Mathematical Programming Study 17 (1982) 28–66.
North-Holland Publishing Company

LAGRANGE MULTIPLIERS AND SUBDERIVATIVES OF OPTIMAL VALUE FUNCTIONS IN NONLINEAR PROGRAMMING*

R.T. ROCKAFELLAR

Department of Mathematics, University of Washington, Seattle, WA 98195, U.S.A.

Received 24 September 1980
Revised manuscript received 25 March 1981

For finite-dimensional optimization problems with locally Lipschitzian equality and inequality constraints and also an abstract constraint described by a closed set, a Lagrange multiplier rule is derived that is sharper is in some respects than the ones of Clarke and Hiriart-Urruty. The multiplier vectors provided by this rule are given meaning in terms of the generalized subgradient set of the optimal value function in the problem with respect to perturbational parameters. Bounds on subderivatives of the optimal value function are thereby obtained and in certain cases the existence of ordinary directional derivatives.

Key words: Lagrange Multipliers, Subgradients, Marginal Values, Nonlinear Programming.

1. Introduction

In this paper we study an optimization problem that depends on parameter vectors $u = (u_1, \dots, u_m) \in \mathbb{R}^m$ and $v = (v_1, \dots, v_d) \in \mathbb{R}^d$:

$$(\mathrm{P}_{u,v})\qquad \begin{array}{l} \text{minimize}\quad f_0(v,x) \text{ over all } x \in \mathbb{R}^n \text{ such that } (v,x) \in D \text{ and} \\ f_i(v,x) + u_i \begin{cases} \leq 0 & \text{for } i = 1, \dots, s, \\ = 0 & \text{for } i = s+1, \dots, m, \end{cases} \end{array}$$

where D is a subset of $\mathbb{R}^d \times \mathbb{R}^n$ and each f_i is a real-valued function on an open set which includes D. We assume that D is closed and f_i is locally Lipschitzian on D (i.e., Lipschitz continuous relative to some $\mathbb{R}^n$-neighborhood of each point of D). Examples where such assumptions are fulfilled include

(a) *the smooth case*: $D = \mathbb{R}^d \times \mathbb{R}^n$ and every f_i of class $\mathscr{C}^1$;

(b) *the convex case*: D closed convex, f_i convex for $i = 0, 1, \dots, s$ and affine for $i = s+1, \dots, m$;

(c) *the mixed smooth-convex case*: $D = \mathbb{R}^d \times C$ with C closed convex, $f_i(v,x)$ of class $\mathscr{C}^1$ with respect to v (the gradient depending continuously on (v,x) rather than just v), as well as convex in x for $i = 0, 1, \dots, s$ and affine in x for $i = s+1, \dots, m$.

Clarke [3] has obtained a Lagrange multiplier rule that unifies the known

*Research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under grant no. 77–3204 at the University of Washington, Seattle.

first-order necessary conditions for optimality in nonlinear programming problems of types (a), (b) or (c) but is applicable as well to the general case of $(P_{u,v})$. This rule has been sharpened by Hiriart-Urruty [18]. Our objective here is to explore the connection between the Lagrange multipliers for the constraints in $(P_{u,v})$ as provided by such a rule, and certain generalized directional derivatives and subgradients of the function

$$p(u, v) = \inf(P_{u,v}) \qquad \text{(global optimal value)}$$

with respect to both u and v.

The function p is everywhere-defined on $\mathbb{R}^m \times \mathbb{R}^d$ under the convention that $p(u, v) = +\infty$ when $(P_{u,v})$ is infeasible, but it can well fail to be differentiable in the ordinary sense at points where it is finite, even in the smooth case (a). Nevertheless, p is of such obvious interest that quite apart from any connection with Lagrange multipliers, there is strong motivation for pushing beyond differentiability to some sort of subdifferential theory of its properties. Generalized derivatives of p have direct significance in sensitivity analysis and in determining criteria for Lipschitzian behavior of p and the like. They also furnish information that might be used in minimizing $p(u, v)$ subject to further constraints on u and v, as can be the task posed in decomposition techniques where $(P_{u,v})$ appears as just a subproblem of a larger problem.

Of even greater importance, though, is the role that generalized derivatives of the optimal value function p can have in answering fundamental questions about the existence, uniqueness and interpretation of Lagrange multiplier vectors, questions which have a bearing on many aspects of theory and computation. This role is well understood in the convex case (cf. [26]) and to some extent also through partial results in the smooth and mixed cases listed above, but it has not been clarified for $(P_{u,v})$ in general.

Roughly speaking on the basis of experience in the special cases which have been tackled, possible rates of change of $p(u, v)$ with respect to u_i should have something to do with possible multiplier values y_i associated with the ith constraint in $(P_{u,v})$. The study of variations with respect to the parameters v_l as well as u_i is approachable by the same idea, because $p(u, v)$ can equally be regarded as the optimal value in the problem:

$$(P'_{u,v}) \qquad \begin{aligned} &\text{minimize} \quad f_0(w, x) \text{ over all } (w, x) \in D \text{ satisfying} \\ &\qquad f_i(w, x) + u_i \begin{cases} \le 0 & \text{for } i = 1, \dots, s, \\ = 0 & \text{for } i = s+1, \dots, m \end{cases} \\ &\qquad -w_l + v_l = 0 \quad \text{for } l = 1, \dots, d. \end{aligned}$$

Multipliers z_l associated with the constraint $-w_l + v_l = 0$ in $(P'_{u,v})$ should be related to some kind of derivative of $p(u, v)$ with respect to v_l, but in view of the equivalence between $(P'_{u,v})$ and $(P_{u,v})$, such multipliers are bound to have close ties with the multipliers y_i.

Altogether then, a duality may be expected between Lagrange multiplier vectors for the constraints in $(\mathrm{P}_{u,v})$ and subdifferential properties of $p(u, v)$. Insofar as this can be formalized, it should afford valuable insight in both directions. The development of a really far-reaching duality beyond the convex case has been hampered, however, by a lack of appropriate mathematical tools and concepts.

Most of the past work on subdifferential properties of the function p has gone into the determination of formulas for the one-sided directional derivatives

$$p'(u, v; h, k) = \lim_{t \downarrow 0} \frac{p(u+th, v+tk) - p(u, v)}{t} \tag{1.1}$$

or bounds on the corresponding upper or lower Dini derivatives, where 'lim' is replaced by 'lim sup' and 'lim inf'. In the convex case (b), $p(u, v)$ is actually convex in (u, v), and $p'(u, v; h, k)$ exists for every (h, k) [26, Sections 28–29]. A theorem of Gol'shtein [15] shows that $p'(u, v; h, k)$ also exists in the mixed case (c) when the set of saddlepoints of the Lagrangian in $(\mathrm{P}_{u,v})$ is nonempty and bounded. This result, proved independently by Hogan [20], generalizes the Mills–Williams marginal value theorem in linear programming [33]. Dini derivatives were studied by Gauvin and Tolle [13] in the smooth case (a) and by Auslender [2] in the somewhat more general situation where only the equality constraints in (a) are $\mathscr{C}^1$. Bounds on Dini derivatives were used by Gauvin and Tolle to demonstrate the existence of $p'(u, v; h, k)$ under certain circumstances [13] and by Gauvin [11] to get a criterion for p to be locally Lipschitzian in the smooth case. The cited results of Gauvin and Tolle [13], Auslender [2] and Gauvin [11], ostensibly treat only parameters of type u_i, but they can be extended to parameters of type v_l using the reformulation of $(\mathrm{P}_{u,v})$ as $(\mathrm{P}'_{u,v})$. For a direct approach to such parameters, cf. [12] and related work of Fiacco and Hutzler [10].

The infinite-dimensional case too has been studied to a certain extent [21–23, 14]. Gollan [14] gives his own definition of Lagrange multipliers for nonsmooth problems, quite different from the Lagrange multipliers of Clarke mentioned earlier, but when his results are applied to classical cases they do not yield derivative bounds as strong as those of Gauvin and Tolle, for instance. Other work on ordinary one-sided derivatives of optimal value functions that should be noted for exceeding the framework in this paper in some respects, although involving significant restrictions in others, is that of Dem'janov et al. [7, 8].

Our objective here is to explore the subdifferential properties of the function p, including extensions of the results cited above, by means of a broader kind of nonlinear analysis that has blossomed from ideas of Clarke [4]. This method of analysis, the pertinent parts of which will be reviewed in Section 2, deals with certain generalized subgradients of p and corresponding 'subderivatives' that are more suited in some ways to the description of functions as irregular as p can be. Smoothness or convexity assumptions on $(\mathrm{P}_{u,v})$ are not required, yet the theory is such that the consequences of such assumptions are readily ascertained.

Subdifferential analysis in this sense has already been applied to optimal value functions like p, although not in such a thorough-going manner as in the present contribution. Clarke himself has employed a mild subderivative condition on p called 'calmness' as a constraint qualification in the derivation of his Lagrange multiplier rule [3]. A result of Gauvin [11] furnishes an outer estimate for the subgradient set $\partial p(u, v)$ in the smooth case (a). This has been carried to certain nonsmooth cases of $(P_{u,v})$, but with smooth equality constraints, by Hiriart-Urruty [7] as part of a more abstract study of marginal values. Clarke and Aubin [6] and Aubin [1] have established for other special cases of $(P_{u,v})$, via some theorems in a Banach space setting accompanied by a number of convexity assumptions, the existence in $\partial p(u, v)$ of certain multiplier vectors—thus, 'inner estimates' for $\partial p(u, v)$. All these results have concerned situations where p is Lipschitzian in a neighborhood of (u, v), and the authors (except for Hiriart-Urruty) have provided conditions on $(P_{u,v})$ that ensure this Lipschitzian behavior. In contrast, exact formulas for $\partial p(u, v)$ in the general case of $(P_{u,v})$ that are valid whether or not p is locally Lipschitzian have been given by Rockafellar [29], but in terms of *limits* of sequences of special multiplier vectors corresponding to saddle-points of the augmented Lagrangian in neighboring problems (P_{u^j,v^j}).

In this paper we derive inner and outer estimates for $\partial p(u, v)$ in terms of Lagrange multiplier vectors that satisfy Clarke's necessary conditions for $(P_{u,v})$ itself (see Section 5). By way of the duality between elements of $\partial p(u, v)$ and 'subderivatives' (see Sections 2–3), we thereby provide for the first time a general interpretation for such multiplier vectors. We also open the route to applying to p various fundamental theorems known about subgradients and subderivatives and we obtain in particular criteria for Lipschitz continuity that go well beyond previous ones. As a by-product, we get a new proof of Clarke's multiplier rule that shows it is valid under somewhat weaker assumptions, and also in a somewhat sharper form, than Clarke's or the version developed by Hiriart-Urruty [18] (see Section 4). We demonstrate that the known bounds on Dini derivatives of p follow from our subgradient estimates, without the restrictions on $(P_{u,v})$ that have been made in the past, and hold actually for Hadamard derivatives (see Section 7). We prove an extension of Golshtein's theorem for the mixed smooth-convex case of $(P_{u,v})$ that requires neither the set of optimal solutions nor the set of multiplier vectors to be compact.

A novel feature of our approach is that no form of implicit function theorem is ever used. At the critical stage we rely instead on our augmented Lagrangian results in [29].

2. Subderivatives and subgradients

The kind of subdifferential analysis initiated by Clarke for nonsmooth, nonconvex functions has in the last several years been expanded and solidified in many ways. The lecture notes [30] can serve as an introduction to the finite-dimensional case with references. There is much to the subject that cannot be told here, but to assist the reader we shall touch on some of the central facts and definitions and do so in the notation of the function p. This will facilitate the applications we wish to make, although for the time being nothing dependent on the special nature of p as an optimal value function will be invoked.

Recall that a function $p : \mathbb{R}^m + \mathbb{R}^d \to (\mathbb{R} \cup \{\pm\infty\}$ is everywhere lower semicontinuous if and only if its epigraph

$$E = \{(u, v, \alpha) \in \mathbb{R}^m \times \mathbb{R}^d \times \mathbb{R} \mid \alpha \geq p(u, v)\} \tag{2.1}$$

is a closed set. In this case the matters we must explain are simpler, but we do not want to be burdened later with having to impose conditions on $(\mathrm{P}_{u,v})$ that imply such *global* lower semicontinuity of its optimal value function. For our purposes all that really is needed is for the epigraph E to be closed relative to some neighborhood (in $\mathbb{R}^m \times \mathbb{R}^d \times \mathbb{R}$) of one of its points $(u, v, p(u, v))$ that happens to be under discussion. This condition is stronger than lower semicontinuity of p just at (u, v), yet not as stringent as requiring lower semicontinuity of p on a neighborhood of (u, v) (in $\mathbb{R}^m \times \mathbb{R}^d$). We shall call it *strict lower semicontinuity of p at* (u, v); it holds if and only if for some $\alpha > p(u, v)$, there is a neighborhood of (u, v) on which the function $\min\{p, \alpha\}$ is lower semicontinuous.

The epigraph point of view and the potential discontinuities of p also force us to be more subtle in speaking of convergence of (u', v') to (u, v). We introduce the notation

$$(u', v') \underset{p}{\to} (u, v) \Leftrightarrow \begin{cases} (u', v') \to (u, v), \\ p(u', v') \to p(u, v), \end{cases} \tag{2.2}$$

in situations where it is really just the convergence of the point $(u', v', p(u', v'))$ in E to $(u, v, p(u, v))$ that counts. Obviously, '$\to_p$' is the same as '$\to$' when p is continuous at (u, v), and in particular whenever p happens to be locally Lipschitzian.

We concentrate henceforth in this section and the next on a point (u, v) where p is finite and strictly lower semicontinuous. Criteria for this in the optimal value case will be given in Propositions 8–10 in Section 5.

Using the notation (2.2), we define the *Clarke derivative* of p at (u, v) with respect to a vector (h, k) as

$$p^\circ(u, v; h, k) = \limsup_{\substack{(u', v') \to_p (u, v) \\ t \downarrow 0}} \frac{p(u' + th, v' + tk) - p(u', v')}{t}. \tag{2.3}$$

Clarke actually considered such derivatives only for locally Lipschitzian functions [4] ('$\to$' in place of '$\to_p$'), but he used them indirectly to develop a notion of 'subgradient' for functions that are merely lower semicontinuous and not necessarily finite-valued. We showed in [28] that the generalized subgradients in question could be characterized thoroughly and directly in terms of slightly more complicated limits than the ones in (2.3), namely the so-called *subderivatives*

$$p^{\uparrow}(u, v; h, k) = \lim_{\epsilon \downarrow 0} \limsup_{\substack{(u', v') \to_p (u, v) \\ t \downarrow 0}} \left[\inf_{\substack{|h'-h| \le \epsilon \\ |k'-k| \le \epsilon}} \frac{p(u' + th', v' + tk') - p(u', v')}{t} \right]. \tag{2.4}$$

The remarkable fact is that $p^{\uparrow}(u, v; h, k)$, as a function of (h, k), is always *convex*, positively homogeneous, lower semicontinuous, not identically $+\infty$ nor identically $-\infty$. Clarke's set of subgradients is given directly as

$$\partial p(u, v) = \{(y, z) \in \mathbb{R}^m \times \mathbb{R}^d \mid y \cdot h + z \cdot k \le p^{\uparrow}(u, v; h, k) \text{ for all } (h, k)\}. \tag{2.5}$$

From this expression and the properties of the subderivative function it follows by general theorems of convex analysis [26, Section 13] that $\partial p(u, v)$ is a closed convex set and

$$\begin{aligned} p^{\uparrow}(u, v; h, k) &= \sup\{y \cdot h + z \cdot k \mid (y, z) \in \partial p(u, v)\} \\ &> -\infty \text{ for all } (h, k), \quad \text{if } \partial p(u, v) \neq \emptyset, \\ p^{\uparrow}(u, v; h, k) &= \pm\infty \text{ for all } (h, k) \quad \text{if } \partial p(u, v) = \emptyset. \end{aligned} \tag{2.6}$$

This formula extends one given by Clarke [4] for his derivatives (2.3) in the locally Lipschitzian case. In that case, $\partial p(u, v)$ is nonempty and compact; conversely, as we proved in [25], if $\partial p(u, v)$ is nonempty and compact, then p is locally Lipschitzian around (u, v) and the derivatives (2.3) coincide. A more general relationship between the two derivatives, established in [28, p. 267], is the following: the two effective domains

$$\operatorname{dom} p^{\circ}(u, v; h, k) = \{(h, k) \mid p^{\circ}(u, v; h, k) < \infty\}, \tag{2.7}$$

$$\operatorname{dom} p^{\uparrow}(u, v; h, k) = \{(h, k) \mid p^{\uparrow}(u, v; h, k) < \infty\}, \tag{2.8}$$

are convex cones containing $(0, 0)$ which have the same interior, and for (h, k) in this interior one has

$$\begin{aligned} \infty > \lim_{\epsilon \downarrow 0} &\left[\limsup_{\substack{(u', v') \to_p (u, v) \\ t \downarrow 0}} \left[\sup_{\substack{|h'-h| \le \epsilon \\ |k'-k| \le \epsilon}} \frac{p(u' + th', v' + tk') - p(u', v')}{t} \right]\right] \\ &= p^{\uparrow}(u, v; h, k) = p^{\circ}(u, v; h, k). \end{aligned} \tag{2.9}$$

With respect to such vectors (h, k), p is said to be *directionally Lipschitzian*. (This concept generalizes Lipschitz continuity in a neighborhood of (u, v), which is the case of $(h, k) = (0, 0)$; then the cones in (2.7) and (2.8) are the whole space, and (2.9) actually holds for all (h, k), with '$\to_p$' identical to '$\to$'.) For the many consequences and uses of the directionally Lipschitzian property, see [28, 27].

Formulas (2.5) and (2.6) underline the complete duality between subderivatives and subgradients. If p is convex, $\partial p(u, v)$ is identical to the subgradient set of convex analysis, while if p is smooth it reduces to the singleton

$\{\nabla p(u, v)\}$ [4, 28]. Indeed, $\partial p(u, v)$ consists of a single vector (y, z) if and only if $p^{\uparrow}(u, v; h, k)$, or equivalently $p^{\circ}(u, v; h, k)$, is linear in (h, k), and in this event p is *strictly* differentiable at (u, v) with $\nabla p(u, v) = (y, z)$:

$$\lim_{\substack{(h', k') \to (h, k) \\ (u', v') \to (u, v) \\ t \downarrow 0}} \frac{p(u' + th', v' + tk') - p(u', v')}{t} = y \cdot h + z \cdot k \tag{2.10}$$

[4, 28]. The implication of this result for our later efforts, incidentally, is that differentiability of p at (u, v) can be deduced from conditions which imply $\partial p(u, v)$ has exactly one element.

In general, bounds on various derivatives of p can be obtained from estimates for $\partial p(u, v)$, and this is the pattern we shall follow. Besides $p^{\circ}(u, v; h, k)$ and $p^{\uparrow}(u, v; h, k)$ we shall consider upper and lower one-sided *Hadamard* derivatives:

$$p^{+}(u, v; h, k) = \limsup_{\substack{(h', k') \to (h, k) \\ t \downarrow 0}} \frac{p(u + th', v + tk') - p(u, v)}{t}, \tag{2.11}$$

$$p_{+}(u, v; h, k) = \liminf_{\substack{(h', k') \to (h, k) \\ t \downarrow 0}} \frac{p(u + th', v + tk') - p(u, v)}{t}. \tag{2.12}$$

Obviously one always has

$$p_{+}(u, v; h, k) \leq p^{\uparrow}(u, v; h, k), \tag{2.13}$$

and if p is directionally Lipschitzian at (u, v) with respect to (h, k), so that (2.9) holds, then also

$$p^{+}(u, v; h, k) \leq p^{\circ}(u, v; h, k). \tag{2.14}$$

The case where equality holds in (2.13) plays an important role in the literature; then we say p is *subdifferentially regular* at (u, v) (cf. [4, 28, 25]).

Note that when $p^{+}(u, v; h, k) = p_{+}(u, v; h, k)$ one has a property stronger than just the existence of $p'(u, v; h, k)$ as defined in (1.1). This is what we will be able to establish in Section 7 in cases where other authors have considered only $p'(u, v; h, k)$, as well as made other restrictions.

In some situations it is crucial to be able to know at least that $\partial p(u, v)$ is nonempty. As recorded already in (2.6), a necessary and sufficient condition for this is the existence of (h, k) such that $p^{\uparrow}(u, v; h, k)$ is finite. We now elaborate the meaning of this.

Proposition 1. *Under the assumption that p is finite and lower semicontinuous at (u, v), one has $\partial p(u, v) \neq \emptyset$ if and only if there exist sequences $t_j \downarrow 0$ and $(u^j, v^j) \to_p (u, v)$ such that for no convergent sequence $(h^j, k^j) \to (h, k)$ does one have*

$$[p(u^j + t_j h^j, v^j + t_j k^j) - p(u^j, v^j)]/t_j \to -\infty.$$

Thus in particular, $\partial p(u, v) \neq \emptyset$ if p is calm at (u, v) in the sense that

$$\liminf_{(u', v') \to (u, v)} \frac{p(u', v') - p(u, v)}{|(u', v') - (u, v)|} > -\infty. \tag{2.15}$$

Proof. Because the function $p^{\uparrow}(u, v; \cdot, \cdot)$ is lower semicontinuous, positively homogeneous and convex, but not identically $+\infty$, it is finite at some point if and only if it is not $-\infty$ at the origin. Therefore, $\partial p(u, v) \neq \emptyset$ if and only if $p^{\uparrow}(u, v; 0, 0) > -\infty$. The first assertion in the proposition merely puts the latter condition in more specific terms using the definition (2.4). The calmness property implies the condition is satisfied with $(u^j, v^j) = (u, v)$ for all j and any sequence $t_j \downarrow 0$.

Calmness of p at (u, v) may be thought of as 'pointwise lower Lipschitz continuity'. It is a concept that has been used to advantage by Clarke in [3].

3. Singular subgradients

In addition to the subgradients discussed so far, we shall find it helpful to speak of as *singular subgradients* of p at (u, v) the elements of the closed convex cone

$$\begin{aligned}\partial^0 p(u, v) &:= \text{polar of the convex cone (2.8)}\\ &= \{(y, z) \mid y \cdot h + z \cdot k \leq 0 \text{ for all } (h, k) \text{ satisfying}\\ &\qquad p^{\uparrow}(u, v; h, k) < \infty\}.\end{aligned} \tag{3.1}$$

It follows from the duality in (2.5) and (2.6) that this set is just the recession cone of $\partial p(u, v)$ [26, Section 13]:

$$\partial^0 p(u, v) = 0^+ \partial p(u, v) = \limsup_{\lambda \downarrow 0} \lambda \partial p(u, v) \quad \text{when } \partial p(u, v) \neq \emptyset. \tag{3.2}$$

Nonzero singular subgradients thus describe directions which can be identified with 'elements of $\partial p(u, v)$ lying at ∞', except that there can be situations where $\partial p(u, v) = \emptyset$ and yet $\partial^0 p(u, v) \neq \emptyset$.

A more geometric description of singular subgradients is possible in terms of Clarke's concept of normal cones to closed sets in Euclidean spaces. Recall from the beginning of Section 2 that when p is finite and *strictly* lower semicontinuous at (u, v), its epigraph E is closed relative to a neighborhood of the point $(u, v, p(u, v))$. The *normal cone* to E at this point is the nonempty closed convex cone

$$N_E(u, v, p(u, v)) = \partial \delta_E(u, v, p(u, v)), \tag{3.3}$$

where δ_E is the indicator function for E. One has

$$\partial p(u, v) = \{(y, z) \mid (y, z, -1) \in N_E(u, v, p(u, v))\}, \tag{3.4}$$

$$\partial^0 p(u, v) = \{(y, z) \mid (y, z, 0) \in N_E(u, v, p(u, v))\}. \tag{3.5}$$

In Clarke's original approach [4], normal cones are first given various direct characterizations, and then (3.4) is taken as the *definition* of the set of subgradients of p at (u, v). As seen from (3.5), the notion of 'singular subgradients' fits neatly into the same picture. The validity of (3.5) stems from the fact that the cone $N_E(u, v, p(u, v))$ and the epigraph of the subderivative function $p^{\uparrow}(u, v; \cdot, \cdot)$ are polar to each other; see [28].

(Incidentally, the assertion made in Section that $p^{\uparrow}(u, v; h, k)$ cannot be identically $-\infty$ as a function of (h, k) follows by duality from the fact that $N_E(u, v, p(u, v))$ cannot consist of just the zero vector. The latter is true because $(u, v, p(u, v))$ is a boundary point of E, and nonzero normal vector always exist at boundary points [25, p. 149].)

Several properties of p can be characterized in terms of singular subgradients, and this will be useful later in seeing the consequences of the estimates that will be given for $\partial^0 p(u, v)$. The following terminology will expedite matters: a cone M (not necessarily convex) will be called *pointed* if the origin cannot be expressed as a sum of nonzero vectors in M. When M is convex (as in the case $M = \partial^0 p(u, v)$), this reduces to the property that M does not contain the negative of any of its nonzero vectors.

Proposition 2. *Under the assumption that p is finite and strictly lower semicontinuous at (u, v), one has p directionally Lipschitzian with respect to (h, k) if and only if for all (h', k') in some neighborhood of (h, k), one has $y \cdot h' + z \cdot k' \leq 0$ for all $(y, z) \in \partial^0 p(u, v)$. Such an (h, k) exists if and only if $\partial^0 p(u, v)$ is pointed.*

Proof. The condition says that (h, k) is an interior point of the polar of $\partial^0 p(u, v)$. Since $\partial^0 p(u, v)$ is the polar of the convex cone (2.8), this means (h, k) belongs to the interior of (2.8). Such vectors (h, k) are the ones with respect to which p is directionally Lipschitzian, as already explained in Section 2. The polar of a closed convex cone has nonempty interior if and only if the cone is pointed.

Proposition 3. *For p to be locally Lipschitzian around (u, v), it is necessary and sufficient that p be finite and strictly lower semicontinuous at (u, v) and have $\partial^0 p(u, v) = \{(0, 0)\}$.*

Proof. This is the case of Proposition 2 where $(h, k) = (0, 0)$. Recall that p is locally Lipschitzian around (u, v) if and only if p is directionally Lipschitzian at (u, v) with respect to $(h, k) = (0, 0)$ [28].

Proposition 4. *Under the assumption that p is finite and strictly lower semicontinuous at (u, v), if $\partial^0 p(u, v)$ is pointed and does not contain any vector of the form $(y, 0)$ with $y \neq 0$, then*

$$\partial_v p(u, v) \subset \{z \mid \exists y \text{ with } (y, z) \in \partial p(u, v)\}, \tag{3.6}$$

$$\partial_v^0 p(u, v) \subset \{z \mid \exists y \text{ with } (y, z) \in \partial^0 p(u, v)\}. \tag{3.7}$$

In particular, (3.6) is valid if p is locally Lipschitzian around (u, v). (Moreover, equality holds in (3.6) and (3.7) if p is subdifferentially regular at (u, v).)

Proof. From a result in [27, p. 350], (3.6) holds (and with equality in the case of subdifferential regularity) when the interior of the convex cone (2.8) contains a vector of form $(0, k)$. The separation theorem for convex sets enables us to translate this condition into the nonexistence of a vector $(y, 0) \neq (0, 0)$ belonging to the polar of the cone (2.8), namely $\partial^0 p(u, v)$ (cf. Proposition 2). The locally Lipschitzian case of (3.6) follows via Proposition 3. There are several ways to get the parallel inclusion (3.7), but the simplest perhaps is to observe that the cited result in [27, p. 350] is a corollary of a theorem that actually yields more when specialized to the case in question: for the function $q = p(u, \cdot)$, one has

$$q^{\uparrow}(v; k) \leq p^{\uparrow}(u, v; 0, k) \quad \text{for all } k, \tag{3.8}$$

(and equality holds in (3.8) when p is subdifferentially regular at (u, v)). Therefore

$$\{(h, k) \mid h = 0,\ q^{\uparrow}(v; k) < \infty\} \supset \{(h, k) \mid p^{\uparrow}(u, v; h, k) < \infty\} \cap [\{0\} \times \mathbb{R}^d]. \tag{3.9}$$

Since the interior of the cone (2.8) contains under our hypothesis a vector of form $(0, k)$, we can take polars on both sides of (3.9) and get

$$[\mathbb{R}^m \times \partial^0 q(v)] \subset \partial^0 p(u, v) + [\mathbb{R}^m \times \{0\}],$$

which is equivalent to (3.7).

Remark. The Lipschitzian case of Proposition 4 was first developed by Clarke, who pointed out that without some condition like subdifferential regularity, there may be no inclusion either way between $\partial p(u, v)$ and $\partial_u p(u, v) \times \partial_v p(u, v)$. See [17, p. 308] for an example of this phenomenon.

4. Lagrange multiplier rule

Our main result about subgradients of p when p is the optimal value function in Section 1 will involve Lagrange multiplier vectors that appear in extended first-order necessary conditions for optimality in $(\mathrm{P}_{u,v})$. This section is devoted to formulating the conditions in question and comparing them to previous contributions. The necessity of the conditions, however, will actually be established in Section 6 as a *consequence* of our estimation theorem, rather than as a preliminary to it.

Henceforth our notation and assumptions are those in Section 1, but we apply freely the general subdifferential theory exposed in Sections 2–3.

Each function f_i, being locally Lipschitzian on an open set containing D, has a nonempty, compact, convex subgradient set $\partial f_i(v, x)$ at every $(v, x) \in D$. We emphasize that this is the subgradient set of convex analysis if f_i is a convex function, and it is just $\{\nabla f_i(v, x)\}$ if f_i is of class $\mathscr{C}^1$. When f_i is mixed smooth-convex as in case (c) of Section 1, it turns out that

$$\partial f_i(v, x) = (\nabla_v f_i(v, x), \partial_x f_i(v, x)) \tag{4.1}$$

(because $f_i^\circ = f_i'$ in this case, as can be verified by direct calculation).

Since D is closed, the indicator function

$$\delta_D(v, x) = \begin{cases} 0, & \text{if } (v, x) \in D, \\ \infty, & \text{if } (v, x) \notin D, \end{cases}$$

is lower semicontinuous everywhere. Its subgradient sets are the normal cones to D:

$$N_D(v, x) = \partial\delta_D(v, x) \quad \text{for each } (v, x) \in D. \tag{4.2}$$

(When D is convex, the vectors $(z, w) \in N_D(v, x)$ are the ones such that $(z, w) \cdot (v', x') \leq (z, w) \cdot (v, x)$ for all $(v', x') \in D$.)

The optimality conditions we shall be concerned with are related to such subgradients, as will be explained below, but they generally take the form of associating with some x which in particular satisfies all the constraints of $(P_{u,v})$ a pair of vectors $y = (y_1, \ldots, y_m)$ and $z = (z_1, \ldots, z_d)$ such that

$$y_i \geq 0 \quad \text{and} \quad y_i[f_i(v, x) + u_i] = 0 \quad \text{for } i = 1, \ldots, s, \tag{4.3}$$

$$(z, 0) \in \partial\left[f_0 + \sum_{i=1}^m y_i f_i + \delta_D\right](v, x). \tag{4.4}$$

For some purposes, we shall need to look at the corresponding degenerate conditions where f_0 does not appear, i.e., where (4.4) is replaced by

$$(z, 0) \in \partial\left[\sum_{i=j}^m y_i f_i + \delta_D\right](v, x). \tag{4.5}$$

We let

$$K(u, v, x) = \text{set of all } (y, z) \text{ satisfying (4.3) and (4.4)},$$
$$K_0(u, v, x) = \text{set of all } (y, z) \text{ satisfying (4.3) and (4.5)}. \tag{4.6}$$

The targeted Lagrange multiplier rule is an assertion that $K(u, v, x) \neq \emptyset$ in certain situations. For immediate comparison with classical conditions, observe that in the smooth case (a) of Section 1, (4.4) reduces to

$$0 = \nabla_x f_0(v, x) + \sum_{i=1}^m y_i \nabla_x f_i(v, x) \quad \text{and} \quad z = \nabla_v f_0(v, x) + \sum_{i=1}^m y_i \nabla_v f_i(v, x), \tag{4.7}$$

while (4.5) reduces to

$$0 = \sum_{i=1}^m y_i \nabla_x f_i(v, x) \quad \text{and} \quad z = \sum_{i=1}^m y_i \nabla_v f_i(v, x). \tag{4.8}$$

Since $(0, 0) \in K_0(u, v, x)$ trivially always, interest in the set $K_0(u, v, x)$ will center on whether it also contains some $(y, z) \neq (0, 0)$. The condition $K_0(u, v, x) = \{(0, 0)\}$ will serve as one kind of constraint qualification. A more subtle constraint qualification that will also play a role can be stated in terms of 'calmness'.

Localizing a definition of Clarke's [3], we say problem $(\mathrm{P}_{u,v})$ is *calm at x*, one of its locally optimal solutions, if there do *not* exist sequences $x^j \to x$ and $(u^j, v^j) \to (u, v)$ with x^j feasible for (P_{u^j,v^j}) such that

$$\frac{f_0(v^j, x^j) - f_0(v, x)}{|(u^j, v^j) - (u, v)|} \to -\infty.$$

Clearly this does hold when p is calm at (u, v) in the sense of (2.15) and x is any (globally) optimal solution to $(\mathrm{P}_{u,v})$. Calmness of p at (u, v), without reference additionally to any point x, is a condition that Clarke calls simply the calmness of problem $(\mathrm{P}_{u,v})$. The exact relationship between this 'global' calmness and our 'local' calmness will be shown later in Proposition 12 (see Section 6).

Theorem 1. *Let x be any locally optimal solution to* $(\mathrm{P}_{u,v})$.

(i) *If* $(\mathrm{P}_{u,v})$ *is calm at x, then* $K(u, v, x) \neq \emptyset$.

(ii) *If* $K_0(u, v, x) = \{(0, 0)\}$, *then* $(\mathrm{P}_{u,v})$ *is indeed calm at x, and moreover* $K(u, v, x)$ *is compact.*

As already remarked, this theorem will not be proved until Section 6, where it will appear chiefly as a sort of corollary of Theorem 2 of Section 5. We have stated it at this early stage in order to put the multiplier sets $K(u, v, x)$ and $K_0(u, v, x)$ in the proper perspective. The rest of this section deals with further clarifications of the nature of these sets. We start by citing a fundamental rule of subdifferential calculus.

Proposition 5. *Let* g_1 *and* g_2 *be extended-real-valued functions on a Euclidean space which are both finite at a point w. If either* g_1 *or* g_2 *is locally Lipschitzian around w, then*

$$\partial(g_1 + g_2)(w) \subset \partial g_1(w) + \partial g_2(w).$$

Moreover, equality holds if either g_1 *or* g_2 *is of class* $\mathscr{C}^1$ *in a neighborhood of w, or if both* g_1 *and* g_2 *are subdifferentially regular at w.*

Proof. This is an immediate consequence of a much broader result obtained in [27, p. 345], except for the business about g_1 or g_2 being of class $\mathscr{C}^1$. If g_2, say, is of class $\mathscr{C}^1$ around w, then g_2 and $-g_2$ are both locally Lipschitzian around w and have $\partial g_2(w) = \{\nabla g_2(w)\}$ and $\partial(-g_2)(w) = \{-\nabla g_2(w)\}$. The basic rule gives both

$$\partial(g_1 + g_2)(w) \subset \partial g_1(w) + \nabla g_2(w)$$

and

$$\partial g_1(w) = \partial(g_1 + g_2 - g_2)(w) \subset \partial(g_1 + g_2)(w) - \nabla g_2(w),$$

and this implies $\partial(g_1 + g_2)(w) = \partial g_1(w) + \nabla g_2(w)$ and finishes the proof.

In the situation at hand, we want to apply Proposition 5 to the expressions in (4.4) and (4.5) along with the elementary rule that (inasmuch as f_i is locally Lipschitzian)

$$\partial(y_i f_i)(v,x) = y_i \partial f_i(v,x) \quad \text{for all } y_i \in \mathbb{R}. \tag{4.9}$$

For this purpose we note that the property of *subdifferential regularity* (see Section 2) holds everywhere for f_i when f_i is convex, of class $\mathscr{C}^1$, or a mixture of the two as in case (c) in Section 1. It holds everywhere for both f_i and $-f_i$ (i.e., for $y_i f_i$ regardless of the sign of y_i) if and only if f_i is of class $\mathscr{C}^1$. It holds for δ_D if and only if D is *tangentially regular* in the sense that at all boundary points of D, the Clarke tangent cone and the classical contingent cone coincide, as is true certainly when D is convex or a 'smooth manifold'; see [4, 27] for more on such properties.

At all events, the strong form of Proposition 5, where equality holds, is thoroughly applicable (together with (4.9)) in cases (a), (b) and (c) of Section 1 and more generally in the following cases of problem $(\mathrm{P}_{u,v})$:

(d) *the subdifferentially regular case*: D tangentially regular, f_i subdifferentially regular for $i = 0, 1, \ldots, s$ and of class $\mathscr{C}^1$ for $i = s+1, \ldots, m$.

(e) *the extended smooth case*: D an arbitrary closed set, every f_i of class $\mathscr{C}^1$.

Clearly (e) subsumes (a), while from the remarks above, (d) subsumes (a), (b) and (c). This allows us to draw an important conclusion.

Proposition 6. *In condition* (4.4) *of the definition of* $K(u,v,x)$, *one has*

$$\partial\Big[f_0 + \sum_{i=1}^m y_i f_i + \delta_D\Big](v,x) \subset \partial\Big[f_0 + \sum_{i=1}^m y_i f_i\Big](v,x) + N_D(v,x)$$
$$\subset \partial f_0(v,x) + \sum_{i=1}^m y_i \partial f_i(v,x) + N_D(v,x). \tag{4.10}$$

Moreover, equality holds in cases (d) *and* (e) *above and hence in particular in the smooth, convex, and mixed smooth-convex cases* (a), (b) *and* (c) *of* $(\mathrm{P}_{u,v})$. *Similarly for condition* (4.5) *of the definition of* $K_0(u,v,x)$.

The second inclusion in Proposition 6 does not depend on the full force of Proposition 5: it is already apparent from an earlier formula of Clarke [4] where g_1 and g_2 are *both* locally Lipschitzian.

Observe that in the mixed smooth-convex case (c), where (4.1) holds and $D = \mathbb{R}^d \times C$, Proposition 6 allows conditions (4.4) and (4.5) to be written instead as

$$0 \in \partial_x\Big[f_0 + \sum_{i=1}^m y_i f_i\Big](v,x) + N_C(x) \quad \text{and } z = \nabla_v f_0(v,x) + \sum_{i=1}^m y_i \nabla_v f_i(v,x), \tag{4.11}$$

$$0 \in \partial_x\Big[\sum_{i=1}^m y_i f_i\Big](v,x) + N_C(x) \quad \text{and} \quad z = \sum_{i=1}^m y_i \nabla_v f_i(v,x). \tag{4.12}$$

Due to convexity in x, these indicate that when $(y, z) \in K(u, v, x)$, the pair (x, y) is (as expected) a saddlepoint of the ordinary Lagrangian for $(\mathrm{P}_{u,v})$ on $C \times [\mathbb{R}^s_+ \times \mathbb{R}^{m-s}]$, and similarly when $(y, z) \in K_0(u, v, x)$, except that then it is the degenerate Lagrangian not involving f_0.

Only in situations where strict inclusions can be encountered in (4.10), and thus never in cases (a), (b), (c), (d) or (e), is the multiplier condition $K(u, v, x) \neq \emptyset$ in Theorem 1 any sharper than the ones of Clarke [3] or Hiriart-Urruty [18]. Clarke's rule corresponds to substituting the largest of the sets in (4.10) for (4.4), while Hiriart-Urruty uses the middle set.

These earlier rules do not actually take the parameter vector v into account, but they can be adapted to yield conditions in the present format simply by posing $(\mathrm{P}_{u,v})$ equivalently as the problem $(\mathrm{P}'_{u,v})$ in Section 1. Conversely, of course, Theorem 1 can be applied with v held fixed and suppressed from consideration. The corresponding multiplier conditions then say nothing about a vector z, and they have in place of (4.4) and (4.5) the relations

$$0 \in \partial_x \Big[f_0 + \sum_{i=1}^m y_i f_i + \delta_D \Big](v, x), \tag{4.13}$$

$$0 \in \partial_x \Big[\sum_{i=1}^m y_i f_i + \delta_D \Big](v, x), \tag{4.14}$$

which again could be elaborated as in Proposition 6. As far as necessary conditions for optimality are concerned, there is no distinction to be made between the two formulations in cases (a) or (c) (where (4.4), (4.5), become (4.7), (4.8), or (4.11), (4.12)). Nor is there any real distinction in the convex case (b), or for that matter in the subdifferentially regular case (d): then (4.13) holds if and only if there exists z such that (4.4) holds (apply the equality clause in Proposition 4 to the functions in question). Generally speaking, however, neither formulation of the conditions directly subsumes the other.

In the smooth case (a), the constraint qualification $K_0(u, v, x) = \{(0, 0)\}$ asserts:

$$\text{there is no } y \neq 0 \text{ satisfying (4.3) with } \sum_{i=1}^m y_i \nabla_x f_i(v, x) = 0. \tag{4.15}$$

This property is equivalent by duality with the *Mangasarian–Fromovitz constraint qualification* [24]:

> the gradients $\nabla_x f_i(v, x)$, $i = s+1, \ldots, m$, are linearly independent, and there is a vector w such that

$$\nabla_x f_i(v, x) \cdot w \begin{cases} < 0 & \text{for } i = 1, \ldots, s \text{ having } f_i(v, x) = 0, \\ = 0 & \text{for } i = s+1, \ldots, m. \end{cases} \tag{4.16}$$

Related conditions for nonsmooth cases of $(\mathrm{P}_{u,v})$ have been introduced by Auslender [2] and Hiriart-Urruty [18, 19]. Our condition $K_0(u, v, x) = \{(0, 0)\}$ is sharper than these in the sense of the inclusions in Proposition 6, but Hiriart-Urruty gives a treatment of equality constraints that is in other respects more refined. On the other hand, Hiriart-Urruty does not prove a multiplier rule based on 'calmness'.

The result in Theorem 1 that the constraint qualification $K_0(u, v, x) = \{(0, 0)\}$ implies calmness at x is new, although in the extended smooth case (e) with D convex it follows in terms of the Mangasarian–Fromovitz qualification via the stability theory of Robinson [31, 32]; c.f. remark of Clarke [3, p. 173].

There is a relationship between $K_0(u, v, x)$ and $K(u, v, x)$ that sheds some further light. Recall that the *recession cone* of the (not necessarily convex) set $K(u, v, x)$ is by definition

$$\begin{aligned} 0^+K(u, v, x) &= \limsup_{\lambda \downarrow 0} \lambda K(u, v, x) \\ &= \{\lim \lambda_j(y^j, z^j) \mid \lambda_j \downarrow 0, (y^j, z^j) \in K(u, v, x)\}. \end{aligned} \tag{4.17}$$

A nonempty set in a Euclidean space is bounded if and only if its recession cone consists of just the zero vector.

Proposition 7. *For any feasible solution x to* $(P_{u,v})$, *the sets $K(u, v, x)$ and $K_0(u, v, x)$ are closed and*

$$0^+K(u, v, x) \subset K_0(u, v, x). \tag{4.18}$$

In cases (d) *and* (e) *above (and hence in particular in the smooth, convex, and mixed smooth-convex cases* (a), (b) *and* (c), *$K(u, v, x)$ and $K_0(u, v, x)$ are also convex. If in addition to this $K(u, v, x)$ is nonempty, then equality holds in* (4.18) *and*

$$K(u, v, x) + K_0(u, v, x) = K(u, v, x). \tag{4.19}$$

Proof. To demonstrate that $K(u, v, x)$ is closed, suppose $(y^j, z^j) \in K(u, v, x)$ and $(y^j, z^j) \to (y, z)$. For all j, one has

$$y_i^j \begin{cases} \geq 0 & \text{for } i = 1, \ldots, s \text{ having } f_i(v, x) = 0, \\ = 0 & \text{for } i = 1, \ldots, s \text{ having } f_i(v, x) < 0, \end{cases}$$

so the same holds for the multipliers $y_i = \lim_j y_i^j$. Also

$$\begin{aligned} (z^j, 0) &\in \partial\Big[f_0 + \sum_{i=1}^m y_i^j f_i + \delta_D\Big](v, x) \\ &= \partial\Big[f_0 + \sum_{i=1}^m y_i f_i + \delta_D + \sum_{i=1}^m (y_i^j - y_i) f_i\Big](v, x). \end{aligned} \tag{4.20}$$

Applying Proposition 5, we get

$$(z^j, 0) \in \partial\Big[f_0 + \sum_{i=1}^m y_i f_i + \delta_D\Big](v, x) + \sum_{i=1}^m (y_i^j - y_i)\partial f_i(v, x).$$

Since $z^j \to z$, $y_i^j - y_i \to 0$, and $\partial f_i(v, x)$ is compact (due to f_i being locally Lipschitzian), it follows that (4.4) holds. Thus $(y, z) \in K(u, v, x)$, and $K(u, v, x)$ is closed. The proof that $K_0(u, v, x)$ is closed is identical.

The proof of the inclusion (4.18) is along similar lines. Suppose $\lambda_j(y^j, z^j) \to (y, z)$, where $\lambda_j \downarrow 0$ and $(y^j, z^j) \in K(u, v, x)$. The critical observation this time is that (4.20) can be written instead in the form

$$\begin{aligned}(\lambda_j z^j, 0) &\in \partial\left(\lambda_j\left[f_0 + \sum_{i=1}^m y_i^j f_i + \delta_D\right]\right)(v, x).\\ &= \partial\left[\sum_{i=1}^m y_i f_i + \delta_D + \lambda_j f_0 + \sum_{i=1}^m (\lambda_j y_i^j - y_i) f_i\right](v, x),\end{aligned}$$

so that by Proposition 5

$$(\lambda_j z^j, 0) \in \partial\left[\sum_{i=1}^m y_i f_i + \delta_D\right](v, x) + \lambda_j \partial f_0(v, x) + \sum_{i=1}^m (\lambda_j y_i^j - y_i)\partial f_i(v, x).$$

Since $\lambda_j z^j \to z$, $\lambda_j y_i^j - y_i \to 0$, $\lambda_j \downarrow 0$ and $\partial f_i(v, x)$ is compact, we get (4.5) in the limit and hence $(y, z) \in K_0(u, v, x)$.

In cases (d) and (e), we know that equality holds in (4.10) and that $\partial f_i(v, x)$ is just a singleton for $i = s+1, \dots, m$. Using this in (4.4), it is easy to verify the convexity of $K(u, v, x)$ and similarly that of $K_0(u, v, x)$, as well as the relation

$$K(u, v, x) + K_0(u, v, x) \subset K(u, v, x). \tag{4.21}$$

(Recall that $(\alpha + \beta)C = \alpha C + \beta C$ when C is a nonempty convex set and $\alpha \geq 0$, $\beta \geq 0$; cf. [26, Section 3].) When $K(u, v, x)$ is convex and nonempty, (4.21) implies $K_0(u, v, x) \subset 0^+K(u, v, x)$ [26, Section 8], whence equality in (4.18) and (4.19).

Remark. In the convex case (b), the condition $K(u, v, x) \neq \emptyset$ is, of course, sufficient for a feasible solution x in $(P_{u,v})$ to be optimal. Indeed, the multiplier relations reduce then to the description of a saddlepoint of the Lagrangian for the equivalent problem $(P'_{u,v})$ in Section 1. Because of this, the set $K(u, v, x)$ is actually the same regardless of which optimal solution x is being considered, and similarly for $K_0(u, v, x)$. Another special result in the convex case, besides the ones noted in Propositions 6 and 7, is the converse of Theorem 1(i): if $K(u, v, x) \neq \emptyset$ for an optimal solution x, then $(P_{u,v})$ is calm at x; in fact p is calm at (u, v). For this, see [26, Sections 28–29].

5. Tameness and subgradient estimates

Our main result will be stated in this section after some preliminaries having to do with lower semicontinuity of the optimal value function p and the existence of solutions to $(P_{u,v})$.

We shall say for a given (u, v) that problem $(P_{u,v})$ is *tame* if there is a set

$A \subset \mathbb{R}^n$ with the property:

$$\begin{aligned}&A \text{ is compact, and for every}\\ &\epsilon > 0 \text{ there exist } \delta > 0 \text{ and } \alpha > p(u, v) \text{ such that}\\ &\text{when } |(u', v') - (u, v)| < \delta \text{ and } p(u', v') < \alpha, \text{ the addition of the}\\ &\text{constraint } \operatorname{dist}(x, A) \le \epsilon \text{ to } (\mathrm{P}_{u',v'}) \text{ would not}\\ &\text{affect the infimum } p(u', v') \text{ in } (\mathrm{P}_{u',v'}).\end{aligned} \tag{5.1}$$

The virtues of this condition are proclaimed in the next three propositions. (Recall the meaning of 'strict' lower semicontinuity, as defined at the beginning of Section 2.) Note that 'tameness' is *not* a constraint qualification like 'calmness', but merely a weak sort of local boundedness assumption on the way the feasible solution set varies with the parameters.

Proposition 8. *Suppose (u, v) is such that $(\mathrm{P}_{u,v})$ is tame in the above sense. Then p is finite at (u, v) and strictly lower semicontinuous at (u, v). Furthermore $(\mathrm{P}_{u,v})$ has at least one optimal solution; indeed, if A is any set with respect to which the definition of tameness is fulfilled, then $(\mathrm{P}_{u,v})$ must have an optimal solution lying in A.*

Proof. Taking $(u', v') = (u, v)$ in (5.1), we see in particular that $p(u, v) < \infty$. Define

$$\beta = \liminf_{(u',v') \to (u,v)} p(u', v') \le p(u, v). \tag{5.2}$$

Select any sequence $\epsilon_j \downarrow 0$ and corresponding sequences of numbers δ_j and α_j. In view of (5.2), a sequence $(u^j, v^j) \to (u, v)$ with $p(u^j, v^j) \to \beta$ exists having actually $|(u^j, v^j) - (u, v)| < \delta_j$ and $p(u^j, v^j) < \alpha_j$. Then for each j, (P_{u^j,v^j}) has feasible solutions which also belong to the set

$$\{x \mid \operatorname{dist}(x, A) \le \epsilon_j\} \tag{5.3}$$

(which is compact because A is compact), and the infimum is unaffected if restricted to such feasible solutions. Since the objective function in (P_{u^j,v^j}) is continuous and the set of all feasible solutions is closed, it follows that (P_{u^j,v^j}) has an optimal solution x^j in the set (5.3). Then $f_0(v^j, x^j) = p(u^j, v^j) \to \beta$ and $\operatorname{dist}(x^j, A) \to 0$. Passing to subsequences if necessary, we can suppose (again because A is compact) that $x^j \to x$, where x is some element of A. The continuity of the functions f_i and the closedness of the set D imply that, since $(u^j, v^j) \to (u, v)$, x is a feasible solution to $(\mathrm{P}_{u,v})$ with $f_0(v, x) = \beta$. We may conclude then from (5.2) that x is optimal and $\beta = p(u, v)$.

Proposition 9. *A sufficient condition for (u, v) to be such that $(\mathrm{P}_{u,v})$ is tame is the existence of $\delta_0 > 0$ and $\alpha_0 > p(u, v)$ with the property: the set of all x' satisfying*

$$\begin{aligned}&\exists (u', v') \text{ with } |(u', v') - (u, v)| \le \delta_0 \text{ such that}\\ &x' \text{ is feasible for } (\mathrm{P}_{u',v'}) \text{ and } f_0(v', x') \le \alpha_0\end{aligned} \tag{5.4}$$

is a bounded set. Indeed, the definition of tameness is then fulfilled with this set as A.

In particular, $(P_{u,v})$ *is tame if it has feasible solutions and D is of the form* $\mathbb{R}^d \times C$ *with C compact. (Then C can serve as the A in the definition of tameness.)*

Proof. Denote the set of x' satisfying (5.4) by A and observe that it is compact. To verify the rest of (5.1) consider any $\epsilon > 0$ and let $\delta = \delta_0$, $\alpha = \alpha_0$. Then for (u', v') with $|(u', v')| - (u, v)| < \delta$ and $p(u', v') < \alpha$, all the feasible solutions x' to $(P_{u',v'})$ with $f_0(v', x') \leq \alpha$ belong to A and therefore satisfy $\operatorname{dist}(x', A) = 0$. Hence the constraint $\operatorname{dist}(x, A) \leq \epsilon$ can be added to $(P_{u',v'})$ with impunity.

Proposition 10. *A necessary and sufficient condition for* (u, v) *to be such that* $(P_{u,v})$ *is tame is the existence of* $\delta_0 > 0$ *and* $\alpha_0 > p(u, v)$ *with the property: there is a bounded mapping* ξ *from the set*

$$\{(u', v') \mid p(u', v') < \alpha_0 \quad \text{and} \quad |(u', v') - (u, v)| < \delta_0\} \tag{5.5}$$

to $\mathbb{R}^n$ *such that for every* (u', v') *in this set,* $\xi(u', v')$ *is an optimal solution to* $(P_{u',v'})$.

Indeed, the definition of tameness is satisfied with respect to a particular compact set A if and only if for some such mapping ξ, *A includes all the cluster points of* $\xi(u', v')$ *as* $(u', v') \to_p (u, v)$ *in the sense of* (2.2). *(These cluster points themselves form a compact set of optimal solutions to* $(P_{u,v})$.)

Proof. If there is such a mapping ξ, and C denotes its set of cluster points of $\xi(u', v')$ as $(u', v') \to_p (u, v)$, then C is a compact set of points x which (by the closedness of D and continuity of f_i) are feasible solutions to $(P_{u,v})$ having $f_0(v, x) = p(u, v)$. Thus C consists of optimal solutions to $(P_{u,v})$, and for any $\epsilon > 0$ there exist $\delta > 0$ and $\alpha > p(u, v)$ such that whenever (u', v') satisfies $|(u', v') - (u, v)| < \delta$ and $p(u', v') < \alpha$, one has $\operatorname{dist}(\xi(u', v'), C) \leq \epsilon$. Since $\xi(u', v')$ is optimal for $(P_{u',v'})$, it follows that for such (u', v') the constraint $\operatorname{dist}(x', C) \leq \epsilon$ can be added to $(P_{u',v'})$ without affecting the infimum in the problem. The same then holds for any compact $A \supset C$; such an A therefore satisfies (5.1).

Conversely, suppose A is a set with property (5.1). Choose any sequence $\epsilon_j \downarrow 0$ (starting with $j = 0$) and corresponding values δ_j and α_j as in (5.1); the latter values can systematically be lowered, if necessary, so that also $\delta_j \downarrow 0$ and $\alpha_j \downarrow p(u, v)$. Let

$$A_j = \{x' \mid \operatorname{dist}(x', A) \leq \epsilon_j\},$$
$$B_j = \{(u', v') \mid p(u', v') < \alpha_j \quad \text{and} \quad |(u', v') - (u, v)| < \delta_j\}.$$

Then A_j is a compact set such that for every $(u', v') \in B_j$ (and in particular for $(u', v') = (u, v)$), problem $(P_{u',v'})$ has feasible solutions in A_j, and over these the infimum of $f_0(v', \cdot)$ is still $p(u', v')$. Since this restricted infimum concerns a continuous function over a certain set that is nonempty and compact (because A_j is compact, D is closed, and every f_i is continuous), it is attained at some point.

Thus when $(u', v') \in B_j$, there is an optimal solution to $(P_{u',v'})$ in A_j (and for $(u', v') = (u, v)$ there is an optimal solution to $(P_{u,v})$ in $\bigcap_j A_j = A$). For each j and each $(u', v') \in B_j$ with $(u', v') \notin B_{j+1}$, select some optimal solution on $(P_{u',v'})$ in A_j and denote it by $\xi(u', v')$; let $\xi(u, v)$ itself denote some optimal solution to $(P_{u,v})$ in A. Then ξ is a well-defined mapping on the set (5.5) (identical to B_0 in the present notation), and $\xi(u', v') \in A_j$ when $(u', v') \in B_j$. This mapping meets all prescriptions: inasmuch as $\epsilon_j \downarrow 0$, $\delta_j \downarrow 0$, and $\alpha_j \downarrow p(u, v)$, all cluster points of $\xi(u', v')$ as $(u', v') \to_p (u, v)$ are contained in $\bigcap_j A_j = A$.

Remark. The tameness condition we have been exploring was inspired in part by a condition introduced by Hiriart-Urruty [17] in a related context. This is clarified by the equivalence in Proposition 10. Hiriart-Urruty's condition is essentially the one in Proposition 10, but stronger in having ordinary topology appear in place of the '$\to_p$' topology.

Other authors who have dealt with this subject have relied on still more stringent assumptions. For instance, to follow the pattern of the papers of Gauvin and Tolle [13], Gauvin [11], Gauvin and Dubeau [12], the multifunction that associates to each (u', v') the set of all feasible solutions to $(P_{u',v'})$ would be assumed to be bounded on an ordinary neighborhood of (u, v). See also earlier work of Evans and Gould [9], Greenberg and Pierskalla [16], on upper and lower semicontinuity properties of optimal value functions.

In our main theorem, which we are now ready to present, 'co' denotes convex hull and 'cl' closure. Again we use the concept of 'pointedness', as defined in Section 2 for cones that are not necessarily convex.

Theorem 2. *Suppose (u, v) is such that $(P_{u,v})$ is tame, and let X be any set of optimal solutions to $(P_{u,v})$ that at least includes whatever optimal solutions to $(P_{u,v})$ happen to lie in A, the set invoked in the definition* (5.1) *of tameness. (In particular, X could be taken to be the set of all optimal solutions to $(P_{u,v})$.) Then*

$$\partial p(u, v) = \operatorname{cl}\operatorname{co}\Big\{\Big(\Big[\bigcup_{x\in X} K(u, v, x)\Big] \cap \partial p(u, v)\Big) + \Big[\bigcup_{x\in X} K_0(u, v, x)\Big] \cap \partial^0 p(u, v)\Big\}, \tag{5.6}$$

$$\partial^0 p(u, v) \supset \operatorname{cl}\operatorname{co}\Big\{\Big[\bigcup_{x\in X} K_0(u, v, x)\Big] \cap \partial^0 p(u, v)\Big\}. \tag{5.7}$$

Equality holds in (5.7) *if* $\bigcup_{x\in X} K(u, v, x) \cap \partial p(u, v) = \emptyset$, *or if the cone*

$$\Big[\bigcup_{x\in X} K(u, v, x)\Big] \cap \partial^0 p(u, v)$$

is pointed; in the latter case $\partial^0 p(u, v)$ too is pointed, and the closure operation is superfluous in both (5.6) *and* (5.7).

Although the proof of Theorem 2 will not be laid out until Section 6, we shall proceed immediately with some corollaries. Consequences about directional derivatives will be saved for Section 7. The reader should note, incidentally, that Theorem 2 and everything that will be based on it remain valid if $K(u, v, x)$ and $K_0(u, v, x)$ are replaced by other sets that at least are sure to include them. In particular, the multiplier conditions (4.4) and (4.5) could be supplanted by the slightly weaker ones of Clarke [3] or Hiriart-Urruty [18] corresponding to the inclusions in Proposition 6.

Corollary 1. *Assuming tameness as in Theorem 2, one has*

$$\partial p(u, v) \subset \operatorname{cl}\operatorname{co}\Big\{\bigcup_{x\in X} K(u, v, x) + \bigcup_{x\in X} K_0(u, v, x)\Big\}. \tag{5.8}$$

If in addition the cone $\bigcup_{x\in X} K_0(u,v,x)$ is pointed, then $\partial^0 p(u,v)$ is pointed and

$$\partial^0 p(u, v) \subset \operatorname{cl}\operatorname{co}\Big\{\bigcup_{x\in X} K_0(u, v, x)\Big\}. \tag{5.9}$$

Corollary 2. *Assuming tameness as in Theorem 2, suppose $K_0(u, v, x) \cap \partial^0 p(u, v) = \{(0, 0)\}$ for all $x \in X$ (as is true certainly if every optimal solution x to $(P_{u,v})$ satisfies the constraint qualification $K_0(u, v, x) = \{(0, 0)\}$). Then p is locally Lipschitzian on a neighborhood of (u, v) and*

$$\partial p(u, v) = \operatorname{cl}\operatorname{co}\Big\{\Big[\bigcup_{x\in X} K(u, v, x)\Big] \cap \partial p(u, v)\Big\}; \tag{5.10}$$

in particular,

$$\partial p(u, v) \subset \operatorname{cl}\operatorname{co}\Big\{\bigcup_{x\in X} K(u, v, x)\Big\}, \tag{5.11}$$

$$\partial_v p(u, v) \subset \operatorname{cl}\operatorname{co}\Big\{z \,\Big|\, \exists y \text{ with } (y, z) \in \bigcup_{x\in X} K(u, v, x)\Big\}. \tag{5.12}$$

This follows via Propositions 3 and 4. It encompasses the estimate of Gauvin [11] for the smooth case (a), namely: under the assumption that $p(u, v) < \infty$ and

$$\begin{aligned}&\{x \mid (u', v') \text{ with } x \text{ feasible in } (P_{u',v'}),\ |(u', v') - (u, v)| \le \delta\}\\ &\textit{is a compact set for some } \delta > 0,\end{aligned} \tag{5.13}$$

if every optimal solution x to $(P_{u,v})$ satisfies the Mangasarian–Fromovitz constraint qualification (4.16), then p is locally Lipschitzian on a neighborhood of (u, v) and (5.11) holds. Corollary 2 also covers the estimate of Gauvin and Dubeau [12], which is (5.12) under the same assumptions. Of course (5.10) is a stronger assertion than (5.12), and Corollary 2 shows that it is valid under far more general circumstances than established previously. Corollary 1, on the other hand, shows that Theorem 2 yields outer estimates for $\partial p(u, v)$ even in cases where p is not locally Lipschitzian around (u, v). This is a new level of result.

Outer estimates for $\partial_v p(u, v)$ more subtle than (5.12) can be produced by combining Proposition 4 directly with Theorem 2. We leave the details to the reader.

Corollary 3. *Assuming tameness as in Theorem 2, one has for the closed convex cone*

$$G = \bigcup_{x \in X} \{(h, k) \mid y \cdot h + z \cdot k \leq 0 \text{ for all } (y, z) \in K_0(u, v, x)\} \tag{5.14}$$

that p is directionally Lipschitzian with respect to every $(h, k) \in \operatorname{int} G$.

Proof. Apply Proposition 2 and Corollary 1.

Corollary 4. *Assuming tameness as in Theorem 2, if* $\partial p(u, v) \neq \emptyset$ *(as is true in particular whenever p is calm at* (u, v)*, cf. Proposition 1), then* $(P_{u,v})$ *has an optimal solution* $x \in X$ *for which there is a vector* $(y, z) \in K(u, v, x)$ *that belongs to* $\partial p(u, v)$.

This result demonstrates that Theorem 2 yields not only 'outer estimates' but 'inner estimates'. Corollary 4 extends a result of Clarke and Aubin [6] for problem $(P_{u,v})$ in the 'almost convex' case, where everything is as in case (b) of Section 1 except that the objective function f_0 is not necessarily convex. It also covers a somewhat more general result of Aubin [1], although the connection in this case takes more effort to establish. The results in question are posed in terms of a problem structure that is different from the one in $(P_{u,v})$, although ultimately encompassed by it. However they also apply to infinite-dimensional problems, in contrast to Corollary 4.

The results in our earlier paper [29] can also be mentioned in conjunction with Corollary 4. These show the existence in $\partial p(u, v)$ of certain limits of multiplier vectors that satisfy higher-order optimality conditions.

Corollary 5. *Under the hypothesis of Theorem 2, if X is a singleton* $\{x\}$*, and for this x the set* $K(u, v, x)$ *is a singleton* $\{(y, z)\}$ *and the constraint qualification* $K_0(u, v, x) = \{(0, 0)\}$ *is satisfied, then p is strictly differentiable at* (u, v) *with* $\nabla p(u, v) = (y, z)$.

Proof. The assumptions imply via Theorem 2 that $\partial p(u, v) = \{(y, z)\}$, and this is equivalent to p being strictly differentiable at (u, v) with $\nabla p(u, v) = (y, z)$, as already noted in Section 2.

Remark. The constraint qualification in Corollary 5 does not have to be postulated separately in cases (a), (b), (c), (d) or (e) of $(P_{u,v})$. In those cases it follows from $K(u, v, x)$ being a singleton; cf. Proposition 7.

Corollary 6. *Suppose there is a mapping* ξ *as described in Proposition 10, and let X be the set of all cluster points of* $\xi(u', v')$ *as* $(u', v') \to_p (u, v)$ *(in the sense of (2.2)). Then the hypothesis of Theorem 2 is satisfied, so the conclusions in Theorem 2 and Corollaries 1, 3 and 4 (and under extra assumptions about X the conclusions in Corollaries 2 and 5) are valid.*

Proof. This follows from Proposition 10.

Corollary 7. *Suppose either that* $(\mathrm{P}_{u,v})$ *is tame and has a unique optimal solution* x, *or that* $(\mathrm{P}_{u,v})$ *has an optimal solution* x *that can be perturbed continuously in the sense of the existence of a mapping* ξ *as in Proposition* 10 *with* $\xi(u',v')\to x$ *as* $(u',v')\to_p(u,v)$. *If* $(\mathrm{P}_{u,v})$ *falls into the subdifferentially regular case* (d) *or extended smooth case* (e) *in Section* 4 *(or in particular one of cases* (a), (b) *and* (c) *of Section* 1), *then*

$$\partial p(u,v)\subset \mathrm{K}(u,v,x) \quad \text{and} \quad \partial^0 p(u,v)\subset \mathrm{K}_0(u,v,x). \tag{5.15}$$

Proof. Either way, we can apply Theorem 2 with $X=\{x\}$ (cf. Corollary 6). Furthermore, the conclusions of Proposition 7 hold in their strongest form. The formulas in Theorem 2 then reduce to (5.15) by virtue of $\partial p(u,v)$ and $\partial^0 p(u,v)$ being closed and convex, with $\partial^0 p(u,v)$ equal to the recession cone of $\partial p(u,v)$ unless $\partial p(u,v)=\emptyset$.

6. The main arguments

We proceed now to derive Theorem 1 from Theorem 2 using a certain characterization of our calmness property, and then to prove Theorem 2 itself by means of a new general result about limits of subgradients.

Proposition 11. *Let* x *be a locally optimal solution to* $(\mathrm{P}_{u,v})$. *Let* $\theta:[0,\infty)\to[0,\infty)$ *be any increasing convex function with* $\theta(0)=0$ *and* $\theta'(0)=0$, *and let* $\epsilon>0$. *Then the parameterized problem*

$$(\tilde{\mathrm{P}}_{u',v'})\qquad \begin{aligned}&\textit{minimize}\quad \tilde f_0(v',x')=f_0(v',x')+\theta(|x'-x|)\ \textit{over all } x' \textit{ satisfying}\\ &(v',x')\in\tilde D=\{(v',x')\in D\mid |x'-x|\le\epsilon\}\ \textit{and}\\ &f_i(v',x')+u_i'\begin{cases}\le 0 & \textit{for } i=1,\dots,s,\\ =0 & \textit{for } i=s+1,\dots,m\end{cases}\end{aligned}$$

in place of $(\mathrm{P}_{u',v'})$ *again satisfies the fundamental assumptions of Section* 1: $\tilde D$ *is again closed and* $\tilde f_0$ *locally Lipschitzian. Moreover, the term* $g(x')=\theta(|x'-x|)$ *in* $\tilde f_0$ *is a finite convex function of* x' *(therefore locally Lipschitzian) which is strictly differentiable at* $x'=x$ *with*

$$\nabla g(x)=0,\qquad g(x)=0,\qquad g(x')>0\quad \textit{for}\quad x'\ne x. \tag{6.1}$$

Furthermore, if $\epsilon<\rho$, *where* ρ *is the radius of a spherical neighborhood of* x *with respect to which the local optimality of* x *holds in* $(\mathrm{P}_{u,v})$, *then* x *is the unique (globally) optimal solution to* $(\tilde{\mathrm{P}}_{u,v})$, *and the optimal value function*

$$\tilde p(u',v')=\inf(\tilde P_{u',v'}) \tag{6.2}$$

has $\tilde p(u,v)=f_0(v,x)$. *If in addition* x *is actually a globally optimal solution to*

$(\mathrm{P}_{u,v})$, *then*

$$\tilde{p}(u,v)=p(u,v), \quad \textit{while } \tilde{p}(u',v')\geq p(u',v') \quad \textit{for all } (u',v')\neq(u,v). \tag{6.3}$$

Proof. All these assertions are elementary, except for the differentiability property of g. The convexity of g allows us to compute $g'(x;h)=\theta'(0)|h|=0$ for all h, from which it follows (cf. [26, Section 23]) that $\partial g(x)=\{0\}$. Then g must be strictly differentiable at 0 with $\nabla g(x)=0$, according to the results cited in Section 2.

Proposition 12. *Let x be a locally optimal solution to $(\mathrm{P}_{u,v})$. For $(\mathrm{P}_{u,v})$ to be calm at x, it is necessary and sufficient that for every function θ as in Proposition 11, one has for all $\epsilon>0$ sufficiently small that the modified optimal value function $\tilde{p}$ in Proposition 11 is calm at (u,v) (in the sense of Proposition 1).*

Proof. The argument will utilize the notation and conclusions of Proposition 11.

Necessity. Suppose $(\mathrm{P}_{u,v})$ is calm at x, and fix any θ as described. If for some $\epsilon\in(0,\rho)$ the function $\tilde{p}$ is not calm at (u,v), there exist for any $\beta\in\mathbb{R}$ points (u',v') arbitrarily near to (u,v) and yielding

$$[\tilde{p}(u',v')-\tilde{p}(u,v)]/|(u',v')-(u,v)|<\beta.$$

Here $\tilde{p}(u,v)=f_0(v,x)$, so the inequality means by the definition of $\tilde{p}$ that

$$[f_0(v',x')+g(x')-f_0(v,x)]/|(u',v')-(u,v)|<\beta$$

for some feasible solution x' to $(\mathrm{P}_{u',v'})$ with $|x'-x|\leq\epsilon$. Thus if there is a sequence of values $\epsilon_j\downarrow 0$ such that the corresponding functions $\tilde{p}$ are not calm at (u,v), we can select for any sequence of values $\beta_j\downarrow-\infty$, corresponding points (u^j,v^j) arbitrarily near to (u,v) and feasible solutions x^j to (P_{u^j,v^j}) with $|x^j-x|\leq\epsilon_j$ and

$$[f_0(v^j,x^j)+g(x^j)-f_0(v,x)]/|(u^j,v^j)-(u,v)|<\beta_j.$$

Then $x^j\to x$ and (since $g\geq 0$)

$$[f_0(v^j,x^j)-f_0(v,x)]/|(u^j,v^j)-(u,v)|\to-\infty. \tag{6.4}$$

In particular, (u^j,v^j) can be selected so as to converge to (u,v), and a contradiction is then obtained to the assumption that $(\mathrm{P}_{u,v})$ is calm at x. Hence under this assumption there cannot exist a sequence of values $\epsilon_j\downarrow 0$ such that $\tilde{p}$ is not calm at (u,v), and this is what we needed to prove.

Sufficiency. Suppose $(\mathrm{P}_{u,v})$ is not calm at x. Then there exist $(u^j,v^j)\to(u,v)$ and $x^j\to x$ such that x^j is feasible in (P_{u^j,v^j}) and (6.4) holds. Let $\delta_j=|(u^j,v^j)-(u,v)|$ and $\epsilon_j=|x^j-x|$; passing to subsequences if necessary, we can suppose that δ_j and ϵ_j are strictly decreasing in j. The line segment joining the points $(\epsilon_j, \delta_j\epsilon_j)$ and $(\epsilon_{j+1}, \delta_{j+1}\epsilon_{j+1})$ in $\mathbb{R}^2$ has slope

$$\lambda_j = (\delta_j\epsilon_j - \delta_{j+1}\epsilon_{j+1})/(\epsilon_j - \epsilon_{j+1})$$

which satisfies

$$\delta_j > \lambda_j > \delta_{j+1}. \tag{6.5}$$

Let $\theta : [0, \infty) \to [0, \infty)$ be the function whose graph is the union of all these segments and $(0, 0)$; then θ is continuous with

$$\theta(\epsilon_j) = \delta_j\epsilon_j \quad \text{for all } j, \qquad \theta(0) = 0.$$

From (6.5) we have $\lambda_j > \lambda_{j+1} > \cdots > 0$; hence θ is actually convex and increasing. Thus θ belongs to the class of functions under consideration, and

$$\theta(|x^j - x|)/|(u^j, v^j) - (u, v)| = \epsilon_j \to 0. \tag{6.6}$$

It will be demonstrated now that for this θ and any $\epsilon \in (0, \rho)$, the function $\tilde{p}$ is not calm at (u, v). Indeed, since x^j is a feasible solution to (P_{u^j,v^j}) with $|x^j - x| = \epsilon_j$, it is a feasible solution to $(\tilde{\mathrm{P}}_{u^j,v^j})$ for j sufficiently large, and then

$$\tilde{p}(u^j, v^j) \le f_0(v^j, x^j) + \theta(|x^j - x|).$$

We also have $\tilde{p}(u, v) = f_0(v, x)$ (because $\epsilon < \rho$), and since (6.4) holds for the chosen sequences, we see via (6.6) that

$$\frac{\tilde{p}(u^j, v^j) - \tilde{p}(u, v)}{|(u^j, v^j) - (u, v)|} \le \frac{f_0(v^j, x^j) - f_0(v, x)}{|(u^j, v^j) - (u, v)|} + \epsilon_j \to -\infty.$$

Thus

$$\liminf_{(u',v') \to (u,v)} \frac{\tilde{p}(u', v') - \tilde{p}(u, v)}{|(u', v') - (u, v)|} = -\infty$$

as claimed.

Proof of Theorem 1 (using Theorem 2). Suppose x is a locally optimal solution to $(\mathrm{P}_{u,v})$. Then for θ and ϵ as in Proposition 11 with $\epsilon < \rho$, x is the unique optimal solution to the modified problem $(\tilde{\mathrm{P}}_{u,v})$. Theorem 2 can be applied to $(\tilde{\mathrm{P}}_{u,v})$ (cf. Proposition 9). Taking $X = \{x\}$, we conclude in particular that

$$\partial\tilde{p}(u, v) \subset \operatorname{cl} \operatorname{co}\{\tilde{K}(u, v, x) + \tilde{K}_0(u, v, x)\}, \quad \text{and if } \tilde{K}_0(u, v, x) \text{ is pointed, also } \partial^0\tilde{p}(u, v) \subset \operatorname{cl} \operatorname{co} \tilde{K}_0(u, v, x), \tag{6.7}$$

where $\tilde{K}(u, v, x)$ and $\tilde{K}_0(u, v, x)$ are the multiplier sets corresponding to $(\tilde{\mathrm{P}}_{u,v})$, i.e., with $\tilde{f}_0$ and $\tilde{D}$ is place of f_0 and D. Actually

$$\tilde{K}(u, v, x) = K(u, v, x) \quad \text{and} \quad \tilde{K}_0(u, v, x) = K_0(u, v, x). \tag{6.8}$$

because

$$\partial\left[\tilde{f}_0 + \sum_{i=1}^{m} y_i f_i + \delta_{\tilde{D}}\right](v, x) = \partial\left[f_0 + \sum_{i=1}^{m} y_i f_i + \delta_D\right](v, x);$$

this is due to the fact that $\tilde{D}$ and D coincide in a neighborhood of (v, x), while $\tilde{f}_0$ and f_0 differ only by a function whose gradient vanishes at (v, x); cf. Proposition 5.

If $K_0(u, v, x) = \{(0,0)\}$, we have $0^+K(u, v, x) = \{(0,0)\}$ (Proposition 7), so $K(u, v, x)$ is not just closed (Proposition 7) but compact. Then too, $\partial^0 p(u, v) = \{(0,0)\}$ by (6.7) and (6.8). Hence according to Proposition 3, $\tilde{p}$ is Lipschitzian around (u, v) and in particular calm at (u, v). Since this holds regardless of the choice of θ, as long as $\epsilon \in (0, \rho)$, it is clear from Proposition 12 that $(P_{u,v})$ must be calm at x.

However, if $(P_{u,v})$ is calm at x, we know from Proposition 12 that when ϵ is sufficiently small, $\tilde{p}$ is calm at (u, v) and therefore by Proposition 1 that $\partial p(u, v) \neq \emptyset$. Then we deduce from (6.7) and (6.8) that $K(u, v, x) \neq \emptyset$. This completes the derivation of Theorem 1.

Next on the agenda is the proof of Theorem 2. The following result is the first step.

Proposition 13. *Suppose (u, v) is such that $(P_{u,v})$ is tame, and let A be a set for which the definition (5.1) of tameness is fulfilled. Then there are numbers $\delta > 0$ and $\alpha > p(u, v)$ and a compact set $\hat{D} \subset D$, such that the replacement of D by $\hat{D}$ does not affect the infimum $p(u', v')$ for any (u', v') satisfying $|(u', v') - (u, v)| < \delta$ and $p(u', v') < \alpha$, nor does it alter the set of optimal solutions x to $(P_{u,v})$ which lie in A or the sets $K(u, v, x)$ and $K_0(u, v, x)$ associated with any such x.*

Proof. Fix any $\epsilon > 0$ and corresponding δ and α as in (5.1). Let

$$\hat{D} = \{(v', x') \in D \mid \operatorname{dist}(x', A) \leq \epsilon \text{ and } |v' - v| \leq \delta\}.$$

Since A is compact and D is closed, $\hat{D}$ is compact. The assertions are then all obvious from (5.1) and the fact that the sets $\{x \mid (v, x) \in \hat{D}\}$ and $\{x \mid (v, x) \in D\}$ agree in a neighborhood of A.

We will also need a new general result about convergence of subgradients.

Proposition 14. *Suppose for $j = 1, 2, \ldots$, that x^j furnishes a finite local minimum of $f + g_j$, where f and g_j are lower semicontinuous functions from $\mathbb{R}^n$ to $(-\infty, \infty]$. If $x^j \to x$, $f(x^j) \to f(x)$ (finite) and $\partial g_j(x_j) \to \{0\}$ (in the sense that for every neighborhood U of 0 one has $\emptyset \neq \partial g_j(x^j) \subset U$ when j is sufficiently large), then $0 \in \partial f(x)$.*

Proof. Suppose $0 \notin \partial f(x)$. Then there exists a vector h such that $f^\uparrow(x; h) < 0$, i.e.,

$$0 > \lim_{\epsilon \downarrow 0} \limsup_{\substack{x' \to_f x \\ t \downarrow 0}} \left[\inf_{|h'-h| \leq \epsilon} \frac{f(x' + th') - f(x')}{t} \right]$$

(cf. the general formula (2.5) for subgradients in terms of subderivatives). In particular, for every $\epsilon > 0$ and sequence $t_j \downarrow 0$ one has

$$0 > \limsup_{j\to\infty}\left[\inf_{|h'-h|\le\epsilon} \frac{f(x^j + t_j h') - f(x^j)}{t_j}\right]. \tag{6.9}$$

The assumption that $f + g_j$ has a local minimum at x^j implies

$$f(x^j + t_j h') + g_j(x^j + t_j h') \ge f(x^j) + g_j(x^j)$$

for any h' once j is large enough, and this can be written

$$[f(x^j + t_j h') - f(x^j)]/t_j \ge -[g_j(x^j + t_j h') - g_j(x^j)]/t_j.$$

Hence by (6.9), for every $\epsilon > 0$ and sequence $t_j \downarrow 0$, one has

$$0 < \liminf_{j\to\infty}\left[\sup_{|h'-h|\le\epsilon} \frac{g_j(x^j + t_j h') - g_j(x^j)}{t_j}\right]. \tag{6.10}$$

Next we use the fact that $\partial g_j(x^j) \to \{0\}$; passing to a subsequence if necessary, it can be supposed that

$$\emptyset \ne \partial g_j(x^j) \subset \{w \mid |w| < \lambda_j\} \quad \text{where } \lambda_j \downarrow 0. \tag{6.11}$$

In particular $\partial g_j(x^j)$ is bounded; hence g_j is locally Lipschitzian around x^j [25, Theorem 4]. Moreover, (6.11) implies that for all $k \in \mathbb{R}^n$

$$\begin{aligned}\lambda_j|k| &> \max\{w\cdot k \mid w \in \partial g_j(x^j)\}\\ &= g_j^0(x^j; k) = \limsup_{\substack{x'\to x^j\\ t\downarrow 0}} \frac{g_j(x' - tk) - g_j(x')}{t}\end{aligned}$$

(recall from Section 2 that $g_j^{\uparrow}(x^j; k)$ reduces to $g_j^0(x^j; k)$ in the locally Lipschitzian case). Thus λ_j serves as a Lipschitz constant for g_j in some neighborhood of x_j, say in a ball of radius δ_j around x^j. Fixing $\epsilon > 0$ arbitrarily, choose the sequence $t_j \downarrow 0$ so that $x^j + t_j h'$ belongs to this neighborhood for all h' satisfying $|h' - h| \le \epsilon$ (it suffices to have $0 < t_j < \delta_j/\epsilon$). Then for all j sufficiently large one has

$$g_j(x^j + t_j h') - g_j(x^j) \le t_j\lambda_j|h'| \quad \text{whenever } |h' - h| \le \epsilon,$$

and hence from (6.10)

$$0 < \liminf_{j\to\infty} \lambda_j(|h| + \epsilon).$$

This contradicts the fact that $\lambda_j \downarrow 0$ and establishes that $0 \in \partial f(x)$ after all.

Proof of Theorem 2. Proposition 13 gives us license to suppose without loss of generality that D is a compact set. It is elementary then that $(P_{u',v'})$ has an optimal solution for every (u', v') such that it has a feasible solution (in particular for (u, v), because $p(u, v) < \infty$ by hypothesis), and that p is lower semicontinuous everywhere and globally bounded below. By means of the reformulation in Section 1, we can also reduce everything to the notationally simpler case where there are no vectors v and z. The reasoning here, as far as the equivalence of the multiplier conditions is concerned, is based on Proposition 5: the reformulation involves the introduction merely of linear constraint functions, and $\partial(g_1+g_2)=\nabla g_1+\partial g_2$ in particular when g_1 is linear.

In this reduced case with D compact, a formula proved in [29, Theorem 2] in terms of the (quadratic) augmented Lagrangian function becomes applicable:

$$\partial p(u)=\operatorname{cl}\operatorname{co}[Y+Y_0] \quad\text{and}\quad \partial^0 p(u)=\operatorname{cl}\operatorname{co} Y_0, \tag{6.12}$$

where

$$Y=\{y \mid \exists u^j \to_p u \text{ and } y^j \text{ an augmented multiplier vector for } (\mathrm{P}_{u^j}),\text{ such that } y^j\to y\}, \tag{6.13}$$

$$Y_0=\{y \mid \exists\, u^j \to_p u,\ \lambda_j\downarrow 0 \text{ and } y^j \text{ an augmented multiplier vector for } (\mathrm{P}_{u^j}),\text{ such that } \lambda_j y^j\to y\}. \tag{6.14}$$

Here y^j is called an 'augmented multiplier vector' for (P_{u^j}) if for all $r_j>0$ sufficiently large, the optimal solutions to (P_{u^j}) are precisely the vectors x^j such that (x^j, y^j) is a (global) saddlepoint of the augmented Lagrangian

$$L(u^j,x,y,r_j)=f_0(x)+\frac{1}{2r_j}\sum_{i=1}^{s}[y_i+r_j(f_i(x)+u_i^j)]_+^2 + \frac{1}{2r_j}\sum_{i=s+1}^{m}[y_i+r_j(f_i(x)+u_i^j)]^2-\frac{1}{2r_j}|y|^2$$

with respect to $x\in D$ and $y\in\mathbb{R}^m$. (We are using the notation that $s_+=\max\{s,0\}$.) Since $\partial p(u)$ and $\partial^0 p(u)$ are closed convex sets, we will be able to derive formulas (5.6) and (5.7) in Theorem 2 by showing that

$$Y\subset\bigcup_{x\in X}K(u,x) \quad\text{and}\quad Y_0\subset\bigcup_{x\in X}K_0(u,x). \tag{6.15}$$

Consider now any sequences $y^j\to y$ and $u^j\to_p u$ such as in the definition (6.13) of Y. Since D is compact and definition (5.1) is fulfilled by A, there is (for j sufficiently large) a sequence of points x^j such that x^j is an optimal solution to (P_{u^j}) and $\operatorname{dist}(x^j, A)\to 0$. Passing to a subsequence if necessary, we can suppose x^j converges to some $x\in A$. Then x is an optimal solution to (P_u) (by the continuity of the functions f_i and the fact that $p(u^j)\to p(u)$), and x belongs to X (the hypothesized set of optimal solutions to (P_u) which includes all those in A). We will demonstrate that $y\in K(u,x)$, and this will establish the first inclusion in (6.15).

To say that (x^j, y^j) is a saddlepoint of $L(u^j,x,y,r_j)$ with respect to $x\in D$ and $y^j\in\mathbb{R}^m$ is to say that

$$f_i(x^j)+u_i^j\le 0,\qquad y_i^j\ge 0,\qquad y_i^j[f_i(x^j)+u_i^j]=0\quad\text{for } i=1,\dots,s,$$
$$f_i(x^j)+u_i^j=0\quad\text{for } i=s+1,\dots,m, \tag{6.16}$$

and that x^j gives the global minimum over D of $L(u^j, y^j, r_j)$. But (6.16) implies that the latter function reduces locally around x^j to

$$f_0 + \sum_{i=1}^{m} y_i^j[f_i + u_i^j] + \frac{r_j}{2}\Big(\sum_{i\in I_0(j)} [f_i + u_i^j]_+^2 + \sum_{i\in I_1(j)} [f_i + u_i^j]^2\Big), \tag{6.17}$$

where

$$\begin{aligned} &I_0(j) = \text{set of all } i \in \{1, \dots, s\} \text{ with } y_i^j = 0,\\ &I_1(j) = \text{set of all other constraint indices.} \end{aligned} \tag{6.18}$$

This reduction makes use of the relation

$$\frac{1}{2r_j}([y_i^j + r_j(f_i + u_i^j)]_+^2 - (y_i^j)^2)$$
$$= \begin{cases} y_i^j[f_i + u_i^j] + \frac{r_j}{2}[f_i + u_i^j]^2 & \text{where } f_i + u_i^j \geq -y_i^j/r_j,\\ -\frac{1}{2r_j}(y_i^j)^2 & \text{where } f_i + u_i^j \leq -y_i^j/r_j. \end{cases} \tag{6.19}$$

(For active inequality constraint indices i with $y_i^j > 0$, one has $-y_i^j/r_j < 0$ but $f_i(x^j) + u_i^j = 0$, so only the first alternative in (6.19) holds in a certain neighborhood of x^j. For all other inequality constraint indices one has $y_i^j = 0$, so that (6.19) simply gives $[f_i + u_i^j]_+^2$.)

From (6.16) we know that in the limit as $j \to \infty$:

$$\begin{aligned} &f_i(x) + u_i \leq 0, \quad y_i \geq 0, \quad y_i[f_i(x) + u_i] = 0 \quad \text{for } i = 1, \dots, s,\\ &f_i(x) + u_i = 0 \quad \text{for } i = s+1, \dots, m. \end{aligned} \tag{6.20}$$

On the other hand, we have seen that x^j gives a local minimum to the function (6.18) over D. This tells us that x^j gives a local minimum to $f + g_j$, where

$$f = f_0 + \sum_{i=1}^{m} y_i f_i + \delta_D, \tag{6.21}$$

$$g_j = \sum_{i=1}^{m} (y_i^j - y_i) f_i + \frac{r_j}{2}\sum_{i=1}^{m} h_{ij}^2, \tag{6.22}$$

$$h_{ij} = \begin{cases} [f_i + u_i^j]_+ & \text{for } i \in I_0(j),\\ f_i + u_i^j & \text{for } i \in I_1(j). \end{cases} \tag{6.23}$$

The functions h_{ij} all vanish at x^j by virtue of the definitions (6.18), and these functions are all locally Lipschitzian. In applying to (6.22) the rules of subdifferential calculus for sums and squares (cf. Proposition 5 and [5, Section 13]), we get

$$\partial g_j(x^j) \subset \sum_{i=1}^{m} (y_i^j - y_i)\partial f_i(x^j) + r_j \sum_{i=1}^{m} h_{ij}(x^j)\partial h_{ij}(x^j),$$

where $h_{ij}(x^j) = 0$; the second sum therefore drops out. But $\lim_j (y_i^j - y_i) = 0$ and $\limsup_j \partial f_i(x^j) \subset \partial f_i(x)$ (because f_i is locally Lipschitzian and $x^j \to x$, cf. [3]).

Hence $\partial g_j(x^j) \to \{0\}$, and we may conclude from Proposition 14 that $0 \in \partial f(x)$ for f as in (6.21). This property along with (6.16) means that $y \in K(u, x)$ as claimed.

The argument is very similar in the case of y^j, u^j and λ_j such as in the definition (6.14) of Y_0. The difference comes in multiplying (6.17) through by λ_j and characterizing x^j accordingly as a local minimizer of $f + g_j$ taken as

$$f = \sum_{i=1}^{m} y_i f_i + \delta_D,$$

$$g_j = \lambda_j f_0 + \sum_{i=1}^{m} (\lambda_j y_i^j - y_i) f_i + \frac{r_j}{2} \sum_{i=1}^{m} h_{ij}^2,$$

with h_{ij} as in (6.3). Again $\partial g_j(x^j) \to \{0\}$, so $0 \in \partial f(x)$ by Proposition 14, and the conclusion is obtained that $y \in K_0(u, x)$.

Thus the second inclusion in (6.15) is valid too, and formulas (5.6) and (5.7) of Theorem 2 are then true in consequence of (6.12), as already explained.

To obtain via (6.12) and (6.15) the final assertions of Theorem 2, about equality holding in (5.7), it will suffice to prove that

$$\partial^0 p(u) = \text{cl co } Y_0 \tag{6.24}$$

if either $Y = \emptyset$ or Y_0 is pointed, and that in the latter case one actually has

$$\partial p(u) = \text{co}[Y + Y_0] \quad \text{and} \quad \partial^0 p(u) = \text{co } Y_0. \tag{6.25}$$

Here we must delve deeper into the argument in [29] by means of which (6.12) was established. The argument was based on representing $\partial p(u)$ and $\partial^0 p(u)$ in terms of the cone

$$N = \{\lambda(y, -1) \mid y \in Y, \lambda > 0\} \cup \{(y, 0) \mid y \in Y_0\} \tag{6.26}$$

in $R^m \times R$ by the formulas

$$\partial p(u) = \{y \mid (y, -1) \in \text{cl co } N\}, \quad \partial^0 p(u) = \{y \mid (y, 0) \in \text{cl co } N\}. \tag{6.27}$$

(See [29, Theorem and its proof].) It was observed also that

$$0 \in Y_0 \supset 0^+ Y := \{y \mid \exists y^j \in Y, \lambda_j \downarrow 0, \text{ with } \lambda_j y^j \to y\}, \tag{6.28}$$

or what amounts to the same thing, that N is closed and nonempty. The statements about (6.24) and (6.25) at the beginning of this paragraph, as well as (6.12) itself, are implied by this representation, as we demonstrate in the following geometric proposition, thereby completing proof of Theorem 2.

Proposition 15. *Let Y and Y_0 be any closed subsets of R^m such that Y_0 is a cone satisfying (6.28), and let N be the closed cone in $R^m \times R$ defined by (6.26). Then*

$$\{y \mid (y, -1) \in \text{cl co } N\} = \text{cl co}[Y + Y_0], \tag{6.29}$$

$$\{y \mid (y, 0) \in \text{cl co } N\} \supset \text{cl co } Y_0. \tag{6.30}$$

Equality holds in (6.30) *if* $Y = \emptyset$, *or if* Y_0 *is pointed; in the latter case* co Y_0 *is itself closed and pointed, as is* co N, *and one actually has*

$$\{y \mid (y, -1) \in \text{cl co } N\} = \text{co}[Y + Y_0], \tag{6.31}$$

$$\{y \mid (y, 0) \in \text{cl co } N\} = \text{co } Y_0. \tag{6.32}$$

Proof. It is trivial from (6.26) that (6.30) always holds, and that it holds with equality when $Y = \emptyset$. Note too that when $Y = \emptyset$, (6.29) and (6.31) hold with both sides empty. We can therefore suppose henceforth that $Y \neq \emptyset$. Then N meets both of the open half-spaces bounded by the hyperplane $H = \{(y, -1) \mid y \in R^m\}$, so co N certainly cannot be separated from H and hence $H \cap \text{ri co } N \neq \emptyset$ [3, Section 11]. This implies

$$H \cap \text{cl co } N = \text{cl}[H \cap \text{co } N]$$

[3, Section 6], or equivalently,

$$\{y \mid (y, -1) \in \text{cl co } N\} = \text{cl}\{y \mid (y, -1) \in \text{co } N\}. \tag{6.33}$$

On the other hand, since Y_0 is a cone containing 0, we find from (6.26) that

$$\{y \mid (y, -1) \in \text{co } N\} = \text{co}[Y + Y_0], \tag{6.34}$$

and of course

$$\{y \mid (y, 0) \in \text{co } N\} = \text{co } Y_0. \tag{6.35}$$

The combination of (6.33) and (6.34) yields (6.29).

We shall demonstrate now that if Y_0 is pointed, then co Y_0 is closed and pointed. Since N, like Y_0, is a closed cone containing the origin, and since N obviously is pointed if and only if Y_0 is pointed, co N too is closed and pointed. Then (6.31) and (6.32) will be seen simply as restatements of (6.34) and (6.35).

Assume Y_0 is pointed. Because Y_0 is a cone in R^m containing the origin, we have (by Carathéodory's theorem [3, Section 17])

$$\text{co } Y_0 = \{y^1 + \cdots + y^{m+1} \mid y^k \in Y_0\}.$$

If co Y_0 were not pointed, we could represent the origin as a sum of nonzero vectors in co Y_0. This would give a representation of the origin as a sum of nonzero vectors in Y_0, contradicting the pointedness of Y_0. Thus co Y_0 is pointed.

Proving that co Y_0 is closed when Y_0 is pointed amounts to proving in the case of the closed cone

$$W = Y \times \cdots \times Y \subset (R^m)^{m+1}$$

and linear transformation

$$A : w = (y^1, \ldots, y^{m+1}) \to y^1 + \cdots + y^{m-1}$$

that

$$\text{if } w \in W \text{ and } A(w) = 0 \text{ imply } w = 0, \quad \text{then } A(W) \text{ is closed.} \tag{6.36}$$

Suppose $A(W)$ were not closed. Then there would exist $w^j \in W$ such that $A(w^j) \to q \notin A(W)$. The sequence $\{w^j\}$ could not have a bounded subsequence, for if so it would have a cluster point w, and then $A(w) = q$. Therefore $|w^j| \to \infty$, and for $\bar{w}^j = w^j/|w^j|$ we would have $\bar{w}^j \in W$ (because W is a cone) and

$$A(\bar{w}^j) = A(w^j)/|w^j| \to 0.$$

Since $|\bar{w}^j| = 1$ and W is closed, the sequence $\{\bar{w}^j\}$ would have a cluster point $\bar{w} \in W$ satisfying $|\bar{w}| = 1$ and $A(\bar{w}) = 0$. This argument verifies (6.36) and finishes the proof of Proposition 15.

Remark. The need for some further conditions on Y_0 in order to ensure equality in (6.30) is demonstrated by

$$Y = \{(y_1, y_2) \in R^2 \mid y_1^2 \le |y_2|\}, \quad Y_0 = \{(y_1, y_2) \in R^2 \mid y_1 = 0\}.$$

In this case one has (6.28) satisfied but $\text{cl co}\, N = \{(y_1, y_2, \eta) \mid \eta \le 0\}$, so that

$$\{(y_1, y_2) \mid (y_1, y_2, 0) \in \text{cl co}\, N\} = R^2, \quad \text{cl co}\, Y_0 = Y_0 = R^1 \times \{0\}.$$

7. Application to generalized directional derivatives

The estimates in Theorem 2 lead to results about the various derivative functions $p^{\uparrow}$, p^0, p^+, p_+ and p' discussed in Section 2. We have already seen one consequence in Corollary 5; there (2.10) holds, and in particular $p'(u, v; h, k) = y \cdot h + z \cdot k$ for all (h, k).

Theorem 3. *With (u, v) and X satisfying the hypothesis of Theorem 2, let (h, k) be a vector belonging to the closed convex cone*

$$G = \bigcap_{x \in X} \{(h, k) \mid y \cdot h + z \cdot k \le 0 \textit{ for all } (y, z) \in K_0(u, v, x)\}. \tag{7.1}$$

If either $p^{\uparrow}(u, v; h, k) < \infty$ or there is at least one $x \in X$ with $K(u, v, x) \ne \emptyset$, one has

$$p^{\uparrow}(u, v; h, k) \le \sup_{x \in X} \left[\sup_{(y, z) \in K(u, v, x)} \{y \cdot h + z \cdot k\} \right], \tag{7.2}$$

(where an empty supremum is interpreted as $-\infty$). This inequality is valid in particular for all $(h, k) \in \text{int}\, G$; in fact for such (h, k), (2.9) holds and one has the further estimates

$$p^{+}(u, v; h, k) \le \inf_{x \in X} \left[\sup_{(y, z) \in K(u, v, x)} \{y \cdot h + z \cdot k\} \right], \tag{7.3}$$

$$p_+(u, v; -h, -k) \le \inf_{x\in X}\left[\sup_{(y,z)\in K(u,v,x)}\{-y\cdot h - z\cdot k\}\right]. \tag{7.4}$$

Proof. The first estimate (7.2) is obtained from the outer estimate (5.8) in Corollary 1 and the formula

$$p^\uparrow(u, v; h, k) = \sup\{y\cdot h + z\cdot k \mid (y, z)\in\partial p(u, v)\},$$

which we know from (2.6) to be correct whenever $p^\uparrow(u, v; h, k) < \infty$ or $\partial p(u, v) \ne \emptyset$. In taking the supremum of $y\cdot h + z\cdot k$ over all (y, z) belonging to the right side of (5.8), we can certainly ignore the 'cl co.' Thus for the sets

$$M = \bigcup_{x\in X} K(u, v, x), \qquad M_0 = \bigcup_{x\in X} K_0(u, v, x), \tag{7.5}$$

we have

$$p^\uparrow(u, v; h, k) \le \sup(y\cdot h + z\cdot k + y^0\cdot h + z^0\cdot k \mid (y, z)\in M, (y^0, z^0)\in M_0\}$$

provided that either $p^\uparrow(u, v; h, k) < \infty$ or $M + M_0 \ne \emptyset$. Here M_0 is actually a cone (not necessarily convex) which contains $(0, 0)$, and G is its polar, so that

$$\sup\{y^0\cdot h + z^0\cdot k \mid (y^0, z^0)\in M_0\} = \begin{cases} 0, & \text{if } (h, k)\in G,\\ \infty, & \text{if } (h, k)\notin G.\end{cases}$$

Thus $M + M_0 \ne \emptyset$ if and only if $M \ne \emptyset$, and for $(h, k)\in G$ the right side of (7.6) reduces to

$$\sup\{y\cdot h + z\cdot k \mid (y, z)\in M\}.$$

In this manner one obtains the validity of (7.2) for all cases having either $p^\uparrow(u, v; h, k) < \infty$ or $M \ne \emptyset$, as asserted.

We have already noted in Corollary 3 that p is directionally Lipschitzian with respect to $(h, k)\in \operatorname{int} G$, which means that (2.9) holds (see Section 2).

To derive (7.3), we initially fix any $x\in X$ and consider the modified problem $(\bar{P}_{u,v})$ in Proposition 11. As long as ϵ is small enough, this has x as its unique optimal solution and again satisfies all our assumptions, including tameness (with respect to $\bar{A} = A\cup\{x\}$). The results obtained so far for $(P_{u,v})$ can therefore be applied to $(\bar{P}_{u,v})$ with $\bar{X} = \{x\}$: for (h, k) belonging to the interior of

$$\bar{G} = \{(h, k) \mid y\cdot h + z\cdot k \le 0 \text{ for all } (y, z)\in K_0(u, v, x)\}, \tag{7.7}$$

one has

$$\bar{p}^\uparrow(u, v; h, k) \le \sup_{(y,z)\in K(u,v,x)}\{y\cdot h + k\cdot z\}, \tag{7.8}$$

and moreover (2.9) holds for $\bar{p}$, so that actually

$$\bar{p}^+(u, v; h, k) \le \bar{p}^\uparrow(u, v; h, k). \tag{7.9}$$

At the same time we have $\bar{p}$ and p related by (6.3) in Proposition 11, and this implies

$$p^+(u, v; h, k) \le \tilde{p}^+(u, v; h, k). \tag{7.10}$$

Putting together (7.8), (7.9) and (7.10), we see that

$$p^+(u, v; h, k) \le \sup_{(y,z)\in K(u,v,x)} \{y \cdot h + z \cdot k\} \quad \text{when } (h, k) \in \operatorname{int} \tilde{G}.$$

Since $\tilde{G} \supset G$, and x was an arbitrary point of X, the truth of (7.3) for all $(h, k) \in \operatorname{int} G$ is immediate from this.

The argument for (7.4) is different. Bear in mind that p is finite and lower semicontinuous at (u, v) under our hypothesis (cf. Proposition 8). Denoting the right side of (7.2) by β, we see that (7.4) can be written in the form

$$\limsup_{\substack{(h',k')\to(h,k)\\ t\downarrow 0}} \frac{p(u, v) - p(u - th', v - tk')}{t} \le \beta, \tag{7.11}$$

while what we know from (7.2) and (2.10) is that

$$\limsup_{\substack{(u',v')\to_p(u,v)\\ (h',k')\to(h,k)\\ t\uparrow 0}} \frac{p(u' + th', v' + tk') - p(u', v')}{t} \le \beta. \tag{7.12}$$

Our task will be to derive (7.11) from (7.12). Let α denote the value of the 'lim sup' in (7.11), and consider any consequences $(h^j, k^j) \to (h, k)$ and $t_j \downarrow 0$ for which it is attained:

$$\lim_{j\to\infty} \frac{p(u, v) - p(u - t_jh^j, v - t_jk^j)}{t} = \alpha. \tag{7.13}$$

We need to show $\alpha \le \beta$, and for this purpose it is enough to look at the case where $\alpha > -\infty$. Passing to subsequences in (7.13) if necessary, we can suppose that

$$\gamma = \lim_{j\to\infty} p(u - t_jh^j, v - t_jk^j)$$

exists. Since $\alpha > -\infty$ and $t_j > 0$ in (7.13) it must be true that $\gamma \le p(u, v)$, yet the opposite inequality must hold too, because p is lower semicontinuous at (u, v). Hence

$$p(u - t_jh^j, v - t_jk^j) \to p(u, v). \tag{7.14}$$

Define $(u^j, v^j) = (u - t_jh^j, v - t_jk^j)$. Then

$$\lim_{j\to\infty} \frac{p(u^j + t_jh^j, v^j + t_jh^j) - p(u^j, v^j)}{t_j} = \alpha$$

by (7.13), and $(u^j, v^j) \to_p (u, v)$ by (7.14). It follows from (7.12) that $\alpha \le \beta$, and this completes our proof.

Remark 1. Inequality (7.4) could also be expressed as

$$p^-(u, v; h, k) \le \sup_{x\in X} [\sup_{(y,z)\in K(u,v,x)} \{y \cdot h + z \cdot k\}], \tag{7.15}$$

where in parallel to the definition (2.11) of $p^+(u, v; h, k)$ one takes

$$p^-(u, v; h, k) = \limsup_{\substack{(h',k')\to(h,k)\\ t\uparrow 0}} \frac{p(u+th', v+tk') - p(u, v)}{t}. \tag{7.16}$$

Remark 2. Inequality (7.3) holds in a more general form, as shown by the proof: for an *arbitrary* set X_0 of optimal solutions to $(P_{u,v})$, if $(h, k) \in \text{int}\, G_0$, where G_0 is the cone obtained when X is replaced by X_0 in (7.1), then (7.3) too is valid with X replaced by X_0.

Corollary 1. *With (u, v) and X satisfying the hypothesis of Theorem 2, suppose that the constraint qualification $K_0(u, v, x) = \{(0, 0)\}$ holds for every $x \in X$. Then for all $(h, k) \in \mathbb{R}^m \times \mathbb{R}^d$ one has (2.9) and*

$$p^+(u, v; h, k) \le \inf_{x\in X} \Big[\sup_{(y,z)\in K(u,v,x)} \{y \cdot h + z \cdot k\}\Big], \tag{7.17}$$

$$p_+(u, v; h, k) \ge \inf_{x\in X} \Big[\inf_{(y,z)\in K(u,v,x)} \{y \cdot h + z \cdot k\}\Big]. \tag{7.18}$$

If in addition $K(u, v, x)$ is a singleton $\{(y(x), z(x))\}$ for each $x \in X$, then the derivatives $p'(u, v; h, k)$ exist, and in fact

$$p^+(u, v; h, k) = p_+(u, v; h, k) = \inf_{x\in X} \{y(x) \cdot h + z(x) \cdot k\}. \tag{7.19}$$

Proof. This is the case of Theorem 3 where G is all of $\mathbb{R}^m \times \mathbb{R}^d$, so that (h, k) and $(-h, -k)$ both always belong to int G. The inner 'sup' in (7.17) and 'inf' in (7.18) coincide, of course, when $K(u, v, x)$ is a singleton.

Corollary 1 generalizes results of Gauvin and Tolle [13], Gauvin [11], Gauvin and Dubeau [12] in the smooth case (a) of $(P_{u,v})$, and of Auslender [2] in the somewhat more general case where the *inequality* constraints need not be smooth. Corollary 1 allows nonsmooth equality constraints too, plus abstract constraints represented by $(v, x) \in D$, and at the same time yields stronger conclusions in terms of Hadamard derivatives instead of just Dini derivatives.

Corollary 2. *With (u, v) and X as in the hypothesis of Theorem 2, and G the cone in (7.1), if* int $G \ne \emptyset$ *one has either $p^+(u, v; h, k) > -\infty$ for all $(h, k) \in \text{int}\, G$ or*

$$p^+(u, v; h, k) = p_+(u, v; h, k) = -\infty \quad \textit{for all } (h, k) \in \text{int}\, G, \tag{7.20}$$

the latter case occurring if and only if $K(u, v, x) = \emptyset$ for some $x \in X$.

Proof. Apply (7.3) and use the fact that $p_+ \leq p^+$.

Corollary 3. *In the case where no parameter vector v (or corresponding multiplier vector z) is being considered, and all the explicit constraints in (P_u) are inequalities (notationally: $s = m$), suppose u is such that (P_u) is tame, and let X be the set of all optimal solutions to (P_u). Then for every strictly negative vector h (i.e., $-h \in \operatorname{int} \mathbb{R}^m_+$) one has*

$$p^+(u; h) \leq \inf_{x \in X} \Big[\sup_{y \in K(u,x)} y \cdot h \Big] \leq 0. \tag{7.21}$$

Proof. Here $K_0(u, x) \subset \mathbb{R}^m_+$ for all $x \in X$, so that the cone G in (7.1) includes $-\mathbb{R}^m_+$. Every strictly negative h therefore belongs to $\operatorname{int} G$, and (7.21) can be obtained as a special case of (7.3).

Our final result treats only a special, but nevertheless very important class of problems. It extends the marginal value theorem of Gol'shtein [5, Section 7] to the case where the set of Kuhn–Tucker pairs associated with $(\mathrm{P}_{u,v})$ is not necessarily compact. Again, conclusions are obtained for Hadamard derivatives rather than just Dini derivatives.

Theorem 4. *In the mixed smooth-convex case* (c) *in Section* 1, *and with (u, v) and X such that the hypothesis of Theorem 2 is satisfied, one has for all (h, k):*

$$p_+(u, v; h, k) \geq \inf_{x \in X} \Big[\sup_{(y,z) \in K(u,v,x)} \{y \cdot h + z \cdot k\} \Big]. \tag{7.22}$$

The set of vectors $y \in \mathbb{R}^m$ satisfying for a given $x \in X$ the complementary slackness conditions (4.3) *and*

$$0 \in \sum_{i=1}^m y_i \partial_x f_i(v, x) + N_C(x),$$

is actually a closed convex cone Y_0 independent of x. The convex cone in (7.1) *takes the form*

$$G = \bigcap_{x \in X} \Big\{ (h, k) \,\Big|\, \sum_{i=1}^m y_i [\nabla_v f_i(x, v) k + h_i] \leq 0 \text{ for all } y \in Y_0 \Big\}, \tag{7.23}$$

and for all $(h, k) \in \operatorname{int} G$ one has

$$p^+(u, v; h, k) = p_+(u, v; h, k) = \min_{x \in X} \Big[\sup_{(y,z) \in K(u,v,x)} \{y \cdot h + z \cdot k\} \Big]. \tag{7.24}$$

Proof. Since (7.22) is trivial if $p_+(u, v; h, k) = +\infty$, we can suppose in proving (7.22) that

$$\infty > p_+(u, v; h, k) = \lim_{j \to \infty} \frac{p(u + t_j h^j, v + t_j k^j) - p(u, v)}{t_j}. \tag{7.25}$$

for certain sequences $t_j \downarrow 0$ and $(h^j, k^h) \to (h, k)$. Let

$$(u^j, v^j) = (u + t_j h^j, v + t_j k^j) \to (u, v).$$

Then $p(u^j, v^j) \to p(u, v)$ by (7.25) and the lower semicontinuity of p at (u, v), the latter being a consequence of the tameness condition in the hypothesis of Theorem 2 (cf. Proposition 8). Thus $(u^j, v^j) \to_p (u, v)$. Introduce next a mapping ξ as in Proposition 10 whose cluster points as $(u', v') \to_p (u, v)$ all belong to the set A invoked in the tameness definition (5.1). Setting $x^j = \xi(u^j, v^j)$ and passing to subsequences of necessary, we get a convergent sequence of optimal solutions x^j to (P_{u^j, v^j}) whose limit is a certain optimal solution $x \in A$ to $(P_{u,v})$. Then $x \in X$, since under the hypothesis of Theorem 2 every optimal solution to $(P_{u,v})$ in A is also in X. Note that $f_0(v^j, x^j) = p(u^j, v^j)$ and $f_0(v, x) = p(u, v)$, and hence

$$\lim_{j\to\infty} \frac{f_0(v^j, x^j) - f_0(v, x)}{t_j} = p_+(u, v; h, k) \tag{7.26}$$

by (7.25). Consider now an arbitrary $(y, z) \in K(u, v, x)$. This satisfies (4.3) and (4.4), but since we are dealing with the mixed case (b) of $(P_{u,v})$, (4.4) can be written as (4.11), or in terms of the function

$$l = f_0 + \sum_{i=1}^{m} y_i f_i$$

even more simply as

$$0 \in \partial_x l(v, x) + N_C(x) \quad \text{and} \quad z = \nabla_v l(v, x). \tag{7.27}$$

Here l inherits from the functions f_i the property of being convex in the second argument, and $\mathscr{C}^1$ in the first argument with gradient depending continuously on *both* arguments. This joint continuity ensures (via the mean value theorem) that actually

$$\lim_{j\to\infty} \frac{l(v + t_j k^j, x^j) - l(v, x^j)}{t_j} = \nabla_v l(v, x) \cdot k = z \cdot k, \tag{7.28}$$

a fact that will be put to use presently. The convexity property of l, on the other hand, allows us to read the first condition in (7.27) as saying that $l(v, \cdot)$ attains it minimum over C at x (cf. [26, Theorem 27.4]). Since x^j is feasible for (P_{u^j, v^j}), and hence in particular $x^j \in C$, it follows from this that

$$l(v, x^j) \geq l(v, x) \tag{7.29}$$

and from (4.3) that

$$f_0(v, x) = f_0(v, x) + \sum_{i=1}^{m} y_i[f_i(v, x) + u_i] = l(v, x) + y \cdot u, \tag{7.30}$$

$$f_0(v^j, x^j) \geq f_0(v^j, x^j) + \sum_{i=1}^{m} y_i[f_i(v^j, x^j) + u_i^j] = l(v^j, x^j) + y \cdot u^j. \tag{7.31}$$

Therefore

$$\begin{aligned} f_0(v^j, x^j) - f_0(v, x) &\geq [l(v^j, x^j) + y \cdot u^j] - [l(u, x) + y \cdot u] \\ &\geq y \cdot (u^j - u) + l(v^j, x^j) - l(v, x^j) \\ &= t_j y \cdot h^j + [l(v + t_j k^j, x^j) - l(v, x^j)]. \end{aligned}$$

Using this estimate in (7.26) and invoking (7.28), we get

$$p_+(u, v; h, k) \geq y \cdot h + z \cdot k.$$

This being true for arbitrary $(y, z) \in K(u, v, x)$, we conclude that

$$p_+(u, v; h, k) \geq \sup_{(y, z) \in K(u, v, x)} \{y \cdot h + z \cdot k\} \tag{7.32}$$

for the particular $x \in X$ which has been constructed, and hence that (7.22) is indeed true.

The rest of the proof of Theorem 4 is mostly a matter of applying Theorem 3, specifically (7.3). The special form for G is readily derived from the fact that condition (4.5) in the definition of $K_0(u, v, x)$ reduces in the present case to (4.12) (cf. also Proposition 6). As pointed out in Section 4 in the remarks following (4.12), the first condition in (4.12), together with (4.3) and the feasibility of x, constitute a certain saddlepoint condition on (x, y). As is well known, the set of saddlepoints of a given function is always a product set; the set of y's corresponding to a given x is independent of the choice of x.

Corollary. *Under the assumptions in Theorem 4, if the set Y_0 consists of just $y = 0$, then p is locally Lipschitzian around (u, v) and (7.24) holds for all (h, k).*

Proof. To say that $Y_0 = \{0\}$ is to say that $K_0(u, v, x) = \{(0, 0)\}$ for each $x \in X$. Then p is locally Lipschitzian by Corollary 2 of Theorem 2 in Section 5.

References

[1] J.P. Aubin, "Further properties of Lagrange multipliers in nonsmooth optimization", *Applied Mathematics and Optimization* 6 (1980) 79–90.

[2] A. Auslender, "Differential stability in nonconvex and non differentiable programming", in: P. Huard, ed., *Point-to-set maps and mathematical programming*, Mathematical Programming Studies, Vol. 10 (North-Holland, Amsterdam, 1979) pp. 29–41.

[3] F.H. Clarke, "A new approach to Lagrange Multipliers", *Mathematics of Operations Research* 1 (1976) 165–174.

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

[5] F.H. Clarke, "Generalized gradients of Lipschitz functionals", *Advances in Mathematics* (1980).

[6] F.H. Clarke and J.P. Aubin, "Shadow prices and duality for a class of optimal control problems", *SIAM Journal on Control and Optimization* 17 (1979) 567–586.

[7] V.E. Dem'janov and V.N. Malozemov, "The theory of nonlinear minimal problems", *Russian Mathematical Surveys* 26 (1971) 57–115.

[8] V.E. Dem'janov and A.B. Pevnyi, "First and second marginal values of mathematical programming problems", *Soviet Mathematics Doklady* 13 (1972) 1502–1506.

[9] J.P. Evans and F.J. Gould, "Stability in nonlinear programming", *Operations Research* 18 (1970) 107–118.

[10] A. Fiacco and W.P. Hutzler, "Extension of the Gauvin–Tolle optimal value differential stability results to general mathematical programs", *Mathematical Programming Study*, to appear.

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

[12] J. Gauvin and F. Dubeau, "Differential properties of the marginal function in mathematical programming", *Mathematical Programming Study*, to appear.

[13] J. Gauvin and J.W. Tolle, "Differential stability in nonlinear programming", *SIAM Journal on Control and Optimization* 15 (1977) 294–311.

[14] B. Gollan, "A general perturbation theory for abstract optimization problems", *Journal of Optimizations Theory and Applications*, to appear.

[15] E.G. Gol'shtein, *Theory of convex programming*, Translations of Mathematical Monographs 36 (American Mathematical Society, Providence RI, 1972).

[16] H.J. Greenberg and W.P. Pierskalla, "Extensions of the Evans–Gould stability results for mathematical programs", *Operations Research* 20 (1972) 143–153.

[17] J.-B. Hiriart-Urruty, "Gradients généralisés de fonctions marginales", *SIAM Journal on Control and Optimization* 16 (1978) 301–316.

[18] J.-B. Hiriart-Urruty, "Refinements of necessary optimality conditions in nondifferentiable programming I", *Applied Mathematics and Optimization* 5 (1979) 63–82.

[19] J.-B. Hiriart-Urruty, "Refinements of necessary optimality conditions in nondifferentiable programming II", to appear.

[20] W.W. Hogan, "Directional derivatives for extremal-value functions with applications to the completely convex case", *Operations Research* 21 (1973) 188–209.

[21] F. Lempio and H. Maurer, "Differential stability in infinite dimensional nonlinear programming", *Applied Mathematics and Optimization*, to appear.

[22] E.S. Levitin, "On the local perturbation theory of mathematical programming in a Banach space", *Soviet Mathematics Doklady* 16 (1975).

[23] E.S. Levitin, "On the perturbation theory of nonsmooth extremal problems with constraints", *Soviet Mathematics Doklady* 16 (1975).

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

[25] R.T. Rockafellar, "Clarke's tangent cones and the boundaries of closed sets in $\mathbb{R}^n$", *Nonlinear Analysis, Theory, Methods and Applications* 3 (1979) 145–154.

[26] R.T. Rockafellar, Convex Analysis (Princeton University Press, Princeton, NJ, 1970).

[27] R.T. Rockafellar, "Directionally Lipschitzian functions and subdifferential calculus", *Proceedings of the London Mathematical Society* 39 (1979) 331–355.

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

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

[30] R.T. Rockafellar, *La théorie des sous-gradients et ses applications: Fonctions convexes et non convexes*, Collection Chaire Aisenstadt (Presses de l'Université de Montréal, Montréal, Québec, 1979).
[31] S.M. Robinson, "Stability theory for systems of inequalities, part I: Linear systems", *SIAM Journal on Numerical Analysis* 12 (1976) 754–769.
[32] S.M. Robinson, "Stability theory for systems of inequalities, part II: Differential nonlinear systems", *SIAM Journal on Numerical Analysis* 13 (1976) 497–513.
[33] A.C. Williams, "Marginal values in linear programming", *SIAM Journal on Applied Mathematics* 11 (1963) 82–99.