dots, u_m) \in \boldsymbol{R}^m$. Let

$$p(u) = \inf(\mathrm{P}_u) \qquad \text{(optimal value)},$$
$$X(u) = \arg\min(\mathrm{P}_u) \qquad \text{(set of optimal solutions)},$$
$$F(u) = \mathrm{feas}(\mathrm{P}_u) \qquad \text{(set of feasible solutions)}.$$

Here p is a well-defined function from $\boldsymbol{R}^m$ to $[-\infty, \infty]$ (under the convention that $\inf(\mathrm{P}_u) = \infty$ when (P_u) is infeasible). To ensure that p is lower semicontinuous everywhere with $p(u) > -\infty$, we assume that the following *inf-boundedness condition* is satisfied:

$$\text{for every } u \in \boldsymbol{R}^m \text{ and } \alpha \in \boldsymbol{R}, \text{ there is a neighborhood } U \text{ of } u \text{ such that the set } \{x \in F(U) \mid f_0(x) \le \alpha\} \text{ is bounded,} \tag{1.1}$$

*Research sponsored in part by the Air Force Office of Scientific Research, Air Force Systems Command, United States Air Force, under grant No. F4960-82-k-0012.

where notationally

$$F(U) := \bigcup_{u' \in U} F(u') = \{x \mid \exists\, u' \in U \text{ with } x \in F(u')\}.$$

This condition implies at the same time that $X(u)$ is a nonempty compact set for every u having $p(u) < \infty$.

The question we want to address is the relationship between the generalized subgradients and subderivatives of p at u and the Lagrange multiplier vectors y that correspond to various optimality conditions on the elements x of $X(u)$. We aim at using this relationship to establish the necessity of certain second-order optimality conditions of a new sort.

Subgradients of arbitrary lower semicontinuous functions, which like p need not be convex, were first defined in a robust manner by Clarke [2]. They were characterized by Rockafellar [15] in terms of certain generalized directional derivatives, called *subderivatives*. At a point $u \in \mathbf{R}^m$ where $p(u) < \infty$, the subderivative of p with respect to a vector $h \in \mathbf{R}^m$ is

$$\begin{aligned} p^{\uparrow}(u; h) &:= \lim_{\epsilon \to 0^+} \Bigg[\limsup_{\substack{t \to 0^+ \\ (u', p(u')) \to (u, p(u))}} \Bigg[\inf_{|h' - h| \le \epsilon} \frac{p(u' + th') - p(u')}{t} \Bigg]\Bigg] \\ &= \inf_{\{h^k\} \to h} \Bigg[\sup_{\substack{\{t_k\} \to 0^+ \\ \{(u^k, p(u^k))\} \to (u, p(u))}} \Bigg[\lim_{k \to \infty} \frac{p(u^k + t_k h^k) - p(u^k)}{t_k} \Bigg]\Bigg]. \end{aligned} \tag{1.2}$$

It was shown in [15] that as a function of h, $p^{\uparrow}(u; h)$ is always lower semicontinuous, *convex* and positively homogeneous. The set of (*generalized*) *subgradients* of p at u is

$$\partial p(u) := \{y \in \mathbf{R}^m \mid p^{\uparrow}(u; h) \ge y \cdot h \quad \text{for all } h \in \mathbf{R}^m\}, \tag{1.3}$$

which is a closed convex set. One has

$$p^{\uparrow}(u; h) = \sup\{y \cdot h \mid y \in \partial p(u)\} \quad \text{if } \partial p(u) \ne \emptyset, \tag{1.4}$$

$$p^{\uparrow}(u; h) = +\infty \text{ or } -\infty \quad \text{for every } h \text{ if } \partial p(u) = \emptyset. \tag{1.5}$$

The theory of such subgradients and subderivatives has undergone much development; see [16] for an exposition. For present purposes we mention only the following facts. If p is convex, then $\partial p(u)$ agrees with the subgradient set of convex analysis. The case where $\partial p(u)$ consists of a solitary vector y is exactly the case where p is *strictly* differentiable [16] at u with $\nabla p(u) = \{y\}$. Finally, a necessary and sufficient condition for p to be Lipschitz continuous (in particular finite) in a neighborhood of u is that $\partial p(u)$ be nonempty and bounded [14].

These relations underscore the fundamental significance of the expressions $p^{\uparrow}(u; h)$, which can be interpreted as *generalized marginal values*. Estimates for such values can be derived via (1.4) from estimates for $\partial p(u)$.

Apart from convex programming and some situations in nonconvex programming where p happens to be a $\mathscr{C}^2$ function, the first strong results relating behavior of p to Lagrange multiplier vectors were obtained by Gauvin [4]. To formulate these and set the stage for other results in this paper, we introduce the functions

$$\begin{aligned} l(x, y) &= f_0(x) + \sum_{i=1}^{m} y_i f_i(x), \\ l_0(x, y) &= \sum_{i=1}^{m} y_i f_i(x). \end{aligned} \tag{1.6}$$

For each $x \in F(u)$, we consider the active index set

$$I(u, x) = \{i \in [1, m] \mid f_i(x) + u_i = 0\} \supset \{s+1, \ldots, m\} \tag{1.7}$$

and the first-order multiplier sets

$$K^1(u, x) = \{y \in \boldsymbol{R}^m \mid \nabla_x l(x, y) = 0 \quad \text{and} \quad y_i \geq 0 \quad \text{for} \quad i \in [1, s], \text{ with } y_i = 0 \text{ if } i \notin I(u, x)\}, \tag{1.8}$$

$$K_0^1(u, x) = \{y \in \boldsymbol{R}^m \mid \nabla_x l_0(x, y) = 0 \quad \text{and} \quad y_i \geq 0 \quad \text{for} \quad i \in [1, s], \text{ with } y_i = 0 \text{ if } i \notin I(u, x)\}. \tag{1.9}$$

The latter is a cone containing 0.

The existence of some y in $K^1(u, x)$ is, of course, a necessary condition for x to be in the optimal solution set $X(u)$, under certain qualifications. We state for reference the following fact (cf. [10], [7, Section 4.10], [18]).

Proposition 1. *For any u and $x \in F(u)$, the following are equivalent:*

(a) *$K^1(u, x)$ is nonempty and bounded;*

(b) *$K_0^1(u, x) = \{0\}$;*

(c) *the Mangasarian–Fromovitz constraint qualification is satisfied, i.e. the gradients $\nabla f_i(x)$, $i = s+1, \ldots, m$, are linearly independent, and there is a vector $w \in \boldsymbol{R}^n$ such that*

$$\nabla f_i(x) \cdot w \begin{cases} < 0 & \text{for } i \in I(u, x),\ i \leq s, \\ = 0 & \text{for } i = s+1, \ldots, m. \end{cases}$$

The subgradient results of Gauvin may now be stated.

Theorem 1 (Gauvin [4]). *Let u be such that $p(u) < \infty$ and $K_0^1(u, x) = \{0\}$ for every $x \in X(u)$. Then p is Lipschitz continuous on a neighborhood of u, the set*

$$\bigcup_{x \in X(u)} K^1(u, x)$$

is compact, and one has

$$\partial p(u) \subset \operatorname{co}\Big[\bigcup_{x \in X(u)} K^1(u, x)\Big], \tag{1.10}$$

so that

$$p^{\uparrow}(u;h) \le \max\{y \cdot h \mid y \in K^1(u,x) \text{ for some } x \in X(u)\}. \tag{1.11}$$

The local Lipschitz continuity in Theorem 1 allows the expression (1.2) for $p^{\uparrow}(u;h)$ to be reduced to a simpler one, the *Clarke derivative*; see [15, 16]. Gauvin actually worked with such derivatives only and assumed merely that $f_i \in \mathscr{C}^1$. Extensions of Theorem 1 can be made to cases where f_i is just locally Lipschitz continuous (not necessarily differentiable), and where an abstract constraint $v \in C(u)$ is present (see Auslender [1], Rockafellar [18]).

Our present goal is in the opposite direction: we want to exploit as far as possible the assumption that $f_i \in \mathscr{C}^2$. Specifically, we look for multiplier sets $K^2(u,x)$ and $K_0^2(u,x)$, included in $K^1(u,x)$ and $K_0^1(u,x)$ but exhibiting second-order properties, such that generalizations of Proposition 1 and Theorem 1 hold. This is a completely new approach to second-order necessary conditions for optimality.

The first step towards this goal is an exact description of $\partial p(u)$ in terms of limits of certain 'augmentable' multiplier vectors associated with the (quadratic) augmented Lagrangian. We appeal here to a formula established in [17]. An extended version of the formula is presented in Section 2; a previous assumption of 'quadratic growth' is avoided, and more information is provided about the cone of *singular subgradients*

$$\partial^0 p(u) = \{y \in \mathbf{R}^m \mid y \cdot h \le 0 \quad \text{for all } h \text{ with } p^{\uparrow}(u;h) < \infty\}. \tag{1.12}$$

This cone is important because it is the recession cone of $\partial p(u)$ when $\partial p(u) \ne \emptyset$, and indeed (cf. [18, Section 3]):

$$\begin{aligned} \partial^0 p(u) = \{0\} &\Leftrightarrow p^{\uparrow}(u;h) < \infty \quad \text{for all } h \\ &\Leftrightarrow \partial p(u) \text{ is nonempty and bounded.} \end{aligned} \tag{1.13}$$

In Sections 3 and 4 the connection between augmentable Lagrange multiplier vectors and the standard kinds of first- and second-order optimality conditions is explored. The new multiplier sets $K^2(u,x)$ and $K_0^2(u,x)$ are introduced in Section 5 and shown to yield results about necessary conditions and subgradient estimates, as proposed.

The reader should know that, while the new second-order necessary conditions derived here help substantially to narrow the estimates which can be given for subderivatives of the optimal value function p, they do not exhaust what can be said towards characterizing local optimality. As far as narrowing the gap between necessity and sufficiency is concerned, and doing so in terms of just the first and second derivatives of the functions f_i at the point in question, the sharpest results so far are those which Ioffe [8, Sections 7, 8] has obtained, following a different approach due to Levitin et al. [9]. In this approach it is the expression

$$\sup\{w \cdot \nabla_x^2 l(x,y) w \mid y \in K^1(u,x)\}$$

whose nonnegativity or positivity for all w in a certain cone is at issue. The corresponding assertions about marginal values, although not fully explored, are oriented toward directional derivatives of p in the Hadamard sense, rather than the subderivitaves studied here. An exception is the recent work of Gollan [5], which does encompass certain second-order estimates for $\partial p(u)$ and $p^{\uparrow}(u; h)$ complementary to ours (see the last part of Section 5).

A final observation: although we consider only a seemingly special form of parameterization in this paper, our results can easily be extended to the case where $f_i(x)$ is replaced by $f_i(v, x)$ with v a general parameter vector in $\boldsymbol{R}^d$. The trick is simply to regard this as minimization jointly in v and x subject to an additional set of equality constraints, namely that each component of v be equal to a preassigned value. See [17] and [18, Section 3] for more details.

2. Subgradients and augmentable multiplier vectors

By the augmented Lagrangian associated with (P_u), we mean the function

$$L_u(x, y, r) = f_0(x) + \sum_{i=1}^{s} \varphi(f_i(x) + u_i, y_i, r) + \sum_{i=s+1}^{m} \psi(f_i(x) + u_i, y_i, r), \tag{2.1}$$

where

$$\psi(f_i(x) + u_i, y_i, r) = y_i[f_i(x) + u_i] + \tfrac{1}{2}r[f_i(x) + u_i]^2, \tag{2.2}$$

$$\varphi(f_i(x) + u_i, y_i, r) = \begin{cases} \psi(f_i(x) + u_i, y_i, r) & \text{if } y_i + r[f_i(x) + u_i] \geq 0, \\ -\dfrac{1}{2r} y_i^2 & \text{if } y_i + r[f_i(x) + u_i] \leq 0. \end{cases} \tag{2.3}$$

The properties of the augmented Lagrangian are of recognized importance in connection with computational methods, but here we shall be occupied with their theoretical significance.

For one thing, it is clear that regardless of the choice of $y \in \boldsymbol{R}^m$ and $r > 0$, one has

$$L_u(x, y, r) \leq f_0(x) \quad \text{for all } x \in F(u) \tag{2.4}$$

and consequently

$$\inf_{x \in C} L_u(x, y, r) \leq p(u) \quad \text{whenever } C \supset F(u). \tag{2.5}$$

Let

$$A(u) = \{y \in \boldsymbol{R}^m \mid \exists\, r > 0 \text{ and neighborhood } U \text{ of } u \text{ such that } \inf_{x \in F(U)} L_u(x, y, r) = p(u) < \infty\}. \tag{2.6}$$

The elements of $A(u)$ will be called *augmentable* Lagrange multiplier vectors.

They allow the constrained minimization in (P_u) to be reduced to an essentially unconstrained minimization in the sense of the proposition below. (Here we generalize results of Rockafellar [14, 15] that correspond to replacing $F(U)$ by all of $\boldsymbol{R}^n$ in (2.6).) Other properties of augmentable multiplier vectors, which indicate their abundance and explain their relationship to more familiar multiplier conditions, will be derived in Sections 3 and 4; for a local characterization of augmentability, see the remark after Proposition 5.

Proposition 2. *Let* $y \in A(u)$ *and let* $B \subset \boldsymbol{R}^m$ *be any bounded set with* $u \in \operatorname{int} B$. *Then* $y_i \geq 0$ *for* $i = 1, \dots, s$, *and for all* $r > 0$ *sufficiently large one has*

$$X(u) = \operatorname*{argmin}_{x \in F(B)} L_u(x, y, r) \subset \operatorname{int} F(B). \tag{2.7}$$

Proof. We note first that

$$L_u(x, y, r) = \min_{u' : x \in F(u')} \{f_0(x) - y \cdot (u' - u) + \tfrac{1}{2} r |u' - u|^2\}. \tag{2.8}$$

This implies for arbitrary $U \subset \boldsymbol{R}^m$ that

$$\inf_{x \in F(U)} L_u(x, y, r) = \inf_{u' \in U} \{p(u') - y \cdot (u' - u) + \tfrac{1}{2} r |u' - u|^2\}, \tag{2.9}$$

with x yielding the minimum on the left if and only if $x \in X(u')$ for some u' yielding the minimum on the right. (Recall that $X(u') \neq \emptyset$ when $p(u') < \infty$.) Taking $U = B$, we see that the desired relation (2.7) is equivalent to

$$\arg\min_{u' \in B} \{p(u') - y \cdot (u' - u) + \tfrac{1}{2} r |u' - u|^2\} = \{u\}, \tag{2.10}$$

since the continuity of the functions f_i ensures

$$\operatorname{int} F(B) \supset F(u) \supset X(u) \quad \text{when } u \in \operatorname{int} B.$$

On the other hand, the condition $y \in A(u)$ translates by (2.9) into the existence of some $r > 0$ and neighborhood U of u such that

$$u \in \operatorname*{argmin}_{u' \in U} \{p(u') - y \cdot (u' - u) + \tfrac{1}{2} r |u' - u|^2\}, \quad p(u) < \infty. \tag{2.11}$$

(Note that this implies $y_i \geq 0$ for $i = 1, \dots, s$, since $p(u')$ is nondecreasing with respect to u'_i, $i = 1, \dots, s$.)

The question therefore boils down to whether (2.11) holding for some $r > 0$ and neighborhood U ensures that (2.10) holds for all r sufficiently large. Choose $\epsilon > 0$ such that $|u' - u| \leq \epsilon$ implies $u' \in U$. It suffices to show that if $p(u)$ is finite and $r_0 > 0$ is such that

$$p(u') \geq p(u) + y \cdot (u' - u) - \tfrac{1}{2} r_0 |u' - u|^2 \quad \text{when } |u' - u| \leq \epsilon, \tag{2.12}$$

then for all $r > 0$ sufficiently large one will have

$$p(u') > p(u) + y \cdot (u' - u) - \tfrac{1}{2} r |u' - u|^2 \quad \text{when } u' \in B, u' \neq u. \tag{2.13}$$

Since p is lower semicontinuous on $\boldsymbol{R}^m$ and B is bounded there is a number $\alpha \in \boldsymbol{R}$ such that

$$p(u') \geq \alpha \quad \text{for all } u' \in B.$$

Also, there is a number $\beta \in \boldsymbol{R}$ such that

$$y \cdot (u' - u) - \tfrac{1}{2} r_0 |u' - u|^2 \leq \beta \quad \text{for all } u' \in \boldsymbol{R}^m.$$

If $r > r_0$ but (2.13) is violated, we would have to have $|u' - u| > \epsilon$ but

$$\begin{aligned} \alpha &\leq p(u) + y \cdot (u' - u) - \tfrac{1}{2} r |u' - u|^2 \\ &\leq p(u) + y \cdot (u' - u) - \tfrac{1}{2} r_0 |u' - u|^2 - \tfrac{1}{2}(r - r_0)|u' - u|^2 \\ &< p(u) + \beta - \tfrac{1}{2}\epsilon(r - r_0), \end{aligned}$$

so that $r < r_0 + (2/\epsilon)[p(u) - \alpha + \beta]$. This shows that (2.13) cannot be violated, when r is sufficiently large.

Proposition 2 reduces to previous results of ours in [12, 13] when a certain quadratic growth condition is satisfied, namely that for $U = \boldsymbol{R}^m$ and some choice of $y \in \boldsymbol{R}^m$ and $r > 0$, the quantity (2.9) is not $-\infty$. This growth condition was also invoked for a formula for $\partial p(u)$ which we gave in [17]. We now present a version of the formula which avoids it and at the same time says more about the cone $\partial^0 p(u)$ in (1.12) and (1.13).

The notion of "lim sup" for multifunctions will be useful. Recall that for a set $M(z)$ depending on a parameter vector z, one defines

$$\limsup_{z' \to z} M(z') = \{w \mid \exists\, z^k \to z,\ w^k \in M(z^k) \text{ with } w^k \to w\}. \tag{2.14}$$

(This limit set is always closed.) As a special case, for any $C \subset \boldsymbol{R}^m$ the set

$$\begin{aligned} 0^+C &:= \limsup_{t \to 0^+} tC \\ &= \{y \mid \exists\, y^k \in C,\ t_k \to 0^+, \text{ with } t_k y^k \to y\} \end{aligned} \tag{2.15}$$

is called the *recession cone* of C. It is a cone (i.e. closed under multiplication by positive scalars), and its constituent rays correspond to the 'direction points' of $\boldsymbol{R}^m$ that can be interpreted as 'points of C at infinity' (cf. [11, Section 8] for the case of C convex). Note that C is nonempty and bounded if and only if $0^+C = \{0\}$.

We shall write the convex hull of a set C as co C. It is elementary that

$$\text{co}[C + D] = \text{co}\, C + \text{co}\, D \quad \text{for all } C \subset \boldsymbol{R}^m,\ D \subset \boldsymbol{R}^m. \tag{2.16}$$

Theorem 2. *Let u be such that $p(u) < \infty$, and define*

$$Y(u) = \limsup_{(u', p(u')) \to (u, p(u))} A(u'), \tag{2.17}$$

$$Y_0(u) = \limsup_{\substack{t \to 0^+ \\ (u', p(u')) \to (u, p(u))}} tA(u'). \tag{2.18}$$

Then $Y(u)$ and $Y_0(u)$ are closed subsets of $\mathbf{R}^m$ such that $Y_0(u)$ is a cone with

$$0 \in Y_0(u) \quad \text{and} \quad 0^+Y(u) \subset Y_0(u). \tag{2.19}$$

One has

$$\partial p(u) = \mathrm{cl}\,\mathrm{co}[Y(u) + Y_0(u)], \quad \partial^0 p(u) \subset \mathrm{cl}\,\mathrm{co}\, Y_0(u), \tag{2.20}$$

and

$$Y(u) = \emptyset \Leftrightarrow \partial p(u) = \emptyset \Rightarrow \{0\} \neq \partial^0 p(u) = \mathrm{cl}\,\mathrm{co}\, Y_0(u). \tag{2.21}$$

Moreover

$$p^{\uparrow}(u; h) = \begin{cases} \sup_{y \in Y(u)} y \cdot h & \text{if } y \cdot h \leq 0 \text{ for every } y \in Y_0(u), \\ +\infty & \text{otherwise}, \end{cases} \tag{2.22}$$

with

$$\mathrm{int}\{h \mid p^{\uparrow}(u; h) < \infty\} = \{h \mid y \cdot h < 0 \quad \text{for all } y \in Y_0(u),\, y \neq 0\}. \tag{2.23}$$

The latter set is nonempty if and only if $Y_0(u)$ is pointed, in which event $\partial^0 p(u)$ is pointed too and one actually has

$$\partial p(u) = \mathrm{co}[Y(u) + Y_0(u)], \quad \partial^0 p(u) = \mathrm{co}\, Y_0(u). \tag{2.24}$$

This is the case in particular when there are no equality constraints in (P_u), *i.e. when $s = m$.*

Here we use the terminology that a cone $C_0 \subset \mathbf{R}^m$, not necessarily convex but containing 0, is *pointed* if 0 cannot be expressed as a sum of nonzero vectors in C_0. (When C_0 is convex, this is equivalent to the condition that if $y \in C_0$ and $y \neq 0$, then $-y \notin C_0$.) Trivially, C_0 is pointed in the degenerate case where $C_0 = \{0\}$, or if $C_0 = \mathbf{R}^m_+$. A dual characterization of pointedness is provided by the next proposition, which will be needed in proving Theorem 2.

Proposition 3. *Suppose C and C_0 are closed subsets of $\mathbf{R}^m$ such that C_0 is a cone with*

$$0 \in C_0 \quad \text{and} \quad 0^+C \subset C_0. \tag{2.25}$$

If C_0 is pointed, then co C_0 *and* co$[C + C_0]$ *are closed. Moreover, C_0 is pointed if* *and only if there exists a vector* $\bar{h}$ *satisfying*

$$y \cdot \bar{h} < 0 \quad \text{for all } y \in C_0 \text{ with } y \neq 0. \tag{2.26}$$

When $\bar{h}$ *does have this property, then*

$$\bar{h} \in \operatorname{int}\{h \mid y \cdot h \leq 0 \quad \text{for all } y \in C_0\} = \operatorname{int}\{h \mid \sup_{y \in C} y \cdot h < \infty\}. \tag{2.27}$$

Proof. Part of this, namely the assertion that co C_0 and co$[C + C_0]$ are closed when C_0 is pointed, is covered by [18, Proposition 15], but an alternative proof will be offered in the course of what follows. Condition (2.26) obviously guarantees that C_0 is pointed. We proceed to show that it implies (2.27) and the closedness of co C_0 and co$[C + C_0]$. The final stage will be an argument that if C_0 is pointed and co C_0 closed, then (2.26) holds for some $\bar{h}$.

If $\bar{h}$ satisfies (2.26) we have

$$0 > \max\{y \cdot \bar{h} \mid y \in C_0, |y| = 1\}$$

(the maximum exists because C_0 is closed). The same is then true for all h in some neighborhood of $\bar{h}$, say defined by $|h - \bar{h}| \leq \epsilon$ for some $\epsilon > 0$. Then

$$\begin{aligned} 0 &> \max\{y \cdot h \mid y \in C_0, |y| = 1, |h - \bar{h}| \leq \epsilon\} \\ &= \max\{y \cdot \bar{h} + \epsilon|y| \mid y \in C_0, |y| = 1\}, \end{aligned}$$

so (since C_0 is a cone)

$$0 > y \cdot \bar{h} + \epsilon|y| \quad \text{for all } y \in C_0,\ y \neq 0. \tag{2.28}$$

By (2.25) this also holds for nonzero $y \in 0^+C$, and the definition (2.15) of the latter cone then implies

$$-\epsilon > \limsup_{\substack{z \in C \\ |z| \to \infty}} |z|^{-1} z \cdot \bar{h}.$$

For some $\rho > 0$, therefore, we have

$$0 > z \cdot \bar{h} + \epsilon|z| \quad \text{for all } z \in C \text{ with } |z| > \rho,$$

from which can be deduced (using the closedness of C) that, for some $\alpha \in \boldsymbol{R}$ sufficiently large,

$$\alpha > y \cdot \bar{h} + \epsilon|y| \quad \text{for all } y \in C. \tag{2.29}$$

In particular, then, if $|h - \bar{h}| \leq \epsilon$ we have

$$\sup_{y \in C} h \cdot y \leq \sup_{y \in C}\{y \cdot \bar{h} + \epsilon|y|\} \leq \alpha.$$

This verifies (2.27).

Continuing with the same vector $\bar{h}$ satisfying (2.28) and (2.29), we demonstrate next that co C_0 and co$[C + C_0]$ are closed. We shall argue first that for any $\eta > 0$

there is a $\zeta > 0$ such that

$$\{y \in \text{co}\, C_0 \,\big|\, |y| \leq \eta\} = \left\{\sum_{j=1}^{m} y^j \,\Big|\, y^j \in C_0, |y^j| \leq \zeta, \left|\sum_{j=1}^{m} y^j\right| \leq \eta\right\}; \quad (2.30)$$

since the set on the right is compact, this will prove co C_0 is closed. Because C_0 is a cone in $\boldsymbol{R}^m$ containing 0, its convex hull can be represented as

$$\text{co}\, C_0 = \left\{\sum_{j=1}^{m} y^j \,\Big|\, y^j \in C_0\right\} \quad (2.31)$$

(Caratheodory's theorem [11, Section 17]). Hence for any $\zeta > 0$ the inclusion $\supset$ is true in (2.30). For the reverse inclusion we must show for arbitrary $\eta > 0$ the existence of $\zeta > 0$ such that

$$\left|\sum_{j=1}^{m} y^j\right| \leq \eta, \quad y^j \in C_0 \Rightarrow |y^j| \leq \zeta. \quad (2.32)$$

We invoke (2.28): the assumptions in (2.32) give us

$$\epsilon \sum_{j=1}^{m} |y^j| \leq -\sum_{j=1}^{m} y^j \cdot \bar{h} \leq \eta |\bar{h}|,$$

so the conclusion in (2.32) is valid for $\zeta = \eta|\bar{h}|/\epsilon$.

To demonstrate in similar fashion the closedness of the set $\text{co}[C + C_0] = \text{co}\, C + \text{co}\, C_0$, we introduce

$$V = \{(v, \lambda) \in \boldsymbol{R}^m \times \boldsymbol{R} \,\big|\, \text{either } \lambda = 0 \text{ and } v \in C_0 \text{ or } \lambda > 0 \text{ and } \lambda^{-1} v \in C\} \quad (2.33)$$

and observe (via Caratheodory's theorem again) the representation

$$\text{co}\, C + \text{co}\, C_0 = \left\{\sum_{j=0}^{2m} v^j \,\Big|\, \exists \lambda_j \text{ with } (v^j, \lambda_j) \in V, \sum_{j=0}^{2m} \lambda_j = 1\right\}. \quad (2.34)$$

Here V is closed, because C_0 and C are closed and (2.25) holds. We shall demonstrate that for arbitrary $\eta > 0$ there is a $\zeta > 0$ such that

$$\{v \in [\text{co}\, C + \text{co}\, C_0] \,\big|\, |v| \leq \eta\} = \quad (2.35)$$

$$= \left\{\sum_{j=0}^{2m} v^j \,\Big|\, \exists \lambda_j \text{ with } (v^j, \lambda_j) \in V, |v^j| \leq \zeta, \sum_{j=0}^{2m} \lambda_j = 1, \left|\sum_{j=0}^{2m} v^j\right| \leq \eta\right\}.$$

This will provide the closedness of co C + co C_0, since the set on the right in (2.35) is compact.

The inclusion $\supset$ in (2.35) is clear from the (2.34), so in order to verify (2.35) we need only establish for arbitrary $\eta > 0$ the existence of a corresponding $\zeta > 0$ such that

$$(v^j, \lambda_j) \in V, \sum_{j=0}^{2m} \lambda_j = 1, \left|\sum_{j=0}^{2m} v^j\right| \leq \eta \Rightarrow |v^j| \leq \zeta. \quad (2.36)$$

Considering elements which satisfy the hypothesis of this desired implication, let us suppose for notational simplicity that (for some index q, $0 < q \le 2m$)

$$\lambda_j \begin{cases} > 0 & \text{for } j = 0, \dots, q, \\ = 0 & \text{for } j = q+1, \dots, 2m. \end{cases}$$

Then by (2.28), (2.29), and the definition (2.33) of V

$$\begin{aligned} &\bar{h} \cdot \lambda_j^{-1} v_j + \epsilon |\lambda_j^{-1} v_j| \quad \text{for } j = 0, \dots, q, \\ &\bar{h} \cdot v^j + \epsilon |v^j| \le 0 \quad \text{for } j = q+1, \dots, 2m, \end{aligned}$$

and therefore

$$\bar{h} \cdot v^j + \epsilon |v^j| \le \lambda_j \alpha \quad \text{for all } j.$$

It follows that

$$\epsilon \sum_{j=0}^{2m} |v^j| \le \sum_{j=0}^{2m} \lambda_j \alpha - \left(\sum_{j=0}^{2m} v^j \right) \cdot \bar{h} \le \alpha + \eta |\bar{h}|,$$

so $|v^j| \le (\alpha + \eta|\bar{h}|)/\epsilon$ for all j. Hence (2.35) is true for $\zeta = (\alpha + \eta|\bar{h}|)/\epsilon$.

We are now at the last step in the proof of Proposition 3. Under the assumption that C_0 is pointed, we must show the existence of an $\bar{h}$ satisfying (2.26). The pointedness of C_0 ensures the pointedness of the convex cone $D = \text{co } C_0$ (cf. (2.31)) and, as we have already determined, the closedness of D. The polar of D is

$$\begin{aligned} D^\circ &= \{h \mid y \cdot h \le 0 \quad \text{for all } y \in D\} \\ &= \{h \mid y \cdot h \le 0 \quad \text{for all } y \in C_0\} \end{aligned} \tag{2.37}$$

and since D is closed, the polar of D° is in turn D:

$$D = \{\bar{y} \mid \bar{y} \cdot h \le 0 \quad \text{for all } h \in D^\circ\}. \tag{2.38}$$

If D° had empty interior, it would not be m-dimensional and would have to be contained in a proper subspace of $\boldsymbol{R}^m$ [11, Section 6]. Thus there would exist some $\bar{y} \ne 0$, $\bar{y} \perp D$, and by (2.38) we would have both $\bar{y}$ and $-\bar{y}$ in D, contrary to D being pointed. Hence there must be some $\bar{h} \in \text{int } D^\circ$. Obviously from (2.37), any such $\bar{h}$ satisfies (2.26).

Proof of Theorem 2. It is evident from (2.17) and (2.18) that $Y(u)$ and $Y_0(u)$ are closed sets, $Y_0(u)$ is a cone, and $0^+ Y(u) \subset Y_0(u)$. As for (2.20) and the fact that $0 \in Y_0(u)$, we proved these in [17, Theorem 2] with the functions f_i merely continuous, but assuming the 'quadratic growth condition' described above, prior to the statement of the theorem, and furthermore with $A(u)$ in (2.17) and (2.18) consisting not of augmentable multiplier vectors as in (2.6), but '*totally*' augmentable ones in the sense of yielding

$$\inf_{x \in \boldsymbol{R}^n} L_u(x, y, r) = p(u) < \infty \quad \text{for some } r > 0. \tag{2.39}$$

Proposition 2 assures us that when y is augmentable we have for arbitrary bounded $B \subset \mathbf{R}^m$ with $u \in \operatorname{int} B$:

$$\inf_{x \in F(B)} L_u(x, y, r) = p(u) < \infty \quad \text{for } r \text{ sufficiently large.} \tag{2.40}$$

The proof of Proposition 2 used the boundedness of B only to guarantee that p is bounded below on B by some $\alpha \in \mathbf{R}$. Certainly, then, if p happens to be bounded below on all of $\mathbf{R}^m$ (a special case of the 'quadratic growth condition'), any augmentable y satisfies (2.41) with $B = \mathbf{R}^m$. Since $F(\mathbf{R}^m) = \mathbf{R}^n$, we see that in this case every augmentable multiplier vector is totally augmentable. Thus when p is bounded below, our earlier result is applicable and allows us to conclude $0 \in Y_0(u)$ and (2.20).

A simple trick reduces the general case to the one where p is bounded below on $\mathbf{R}^m$: set $\alpha = p(u) - 1$ and replace f_0 by $\tilde{f}_0 = \max\{f_0, \alpha\}$. This replaces p by a new optimal value function $\tilde{p} \geq \alpha$ and the sets $A(u)$ in (2.16) and (2.17) by sets $\tilde{A}(u)$, in terms of which the desired formula is valid. Now since p was lower semicontinuous at u, and $p(u) > \alpha$, we have $p(u') > \alpha$ for all u' in some open neighborhood U_0 for u; then for all $x \in F(U_0)$ we have $f_0(x) > \alpha$, so $\tilde{f}_0(x) = f_0(x)$. This reveals that for all $u' \in U_0$, both $\tilde{p}(u') = p(u')$ and $\tilde{A}(u') = A(u')$. The formula in terms of $\tilde{p}$ and $\tilde{A}$ is thereby identical to the one asserted in terms of p and A. Thus we are sure in general that (2.20) holds and $0 \in Y_0(u)$.

The proof of the cited result [17, Theorem 2] actually established that the normal cone to the epigraph of p at $(u, p(u))$ is

$$\operatorname{cl} \operatorname{co}\{(y, \lambda) \mid \text{either } \lambda = 0 \text{ and } y \in Y_0(u), \text{ or } \lambda > 0 \text{ and } \lambda^{-1} y \in Y(u)\}.$$

On the other hand, from the general theory of subgradients the normal cone is

$$\{(y, \lambda) \mid \text{either } \lambda = 0 \text{ and } y \in \partial^0 p(u), \text{ or } \lambda > 0 \text{ and } \lambda^{-1} y \in \partial p(u)\} \tag{2.41}$$

[16, 18]. Obviously if $Y(u) = \emptyset$, we may conclude $\partial^0 p(u) = \operatorname{cl} \operatorname{co} Y_0(u)$. Since $Y(u) + Y_0(u) = \emptyset$ if and only if $Y(u) = \emptyset$ (recall $0 \in Y_0(u)$), while $\partial^0 p(u) \neq \{0\}$ when $\partial p(u) = \emptyset$ (inasmuch as $(u, p(u))$ is a boundary point of the epigraph of p, so the normal cone (2.41) cannot be just the zero vector [14, p. 149]), we may conclude the validity of (2.21). Then (2.22) follows by the fundamental formulas (1.4), (1.5), (1.12) (using the convection that $\sup \emptyset = -\infty$).

The inclusion $\subset$ in (2.23) is immediate from (2.22). Applying Proposition 3 to $C = Y(u)$ and $C_0 = Y_0(u)$, we obtain the reverse inclusion, again via (2.22), as well as the assertions in the final sentence of Theorem 2, except concerning $\partial^0 p(u)$. As for the latter, if $\bar{h}$ belongs to the set in (2.23), then by (1.12)

$$y \cdot \bar{h} < 0 \quad \text{for all } y \in \partial^0 p(u).$$

This implies by Proposition 3 that $\partial^0 p(u)$ is pointed (the case of $C_0 = \partial^0 p(u)$ in Proposition 3).

If there are no equality constraints, we have in (2.17) and (2.18) that $A(u') \subset$

$\boldsymbol{R}^m_+$ (cf. Proposition 2) and therefore $Y(u) \subset \boldsymbol{R}^m_+$, $Y_0(u) \subset \boldsymbol{R}^m_+$. In particular $Y_0(u)$ is pointed in this case.

Theorem 3. *Let u be such that $p(u) < \infty$. The condition $Y_0(u) = \{0\}$ is then a necessary and sufficient condition for p to be Lipschitz continuous on a neighborhood of u. In that case $Y(u)$ is a nonempty compact set, and one has*

$$\partial p(u) = \operatorname{co} Y(u), \tag{2.42}$$

$$p^{\uparrow}(u; h) = \max_{y \in Y(u)} y \cdot h = \limsup_{\substack{y' \in A(u') \\ u' \to u \\ p(u') \to p(u)}} y' \cdot h. \tag{2.43}$$

Proof. From general theory we know that p is Lipschitz continuous on a neighborhood of u if and only if $\partial p(u)$ is nonempty and bounded, or equivalently, $\partial^0 p(u) = \{0\}$. If the latter holds, then $Y_0(u) = \{0\}$ by (2.20) in Theorem 2. On the other hand, if $Y_0(u) = \{0\}$ (in which case $Y_0(u)$ is pointed), we have by (2.24) both $\partial^0 p(u) = \{0\}$ and $\partial p(u) = \operatorname{co} Y(u)$; moreover $Y(u) \neq \emptyset$ by (2.21) and $0^+ Y(u) = \{0\}$ by (2.10). The latter implies $Y(u)$ is compact, so that (2.22) reduces to the first equation in (2.43). The second equation in (2.43) then follows from the definition (2.17) and (2.18) of $Y(u)$ and $Y_0(u)$: Since $Y_0(u) = \{0\}$, every sequence $\{y^k\}$ having $y^k \in A(u^k)$ for some sequence $\{u^k\}$ with $u^k \to u$, $p(u^k) \to p(u)$, must be a bounded sequence whose cluster points all lie in $Y(u)$.

Remark. Nothing in this section has made use of the smoothness of the functions f_i. They could merely be continuous functions on $\boldsymbol{R}^n$. An abstract constraint $x \in C$ (where C is a nonempty closed subset of $\boldsymbol{R}^n$). could also be added, the description of the feasible set $F(u)$ being altered accordingly. In this general setting, Proposition 2, Theorem 2 and its corollary still hold with the same wording.

The smoothness of f_i enters, of course, in trying to relate augmentable Lagrange multiplier vectors (and their limits) to vectors that satisfy first and second-order optimality conditions in differential form, as we do in the next two sections. It is then that the connection between the preceding theorem and Theorem 1 (Gauvin) will become clear (Theorem 4).

3. Estimates in terms of first-order conditions

Estimates of the subgradient set $\partial p(u)$ and subderivatives $p^{\uparrow}(u; h)$ can be generated from Theorem 2 by way of estimates for the sets $Y(u)$ and $Y_0(u)$. With this purpose in mind, we now explore connections between the vectors in $Y(u)$ and $Y_0(u)$ and various Lagrange multipliers that appear in classical optimality conditions.

We start with a result about the relationship between augmentable Lagrange multiplier vectors and the first-order conditions already mentioned in Section 1. (The result only requires $f_i \in \mathscr{C}^1$.) This will lead to a generalization of Theorem 1.

Proposition 4. *For any $u \in \mathbf{R}^m$ with $p(u) < \infty$, one has*

$$A(u) \subset Y(u) \subset \bigcup_{x \in X(u)} K^1(u, x), \quad Y_0(u) \subset \bigcup_{x \in X(u)} K_0^1(u, x). \tag{3.1}$$

In the convex programming case (where f_i is convex for $i = 0, 1, \dots, s$ and affine for $i = s+1, \dots, m$), it is actually true that

$$A(u) = Y(u) = K^1(u, x) \quad \text{and} \quad Y_0(u) = K_0^1(u, x) \quad \text{for all } x \in X(u). \tag{3.2}$$

Proof. Trivially $A(u) \subset Y(u)$ by definition (2.16). Whenever $y \in A(u)$ and $x \in X(u)$, we have (by Proposition 2, when r is sufficiently large) that $L_u(\cdot, y, r)$ has a local minimum at x; this minimum is equal to $p(u)$ by (2.6). Then since $L_u(x, y', r) \le f_0(x) = p(u) = L_u(x, y, r)$ for all y' by (2.4) the function $L_u(x, \cdot, r)$ also has a local maximum at y, and it follows that

$$\nabla_x L_u(x, y, r) = 0 \quad \text{and} \quad \nabla_y L_u(x, y, r) = 0. \tag{3.3}$$

In terms of the notation

$$\eta_i = \begin{cases} \max\{f_i(x) + u_i, -y_i/r\} & \text{for} \quad i = 1, \dots, s, \\ f_i(x) + u_i & \text{for} \quad i = s+1, \dots, m, \end{cases} \tag{3.4}$$

One has

$$\begin{aligned} \nabla_x L_u(x, y, r) &= \nabla f_0(x) + \sum_{i=1}^m (y_i + r\eta_i) f_i(x), \\ \nabla_y L_u(x, y, r) &= (\eta_1, \dots, \eta_m) \end{aligned} \tag{3.5}$$

so (3.3) is equivalent to having $x \in F(u)$ and $y \in K^1(u, x)$. Thus

$$A(u) \subset K^1(u, x) \quad \text{when} \quad x \in X(u). \tag{3.6}$$

Next we recall from the inf-boundedness condition (1.1) assumed at the beginning of Section 1 that for any bounded neighborhood U of u the set $\{x \in F(U) \mid f_0(x) \le p(u) + 1\}$ is bounded. If $y \in Y(u)$, there exist by definition (2.18) sequences $y^k \to y$ and $u^k \to u$ such that $p(u^k) \to p(u)$ and $y^k \in A(u^k)$. Since $p(u) < \infty$, we have $p(u^k) < \infty$ (at least for k sufficiently large), so $X(u^k) \ne \emptyset$. Taking arbitrary $x^k \in X(u^k)$, we have $A(u^k) \subset K^1(u^k, x^k)$ by (3.6), hence $y^k \in K^1(u^k, x^k)$. Moreover $f_0(x^k) = p(u^k)$, so that for k sufficiently large we not only have $u^k \in U$ (implying $x^k \in X(U) \subset F(U)$) but $f_0(x^k) \le p(u) + 1$, i.e. x^k belongs to the bounded set $\{x \in F(U) \mid f_0(x) \le p(u) + 1\}$. The sequence $\{x^k\}$ is therefore bounded and can be assumed to converge to some x. Then by the continuity of f_i and ∇f_i we have, since $x^k \in F(u^k)$ and $y^k \in K^1(u^k, x^k)$, that $x \in F(u)$ and

$y \in K^1(u, x)$; furthermore $f_0(x) = \lim f_0(x^k) = \lim p(u^k) = p(u)$, so actually $x \in X(u)$. Thus for each $y \in A(u)$ there exists $x \in X(u)$ with $y \in K^1(u, x)$, which is the assertion of the first inclusion in (3.1).

The second inclusion in (3.1) has a parallel proof. If $y \in Y_0(u)$, there exist by definition (2.18) sequences $t_k y^k \to y$, and $u^k \to u$ such that $t_k \to 0^+$, $p(u^k) \to p(u)$ and $y^k \in A(u^k)$. Again we can find $x^k \in X(u^k)$ such that $y^k \in K^1(u^k, x^k)$ and $x^k \to x \in X(u)$. This time, however, we have

$$0 = t_k \nabla_x l(x^k, y^k) = t_k \nabla f_0(x^k) + \sum_{i=1}^{m} (t_k y_i^k) \nabla f_i(x^k) \to \sum_{i=1}^{m} y_i \nabla f_i(x) \tag{3.7}$$

and consequently $y \in K_0^1(u, x)$.

In the case of convex programming, $L_u(x, y, r)$ is convex in x, as well as concave in y (this follows from (2.8), cf. [13]), so the conditions $x \in F(u)$ and $y \in K^1(u, x)$, which we have seen to be equivalent to (3.3), imply (x, y) is a (global) saddle point of $L_u(\cdot, \cdot, r)$:

$$\inf_{x' \in \mathbf{R}^n} L_u(x', y, r) = L_u(x, y, r) = \sup_{y' \in \mathbf{R}^m} L_u(x, y', r).$$

Since

$$\sup_{y' \in \mathbf{R}^m} L_u(x, y', r) = \begin{cases} f_0(x) & \text{if } x \in F(u) \\ \infty & \text{if } x \notin F(u) \end{cases} \geq p(u),$$

This implies $y \in A(u)$. Thus $K^1(u, x) \subset A(u)$ when $x \in X(u)$, and in view of (3.6) we must indeed have $A(u) = Y(u) = K^1(u, x)$ as asserted in (3.2).

We work now towards verifying the second part of (3.2). Let L_u^0 be the function obtained by deleting the f_0 term from the augmented Lagrangian. Thus

$$L_u(x, y, r) = f_0(x) + L_u^0(x, y, r) \tag{3.8}$$

where $L_u^0(x, y, r)$ is convex in x, concave in y, and

$$L_u^0(x, y, r) = \min_{u' : x \in F(u')} \{-y \cdot (u' - u) + \tfrac{1}{2} r |u' - u|^2\}, \tag{3.9}$$

$$\sup_{y \in \mathbf{R}^m} L_u^0(x, y, r) = \begin{cases} 0 & \text{if } x \in F(u), \\ \infty & \text{if } x \notin F(u). \end{cases} \tag{3.10}$$

Fix any $\bar{r} > 0$ and define

$$A^0(u) = \{y \in \mathbf{R}^m \mid \inf_{x \in \mathbf{R}^n} L_u^0(x, y, \bar{r}) = 0\}. \tag{3.11}$$

We claim that

$$K_0^1(u, x) = A^0(u) \quad \text{for all } x \in F(u). \tag{3.12}$$

Indeed, from (3.9) and (3.10) we know that

$$0 \geq \inf_{x\in \boldsymbol{R}^n} [\sup_{y\in \boldsymbol{R}^m} L_u^0(x, y, \bar{r})] \geq \sup_{y\in \boldsymbol{R}^m} \inf_{x\in \boldsymbol{R}^n} L_u^0(x, y, \bar{r}) \geq \inf_{x\in \boldsymbol{R}^n} L_u^0(x, 0, \bar{r}) = 0.$$

Hence 0 is the saddle value of $L_u^0(\cdot, \cdot, r)$ on $\boldsymbol{R}^n \times \boldsymbol{R}^m$, and the saddle points are the pairs (x, y) such that

$$\sup_{\boldsymbol{R}^m} L_u^0(x, \cdot\,, \bar{r}) = 0 = \inf_{\boldsymbol{R}^n} L_u^0(\,\cdot\,, y, \bar{r}), \tag{3.13}$$

i.e. from (3.10) and (3.12) the elements of $F(u) \times A^0(u)$. They are also the pairs (x, y) satisfying

$$\nabla_x L_u^0(x, y, \bar{r}) = 0 \quad \text{and} \quad \nabla_y L_u^0(x, y, \bar{r}) = 0. \tag{3.14}$$

As in the case of L_u, the gradients of L_u^0 are described by (3.4) and (3.5), without the ∇f_0 term, so by condition (3.14) the saddle points must be the pairs (x, y) such that

$$\max\{f_i(x) + u_i, y_i/r\} = 0 \quad \text{for} \quad i = 1, \dots, s,$$
$$f_i(x) + u_i = 0 \quad \text{for} \quad i = s+1, \dots, m,$$
$$\sum_{i=1}^{m} y_i \nabla f_i(x) = 0,$$

or in other words, such that $x \in F(u)$ and $y \in K_0^1(u, x)$. We deduce that

$$F(u) \times A^0(u) = \{(x, y) \mid x \in F(u),\ y \in K_0^1(u, x)\},$$

which means (3.12).

Applying (3.12) to the second inclusion in (3.1), we see that $Y_0(u) \subset K_0^1(u, x)$ for all $x \in X(u)$, and that to establish equality, as asserted in the second part of (3.2), it will suffice to show $A^0(u) \subset Y_0(u)$. Accordingly, we consider any $y \in A^0(u)$ and try to find sequences $\{t_k\}$, $\{y^k\}$ and $\{u^k\}$ such that

$$t_k y^k \to y, \quad y^k \in A(u^k), \quad t_k \to 0^+, \quad u^k \to u, \quad p(u^k) \to p(u). \tag{3.15}$$

Success in this matter will finish the proof of Proposition 4.

We can make use of what has already been proved in the first part of (3.2) in the following way:

$$K^1(u, x) = Y(u) \subset Y(u) + Y_0(u) \subset K^1(u, x) + K_0^1(u, x) = K^1(u, x),$$

so that $Y(u) + Y_0(u) = K^1(u, x)$ (closed convex) and by formula (2.20) in Theorem 2:

$$\partial p(u) = K^1(u, x) = A(u).$$

This having been established for arbitrary u with $p(u) < \infty$, we can apply it to the proposed elements u^k in (3.15):

$$y^k \in A(u^k) \Leftrightarrow y^k \in \partial p(u^k), \quad \text{when} \quad p(u^k) < \infty. \tag{3.16}$$

Fixing any $y \in A^0(u)$, and letting

$$q(u') = -y \cdot (u' - u) + \tfrac{1}{2}\bar{r}|u' - u|^2, \qquad D = \{u' \mid F(u') \neq \emptyset\} = \{u' \mid p(u') < \infty\}, \tag{3.17}$$

we observe from (3.9) that

$$0 = \inf_{x \in \mathbf{R}^n} L^0(x, y, \bar{r}) = \inf_{x \in \mathbf{R}^n} \min_{u' : x \in F(u')} q(u') = \inf_{u' \in D} q(u').$$

The set D is convex, the function q is strictly convex, and $q(u) = 0$, $u \in D$. Therefore

$$q(u') \begin{cases} > 0 & \text{if } p(u') < \infty \text{ but } u' \neq u, \\ = 0 & \text{if } u' = u. \end{cases} \tag{3.18}$$

Since we are dealing with convex programming, the optimal value function p is itself convex [11], as well as (under our inf-boundedness assumption (1.1)) lower semicontinuous, finite at u. It follows that for $k = 1, 2, \ldots$ there is a unique

$$u^k \in \operatorname*{argmin}_{u' \in \mathbf{R}^m} \{p(u') + kq(u')\}. \tag{3.19}$$

This satisfies

$$p(u^k) \leq p(u^k) + kq(u^k) \leq p(u) + kq(u) = p(u) \tag{3.20}$$

and the subgradient condition

$$0 \in \partial p(u^k) + k\partial q(u^k) = \partial p(u^k) + k[-y + \bar{r}(u^k - u)]. \tag{3.21}$$

Let

$$y^k = k[y - \bar{r}(u^k - u)], \quad t_k = k^{-1}.$$

Then $t_k \to 0^+$, $y^k \in \partial p(u^k)$ by (3.21), and $t_k y^k = y\bar{r}(u^k - u)$. We need only show that $u^k \to u$, for then $t_k y^k \to y$ and $p(u^k) \to p(u)$ (by (3.20) and the lower semicontinuity of p), and in view of (3.16) we will know that the chosen sequences meet the prescription (3.15) for $y \in Y_0(u)$. The proof that $u^k \to u$ rests on (3.18). Inasmuch as p is a proper convex function on $\mathbf{R}^m$, there is an affine function a satisfying

$$a(u') \leq p(u') \quad \text{for all } u' \in \mathbf{R}^m.$$

By (3.19) and (3.20)

$$u^k \in \{u' \mid p(u') + kq(u') \leq p(u)\} \subset \{u' \mid a(u') + q(u') \leq p(u)\},$$

the latter set being some closed ball B (due to the form of q in (3.17)); let

$$\beta = \min_{u' \in B} p(u'),$$

a quantity which is finite because p is lower semicontinuous. We see that $\{u^k\}$ is a bounded sequence which by (3.20) satisfies

$$kq(u^k) \le p(u) - p(u^k) \le p(u) - \beta \quad \text{for all } k$$

and hence by (3.18) can have no cluster point other than u. Thus $u^k \to u$ and the proof of Proposition 4 is complete.

Corollary 1. *For any $u \in \boldsymbol{R}^m$ with $p(u) < \infty$, one has $\partial p(u) \supset A(u)$ and*

$$p^{\uparrow}(u; h) \ge \sup_{y \in A(u)} y \cdot h. \tag{3.22}$$

Proof. Since $A(u) \subset Y(u)$ and $0 \in Y_0(u)$, we have $A(u) \subset Y(u) + Y_0(u)$. Specialize formulas (2.20) and (2.22) of Theorem 2 accordingly.

Corollary 2. *In the convex programming case, one has for any u with $p(u) < \infty$ and any $x \in X(u)$ that*

$$\partial p(u) = A(u) = K^1(u, x), \quad \partial^0 p(u) = K_0^1(u, x), \tag{3.23}$$

$$p^{\uparrow}(u; h) = \sup_{y \in K^1(u,x)} y \cdot h \quad \textit{for all } h \textit{ if } K^1(u, x) \ne \emptyset, \tag{3.24}$$

$$p^{\uparrow}(u; h) = -\infty \quad \textit{if } p^{\uparrow}(u; h) < \infty \textit{ but } K^1(u, x) = \emptyset, \tag{3.25}$$

$$p^{\uparrow}(u, h) < \infty \Leftrightarrow y \cdot h \le 0 \quad \textit{for all } y \in K_0^1(u, x). \tag{3.26}$$

Proof. Again specialize formulas (2.20) and (2.22) of Theorem 2, this time in terms of (3.2) and the fact that $K^1(u, x)$ and $K_0^1(u, x)$ are *polyhedral convex* sets with

$$K_0^1(u, x) = 0^+ K^1(u, x) \quad \text{when } K^1(u, x) \ne \emptyset. \tag{3.27}$$

Polyhedral convexity provides a decomposition

$$K^1(u, x) = C + 0^+ K^1(u, x)$$

for some compact convex set C, and this is why (3.26) is valid, rather than just (2.23) with $K_0^1(u, x)$ in place of $Y_0(u)$.

Corollary 3. *In the convex programming case, consider any u with $p(u) < \infty$ and any $x \in X(u)$. Then for each $h = (h_1, \dots, h_m) \in \boldsymbol{R}^m$, $p^{\uparrow}(u; h)$ is the optimal value in the linear programming problem*

$$\textit{minimize} \quad \nabla f_0(x) \cdot w \quad \textit{over all } w \in \boldsymbol{R}^n,$$

$$\textit{such that} \quad \nabla f_i(x) \cdot w + h_i \begin{cases} \le 0 & \textit{for } i \in I(u, x),\ 1 \le i \le s, \\ = 0 & \textit{for } i = s+1, \dots, m. \end{cases} \tag{3.28}$$

Proof. The linear programming problem in question is the one dual to

$$\begin{aligned} &\text{maximize} \quad y \cdot h \quad \text{over all } y \in \boldsymbol{R}^m, \\ &\text{such that} \quad -\sum_{i=1}^{m} y_i \nabla f_i(x) = \nabla f_0(x), \\ &\qquad y_i \begin{cases} \leq 0 & \text{for } i \in I(u,x),\ 1 \leq i \leq s, \\ = 0 & \text{for } i \not\in I(u,x),\ 1 \leq i \leq s. \end{cases} \end{aligned} \tag{3.29}$$

The feasible set in (3.29) is, of course, just $K^1(u,x)$, and the optimal value is, according to Corollary 2, $p^{\uparrow}(u;h)$ in every case except the one where $K^1(u,x) = \emptyset$ but h does not satisfy $y \cdot h \leq 0$ for all $y \in K_0^1(u,x)$. In that case the supremum in (3.29) is $-\infty$, while $p^{\uparrow}(u;h) = \infty$ (cf. (3.26)). But that is also just the case where the linear programming duality theorem fails because (3.28), as well as (3.29), is infeasible. The the infimum in (3.29) is ∞. In all cases, therefore, $p^{\uparrow}(u;h)$ agrees with the infimum in (3.28).

Theorem 4. *Let u be such that $p(u) < \infty$, and define*

$$Y^1(u) = \bigcup_{x \in X(u)} K^1(u,x), \quad Y_0^1(u) = \bigcup_{x \in X(u)} K_0^1(u,x). \tag{3.30}$$

Then $Y^1(u)$ and $Y_0^1(u)$ are closed subsets of $\boldsymbol{R}^m$ such that $Y_0^1(u)$ is a cone with

$$0 \in Y_0^1(u) \quad \textit{and} \quad 0^+ Y^1(u) \subset Y_0^1(u). \tag{3.31}$$

One always has

$$\partial p(u) \subset \operatorname{cl} \operatorname{co}[Y^1(u) + Y_0^1(u)] \tag{3.32}$$

and

$$Y^1(u) = \emptyset \Rightarrow \partial p(u) = \emptyset \Rightarrow \{0\} \neq \partial^0 p(u) \subset \operatorname{cl} \operatorname{co} Y_0^1(u). \tag{3.33}$$

Moreover

$$p^{\uparrow}(u;h) \leq \begin{cases} \sup_{y \in Y^1(u)} y \cdot h & \textit{if } y \cdot h \leq 0 \textit{ for all } y \in Y_0^1(u), \\ \infty & \textit{otherwise}, \end{cases} \tag{3.34}$$

with

$$\operatorname{int}\{h \mid p^{\uparrow}(u;h) < \infty\} \supset \{h \mid h \cdot y < 0 \quad \textit{for all } y \in Y_0^1(u)\}. \tag{3.35}$$

The latter set is nonempty if and only if $Y_0^1(u)$ is pointed, in which event $\partial^0 p(u)$ is pointed too and actually

$$\partial p(u) \subset \operatorname{co}[Y^1(u) + Y_0^1(u)], \quad \partial^0 p(u) \subset \operatorname{co} Y_0^1(u). \tag{3.36}$$

This is the case in particular when there are no equality constraints in (P_u), i.e. when $s = m$.

Proof. The closedness of $Y^1(u)$ and $Y_0^1(u)$ is an elementary consequence of the continuity of the functions f_i and ∇f_i and the compactness of $X(u)$. The fact that $0 \in Y_0^1(u)$ is trivial: $0 \in K_0^1(u, x)$ for every x. If $y \in 0^+ Y^1(u)$, there exist by definition sequences $\{t_k\}$ and $\{y^k\}$ such that $t_k \to 0^+$, $y^k \in Y^1(u)$, $t_k y^k \to y$. Then for each y^k there is an $x^k \in X(u)$ such that $y^k \in K^1(u, x^k)$. Since $X(u)$ is compact, we can suppose $x^k \to x \in X(u)$. For all k we have (by the definition of $K^1(u, x^k)$)

$$\max\{f_i(x^k) + u_i, -y_i^k\} = 0 \quad \text{for } i = 1, \ldots, s,$$
$$\nabla f_0(x^k) + \sum_{i=1}^m y_i^k \nabla f_i(x^k) = 0.$$

Multiplying the second equation by t_k and taking the limit as $k \to \infty$, we get

$$\max\{f_i(x) + u_i, -y_i\} = 0 \quad \text{for } i = 1, \ldots, s$$
$$\sum_{i=1}^m y_i \nabla f_i(x) = 0,$$

which means precisely that $y \in K_0^1(u, x)$. Thus $y \in Y_0^1(u)$.

The rest of the proof of Theorem 4 is merely a matter of applying Theorem 2 and Proposition 3 to $Y^1(u)$ and $Y_0^1(u)$, using the fact that

$$Y(u) \subset Y^1(u) \quad \text{and} \quad Y_0(u) \subset Y_0^1(u) \tag{3.37}$$

by Proposition 4. Obviously if there are no equality constraints we have $Y_0^1(u) \subset \boldsymbol{R}_+^m$, so $Y_0^1(u)$ is pointed.

Theorem 4 generalizes Theorem 1. Indeed, Theorem 1 is the corollary of Theorem 4 for the case where $Y_0^1(u) = \{0\}$. Then in particular, $Y_0^1(u)$ is pointed, and the set $\{h \mid p^{\uparrow}(u; h) < \infty\}$ therefore has a nonempty interior, p is 'directionally Lipschitzian' at u and special formulas hold for $p^{\uparrow}(u; \cdot)$ (see [15, 16]).

Theorem 4 can be extended to the situation where the functions f_i are not of class $\mathscr{C}^1$ but just locally Lipschitzian, and where an abstract constraint is present (see [18]).

4. Estimates in terms of standard second-order conditions

The basic formulas in Theorem 2 lead to the estimates in Theorem 1 in terms of the multiplier vectors in $K^1(u, x)$ and $K_0^1(u, x)$, and the question is whether sharper estimates can be obtained by taking second-order optimality conditions into account. A positive answer will be provided in the next section, but not quite in terms of second-order conditions as traditionally formulated. Multiplier sets $K^2(u, x)$ and $K_0^2(u, x)$ will be introduced which not only are smaller in general than $K^1(u, x)$ and $K_0^1(u, x)$ and can be substituted for them in Theorems 1 and 4, but also correspond to a new sort of theory of second-order necessary conditions for local optimality.

The relationship between augmentable Lagrange multiplier vectors and the classical kinds of second-order conditions must be studied first. For any $u \in \boldsymbol{R}^m$ and $x \in F(u)$ we consider the linearized constraint system

$$\nabla f_i(x) \cdot w \begin{cases} \leq 0 & \text{for } i \in I(u,x),\ 1 \leq i \leq s, \\ = 0 & \text{for } i = s+1, \dots, m, \end{cases} \tag{4.1}$$

and the polyhedral convex cone

$$W(u,x) = \{w \in \boldsymbol{R}^n \mid (4.1) \text{ holds and } \nabla f_0(x) \cdot w \leq 0\}. \tag{4.2}$$

Note that for any $y \in K^1(u,x)$ and the index sets

$$\begin{aligned} I_0(u,x,y) &= \{i \in I(u,x) \mid 1 \leq i \leq s,\ y_i = 0\}, \\ I_1(u,x,y) &= I(u,x) \setminus I_0(u,x,y), \end{aligned} \tag{4.3}$$

one has the characterization of $W(u,x)$ as the set of all $w \in \boldsymbol{R}^n$ satisfying the system

$$\nabla f_i(x) \cdot w \begin{cases} \leq 0 & \text{for all } i \in I_0(u,x,y), \\ = 0 & \text{for all } i \in I_1(u,x,y). \end{cases} \tag{4.4}$$

Denoting by $\nabla^2 f_i(x)$ the Hessian of f_i at x, so that

$$\nabla_x^2 l(x,y) = \nabla^2 f_0(x) + \sum_{i=1}^{m} y_i \nabla^2 f_i(x) \tag{4.5}$$

(recall the definition (1.6) of l), we define for $x \in F(u)$ the sets

$$N(u,x) = \{y \in K^1(u,x) \mid w \cdot \nabla_x^2 l(x,y) w \geq 0 \quad \text{for all } w \in W(u,x)\}, \tag{4.6}$$

$$S(u,x) = \{y \in K^1(u,x) \mid w \cdot \nabla_x^2 l(x,y) w > 0 \text{ for all nonzero } w \in W(u,x)\}. \tag{4.7}$$

The notation is suggested by the well-known fact that the existence of some $y \in S(u,x)$ is always *sufficient* for x to be a locally optimal solution to (P_u), while under certain constraint qualifications (!) the existence of some $y \in N(u,x)$ is *necessary* for x to be a locally optimal solution to (P_u). (The constraint qualifications in question are rather stringent, however. More will be said on this issue in Section 5.)

Proposition 5. *For any $u \in \boldsymbol{R}^m$ with $p(u) < \infty$, one has*

$$\bigcap_{x \in X(u)} S(u,x) \subset A(u) \subset \bigcap_{x \in X(u)} N(u,x). \tag{4.8}$$

In fact if

$$\bigcap_{x \in X(u)} S(u,x) \neq \emptyset, \tag{4.9}$$

then

$$\operatorname{cl} \bigcap_{x\in X(u)} S(u,x) = \operatorname{cl} A(u) = \bigcap_{x\in X(u)} N(u,x). \tag{4.10}$$

Proof. A simplified representation for the augmented Lagrangian L_u will assist us. Suppose $x \in F(u)$ and $y \in K^1(u, x)$. Looking back to the definition of L_u at the beginning of Section 2, we see from the continuity of f_i that any inequality constraints which are inactive at x are inactive in an open neighborhood of x, and for all x' in such a neighborhood:

$$\begin{aligned} L_u(x', y, r) &= f_0(x') + \sum_{i\in I_0(u,x,y)} \tfrac{1}{2} r[f_i(x') + u_i]_+^2 \\ &\quad + \sum_{i\in I_1(u,x,y)} (y_i[f_i(x') + u_i] + \tfrac{1}{2} r[f_i(x') + u_i]^2) \\ &= y \cdot u + f_0(x') + \sum_{i=1}^{m} y_i f_i(x') + r q_u(x', y) \end{aligned} \tag{4.11}$$

where

$$[\alpha]_+ = \max\{0, \alpha\} \quad \text{for} \quad \alpha \in \boldsymbol{R}, \tag{4.12}$$

$$\begin{aligned} q_u(x', y) &= \frac{1}{2}\Big(\sum_{i\in I_0(u,x,y)} [f_i(x') + u_i]_+^2 + \sum_{i\in I_1(u,x,y)} [f_i(x') + u_i]^2\Big) \\ &= \frac{1}{2}\Big(\sum_{i\in I_0(u,x,y)} [f_i(x') - f_i(x)]_+^2 + \sum_{i\in I_1(u,x,y)} [f_i(x') - f_i(x)]^2\Big) \end{aligned} \tag{4.13}$$

and $f_0(x') + \sum_{i=1}^m y_i f_i(x') = l(x', y)$, with $\nabla_x l(x, y) = 0$. Therefore

$$\begin{aligned} L_u(x', y, r) &= f_0(x) + \tfrac{1}{2}(x' - x)\nabla_x^2 l(x, y)(x' - x) \\ &\quad + \tfrac{1}{2} r \sum_{i\in I_0(u,x,y)} \max{}^2\{0, \nabla f_i(x) \cdot (x' - x)\} \\ &\quad + \tfrac{1}{2} r \sum_{i\in I_1(u,x,y)} [\nabla f_i(x) \cdot (x' - x)]^2 + o(|x' - x|^2) \end{aligned} \tag{4.14}$$

when $x \in F(u)$ and $y \in K^1(u, x)$.

Consider now any $x \in X(u)$ and $y \in A(u)$. It was demonstrated in the proof of Proposition 4 that such an (x, y) is a local saddle point of $L_u(\cdot\,, \cdot\,, r)$ for r sufficiently large, and that this implies $y \in K^1(u, x)$. We thus have (4.14) at our disposal, with $L_u(x', y, r) \geq L_u(x, y, r)$ for x' near enough to x. Setting $x' = x + tw$ and considering what happens as $t \to 0^+$, we deduce

$$0 \leq w \cdot \nabla_x^2 l(x, y) w + r\Big(\sum_{i\in I_0(u,x,y)} [\nabla f_i(x) \cdot w]_+^2 + \sum_{i\in I_1(u,x,y)} [\nabla f_i(x) \cdot w]^2\Big) \tag{4.15}$$

for all $w \in \boldsymbol{R}^n$. In this inequality the r term vanishes exactly when $w \in W(u, x)$, so for such w we must have $0 \leq w \cdot \nabla_x^2 l(x, y) w$. Thus $y \in N(u, x)$, and the second of the inclusions in (4.8) is valid.

Next consider a vector

$$y \in \bigcap_{x \in X(u)} S(u, x). \tag{4.16}$$

We shall demonstrate first that for each $x \in X(u)$ the function $L_u(\cdot, y, r)$ has a local minimum at x, provided r is sufficiently large; then we shall show that this property implies $y \in A(u)$. Consider any $x \in X(u)$ and observe from the definition (4.7) of y being in $S(u, x)$ that the compact set

$$D = \{w \in \mathbf{R}^n \mid w \cdot \nabla_x^2 l(x, y) w \leq 0, \quad |w| = 1\}$$

does not meet $W(u, x)$ and therefore has

$$0 < \min_{w \in D} \Big\{ \sum_{i \in I_0(u, x, y)} [\nabla f_i(x) \cdot w]_+^2 + \sum_{i \in I_1(u, x, y)} [\nabla f_i(x) \cdot w]^2 \Big\}.$$

Denote this minimum by δ and let

$$\alpha = \min_{w \in D} \{w \cdot \nabla_x^2 l(x, y) w\} \leq 0.$$

For any $r > 0$ large enough that $\alpha + \frac{1}{2} r\delta > 0$, it is clear that strict inequality holds in (4.15) when $w \in D$, but it also holds trivially when $w \notin D$, $|w| = 1$, because $w \cdot \nabla_x^2 l(x, y) w > 0$ for such vectors w. Since the right side of (4.15) is positively homogeneous of degree 2 as a function of w, we conclude that when r is sufficiently large, strict inequality holds in (4.15) for every $w \neq 0$. It follows then from the second-order expansion (4.14) that when r is sufficiently large, $L_u(\cdot, y, r)$ has a local minimum at x, and this minimum value is $f_0(x)$, i.e. $p(u)$ (because $x \in X(u)$).

We can now associate with each $x \in X(u)$ an open neighborhood V_x and value $r_x > 0$ such that

$$L_u(x', y, r) \geq p(u) \quad \text{for all } x' \in V_x \text{ when } r \geq r_x.$$

Since $X(u)$ is compact, it is covered by finitely many such neighborhoods V_x. Taking V^* to be the union of this finite covering, and r^* to be the maximum of the corresponding values r_x (finitely many), we see that V^* is an open set containing $X(u)$, such that

$$L(x', y, r) \geq p(u) \quad \text{for all } x' \in V^* \text{ when } r \geq r^*. \tag{4.17}$$

In order to prove $y \in A(u)$, we want to demonstrate from this that there exist a neighborhood U of u and a value $\bar{r} > 0$ with

$$L(x', y, r) \geq p(u) \quad \text{for all } x' \in F(U) \text{ when } r \geq \bar{r}. \tag{4.18}$$

We shall assume the opposite and argue to a contradiction: suppose (4.18) is not true for any U or $\bar{r}$. Then there exist sequences $\{x^k\}$, $\{u^k\}$, $\{r^k\}$ with

$$u^k \to u,\ r^* \le r_k \to \infty,\ x^k \in F(u^k), \quad L_u(x^k, y, r_k) < p(u). \tag{4.19}$$

Recalling that

$$\begin{aligned} L_u(x^k, y, r_k) &= f_0(x^k) + \min_{u': x^k \in F(u')} \{y \cdot u' + \tfrac{1}{2} r_k |u' - u|^2\} \\ &\ge f_0(x^k) + \min_{u' \in R^n} \{y \cdot u' + \tfrac{1}{2} r_k |u' - u|^2\} \\ &= f_0(x^k) - |y|^2/2r_k, \end{aligned}$$

we find that

$$f_0(x^k) < p(u) + |y|^2/2r_k \to p(u). \tag{4.20}$$

Certainly then for all k sufficiently large

$$x^k \in \{x' \in F(B) \mid f_0(x') \le p(u) + 1\} \quad \text{with } B = \{u' \mid |u' - u| \le 1\},$$

and by our inf-boundedness assumption (1.1) this implies $\{x^k\}$ is bounded. We can therefore suppose x^k converges to some x; then $f_0(x) \le p(u)$ by (4.20), while also $x \in F(u)$ because $x^k \in F(u^k)$. Hence $x \in X(u)$. However, we also have $x^k \notin V^*$ for all k by (4.17) and (4.18), and inasmuch as V^* is open this implies $x \notin V^*$. We have reached a contradiction, because $X(u) \subset V^*$. Thus it is true after all that $y \in A(u)$, and the first inclusion in (4.8) is correct.

The rest of Proposition 5 is easy. Taking again any y as in (4.16) and considering any y' belonging to the intersection on the right in (4.8), we observe that for every $x \in X(u)$ and $t \in [0, 1)$:

$$(1-t)y + ty' \in S(u, x)$$

(cf. definitions (4.6) and (4.7)). Thus $(1-t)y + ty'$ belongs to the intersection on the left in (4.8) as $t \to 1^-$, and the limit y' therefore belongs to the closure of the intersection. This shows that

$$\bigcap_{x \in X(u)} N(u, x) \subset \operatorname{cl} \bigcap_{x \in X(u)} S(u, x)$$

when (4.9) holds, and in combination with (4.8) this inclusion yields (4.10).

Remark. The proof of Proposition 5 actually establishes a stronger result. For $x \in F(u)$, define

$$A(u, x) = \{y \in K^1(u, x) \mid \exists\ r > 0 \quad \text{such that } l(\cdot, y) + rq_u(\cdot, y) \text{ has a local min at } x\}, \tag{4.21}$$

where q_u is the penalty function in (4.13). (The condition $y \in K^1(u, x)$ by itself merely implies that $l(\cdot, y) + rq_u(\cdot, y)$ has a stationary point at x, since

$\nabla_x q_u(x, y) = 0$.) Then

$$S(u,x) \subset A(u,x) \subset N(u,x), \quad \text{with cl } S(u,x) = \text{cl } A(u,x) = N(u,x) \text{ when } S(u,x) \neq \emptyset, \tag{4.22}$$

and (under our inf-boundedness condition)

$$A(u) = \bigcap_{x \in X(u)} A(u,x). \tag{4.23}$$

A vector $y \in A(u,x)$ may be called *locally augmentable* at x. The set $A(u,x)$, like $S(u,x)$ and $N(u,x)$, is convex. The existence of some $y \in A(u,x)$ is sufficient for an $x \in F(u)$ to be *locally* optimal in (P_u). Indeed, the pairs (x, y) such that $x \in F(u)$ and $y \in A(u,x)$, are precisely the *local* saddle points of the augmented Lagrangian $L_u(\cdot\,,\cdot\,,r)$ for various values of $r > 0$.

Corollary. *Let u be such that $p(u) < \infty$. Then the multiplier sets $Y(u)$ and $Y_0(u)$ of Theorem 2 are estimated by*

$$\limsup_{\substack{u' \to u \\ p(u') \to p(u)}} \Big[\bigcap_{x' \in X(u')} S(u',x')\Big] \subset Y(u) \subset \limsup_{\substack{u' \to u \\ p(u') \to p(u)}} \Big[\bigcap_{x' \in X(u')} N(u',x')\Big], \tag{4.24}$$

$$\limsup_{\substack{t \to 0^- \\ u' \to u \\ p(u') \to p(u)}} \Big[\bigcap_{x' \in X(u')} tS(u',x')\Big] \subset Y_0(u) \subset \limsup_{\substack{t \to 0^+ \\ u' \to u \\ p(u') \to p(u)}} \Big[\bigcap_{x' \in X(u')} tN(u',x')\Big]. \tag{4.25}$$

This corollary enables us to describe a situation where the subgradients and subderivatives of the optimal value function p are completely expressible in terms of classical multiplier vectors. For this purpose we need to introduce in association with the set $N(u,x)$ in (4.6) the polyhedral convex cone

$$N_0(u,x) = \{y \in K_0^1(u,x) \mid w \cdot \nabla_x^2 l_0(x,y) w \geq 0 \quad \text{for all } w \in W_0(u,x)\}, \tag{4.26}$$

where of course (by the definition (1.6) of l_0)

$$\nabla_x^2 l_0(x,y) = \sum_{i=1}^m y_i \nabla^2 f_i(x).$$

It is elementary that

$$0^+ N(u,x) = N_0(u,x) \quad \text{when } N(u,x) \neq \emptyset. \tag{4.27}$$

Theorem 5. *For some u such that $p(u) < \infty$, assume*

$$X(u) = \{x\} \quad (\textit{unique optimal solution}), \quad S(u,x) \neq \emptyset \tag{4.28}$$

and the semicontinuity properties

$$\limsup_{\substack{(u',x') \to (u,x) \\ x' \in F(u')}} N(u',x') \subset N(u,x), \tag{4.29}$$

$$\limsup_{\substack{t\to 0^+\\(u',x')\to(u,x)\\x'\in F(u')}} tN(u',x')\subset N_0(u,x). \tag{4.30}$$

Then

$$\partial p(u)=N(u,x)\neq\emptyset,\quad \partial^0 p(u)=N_0(u,x), \tag{4.31}$$

and

$$p^{\uparrow}(u;h)=\begin{cases}\sup\{y\cdot h \mid y\in N(u,x)\}\\ \quad \text{if } y\cdot h\le 0 \text{ for every } y\in N_0(u,x),\\ \infty \quad \text{otherwise.}\end{cases} \tag{4.32}$$

Proof. Our inf-boundedness assumption (1.1) implies that whenever $x^k\in X(u^k)$ for a sequence $\{u^k\}$ with $u^k\to u$ and $p(u^k)\to p(u)$, the sequence $\{x^k\}$ is bounded and has all of its cluster points in $X(u)$ (see the proof of Proposition 2, for instance). Therefore

$$\limsup_{\substack{u'\to u\\p(u')\to p(u)}}\Big[\bigcap_{x'\in X(u')} N(u',x')\Big]\subset N(u,x)$$

by (4.29). It follows from (4.27) that

$$S(u,x)\subset Y(u)\subset N(u,x).$$

Then since $Y(u)$ is a closed set (Theorem 2), we have from (4.28) and Proposition 4 that

$$\emptyset\neq Y(u)=N(u,x)\quad \text{(convex)}. \tag{4.33}$$

We calculate next from (4.30) and the second inclusion in (4.25) that

$$\limsup_{\substack{t\to 0^+\\(u',x')\to(u,x)\\x'\in X(u')}} tN(u',x')\subset N_0(u,x). \tag{4.34}$$

The set $N(u,x)$ being closed and convex, we have

$$N_0(u,x)=0^+Y(u)\subset Y_0(u).$$

Combining this with (4.34) we get

$$0^+Y(u)=Y_0(u)=N_0(u,x). \tag{4.35}$$

Since $Y(u)$ is a nonempty closed convex set by (4.33), we have $Y(u)+0^+Y(u)=Y(u)$ (cf. [11, Section 8]) and consequently $Y(u)=Y(u)+Y_0(u)$ in view of (4.35), so that $\partial p(u)=Y(u)$ in Theorem 2. This yields via (4.33) the first formula in (4.31). We note next that for any nonempty closed *convex* set $C\subset \boldsymbol{R}^m$,

$$0^+C=\{y\in\boldsymbol{R}^m \mid y\cdot h\le 0\quad \text{for all } h \text{ with } \sup_{z\in C} z\cdot h<\infty\}$$

(see [11, Section 13]). Recalling (1.4) and the definition (1.12) of $\partial^0 p(u)$, we see that for $C = \partial p(u)$ this reduces to

$$0^+\partial p(u) = \partial^0 p(u) \quad \text{when} \quad \partial p(u) \neq \emptyset. \tag{4.36}$$

Since $\partial p(u) = Y(u)$ in the present case, we are able to conclude from (4.35) that the second formula in (4.33) is valid too. Formula (4.32) then follows immediately from (1.4) and (1.12).

Example. Consider the inequality-constrained case of (P_u) where $m = s = 2$ and for $x = (x_1, x_2) \in \boldsymbol{R}^2$:

$$f_0(x) = x_1^2 - x_2^2, \qquad f_1(x) = -x_1 + x_2^2, \qquad f_2(x) = x_1 + x_2^2,$$
$$\nabla f_0(x) = (2x_1, -2x_2), \qquad \nabla f_1(x) = (-1, 2x_2), \qquad \nabla f_2(x) = (1, 2x_2),$$
$$\nabla^2 f_0(x) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}, \qquad \nabla^2 f_1(x) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}, \qquad \nabla^2 f_2(x) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}.$$

We shall analyze the behavior of the second-order multiplier set $N(u, x)$ for $u = (u_1, u_2)$ near $\bar{u} = (0, 0)$ and verify by way of Theorem 5 that

$$\begin{aligned} &\partial p(\bar{u}) = \{(y_1, y_2) \mid y_1 = y_2 \geq \tfrac{1}{2}\}, \\ &\partial^0 p(\bar{u}) = \{(y_1, y_2) \mid y_1 = y_2 \geq 0\}, \\ &p^{\uparrow}(\bar{u}; h) = \begin{cases} \tfrac{1}{2}(h_1 + h_2) & \text{for } h = (h_1, h_2) \text{ with } h_1 + h_2 \leq 0, \\ \infty & \text{for } h = (h_1, h_2) \text{ with } h_1 + h_2 > 0. \end{cases} \end{aligned} \tag{4.37}$$

Since

$$\begin{aligned} F(u) &= \{x \mid -x_1 + x_2^2 + u_1 \leq 0 \quad \text{and} \quad x_1 + x_2^2 + u_2 \leq 0\} \\ &= \{x \mid u_1 \leq x_1 \leq -u_2,\ x_2^2 \leq \min\{x_1 - u_1, -x_1 - u_2\}\}, \end{aligned}$$

it is clear that

$$p(u) < \infty \Leftrightarrow F(u) \neq \emptyset \Leftrightarrow u_1 \leq -u_2, \tag{4.38}$$
$$F(u) = \{(u_1, 0)\} = \{(-u_2, 0)\} = X(u) \quad \text{when} \quad u_1 = -u_2. \tag{4.39}$$

For any $x \in F(u)$, we have

$$\begin{aligned} K^1(u, x) = \{y \in \boldsymbol{R}_+^2 \mid &\nabla_x l(x, y) = 0, \\ &y_1(-x_1 + x_2^2 + u_1) = 0 = y_2(x_1 + x_2^2 + u_2)\}, \\ K_0^1(u, x) = \{y \in \boldsymbol{R}_+^2 \mid &\nabla_x l_0(x, y) = 0, \\ &y_1(-x_1 + x_2^2 + u_2) = 0 = y_2(x_1 + x_2^2 + u_2)\} \end{aligned}$$

where

$$\nabla_x l(x, y) = (2x_1 - y_1 + y_2, 2x_2(-1 + y_1 + y_2)),$$
$$\nabla_x l_0(x, y) = (-y_1 + y_2, 2x_2(y_1 + y_2)).$$

Since we are only interested in u near $\bar{u} = (0, 0)$ such that $F(u) \neq \emptyset$, we can assume

$$-\tfrac{1}{2} < u_1 \leq -u_2 < \tfrac{1}{2}. \tag{4.40}$$

Direct calculation determines that under this restriction

$$K_1(u,x) = \begin{cases} \{(\tfrac{1}{2}(1+u_1-u_2), \tfrac{1}{2}(1+u_2-u_1))\} \\ \qquad \text{when } u_1 < -u_2 \text{ and } x = (\tfrac{1}{2}(u_1-u_2), \pm[\tfrac{1}{2}|u_1+u_2|]^{1/2}) \\ \{(2u_1, 0)\} \qquad \text{when } 0 \le u_1 < -u_2,\ x = (u_1, 0), \\ \{(0, 2u_2)\} \qquad \text{when } u_1 < -u_2 \le 0,\ x = (-u_2, 0), \\ \{(\tfrac{1}{2}(v+u_1-u_2), \tfrac{1}{2}(v+u_2-u_1)) \mid v \ge |u_1-u_2|\} \\ \qquad \text{when } u_1 = -u_2,\ x = (u_1, 0) = (-u_2, 0), \\ \emptyset \qquad \text{otherwise for } x \in F(u) \text{ (under (4.40))}, \end{cases} \tag{4.41}$$

$$K_0^1(u,x) = \begin{cases} \{(v,v) \mid v \ge 0\} & \text{when } u_1 = -u_2,\ x = (u_1, 0) = (-u_2, 0), \\ \{(0,0)\} & \text{otherwise for } x \in F(u) \text{ (under (4.40))}. \end{cases} \tag{4.42}$$

In particular for $\bar{u} = (0,0)$ we have

$$\begin{aligned} &F(\bar{u}) = X(\bar{u}) = \{\bar{x}\}, \quad \text{where } \bar{x} = (0,0), \\ &K^1(\bar{u},\bar{x}) = K_0^1(\bar{u},\bar{x}) = \{(y_1,y_2) \mid y_1 = y_2 \ge 0\}. \end{aligned} \tag{4.43}$$

Note incidentally that since $K_0^1(\bar{u},\bar{x})$ is not just $(0,0)$, the Mangasarian–Fromovitz constraint qualification is not satisfied for $(\mathrm{P}_{\bar{u}})$.

Next we check the elements of $K^1(u,x)$ and $K_0^1(u,x)$ to see if they actually belong to $N(u,x)$ or $N_0(u,x)$. The hessian matrices are

$$\nabla_x^2 l(x,y) = \begin{bmatrix} 2 & 0 \\ 0 & 2(y_1+y_2-1) \end{bmatrix}, \quad \nabla_x^2 l_0(x,y) = \begin{bmatrix} 0 & 0 \\ 0 & 2(y_1+y_2) \end{bmatrix}.$$

In the first case of (4.41), $W(u,x) = \{(0,0)\}$, so $K^1(u,x) = N(u,x)$ trivially. In the second and third cases of (4.41), $W(u,x) = \{0\} \times \boldsymbol{R}$ and the hessian fails the second derivative condition by virtue of (4.40). (These cases correspond to local *maxima*.) In the fourth case of (4.41) we likewise have $W(u,x) = \{0\} \times \boldsymbol{R}$, and the second derivative condition is satisfied if also $y_1 + y_2 \ge 1$. In the first case of (4.42), the hessian is positive semidefinite, so $K_0^1(u,x) = N_0(u,x)$. Thus

$$N(u,x) = \begin{cases} \{(\tfrac{1}{2}(1+u_1-u_2), \tfrac{1}{2}(1+u_2-u_1))\} \\ \qquad \text{when } u_1 < -u_2,\ x = (\tfrac{1}{2}(u_1-u_2), \pm[\tfrac{1}{2}(u_1+u_2)]^{1/2}) \\ \{(\tfrac{1}{2}(v+u_1-u_2), \tfrac{1}{2}(v+u_2-u_1)) \mid v \ge 1\} \\ \qquad \text{when } u_1 = -u_2,\ x = (u_1, 0) = (-u_2, 0), \\ \emptyset \qquad \text{otherwise for } x \in F(u) \text{ (under (4.40))}, \end{cases} \tag{4.44}$$

$$N_0(u,x) = \begin{cases} \{(v,v) \mid v \ge 0\} & \text{when } u_1 = -u_2,\ x = (u_1, 0) = (-u_2, 0), \\ \{(0,0)\} & \text{otherwise for } x \in F(u) \text{ (under (4.40))}, \end{cases} \tag{4.45}$$

and in particular for $\bar{u} = ((0, 0))$, $\bar{x} = (0, 0)$:

$$\begin{aligned} N(\bar{u}, \bar{x}) &= \{(y_1, y_2) \mid y_1 = y_2 \geq \tfrac{1}{2}\} \neq K^1(\bar{u}, \bar{x}), \\ N_0(\bar{u}, \bar{x}) &= \{(y_1, y_2) \mid y_1 = y_2 \geq 0\} = K_0^1(\bar{u}, \bar{x}). \end{aligned} \tag{4.46}$$

Furthermore

$$S(\bar{u}, \bar{x}) = \{(y_1, y_2) \mid y_1 = y_2 > \tfrac{1}{2}\}. \tag{4.47}$$

From (4.43)–(4.47) we see that the assumptions in Theorem 5 are fulfilled, and that (4.43) is the conclusion.

5. Estimates in terms of new second-order conditions

We have seen in Theorem 5 that subgradients and subderivatives of the optimal value function p can sometimes be characterized in terms of the multiplier sets

$$\begin{aligned} N(u, x) &= \{y \in K^1(u, x) \mid w \cdot \nabla_x^2 l(x, y) w \geq 0, \quad \forall\, w \in W(u, x)\}, \\ N_0(u, x) &= \{y \in K_0^1(u, x) \mid w \cdot \nabla_x^2 l_0(x, y) w \geq 0, \quad \forall\, w \in W(u, x)\}, \end{aligned} \tag{5.1}$$

where

$$w \in W(u, x) \Leftrightarrow \nabla f_i(x) \cdot w \begin{cases} \leq 0 & \text{for } i = 0 \text{ and } i \in I(u, x),\ i \leq s, \\ = 0 & \text{for } i = s+1, \dots, m. \end{cases}$$

This characterization depended, however, on N and N_0 having certain limit properties (4.29) and (4.30), which unfortunately can fail in some situations. Although

$$\limsup_{\substack{(u', x') \to (u, x) \\ x' \in F(u')}} K^1(u', x') = K^1(u, x), \tag{5.2}$$

$$\limsup_{\substack{t \to 0^- \\ (u', x') \to (u, x) \\ x' \in F(u')}} tK^1(u', x') \subset K_0^1(u, x), \tag{5.3}$$

by virtue of the continuity of the functions f_i and ∇f_i, the same is not true of N and N_0. It appears that $N(u, x)$ and $N_0(u, x)$ are sometimes too small to contain all the needed limits.

Another deficiency is that the condition $N_0(u, x) = \{0\}$ does not guarantee that $N(u, x) \neq \emptyset$ and therefore cannot serve as a constraint qualification in the manner that the condition $K_0^1(u, x) = \{0\}$ does in Proposition 1. Moreover, in general

$$Y(u) \not\subset \bigcup_{x \in X(u)} N(u, x), \quad Y_0(u) \not\subset \bigcup_{x \in X(u)} N_0(u, x),$$

so Theorem 4 topples when $N(u,x)$ and $N_0(u,x)$ are substituted for $K^1(u,x)$ and $K_0^1(u,x)$. These facts point to the need for some sort of enlargement of $N(u,x)$ and $N_0(u,x)$, if a second-order refinement of the results concerning $K^1(u,x)$ and $K_0^1(u,x)$ is to be developed.

The definition of the sets which we shall presently introduce for this purpose as $K^2(u,x)$ and $K_0^2(u,x)$ depends on the concept of a *sequence of subspaces* M^k *of* $\boldsymbol{R}^n$ *converging to a subspace* M as $k \to \infty$. This means

$$\operatorname{dist}(z, M^k) \to \operatorname{dist}(z, M) \quad \text{for each } z \in \boldsymbol{R}^n. \tag{5.4}$$

An equivalent statement is that $M^k \cap B$ converges to $M \cap B$ in the Hausdorff metric, where B is the closed unit ball. See Salinetti and Wets [19], Wets [20], for more on such convergence and its characterizations. Every sequence of subspaces of $\boldsymbol{R}^n$ has a subsequence which is convergent in this sense.

Our interest will center on sequences of subspaces of the form

$$M^k = \{w \in \boldsymbol{R}^n \mid \nabla f_i(x^k) \cdot w = 0, \quad \forall\, i \in I(u,x)\} \tag{5.5}$$

associated with u, $x \in F(u)$, and sequences of points $x^k \to x$. We set

$$\mathcal{M}(u,x) = \{M(\text{subspace}) \subset \boldsymbol{R}^n \mid \exists M^k \quad \text{as in (5.5) with } x^k \to x,\ M^k \to M\}. \tag{5.6}$$

Clearly $\mathcal{M}(u,x)$ contains the subspace

$$Z(u,x) = \{w \in \boldsymbol{R}^n \mid \nabla f_i(x) \cdot w = 0, \quad \forall i \in I(u,x)\}, \tag{5.7}$$

and

$$Z(u,x) \subset W(u,x) \quad \text{when } K^1(u,x) \neq \emptyset.$$

Every other element of $\mathcal{M}(u,x)$ is included in $Z(u,x)$ and has dimension at least $n - |I(u,x)|$, where $I(u,x)$ denotes the number of indices in $I(u,x)$. More will be said about the nature of the collection $\mathcal{M}(u,x)$ in Proposition 7 below.

We define

$$\begin{aligned} K^2(u,x) = \{y \in K^1(u,x) \mid \exists\, M \in \mathcal{M}(u,x) \\ \text{with } w \cdot \nabla_x^2 l(x,y) w \geq 0 \text{ for all } w \in M\}, \\ K_0^2(u,x) = \{y \in K_0^1(u,x) \mid M \in \mathcal{M}(u,x) \\ \text{with } w \cdot \nabla_x^2 l_0(x,y) w \geq 0 \text{ for all } w \in M\}. \end{aligned} \tag{5.8}$$

Obviously, then, for any u and $x \in F(u)$ we have

$$K^2(u,x) \supset N(u,x), \quad \text{and} \quad K_0^2(u,x) \supset N_0(u,x) \tag{5.9}$$

when $K^1(u,x) \neq \emptyset$ (hence when $N(u,x) \neq \emptyset$).

The multiplier sets $K^2(u,x)$ and $K_0^2(u,x)$ need not be convex, however.

Proposition 6. *Let* u *be such that* $p(u) < \infty$. *For any* $x \in F(u)$, $K^2(u,x)$ *and*

$K_0^2(u,x)$ are closed subsets of $\boldsymbol{R}^m$ such that $K_0^2(u,x)$ is a cone with

$$0 \in K_0^2(u,x) \quad \text{and} \quad 0^+K^2(u,x) \subset K_0^2(u,x). \tag{5.10}$$

One has the semicontinuity properties

$$\limsup_{\substack{(u',x')\to(u,x)\\ x'\in F(u')}} K^2(u',x') = K^2(u,x), \tag{5.11}$$

$$\limsup_{\substack{(u',x')\to(u,x)\\ x'\in F(u')}} K_0^2(u',x') = K_0^2(u,x), \tag{5.12}$$

$$\limsup_{\substack{t\to 0^+\\ (u',x')\to(u,x)\\ x'\in F(u')}} tK^2(u',x') \subset K_0^2(u,x). \tag{5.13}$$

Furthermore, the sets $Y(u)$ and $Y_0(u)$ in Theorem 2 satisfy

$$Y(u) \subset \bigcup_{x\in X(u)} K^2(u,x), \qquad Y_0(u) \subset \bigcup_{x\in X(u)} K_0^2(u,x). \tag{5.14}$$

Proof. The closedness of $K^2(u,x)$ is implicit in formula (5.11), which we therefore proceed to verify. The inclusion $\supset$ is trivial in (5.11) (consider constant sequences).

In order to prove the inclusion $\subset$ in (5.11), we first need to observe that $K^2(u,x)$ and $K_0^2(u,x)$ would not be altered if the collection $\mathcal{M}(u,x)$ in their definition (5.8) were replaced by

$$\begin{aligned}\mathcal{M}'(u,x) = \{M(\text{subspace}) \subset \boldsymbol{R}^n \mid \exists (u^k,x^k)\to(u,x) \\ \text{with } x^k\in F(u^k) \text{ and } Z(u^k,x^k)\to M\},\end{aligned} \tag{5.15}$$

where $Z(u,x)$ is the subspace in (5.7). Indeed, if $M \in \mathcal{M}(u,x)$ and $M^k \to M$ with M^k as in (5.5) and $x^k \to x$, we can set $u_i^k = -f_i(x^k)$ for $i \in I(u,x)$ and $u_i^k = u_i$ for all other i (namely with $f_i(x)+u_i<0$) to get $u^k\to u$ and have, for all k sufficiently large, $x^k \in F(u^k)$ and $I(u^k,x^k) = I(u,x)$; then $M^k = Z(u^k,x^k)$. Thus

$$\mathcal{M}(u,x) \subset \mathcal{M}'(u,x). \tag{5.16}$$

On the other hand, if $M' \in \mathcal{M}'(u,x)$ and $Z(u^k,x^k)\to M'$ with $(u^k, x^k)\to(u,x)$, $x^k\in F(u^k)$, we must have $I(u^k,x^k)\subset I(u,x)$ for all k sufficiently large, because if $f_i(x^k)+u_i^k=0$ for infinitely many values of k it must also be true by continuity that $f_i(x)+u_i=0$. Then, taking M^k to be the subspace in (5.5), we have $M^k\subset Z(u^k,x^k)$. Passing to subsequences if necessary, we can assume that M^k converges to a subspace M; then M belongs to $\mathcal{M}(u,x)$, and $M\subset M'$, inasmuch as $Z(u^k,x^k)\to M'$. Therefore

$$\text{every } M'\in\mathcal{M}'(u,x) \text{ includes some } M\in\mathcal{M}(u,x). \tag{5.17}$$

Our claim that the definitions (5.8) would be unaffected by a substitution of $\mathcal{M}'(u, x)$ for $\mathcal{M}(u, x)$ is correct, in view of (5.16) and (5.17).

Suppose now that $y^j \in K^2(u^j, x^j)$ and $(u^j, x^j, y^j) \to (u, x, y)$ with $x^j \in F(u^j)$. To finish the verification of (5.11), we need to show that $y \in K^2(u, x)$. Certainly $y \in K^1(u, x)$ by (5.2), and for each j there is by the definition of $K^2(u^j, x^j)$ a subspace $M^j \in \mathcal{M}(u^j, x^j)$ such that

$$w \cdot \nabla_x^2 l(x^j, y^j) w \geq 0 \quad \text{for all } w \in M^j. \tag{5.18}$$

From (5.16) we actually have $M^j \in \mathcal{M}'(u^j, x^j)$; thus there exist sequences $(u^{jk}, x^{jk} \to_k (u^j, x^j)$ with $x^{jk} \in F(u^{jk})$ and $Z(u^{jk}, x^{jk}) \to_k M^j$. Then $Z(u^{jk}, x^{jk}) \cap B$ converges to $M^j \cap B$ in the Hausdorff metric (where $B = \{w \in \mathbf{R}^n \mid |w| \leq 1\}$), so for

$$\epsilon_{jk} = -\min\{w \cdot \nabla^2 l(x^j, y^j) w \mid w \in Z(u^{jk}, x^{jk}) \cap B\} \geq 0$$

we have $\epsilon_{jk} \to_k 0$. Diagonalizing, we can choose for each j an index k_j in such a way that for

$$\tilde{u}^j = u^{jk_j}, \qquad \tilde{x}^j = x^{jk_j}, \qquad \tilde{\epsilon}_j = \epsilon_{jk_j},$$

we get $(\tilde{u}^j, \tilde{x}^j) \to (u, x)$ and $\tilde{\epsilon}_j \to 0$. Then $\tilde{x}^j \in F(\tilde{u}^j)$ and

$$w \cdot \nabla^2 l(x^j, y^j) w \geq -\tilde{\epsilon}_j \quad \text{for all } w \in Z(\tilde{u}^j, \tilde{x}^j) \cap B. \tag{5.19}$$

Passing to a subsequence if necessary, we can assume that $Z(\tilde{u}^j, \tilde{x}^j)$ converges to some subspace M' as $j \to \infty$. Then $M' \in \mathcal{M}'(u, x)$ by definition (5.15). Furthermore, $Z(\tilde{u}^j, \tilde{x}^j) \cap B$ converges to $M' \cap B$ in the Hausdorff metric, and it follows therefore from (5.19) and the continuity of $\nabla_x^2 l$ that $w \cdot \nabla_x^2 l(x, y) w \geq 0$ for all $w \in M' \cap B$, hence for all $w \in M'$. By (5.17) there exists $M \in \mathcal{M}(u, x)$ with $M \subset M'$. Thus for a certain $M \in \mathcal{M}(u, x)$ we have $w \cdot \nabla_x^2 l(x, y) w \geq 0$ for all $w \in M$, and we may conclude that $y \in K^2(u, x)$, as we wanted. This proves (5.11).

The verification of (5.12) proceeds along identical lines and yields the closedness of $K_0^2(u, x)$. For (5.13), the argument differs only in having, not $y^j \to y$, but $t_j y^j \to y$ with $t_j \downarrow 0$; then $\nabla_x^2 l(x^j, y^j)$ is replaced in (5.19) by

$$t_j \nabla_x^2 l(x^j, y^j) = t_j \nabla^2 f_0(x^j) + \sum_{i=1}^m t_j y_i^j \nabla^2 f_i(x^j) \to \nabla_x^2 l_0(x, y).$$

The case of constant sequences $\{(u^j, x^j)\}$ in (5.13) yields the inclusion in (5.10).

The only thing left to prove in Proposition 6 is (5.14). But this is immediate from the definitions of $Y(u)$ and $Y_0(u)$ in Theorem 2, the relations

$$A(u') \subset N(u', x') \subset K^2(u', x') \quad \text{when } x' \in X(u') \tag{5.20}$$

(cf. Proposition 5 and definition (5.8)), and the limit inclusions (5.11) and (5.13).

Corollary. *The limit conditions* (4.28) *and* (4.29) *in the hypothesis of Theorem* 5 *are satisfied if*

$$\emptyset \neq K^2(u,x) \subset N(u,x) \quad \text{and} \quad K_0^2(u,x) \subset N_0(u,x) \quad \text{for each } x \in X(u). \tag{5.21}$$

Proof. This is obvious from (5.11), (5.13), and the general inclusion $N(u',x') \subset K^2(u',x')$.

The generalization of Theorems 1 and 4 which we have been aiming at can now be stated.

Theorem 6. *Let u be such that $p(u) < \infty$, and define*

$$Y^2(u) = \bigcup_{x \in X(u)} K^2(u,x), \qquad Y_0^2(u) = \bigcup_{x \in X(u)} K_0^2(u,x). \tag{5.22}$$

Then $Y^2(u)$ and $Y_0^2(u)$ are closed subsets of $\mathbf{R}^m$ such that $Y_0^2(u)$ is a cone with

$$0 \in Y_0^2(u) \quad \text{and} \quad 0^+Y^2(u) \subset Y_0^2(u). \tag{5.23}$$

One always has

$$\partial p(u) \subset \operatorname{cl} \operatorname{co}[Y^2(u) + Y_0^2(u)] \tag{5.24}$$

and

$$Y^2(u) = \emptyset \Rightarrow \partial p(u) = \emptyset \Rightarrow \{0\} \neq \partial^0 p(u) \subset \operatorname{cl} \operatorname{co} Y_0^2(u). \tag{5.25}$$

Moreover

$$p^{\uparrow}(u;h) \leq \begin{cases} \sup_{y \in Y^2(u)} y \cdot h & \text{if } y \cdot h \leq 0 \text{ for all } y \in Y_0^2(u) \\ \infty & \text{otherwise,} \end{cases} \tag{5.26}$$

with

$$\operatorname{int}\{h \mid p^{\uparrow}(u;h) < \infty\} \supset \{h \mid y \cdot h < 0 \quad \text{for all } y \in Y_0^2(u)\}. \tag{5.27}$$

The latter set is nonempty if and only if $Y_0^2(u)$ is pointed, in which case $\partial^0 p(u)$ is pointed too and actually

$$\partial p(u) \subset \operatorname{co}[Y^2(u) + Y_0^2(u)], \qquad \partial^0 p(u) \subset \operatorname{co} Y_0^2(u). \tag{5.28}$$

This is the case in particular when there are no equality constraints in (P_u), *i.e. when $s = m$.*

Proof. The assertions about the nature of $Y^2(u)$ and $Y_0^2(u)$ are supported by Proposition 6 and the compactness of $X(u)$. To deduce the remainder of the result from Theorem 2, simply apply Proposition 3 using the inclusions $Y(u) \subset Y^2(u)$ and $Y_0(u) \subset Y_0^2(u)$ furnished by (5.14).

Corollary. *Let u be such that $p(u) < \infty$, and $K_0^2(u, x) = \{0\}$ for every $x \in X(u)$. Then p is Lipschitz continuous on a neighborhood of u, the set*

$$\bigcup_{x \in X(u)} K^2(u, x)$$

is compact, and one has

$$\emptyset \neq \partial p(u) \subset \mathrm{co}\Big[\bigcup_{x \in X(u)} K^2(u, x)\Big],$$

so that

$$-\infty < p^{\uparrow}(u; h) \leq \max\{y \cdot h \mid y \in K^2(u, x) \quad \textit{for some}\, x \in X(u)\}.$$

Proof. These conclusions follow from Theorem 6 when $Y_0^2(u) = \{0\}$ (cf. the facts about Lipschitz continuity cited in Section 1).

Since $K^2(u, x) \subset K^1(u, x)$ and $K_0^2(u, x) \subset K_0^1(u, x)$, the preceding corollary sharpens Theorem 1 (Gauvin), just as Theorem 6 more generally sharpens the first-order facts in Theorem 4.

Example. A simple illustration of the sharper nature of Theorem 6 and its corollary is provided by the problem.

$$\begin{aligned} &\text{minimize} && f_0(x_1, x_2, x_3) = x_1^4 + x_2^4 + x_3^4 + x_1^2 - x_2^2 - x_3^2, \\ &\text{subject to} && 0 = f_1(x_1, x_2, x_3) + u_1 = x_1^2 - x_2^2 - x_3^2 + u_1. \end{aligned} \tag{5.29}$$

This satisfies our inf-boundedness assumption (1.1) and has, for $u_1 = 0$, a unique optimal solution $\bar{x} = (0, 0, 0)$. Moreover $\nabla f_0(\bar{x}) = \nabla f_1(\bar{x}) = (0, 0, 0)$ so

$$K^1(0, \bar{x}) = K_0^1(0, \bar{x}) = \boldsymbol{R}^1.$$

Since $I(0, \bar{x}) = \{1\}$, the subspaces in the collection $\mathcal{M}(0, \bar{x})$ all have dimension 2 or more (actually $\mathcal{M}(0, \bar{x})$ consists of all such subspaces of $\boldsymbol{R}^3$), but $\nabla_x^2 l(\bar{x}, y) = 2(y-1)J$ and $\nabla_x^2 l_0(\bar{x}, y) = 2yJ$, where

$$J = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

There does not exist a subspace M of $\boldsymbol{R}^3$ with $\dim M \geq 2$, such that $w \cdot Jw \geq 0$ for all $w \in M$. Therefore

$$K^2(0, \bar{x}) = \{1\}, \quad K_0^2(0, \bar{x}) = \{0\},$$

and the corollary above is applicable: $\emptyset \neq \partial p(0) \subset K^2(0, \bar{x})$, hence $\partial p(0) = \{1\}$. Then in fact p must be strictly differentiable at 0 with derivative $p'(0) = 1$ (cf. Section 1).

Note that this situation is not covered by Theorem 5, since $S(0,\bar{x})=\emptyset$. Another interesting feature is that the feasible set $F(0)$ consists of the points (x_1,x_2,x_3) satisfying

$$x_3=\pm[x_1^2+x_2^2]^{1/2}.$$

The behavior of this set at $\bar{x}=(0,0,0)$ does not lend itself to characterization by means of any constraint qualification which implies 'tangential convexity' at $\bar{x}$ (as does the Mangasarian–Fromovitz condition, cf. Proposition 1). Yet the condition $K_0^2(0,\bar{x})=\{0\}$ is able here to act as a constraint qualification and provide strong information.

Before turning to the question of how all this is related to second-order necessary conditions for optimality, we prove an auxiliary result which, in some cases, furnishes a complete description of the subspace collection $\mathcal{M}(u,x)$ used in defining $K^2(u,x)$ and $K_0^2(u,x)$. For each vector $z\in\boldsymbol{R}^n$, let

$$\begin{aligned}M(u,x,z)=\{w\in Z(u,x)\mid \exists\, w'\in\boldsymbol{R}^n\\ \text{with } w\cdot\nabla^2 f_i(x)z+w'\cdot\nabla f_i(x)=0,\ i\in I(u,x)\}\end{aligned}\tag{5.30}$$

where $Z(u,x)$ is the subspace in (5.7). Clearly $M(u,x,z)$ is a subspace included in $Z(u,x)$, and $M(u,x,z)=Z(u,x)$ for $z=0$.

Proposition 7. *Suppose that $x\in F(u)$ and the following condition holds: if η_i and η_i' are coefficients such that, for some vector $z\neq 0$,*

$$\sum_{i\in I(u,x)}\eta_i\nabla f_i(x)=0\quad and\quad\Big[\sum_{i\in I(u,x)}\eta_i\nabla^2 f_i(x)\Big]z+\sum_{i\in I(u,x)}\eta_i'\nabla f_i(x)=0,\tag{5.31}$$

then $\eta_i=0$ for all $i\in I(u,x)$.

Then $\mathcal{M}(u,x)$ consists of all the subspaces $M(u,x,z)$, as z ranges over $\boldsymbol{R}^n$.

Proof. We already know that $\mathcal{M}(u,x)$ contains the subspace $Z(u,x)=M(u,x,0)$, and that every subspace in $\mathcal{M}(u,x)$ has dimension at least $n-|I(u,x)|$. We shall demonstrate first that every subspace in $\mathcal{M}(u,x)$, other than $Z(u,x)$, is included in a subspace of form $M(u,x,z)$ with $z\neq 0$, and second that when (5.31) holds, every such subspace $M(u,x,z)$ has dimension $n-|I(u,x)|$. Since two subspaces, one included in the other, must be equal if they have the same dimension, this will establish the result.

Consider a convergent sequence of subspaces M^k of form (5.5) with $x^k\to x$, such that the limit of M^k is properly included in $Z(u,x)$, and therefore $x^k\neq x$ except perhaps for finitely many indices k. Passing to subsequences if necessary, we can suppose that the vectors $z^k=(x^k-x)/|x^k-x|$ converge to some z (with $|z|=1$). Each $w\in\lim_k M^k$ is a limit of vectors $w^k\in M^k$; for every $i\in I(u,x)$ one has $w \cdot \nabla f_i(x) = 0$ and $w^k \cdot \nabla f_i(x^k) = 0$, so

$$0 = w^k \cdot [\nabla f_i(x^k) - \nabla f_i(x)]/|x^k - x| + [(w^k - w)/|x^k - x|] \cdot \nabla f_i(x). \qquad (5.32)$$

Let Λ be the subspace of $\boldsymbol{R}^m$ consisting of all vectors $\lambda = (\lambda_1, \dots, \lambda_m)$ such that there exists $w' \in \boldsymbol{R}^n$ with

$$\lambda_i = w' \cdot \nabla f_i(x) \quad \text{for all } i \in I(u, x).$$

According to (5.32), each of the vectors λ^k defined by

$$\lambda_i^k = w^k \cdot [\nabla f_i(x^k) - \nabla f_i(x)]/|x^k - x| \quad \text{for } i = 1, \dots, m$$

belongs to Λ. Then so does the vector $\lambda = \lim_k \lambda^k$, whose components are $\lambda_i = w \cdot \nabla^2 f_i(x) z$. Therefore $w \in M(u, x, z)$, and it follows that $\lim_k M^k \subset M(u, x, z)$.

Our next task is to show that $\dim M(u, x, z) = n - |I(u, x)|$ if $z \neq 0$. Let

$$M^*(u, x, z) = \{(w, w') \in \boldsymbol{R}^{2n} \mid w \cdot \nabla f_i(x) = 0 \quad \text{for all } i \in I(u, x), \quad \text{and} \\ w \cdot \nabla^2 f_i(x) z + w' \nabla f_i(x) = 0 \quad \text{for all } i \in I(u, x)\}.$$

Clearly $M(u, x, z)$ is the image of the subspace $M^*(u, x, z)$ under the projection $(w, w') \to w$. The orthogonal complements of these spaces therefore satisfy

$$M(u, x, z)^\perp = \{v \mid (v, 0) \in M^*(u, x, z)^\perp\}. \qquad (5.33)$$

Moreover, $M^*(u, x, z)^\perp$ is the set of all vectors $(v, v') \in \boldsymbol{R}^{2n}$ of the form

$$\sum_{i \in I(u,x)} \eta_i (\nabla^2 f_i(x) z, \nabla f_i(x)) + \sum_{i \in I(u,x)} \eta_i' (\nabla f_i(x), 0) \\ = \Big(\Big[\sum_{i \in I(u,x)} \eta_i \nabla^2 f_i(x) \Big] z + \sum_{i \in I(u,x)} \eta_i' \nabla f_i(x), \sum_{i \in I(u,x)} \eta_i \nabla f_i(x) \Big),$$

for arbitrary coefficients η_i and η_i'. Since

$$\Big\{ \sum_{i \in I(u,x)} \eta_i' \nabla f_i(x) \mid \eta_i' \in \boldsymbol{R} \} = Z(u, x)^\perp, \qquad (5.34)$$

we obtain from (5.33) that $M(u, x, z)^\perp$ consists of all vectors of the form

$$\Big[\sum_{i \in I(u,x)} \eta_i \nabla^2 f_i \Big] z + v \quad \text{with } \textstyle\sum_{i \in I(u,x)} \eta_i \nabla f_i(x) = 0, \ v \in Z(u, x)^\perp.$$

In accordance with this expression, let

$$G = \Big\{ \eta \in \boldsymbol{R}^m \mid \eta_i = 0 \quad \text{for} \quad i \notin I(u, x), \quad \text{and } \textstyle\sum_{i=1}^m \eta_i \nabla f_i(x) = 0 \Big\},$$

$$T(\eta, v) = \Big[\sum_{i=1}^m \eta_i \nabla^2 f_i(x) \Big] z + v.$$

Then G is a subspace of $\boldsymbol{R}^m$, T is a linear transformation from $\boldsymbol{R}^m \times \boldsymbol{R}^n$ into $\boldsymbol{R}^n$, and the image of $G \times Z(u,x)^{\perp}$ under T is $M(u,x,z)^{\perp}$. Furthermore, the only vector (η, v) in $G \times Z(u,x)^{\perp}$ with $T(\eta, v) = 0$ is $(\eta, v) = (0,0)$; this is the substance of condition (5.31). Hence T is one-to-one on $G \times Z(u,x)^{\perp}$, and we have

$$\begin{aligned}\dim M(u,x,z)^{\perp} &= \dim[G \times Z(u,x)^{\perp}],\\ &= \dim G + \dim Z(u,x)^{\perp}.\end{aligned} \tag{5.35}$$

Now consider

$$G_0 = \{\eta \in \boldsymbol{R}^m \mid \eta_i = 0 \quad \text{for} \quad i \not\in I(u,x)\},$$

$$T_0(\eta) = \sum_{i=1}^{m} \eta_i \nabla f_i(x).$$

Again G_0 is a subspace of $\boldsymbol{R}^m$, and T_0 is a linear transformation from $\boldsymbol{R}^m$ into $\boldsymbol{R}^n$. The image of G_0 under T_0 is $Z(u,x)^{\perp}$ by (5.34), while the set of $\eta \in G_0$ with $T_0(\eta) = 0$ is G. Therefore

$$\dim G + \dim Z(u,x)^{\perp} = \dim G_0 = |I(u,x)|,$$

and we deduce from (5.35) that

$$\dim M(u,x,z) = n - \dim M(u,x,z)^{\perp} = n - |I(u,x)|,$$

as was our goal.

Our next job is to show that the sets $K^2(u,x)$ and $K_0^2(u,x)$ are not only of value in the study of $\partial p(u)$, but also furnish a second-order necessary condition for local optimality in (P_u).

Theorem 7. *Suppose x is a locally optimal solution to (P_u) such that $K_0^2(u,x) = \{0\}$. Then the multiplier set $K^2(u,x)$ is nonempty and compact.*

Proof. A function g of class $\mathscr{C}^2$ can be constructed with the properties that $g(x) = 0$, $\nabla g(x) = 0$, $\nabla^2 g(x) = 0$, and

$$g(x') > f_0(x) - f_0(x') \quad \text{for all } x' \in F(u),\ x' \neq x.$$

Let $\bar{f}_0 = f_0 + g$, and let $\bar{p}$ be the optimal value function obtained in place of p by this modification. The modified problem $(\bar{\mathrm{P}}_u)$ has x as its unique optimal solution, and it has the same multiplier sets at x as does (P_u), namely $K^2(u,x)$ and $K_0^2(u,x)$, because these sets do not depend on f_0 beyond the values of $\nabla f_0(x)$ and $\nabla^2 f_0(x)$, which are the same for $\bar{f}_0$ as for f_0. Since $K_0^2(u,x) = \{0\}$, we know by the corollary to Theorem 6 that $\emptyset \neq \partial \bar{p}(u) \subset K^2(u,x)$ (compact), and this is all we needed to prove.

An alternative to the constraint qualification $K_0^2(u,x) = \{0\}$ in Theorem 7, but yielding a weaker conclusion, is a property we have introduced in [18] in connection with first-order necessary conditions for nonsmooth problems. We say that (P_u) *is calm at* x, one of its *locally* optimal solutions, if there do *not* exist sequences $u^j \to u$, $x^j \to x$, such that $x^j \in F(u^j)$ and

$$[f_0(x^j) - f_0(x)]/|u^j - u| \to -\infty.$$

Theorem 8. *Suppose x is a locally optimal solution to (P_u) such that (P_u) is calm at x. Then the multiplier set $K^2(u, x)$ is nonempty (and closed, but not necessarily compact).*

Proof. Calmness of (P_u) at x implies by [18, Proposition 12] that for g constructed as in the proof of Theorem 7 and having the form $g(x') = \theta(|x'|)$ for θ convex on $\boldsymbol{R}^1$, the modified optimal value function $\tilde{p}$ satisfies

$$\liminf_{u' \to u} \frac{\tilde{p}(u') - \tilde{p}(u)}{|u' - u|} > -\infty.$$

As shown in [18, Proposition 1], this ensures that $\partial\tilde{p}(u) \neq \emptyset$. But since the modified problem $(\tilde{\mathrm{P}}_u)$ has x as its unique optimal solution, with corresponding multiplier sets $K^2(u, x)$ and $K_0^2(u, x)$, the same as in (P_u), we have by Theorem 6

$$\partial\tilde{p}(u) \subset \operatorname{cl} \operatorname{co}[K^2(u, x) + K_0^2(u, x)].$$

Therefore $K^2(u, x) \neq \emptyset$.

The traditional approach to second-order necessary conditions for optimality in (P_u), and for that matter first-order conditions, relies on the existence of certain sequences or arcs which approach a locally optimal solution x from elsewhere in the feasible set $F(u)$ (cf. Hestenes [6], [7], Fiacco and McCormick [3]). In our approach, perturbations of u as well as x play a role. To provide a closer comparison between the results obtainable by the two approaches, we consider another kind of constraint qualification.

A well-known condition under which $K^1(u, x)$ must be nonempty when x is locally optimal in (P_u) is the following (cf. Hestenes [6]): for every nonzero $w \in \boldsymbol{R}^n$ satisfying

$$\nabla f_i(x) \cdot w \begin{cases} \leq 0 & \text{for } i \in I(u, x),\ i \leq s, \\ = 0 & \text{for } i = s+1 \dots, m, \end{cases}$$

there is a sequence $x^k \to x$ in $F(u)$, $x^k \neq x$, with $(x^k - x)/|x^k - x| \to w$. This is appropriately called the *first-order tangential constraint qualification*. Now let

$$I^*(u, x) = \{i \in I(u, x) \mid \nabla f_i(x) \cdot w = 0 \quad \text{for all } w \in W(u, x)\}. \tag{5.36}$$

We shall say that the *second-order tangential constraint qualification* is satisfied if for every nonzero w satisfying

$$\nabla f_i(x) \cdot w \begin{cases} < 0 & \text{for } i \in I(u, x) \setminus I^*(u, x), \\ = 0 & \text{for } i \in I^*(u, x), \end{cases} \tag{5.37}$$

there is a sequence $x^k \to x$, $x^k \neq x$, with $f_i(x^k) + u_i = 0$ for all $i \in I^*(u, x)$ and $(x^k - x)/|x^k - x| \to w$. This is quite similar to the second-order constraint qualification of Hestenes [6], with respect to x and a multiplier vector $y \in K^1(u, x)$, namely that for each nonzero w satisfying

$$\nabla f_i(x) \cdot w \begin{cases} \leq 0 & \text{for } i \in I_0(u, x, y), \\ = 0 & \text{for } i \in I_1(u, x, y), \end{cases} \tag{5.38}$$

where $I_0(u, x, y)$ and $I_1(u, x, y)$ are the index sets defined in (4.3), there is a sequence $x^k \to x$, $x^k \neq x$, with

$$f_i(x^k) + u_i \begin{cases} \leq 0 & \text{for } i \in I_0(u, x, y), \\ = 0 & \text{for } i \in I_1(u, x, y), \end{cases} \tag{5.39}$$

and $(x^k - x)/|x^k - x| \to w$. Under the latter condition, y must belong to $N(u, x)$ (cf. [6, p. 37]). (Fiacco and McCormick [3] make a stronger assumption in terms of w being tangent to an arc of class $\mathscr{C}^2$ in $F(u)$; they have equality in place of the inequalities in (5.38) and (5.39) and correspondingly they conclude only that $w \cdot \nabla_x^2 l(x, y) w \geq 0$ for a smaller set of vectors w.)

A peculiarity of Hestenes' second-order constraint qualification is that the system (5.38) does not actually depend on y. The vectors which satisfy it are precisely the ones in $W(u, x)$, as already noted at the beginning of Section 4. It follows that no assumption on the gradients in (5.38), such as a generalized Mangasarian–Fromovitz condition of some sort which could assist in verification, can possibly distinguish between different forms of (5.39) associated with different vectors $y \in K^1(u, x)$. In practice, therefore, one might just as well settle for the best version of Hestenes' condition that can be formulated without reference to any particular y. This is the motivation behind the second-order tangential constraint qualification formulated above. Indeed, when $K^1(u, x) \neq \emptyset$ one has

$$\begin{aligned} I^*(u, x) &= \bigcup \{I_1(u, x, y) \mid y \in K^1(u, x)\} \\ &= I_1(u, x, y) \qquad \text{for arbitrary } y \in \operatorname{ri} K^1(u, x) \end{aligned} \tag{5.40}$$

(where ri C denotes the relative interior of a convex set C [11, Section 6]), as can readily be seen by way of the lemma of Farkas.

Theorem 9. *Suppose x is a locally optimal solution to (P_u) at which the second-order tangential constraint qualification is satisfied. Then*

$$N(u, x) = K^2(u, x) = K^1(u, x). \tag{5.41}$$

Proof. We may assume $K^1(u, x) \neq \emptyset$, for otherwise (5.41) holds with all sets empty, cf. (5.9). The vectors w satisfying (5.51) are then precisely the ones belonging to the relative interior ri $W(u, x)$ of the polyhedral cone $W(u, x)$ (cf.

[11, Section 6]). Consider any such w and a corresponding sequence $x^k \to x$ with $f_i(x^k) + u_i = 0$ and all $i \in I^*(u, x)$ and $(x^k - x)/|x^k - x| \to w$, such as is guaranteed by our constraint qualification. For $i \in I(u, x) \setminus I^*(u, x)$, we have by (5.37) and the mean value theorem

$$\lim_{k\to\infty} [f_i(x^k) - f_i(x)]/|x^k - x| < 0,$$

where $f_i(x) = -u_i$ and therefore $f_i(x^k) + u_i < 0$ for large k. Then $x^k \in F(u)$, so $f_0(x^k) \geq f_0(x)$ by the local optimality of x. Now consider any $y \in K^1(u, x)$. Since $W(u, x)$ has the alternative description as the set of w satisfying $\nabla f_i(x) \cdot w \leq 0$ for $i \in I_0(u, x, y)$ and $\nabla f_i(x) \cdot w = 0$ for $i \in I_1(u, x, y)$ (cf. beginning of Section 4), we have $y_i = 0$ for all $i \in I(u, x) \setminus I^*(u, x)$. Hence $y_i f_i(x^k) = y_i f_i(x) = -y_i u_i$ for $i = 1, \ldots, m$, so that

$$l(x^k, y) = f_0(x^k) \geq f_0(x) = l(x, y) \quad \text{for large } k.$$

Recalling that $\nabla_x l(x, y) = 0$ (because $y \in K^1(u, x)$), we calculate

$$0 \leq \lim_{k\to\infty} \frac{l(x^k, y) - l(x, y) - \nabla_x l(x, y) \cdot (x^k - x)}{|x^k - x|^2} = \frac{1}{2} w \cdot \nabla_x^2 l(x, y) w.$$

This being true for arbitrary $w \in \operatorname{ri} W(u, x)$, it also holds for all $w \in \operatorname{cl}(\operatorname{ri} W(u, x)) = W(u, x)$. Thus $y \in N(u, x)$. We have shown $K^1(u, x) \subset N(u, x)$, and the equalities in (5.41) now follow at once from the general inclusion in (5.9).

Corollary. *Suppose x is a locally optimal solution to* (P_u) *such that for all x' in some neighborhood of x, the matrix $J(x')$ whose rows are the vectors $\nabla f_i(x')$ for $i \in I^*(u, x)$ has constant rank. (This is true in particular if the vectors $\nabla f_i(x)$ for $i \in I^*(u, x)$ are linearly independent.) Then the conclusion* (5.41) *of Theorem 9 is valid.*

Proof. Let the rank in question be d, and let I' be any subset of $I^*(u, x)$ such that the vectors $\nabla f_i(x)$ for $i \in I'$ are linearly independent. By a classical theorem in advanced calculus (based on the implicit function theorem), there is a neighborhood of x in which each f_i for $i \notin I'$ can be expressed as a $\mathscr{C}^2$ function of the f_i's for $i \in I'$. Then for each w satisfying $\nabla f_i(x) \cdot w = 0$ for all $i \in I'$ there is an arc $a(t)$ of class $\mathscr{C}^2$ with $a(0) = x$, $a'(0) = w$, and $f_i(a(t)) = f_i(x)$ for all $i \in I'$, hence for all $i \in I^*(u, x)$. In particular, the hypothesis of Theorem 9 is satisfied in this case.

Theorem 9 and its corollary can be interpreted as saying that the main results effectively obtainable by the traditional 'tangential' approach to second-order necessary conditions are covered, in a sense, by the ones in Theorems 7 and 8 in terms of $K^2(u, x)$. Moreover the latter are definitely more general to the extent of being able to handle situations where it does not turn out that *every* vector y

satisfying the first-order conditions automatically satisfies the second-order conditions.

There is, however, a class of second-order results which is not covered by our approach, namely those where a possibly different y with certain properties is associated with each $w \in W(u, x)$ (cf. Hestenes [7], Ioffe [8], Gollan [5]). The results of Gollan [5] do include an estimate for $\partial p(u)$ complementary to the one given here. To formulate this, let

$$\hat{W}(u, x) = \{w \in \boldsymbol{R}^n \mid \nabla f_i(x) \cdot w = 0, \forall\, i \in I(u, x)\},$$
$$\hat{K}^2(u, x, w) = \{y \in K^1(u, x) \mid w \cdot \nabla_x^2 l(x, y) w \geq 0\},$$
$$\hat{K}_0^2(u, x, w) = \{y \in K_0^1(u, x) \mid w \cdot \nabla_x^2 l_0(x, y) w \geq 0\}.$$

Gollan shows that for any choice of $w(x) \in W(u, x)$ for each $x \in X(u)$, one has

$$\partial p(u) \subset \operatorname{cl} \operatorname{co}\Big[\bigcup_{x \in X(u)} \hat{K}^2(u, x, w(x)) + \bigcup_{x \in X(u)} \hat{K}_0^2(u, x, w(x))\Big].$$

References

[1] A. Auslender, "Differential stability in nonconvex and nondifferentiable programming", *Mathematical Programming Study* 10 (1978).

[2] F.H. Clarke, "Generalized gradients and applications", *Transactions of the American Mathematical Society* 205 (1975) 247–262.

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

[4] J. Gauvin, "The generalized gradient of a marginal value function in mathematical programming", *Mathematics of Operations Research* 4 (1979) 458–463.

[5] B. Gollan, "On the marginal function in nonlinear programming", *Mathematics of Operations Research*, forthcoming.

[6] M.R. Hestenes, *Calculus of variations and optimal control theory* (Wiley and Sons, New York, 1968).

[7] M.R. Hestenes, *Optimization theory: The finite-dimensional case* (Wiley and Sons, New York, 1975).

[8] A.D. Ioffe, "Necessary and sufficient conditions for a local minimum. 3: Second-order conditions and augmented duality", *SIAM Journal on Control and Optimization* 17 (1979) 266–288.

[9] E.S. Levitin, A.A. Miljutin and N.P. Osmolovskii, "On conditions for a local minimum in a problem with constraints", in: B.S. Mitjagin, ed., *Mathematical economics and functional analysis* (Nauka, Moscow, 1974) pp. 139–202 (in Russian).

[10] O.L. Mangasarian and S. Fromovitz, "The Fritz John necessary conditions in the presence of equality and inequality constraints", *Journal of Mathematical Analysis and Applications* 17 (1967) 37–47.

[11] R.T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, NJ, 1970).

[12] R.T. Rockafellar, "Augmented Lagrange multiplier functions and duality in nonconvex programming", *SIAM Journal on Control* 12 (1974) 268–285.

[13] R.T. Rockafellar, Solving a nonlinear programming problem by way of a dual problem", *Symposia Mathematica* 19 (1976) 135–160.

[14] R.T. Rockafellar, "Clarke's tangent cones and the boundaries of closed sets in $\boldsymbol{R}^n$", *Nonlinear Analysis: Theory, Methods and Applications* 3 (1979) 45–154.

[15] R.T. Rockafellar, "Generalized directional derivatives and subgradients of nonconvex functions", *Canadian Journal of Mathematics* 32 (1980) 257–280.

[16] R.T. Rockafellar, *The theory of subgradients and its applications: Convex and nonconvex functions* (Halderman Verlag, West Berlin, 1981).

[17] R.T. Rockafellar, "Proximal subgradients, marginal values and augmented Lagrangians in nonconvex optimization", *Mathematics of Operations Research* 6 (1981) 424–436.

[18] R.T. Rockafellar, "Lagrange Multipliers and subderivatives of optimal value functions in nonlinear programming", *Mathematical Programming Study* 17 (1982) 28–66.

[19] G. Salinetti and R.J.B. Wets, "On the convergence of convex sets in finite dimensions", *SIAM Review* 21 (1979) 18–33.

[20] R.J.B. Wets, "On a compactness theorem for epiconvergent sequences of functions", forthcoming.