AUGMENTED LAGRANGIANS AND MARGINAL VALUES IN PARAMETERIZED OPTIMIZATION PROBLEMS

R. Tyrrell Rockafellar

February 1981
WP-81-19

Research sponsored in part by the Air Force Office of Scientific Research, Air Force Systems Command, United States Air Force, under grant number 77-3204.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

ABSTRACT

When an optimization problem depends on parameters, the minimum value in the problem as a function of the parameters is typically far from being differentiable. Certain subderivatives nevertheless exist and can be intepreted as generalized marginal values. In this paper such subderivatives are studied in an abstract setting that allows for infinite dimensionality of the decision space. By means of the notion of proximal subgradients, a new general formula of subdifferentiation is established which provides an upper bound for the marginal values in question and a very broad criterion for local Lipschitz continuity of the optimal value function. Augmented Lagrangians are introduced and shown to lead to still sharper estimates in terms of special multiplier vectors. This approach opens a way to taking higher-order optimality conditions into account in such estimates.

AUGMENTED LAGRANGIANS AND MARGINAL VALUES
IN PARAMETERIZED OPTIMIZATION

INTRODUCTION

An enormous variety of optimization problems can be posed in the form

$$(1.1)\qquad \text{minimize } F(u,x) \text{ over all } x \in X \quad ,$$

where X is some linear topological space (locally convex and separated), u is a parameter vector ranging over another such space U, and F is an *extended-real-valued* function on $U \times X$. For example, a nonlinear programming problem

$$\text{minimize } f_0(x) \text{ over all } x \in C \text{ satisfying}$$

$$(1.2)\qquad f_i(x) + u_i \begin{cases} \leq 0 & \text{for } i = 1, .., s, \\ = 0 & \text{for } i = s+1, .., m, \end{cases}$$

where $C = X$ and $f_i : X \to R$, can be represented in terms of $u = (u_1, .., u_m)$ and

$$(1.3)\qquad F(u,x) = \begin{cases} f_0(x) & \text{if } x \text{ is feasible in (1.2)} \quad , \\ +\infty & \text{if } x \text{ is not feasible.} \end{cases}$$

The abstract formulation (1.1) is illuminating because it applies equally well to problems and parameterizations quite beyond the nonlinear programming framework (1.2), and because it directs our attention to the fundamental difficulties in studying the optimal value function

$$(1.4) \qquad p(u) = \inf_{x \in X} F(u,x) \quad .$$

These difficulties revolve around the fact that no amount of smoothness assumed on the data in the problem, such as smoothness of the functions f_i in (1.2), is enough to imply that p is differentiable. Even if F itself were finite everywhere and smooth, differentiability of p could fail. Yet this negative observation cannot be the end of the story, because p is an extremely important function in many applications. Some understanding, however imperfect, of its "subdifferential" properties is essential.

Progress has been made in various directions over the years, but recently there have been redoubled efforts in terms of a generalized theory of differentiation founded by Clark [1]. For the abstract case of (1.4) specifically there are results of Clarke [1] and Hiriart-Urruty [2], and when F represents a nonlinear (possibly nonconvex) programming problem as in (1.3), there are more detailed analyses of Gauvin [3], Gauvin and Dubeau [4], and Rockafellar [5], [6], [7]. Here we shall prove a new theorem for the abstract case and show how augmented Lagrangian functions can be introduced and utilized to get improvements. The importance of augmented Lagrangians as a theoretical took for such purposes has already been demonstrated in work in finite-dimensional nonlinear programming [5], [6], [7]. But the fact that the same idea can be pursued more generally, and may even open a new route to the study of higher-order optimality conditions for problems that can be put in the form (1.1), has not previously been pointed out.

To set the stage we make the blanket assumptions that F is lower semicontinuous, $F(u,x) > -\infty$ everywhere, and

(1.5) $$\begin{cases} \text{for every } \bar{u} \in \mathcal{U} \text{ and } \alpha \in R \text{ there is a neighborhood } U \text{ of } \bar{u} \\ \text{and a compact set } K \subset X \text{ such that} \\ \qquad u \in U, F(u,x) \leq \alpha \Rightarrow x \in K \quad . \end{cases}$$

This is relatively painless and has the virtue of ensuring that p is a lower semicontinuous function on $\mathcal{U}$ with $p(u) > -\infty$ everywhere. It implies further that the optimal solution multifunction $X: \mathcal{U} \rightrightarrows X$ defined by

(1.6) $$X(u) = \arg\min_{x \in X} F(u,x)$$

is upper semicontinuous, nonempty-valued where $p < \infty$, and locally has values uniformly contained in a compact set.

Clarke founded his theory of generalized differentiation on a concept of "subgradient" and showed that for locally Lipschitzian functions on Banach spaces, subgradients are dual to certain special directional derivatives. We extended this duality in [8] to the non-Lipschitzian case through an appropriate definition of "subderivatives" slightly more complicated than the expressions considered by Clarke. These are the sort of derivatives needed in dealing with the optimal value function p, since although lower semicontinuity is no real problem, we cannot suppose *a priori* that p is locally Lipschitzian. Indeed, we hold the hope of developing by subdifferential theory useful conditions that *imply* p is locally Lipschitzian.

Let u be a point where p(u) is finite. For each $h \in \mathcal{U}$, let $\mathcal{N}(h)$ denote the collection of all neighborhoods of h. The (*upper*) *subderivative* of p at u is the quantity

(1.7) $$p^{\uparrow}(u;h) = \sup_{U \in \mathcal{N}(h)} \Big[\limsup_{\substack{u' \to u \\ p(u') \to p(u) \\ t \downarrow 0}} \Big[\inf_{h' \in U} \frac{p(u'+th') - p(u')}{t} \Big] \Big]$$

This limit may initially seem rather peculiar, not to mention complicated, but it emerges as fundamental in so many ways that the reader would do well to reflect carefully on its meaning. Bolstered by the mathematical evidence already compiled of the robustness of this definition in application to a large number of situations, one is tempted to suggest that these subderivatives are just what should be put in mind when the subject of "marginal values" in the parameterized problem (1.1) comes up.

Some of the properties of subderivatives are quite surprising. As a function of h, $p^{\uparrow}(u;h)$ is lower semicontinuous and *sublinear* (convex and positively homogeneous), not identically $+\infty$. If h is such that

$$(1.8) \qquad \inf_{U \in N(h)} \Big[\limsup_{\substack{u' \to u \\ p(u') \to p(u)}} \Big[\sup_{h' \in U} \frac{p(u'+th')}{t} \Big] \Big] < \infty \quad ,$$

p is said to be *directionally Lipschitzian* at u with respect to h; Lipschitz continuity of p in a neighborhood of u corresponds to h = 0. It turns out that if (1.8) holds for any h at all, then the set

$$(1.9) \qquad D(u) = \{h \mid p^{\uparrow}(u;h) < \infty\}$$

has a nonempty interior, and for *every* $h \in \operatorname{int} D(\ )$, (1.8) holds and the limits in (1.7) and (1.8) coincide. When the space U is finite-dimensional, this conclusion holds even without the prior assumption that (1.7) is satisfied by at least one h. Note that in these cases where (1.7) and (1.8) give the same value, there is a certain uniformity in the behavior of the difference quotient with respect to the way h is approached, and in fact $p^{\uparrow}(u;h)$ is then continuous locally in h. See [8] for the proofs of these assertions.

For the dual concepts, we need to refer to the space U^* of continuous linear functionals on U; we write $\langle y,h \rangle$ for the pairing of elements $y \in U^*$ and $h \in U$. The *subgradient* set of p at u is

(1.10) $\partial p(u) = \{y \in U^* | \langle y,h \rangle \leq p^{\uparrow}(u,h), \forall h \in U\}$,

and the *singular subgradient* set is

(1.11) $\partial^\circ p(u) = \{y \in U^* | \langle y,h \rangle \leq 0, \forall h \text{ with } p^{\uparrow}(u,h) < \infty\}$.

These are closed convex sets, and the second is obviously the polar of the cone D(u). The basic properties of the subderivative function imply that either $\partial p(u) \neq \emptyset$ and

(1.12) $p^{\uparrow}(u;h) = \sup \{\langle y,h \rangle \mid y \in \partial p(u)\} > \infty, \forall h$,

or $\partial p(u) = \emptyset$ and

(1.13) $$p^{\uparrow}(u,h) = \begin{cases} -\infty \text{ for } y \in D(u) , \\ +\infty \text{ for } y \notin D(u) . \end{cases}$$

The case where $\partial p(u)$ consists of just a single element y corresponds by (1.12) to a strong form of differentiability of p at u with $\nabla p(u) = y$. When p is convex, as is true under (1.4) when F is convex, $\partial p(u)$ is the usual subgradient set of convex analysis. Again we refer the reader elsewhere [1], [8], for the details.

The relationship between $\partial p(u)$ and $\partial^\circ p(u)$ is quite simple. Obviously from (1.10) and (1.11), one has

$$\partial^\circ p(u) = \{y^\circ | y+ty^\circ \in \partial p(u), \forall y \in \partial p(u), t>0\} \text{ if } \partial p(u) \neq \emptyset .$$

Thus the rays in $\partial^\circ p(u)$ represent the "points of $\partial p(u)$ lying at ∞," except that there can be such "points" even when $\partial p(u)$ is empty, as in (1.13). In any event, $\partial^\circ p(u)$ is a sort of measure of the unboundedness of $\partial p(u)$. When the space U is finite-dimensional (which is the case we will mainly be occupied with, although the decision space X will be allowed to remain either finite or infinite-dimensional), $\partial^\circ p(u)$ consists of just the zero vector if and only if $\partial p(u)$ is nonempty and compact (see [6, Prop.3]), and this is in turn equivalent (by the facts cited in the preceding paragraph) to p being finite and Lipschitz continuous on a neighborhood of u. More generally, estimates of $\partial^\circ p(u)$ can provide information about directionally Lipschitzian behavior of p at u.

Estimates of $\partial p(u)$ and $\partial^\circ p(u)$ are both of interest, therefore, in connection with bounds on the subderivatives $p^\uparrow(u;h)$. Outer estimates and corresponding upper bounds will be the theme of the rest of this article.

2. PROXIMAL SUBGRADIENTS AND A SUBDIFFERENTIATION FORMULA

A special technique has recently been developed for analyzing subgradients in the finite-dimensional case. While an infinite-dimensional generalization of some sort may be possible, none has yet been worked out. This technique involves lower quadratic supports to a function, and when applied to the optimal value function p for the nonlinear programming model (1.3), it is intimately connected with the theory of augmented Lagrangians [5]. Although augmented Lagrangian functions have been studied for nonlinear programming problems with infinite-dimensional parameter vectors u (cf. [9]), we shall limit ourselves here, because of the technique in question, to finitely many parameters and *assume henceforth that*

$$(2.1) \qquad U = R^m = U^*, \quad u = (u_1,\ldots,u_m), \quad y = (y_1,\ldots,y_m) \quad .$$

A vector y is called a *proximal subgradient* of p at u (a point where p is finite) if for some $r > 0$ sufficiently large and some $\varepsilon > 0$,

$$(2.2) \qquad p(u') \geqslant p(u) + \langle y, u'-u \rangle - \frac{r}{2}|u'-u|^2 \text{ when } |u'-u| < \varepsilon \ .$$

(Here $|\cdot|$ denotes the Euclidean norm). A condition that can be seen to be equivalent is the following: there is a function q of class C^2 on a neighborhood of u such that $q \leqslant p$, $q(u) = p(u)$, and $\nabla q(u) = y$. Let

(2.3) $\partial^* p(u) = \{y \mid y \text{ is a proximal subgradient at } u\}$.

Working from a result of Clarke [1, p.254] about normal cones to closed sets, we demonstrated in [5] that not only is $\partial^* p(u) \subset \partial p(u)$, but more significantly, the multifunction $\partial^* p$ serves completely to determine ∂p and $\partial^\circ p$ as follows: for the sets

(2.4)
$$Y(u) = \{y \mid \exists\, y^k \to y \text{ with } y^k \in \partial^* p(u^k), u^k \to u, p(u^k) \to p(u)\} ,$$
$$Y_0(u) = \{y \mid \exists\, \lambda_k y^k \to y \text{ with } \lambda_k \downarrow 0, y^k \in \partial^* p(u^k), u^k \to u, p(u^k) \to p(u)\},$$

one has (denoting the closure of a set by "cl" and the convex hull by "co")

(2.5) $$\partial p(u) = \text{cl co } [Y(u) + Y_0(u)], \quad \partial^\circ p(u) = \text{cl co } Y_0(u) .$$

Dual to these expressions there are, by the relations explained in §1, corresponding formulas for $p^\uparrow(u;h)$ and $\text{cl}\, D(u)$, but of particular note is the estimate

(2.6) $$p^\uparrow(u,h) \leq \limsup_{\substack{y^k \in \partial^* p(u^k) \\ u^k \to u \\ p(u^k) \to p(u)}} \langle y^k, h \rangle .$$

This is "tight" in a sense we shall not go into here.

In the results we now state and prove, subgradients and subderivatives of F are used to estimate those of p. The definitions of such things for F are the obvious analogue of those for p and involve the natural pairing of $R^n \times X$ with $R^n \times X^*$.

THEOREM 1.

Assuming (1.5) *and* (2.1), *consider any* u *where* p *is finite, and let*

$$(2.7)\qquad \begin{aligned} M(u) &= \{y \mid \exists x\in X(u) \text{ with } (y,0)\in\partial F(u,x)\}\ , \\ M_0(u) &= \{y \mid \exists x\in X(u) \text{ with } (y,0)\in\partial^\circ F(u,x)\}\ . \end{aligned}$$

Then

$$(2.8)\qquad \partial p(u) \subset \text{cl co}[M(u)+M_0(u)], \quad \partial^\circ p(u) \subset \text{cl co}\, M_0(u)\ .$$

Proof:

Resting our argument on (2.5), we are obliged only to demonstrate that

$$(2.9)\qquad Y(u) \subset M(u), \quad Y_0(u) \subset M_0(u)\ .$$

For the first inclusion, choose any $y\in Y(u)$ and corresponding sequences of elements y^k, u^k, as in the definition (2.4) of $Y(u)$. Since $p(u^k) \to p(u)$, we have (at least for k large enough) that $p(u^k)$ is finite and hence by our blanket assumption (1.5) that $X(u^k) \neq \emptyset$. Choose any $x^k\in X(u^k)$ and recall that (1.5) implies the multifunction X is upper semicontinuous and maps some neighborhood of u into a compact set. From this it can be supposed, passing to a subsequence if necessary, that x^k converges to some $x\in X(u)$. (Without some restriction on the topology of X, we should really at this stage employ the language of nets or filters, rather than speak of sequences and subsequences, but this would affect nothing essential.) For some $r_k > 0$ and $\varepsilon_k > 0$, we have by definition of the relation $y^k \in \partial^* p(u^k)$, the value $p(u^k)$ and set $X(u^k)$, that

$$(2.10)\qquad F(u',x') \geq F(u^k,x^k) + \langle y^k, u'-u^k\rangle - \frac{r_k}{2}|u'-u^k|^2$$
$$\text{for all } x'\in X \text{ and } u'\in R^m \text{ satisfying } |u'-u^k| < \varepsilon_k\ ,$$

where

$$(2.11)\qquad F(u^k,x^k) = p(u^k) \to p(u) = F(u,x)\ .$$

If X were finite-dimensional, we could conclude from (2.10) that $(y^k,0)\in\partial^*F(u^k,x^k)$ and hence in the limit, via the formula for ∂F analogous to (2.5) for ∂p, that $(y,0)\in\partial F(u,x)$ and consequently $y\in M(u)$. The infinite-dimensional case of X requires a more direct approach, however, to establish that $(y,0)\in\partial F(u,x)$.

Suppose it were true that $(y,0)\notin\partial F(u,x)$. Then by the duality between subgradients and subderivatives there would have to exist $(h,w)\in R^m \times X$ with

$$\langle y,h\rangle = \langle (y,0),(h,w)\rangle > F^{\uparrow}(u,x;h,w)$$

$$= \sup_{\substack{U\in N(h)\\ W\in N(w)}} \Big[\limsup_{\substack{(u',x')\to(u,x)\\ F(u',x')\to F(u,x)\\ t\downarrow 0}} \Big[\inf_{\substack{h'\in U\\ w'\in W}} \frac{F(u'+th',x'+tw')-F(u',x')}{t}\Big]\Big]$$

$$\geqslant \sup_{\substack{U\in N(h)\\ W\in N(w)}} \Big[\limsup_{\substack{k\to\infty\\ t\downarrow 0}} \Big[\inf_{\substack{h'\in U\\ w'\in W}} \frac{F(u^k+th',x^k+tw')-F(u^k,x^k)}{t}\Big]\Big] .$$

Then for every $U\in N(h)$ and sequence $t_k\downarrow 0$, we would in particular have (using (2.11) and the definition of p)

$$\langle y,h\rangle > \limsup_{k\to\infty} \Big[\inf_{\substack{h'\in U\\ w'\in X}} \frac{F(u^k+t_kh',x^k+t_kw')-F(u^k,x^k)}{t_k}\Big] \tag{2.12}$$

$$= \limsup_{k\to\infty} \Big[\inf_{h'\in U} \frac{p(u^k+t_kh')-p(u^k)}{t_k}\Big] .$$

But on the other hand we know

$$p(u') \geqslant p(u^k) + \langle y^k,u'-u^k\rangle - \frac{r_k}{2}|u'-u^k|^2 \tag{2.13}$$

when

$$|u'-u^k| < \varepsilon_k .$$

Taking the arbitrary neighborhood U in (2.12) to be of the form

$$(2.14) \qquad U = \{h' \mid |h'-h| \leqslant \varepsilon\} \quad \text{for some } \varepsilon > 0 \quad ,$$

we may select the arbitrary sequence $t_k \downarrow 0$ in such a manner that $t_k r_k \downarrow 0$ and

$$|(u^k+t_k h') - u^k| < \varepsilon_k \text{ when } h' \in U \quad ,$$

as is obviously possible simply by requiring $t_k < \varepsilon_k/\varepsilon$. Then (2.13) implies for $u' = u^k + t_k h'$ that

$$\langle y^k, h'\rangle - \frac{t_k r_k}{2} |h'|^2 \leqslant \frac{p(u^k+t_k h') - p(u^k)}{t_k} \text{ for all } k,$$

or by taking the infimum over both sides subject to $h' \in U$, that

$$(2.15) \qquad \langle y^k, h\rangle - \varepsilon|y^k| - \frac{t_k r_k \varepsilon^2}{2} \leqslant \inf_{h' \in U} \frac{p(u^k+t_k h') - p(u^k)}{t_k}$$

and in the limit

$$\langle y, h\rangle - \varepsilon|y| \leqslant \limsup_{k \to \infty} \left[\inf_{h' \in U} \frac{p(u^k+t_k h') - p(u^k)}{t_k}\right] .$$

Since there was free choice of ε in (2.14), this inequality leads to a contradiction with (2.12). Therefore it is impossible that $(y,0) \notin \partial F(u,x)$, and the proof of the first inclusion in (2.9) is complete.

The proof of the second inclusion is identical in the finite-dimensional case and only a little different when X is infinite-dimensional. In the latter case, the relation $(y,0) \notin \partial^\circ F(u,x)$, which must be proved impossible, means that there exists $(h,w) \in R^m \times X$ with

$$F^\uparrow(u,x;h,w) < \infty \text{ but } \langle (y,0),(h,w)\rangle > 0 \quad .$$

Keeping to the earlier pattern, one can deduce from $F^\uparrow(u,x;h,w) < \infty$ that

$$(2.16) \qquad \limsup_{k\to\infty} \left[\inf_{h'\in U} \frac{p(u^k+t_k h') - p(u^k)}{t_k}\right] < \infty$$

for every $U \in N(h)$ and sequence $t_k \downarrow 0$. On the other hand, through appropriate choice of t_k and U we still have (2.15), and multiplying this through by λ_k (where $\lambda_k \downarrow 0$ and $\lambda_k y^k \to y$ as in the definition of $Y_0(u)$) and taking the limit as $k \to \infty$ we get from (2.16) that

$$\langle y,h\rangle - \varepsilon|y| \leqslant 0 \quad .$$

This being valid for arbitrary $\varepsilon > 0$, we see a contradiction to the starting inequality $\langle (y,0),(h,w)\rangle > 0$, and the second inclusion in (2.9) is thereby confirmed.

COROLLARY 1.

Under the hypothesis of Theorem 1, one also has

$$(2.17) \qquad \partial p(u) \subset \text{cl co } [M(u)+M_0'(u)] \quad ,$$

where

$$M_0'(u) = \{y \mid \exists x \in X(u) \text{ with } \partial F(u,x) = \emptyset \text{ and } (y,0) \in \partial^\circ F(u,x)\} \quad .$$

Proof:
Since $\partial F(u,x) + \partial^\circ F(u,x) = \partial F(u,x)$, all the information represented by $\partial^\circ F(u,x)$ is already embodied in $\partial F(u,x)$ when $\partial F(u,x) \neq \emptyset$. In fact

$$(2.19) \qquad M(u) + M_0(u) = M(u) + M_0'(u) \quad .$$

COROLLARY 2.

Under the hypothesis of Theorem 1, if X(u) *consists of a single element* x *one has*

$$(2.20) \qquad \partial p(u) \subset \{y \mid (y,0) \subset \partial F(u,x)\} \ , \ \partial^\circ p(u) \subset \{y \mid (y,0) \in \partial^\circ F(u,x)\} \quad .$$

Proof. As in Corollary 1, we use the fact that $\partial F(u,x) + \partial^\circ F(u,x) = \partial F(u,x)$. Since $\partial F(u,x)$ and $\partial^\circ F(u,x)$ are closed convex sets, the "cl" and "co" operations can be omitted.

COROLLARY 3.

Under the hypothesis of Theorem 1, if

$$(2.21) \qquad \nexists x \in X(u) \textit{ and } y \neq 0 \textit{ with } (y,0) \in \partial^\circ F(u,x) \quad ,$$

then p *is Lipschitz continuous on a neighborhood of* u.

Proof:

In this case we have $\partial^\circ p(u) = \{0\}$ by the second inclusion in (2.20). Then p is Lipschitzian around u, as explained in §1.

Corollary 3 provides a far more general criterion for Lipschitz continuity than has previously been available. An elementary fact that has long been recognized (e.g., Clarke [10]) is the following: if for some neighborhood U of u there is a set $S \subset X$ such that

$$(2.22) \qquad \begin{cases} \text{the functions } F(\cdot,x) \text{ for } x \in S \text{ are all Lipschitzian} \\ \text{on } U \text{ with respect to a common Lipschitz constant } \lambda \end{cases} ,$$

and

$$(2.23) \qquad X(u') \subset S \text{ for all } u' \in U \quad ,$$

then p is Lipschitzian on U (with the same constant λ). In contrast, Corollary 3 makes no demands on the properties of F over an entire set of the form $U \times S$ but only at the crucial points (u',x) with $x \in X(u)$. Nor does it even require $F(u',x)$ to be Lipschitzian in u' on a neighborhood of u when $x \in X(u)$.

For example, if u' and x are simply real variables and

$$F(u',x) = \begin{cases} 0 \text{ if } u' + x \leq 0 \quad , \\ \infty \text{ if } u' + x > 0 \quad , \end{cases}$$

one has for $(u,x) = (0,0)$ that $\partial^\circ F(0,0) = \{(t,t) \mid t \geq 0\}$. It is true then that there is no $y \neq 0$ with $(y,0) \in \partial^\circ F(0,0)$, yet $F(u',0)$ is not even finite on an entire neighborhood of $u' = 0$, much less Lipschitzian on such a neighborhood.

Although distant from a discussion of "marginal values" in parametric optimization, there is another consequence of Theorem 1 that is well worth recording for the sake of other applications. This concerns the calculation of normal cones, which can be defined as follows: for a closed set H and its indicator function

$$\delta_H(u) = \begin{cases} 0 \text{ if } u \in H \quad , \\ \infty \text{ if } u \notin H \quad , \end{cases}$$

the *normal cone* to H at a point $u \in H$ is

$$N_H(u) = \partial \delta_H(u) \quad (= \partial^\circ \delta_H(u)) \quad .$$

COROLLARY 4. *Let* G *be a nonempty closed subset of* $R^m \times X$, *and let*

$$H = \{u \in R^m \mid \exists x \in X \text{ with } (u,x) \in G\} \quad ,$$

$$X(u) = \{x \in X \mid (u,x) \in G\} \quad .$$

Suppose that for each $\bar{u} \in H$ *there is a neighborhood* U of $\bar{u}$ *and a compact set* $K \subset X$ *such that* $X(u) \subset K$ *for all* $u \in U$. *Then* H *is closed and*

$$N_H(u) \subset \text{cl co} \left\{ \bigcup_{x \in X(u)} N_G(x,u) \right\} \textit{ for all } u \in H \quad .$$

Proof: Simply take F in Theorem 1 to be the indicator δ_G. The compactness condition in the corollary is the corresponding version of (1.5).

COROLLARY 5. *Under the assumptions in Corollary 4, suppose* u *is a point of* C *such that*

$$\exists x \in X(u) \text{ and } y \neq 0 \text{ with } (y,0) \in N_G(u,x) \quad .$$

Then $0 \in \operatorname{int} C$.

Proof:

This specializes Corollary 3 to the case treated in corollary 4.

We now state the dual form of Theorem 1.

THEOREM 2.

Under the hypothesis of Theorem 1, one has

$$(2.24) \qquad p^{\uparrow}(u;h) \leqslant \sup_{x \in X(u)} \inf_{w \in X} F^{\uparrow}(u,x;h,w) \text{ for all } h.$$

Proof:

Suppose first that $\partial p(u) \neq \emptyset$, so that (1.12) is valid. The estimate already obtained for $\partial p(u)$, which we take in the form in Corollary 1 above, then says

$$p^{\uparrow}(u;h) \leqslant \sup \{\langle y+y^{\circ},h\rangle \mid y \in M(u), y^{\circ} \in M'_0(u)\} \quad .$$

Noting that $M'_0(u)$ is closed under multiplication by positive scalars and letting

$$(2.25) \qquad \begin{aligned} A &= \{h \mid \langle y^{\circ},h\rangle \leqslant 0 \text{ for all } y^{\circ} \in M'_0(u)\} \\ &= \bigcap_{\substack{x \in X(u) \\ \partial F(u,x) = \emptyset}} \{h \mid \langle y,h\rangle \leqslant 0 \text{ whenever } (y,0) \in \partial F(u,x)\} \ , \end{aligned}$$

we can translate this into

$$(2.26) \qquad p^{\uparrow}(u;h) \leqslant \sup_{\substack{x \in X(u) \\ \partial F(u,x) \neq \emptyset}} \sup_{(y,0) \in \partial F(u,x)} \langle y,h\rangle \text{ for all } h \in A \quad .$$

The definition of $\partial F(u,x)$ entails that when $(y,0) \in \partial F(u,x)$

we have

$$\langle (y,0),(h,w)\rangle \leq F^{\uparrow}(u,x;h,w) \text{ for all } (h,w) \quad .$$

Therefore

$$(2.27) \qquad \sup_{(y,0)\in\partial F(u,x)} \langle y,h\rangle \leq \inf_{w\in X} F^{\uparrow}(u,x;h,w) \text{ for all } h,$$

an estimate which in due course will be employed in (2.26).

Let us next analyze the set A in (2.25) a bit further. For $x\in X(u)$ with $\partial F(u,x) = \emptyset$ we have by the analogue of (1.13) for F that

$$(2.28) \qquad F^{\uparrow}(u,x;h,w) = \begin{cases} -\infty \text{ for } (h,w)\in E(u,x) \ , \\ +\infty \text{ for } (h,w)\notin E(u,x) \ , \end{cases}$$

where $E(u,x)$ is a certain nonempty convex cone whose polar is $\partial^{\circ}F(u,x)$. The polarity implies that when $(y,0)\in\partial^{\circ}F(u,x)$ we have

$$\langle (y,0),(h,w)\rangle \leq 0 \text{ for all } (h,w)\in E(u,x) \ ,$$

so that

$$\{h \mid \exists w \text{ with } (h,w)\in E(u,x)\} < \{h \mid \langle y,h\rangle \leq 0, \forall (y,0)\in\partial^{\circ}F(u,x)\} \quad .$$

Hence (2.28) yields

$$\inf_{w\in X} F^{\uparrow}(u,x;h,w) = \begin{cases} -\infty \text{ if } \exists w \text{ with } (h,w)\in E(y,x) \ , \\ +\infty \text{ otherwise} \end{cases}$$

$$\geq \begin{cases} -\infty \text{ if } \langle y,h\rangle \leq 0, \forall (y,0)\in\partial^{\circ}F(u,x) \ , \\ +\infty \text{ otherwise} \quad . \end{cases}$$

Consequently

$$\sup_{\substack{x \in X(u) \\ \partial F(u,x)=\emptyset}} \inf_{w \in X} F^{\uparrow}(u,x;h,w) \geqslant \begin{cases} -\infty & \text{if } h \in A \\ +\infty & \text{if } h \notin A \end{cases} .$$

Substituting (2.27) into (2.26) for $x \in X(u)$ with $\partial F(u,x) \neq \emptyset$ and using (2.29), we extract from (2.26) the desired estimate (2.24).

3. THE ROLE OF AUGMENTED LAGRANGIANS

While Theorems 2 and 3 have much to say about the optimal value function p, they do not go far enough in one important respect. They really are first-order results only. The vectors y such that $(y,0) \in \partial F(u,x)$ or $(y,0) \in \partial^{\circ} F(u,x)$ do help characterize the optimality of an $x \in X(u)$, but there may be more of them than are needed or relevant. It would be nice if one could pare the set down by considering second-order properties, for instance, but this is difficult to do directly in the context of the function F even in situations like nonlinear programming. Some of the trouble comes from the fact that F itself may not be the best vehicle for expressing the optimality conditions in question. Often some kind of Lagrangian does the job better.

In nonlinear programming with smooth objective and constraints, second-order estimates of $\partial p(u)$ can be derived by way of the usual (quadratic-type) augmented Lagrangian function; see [7]. What we propose to do here is to trace the general chain of reasoning and demonstrate that in principle, at least, it provides a method of taking higher-order conditions into account in estimates for $\partial p(u)$. As a matter of fact, it may even assist in the discovery of the form those conditions might take.

The first step is the definition of the augmented Lagrangian in the general framework of problem (1.1): for each $u \in R^m$, $x \in X$, $y \in R^m$ and $r > 0$, set

$$(3.1) \qquad L_{u,r}(x,y) = \inf_{u' \in R^m} \{F(u,'x) - \langle y, u'-u \rangle + \frac{r}{2} |u'-u|^2\} .$$

To get very far with this concept of augemented Lagrangian, it would be necessary in a given case to be able to calculate the infimum in closed form. We do not pretend that is easy, although some powerful results in convex analysis can be brought to bear when $F(\cdot,x)$ is convex on R^m for each $x \in X$, say. Yet there are some highly significant situations where the calculation is elementary, and there could be others.

The nonlinear programming model (1.2),(1.3), offers the prime example; note that in that case $F(u,x)$ is indeed convex in u for fixed x, regardless of any nonconvexity of the functions f_i. Formula (3.1) then yields (see [11]):

$$(3.3)\qquad L_{u,r}(x,y) = \begin{cases} f_0(x) + \sum_{i=1}^{s} \phi_i(f_i(x)+u_i,y_i,r) + \sum_{i=s+1}^{m} \psi_i(f_i(x)+u_i,y_i,r) & \text{if } x\in C, \\ +\infty \text{ if } x\notin C, & \end{cases}$$

where

$$\psi_i(f_i(x)+u_i,y_i,r) = y_i\left[f_i(x)+u_i\right] + \frac{r}{2}\left[f_i(x)+u_i\right]^2$$

$$(3.4)\qquad \phi_i(f_i(x)+u_i,y_i,r) = \begin{cases} \psi_i(f_i(x)+u_i,y_i,r) & \text{if } f_i(x)+u_i \geq -y_i/r, \\ -y_i^2/2r & \text{otherwise} \end{cases}$$

The valuable computational and theoretical properties of this function are well known. It is easy to see that $L_{u,r}(x,y)$ is always nondecreasing in r and concave in y. If $F(u,x)$ is convex in u, it can be shown that $L_{u,r}(x,y)$ is continuously differentiable in y, except when u,r,x, are such that it is identically $+\infty$ in y.

A mild assumption will simplify the general discussion that follows:

(3.5) p majorizes some quadratic function on R^m, or equivalently, there exist $\tilde{u} \in R^m$, $\tilde{y} \in R^m$, $\tilde{r} > 0$, such that $\inf_{x \in X} L_{\tilde{u},\tilde{r}}(x,\tilde{y}) > -\infty$

(The equivalence asserted in (3.5) is immediate from the definitions of p and $L_{u,r}$ in terms of F). Clearly (3.5) is quite a mild assumption in situations where only local properties are really at stake, as here. It is satisfied trivially if p is bounded below, i.e., if F is bounded below, and in conjuction with our blanket assumption (1.5) this could always be arranged by some innocuous modification of the values of F(u,x) when $|u|$ is large.

The key to using the augmented Lagrangian in the study of subdifferential properties of p is the following connection with the proximal subgradients of p considered in (2.2),(2.3). Recall that (x,y) is said to be a *saddle point* of $L_{u,r}$ if

(3.6) $$L_{u,r}(x',y) \geq L_{u,r}(x,y) \geq L_{u,r}(x,y') \text{ for all } x' \in X,\ y' \in R^m.$$

Let

(3.7) $$S(u) = \{(x,y) \mid \exists\, r > 0 \text{ with } (x,y) \text{ a saddle point of } L_{u,r}\}.$$

THEOREM 3. *Assume* (1.5),(2.1) *and* (3.5), *and consider any* u *with* p(u) < ∞. *One has*

(3.8) $$(x,y) \in S(u) \Leftrightarrow x \in X(u) \text{ and } y \in \partial^* p(u).$$

Furthermore, the formulas

(3.9) $$\partial p(u) = \text{cl co}[Y(u) + Y_0(u)],\quad \partial^\circ p(u) = \text{cl co } Y_0(u),$$

hold with

$$(3.10)\qquad Y(u) = \{y \mid \exists x \in X(u) \text{ and } (x^k,y^k) \to (x,y) \text{ with } (x^k,y^k) \in S(u^k),\ u^k \to u,\ p(u^k) \to p(u)\},$$

$$(3.11)\qquad Y_0(u) = \{y \mid \exists x \in X(u) \text{ and } (x^k,\lambda_k y^k) \to (x,y) \text{ with } \lambda_k \downarrow 0, (x^k,y^k) \in S(u^k),\ u^k \to u,\ p(u^k) \to p(u)\}.$$

Proof.

Condition (3.5) ensures that when $y \in \partial^* p(u)$, as defined by property (2.2) holding for some r and ε, then simply by choosing r somewhat larger if necessary, one can have the same property globally (see [11]):

$$(3.12)\qquad p(u') \geq p(u) + \langle y, u'-u \rangle - (r/2)|u'-u|^2 \text{ for all } u' \in R^m .$$

Since $p(u) \leq F(u,x)$ for all $x \in X$, and equality holds if and only if $x \in X(u)$, we see that the two conditions $y \in \partial^* p(u)$ and $x \in X(u)$ are equivalent to

$$(3.13)\qquad F(u',x') \geq F(u,x) + \langle y, u'-u \rangle - r/2|u'-u|^2 \quad \text{for all } u' \in R^m,\ x' \in X ,$$

or even better,

$$(3.14)\qquad L_{u,r}(x',y) \geq F(u,x) \text{ for all } x' \in X ,$$

Since on the other hand it is always true from the definition of $L_{u,r}$ that

$$(3.15)\qquad L_{u,r}(x,y') \leq F(u,x) \quad \text{for all } y' \in R^m$$

(take y' in place of y in (3.1) and consider u' = u), we see that (3.14) is equivalent to the saddle point condition $(x,y) \in S(u)$. This proves (3.8).

Already in (2.5) we cited formula (3.9) as valid with Y(u) and $Y_0(u)$ expressed by (2.4), and the job now is to verify that these expressions are equivalent to (3.10) and (3.11). This is easy. For a sequence $u^k \to u$ with $p(u^k) \to p(u) < \infty$, we have for k sufficiently large that $p(u^k) < \infty$ and hence $X(u^k) \neq \emptyset$. By the properties of the multifunction X mentioned in §1 as consequences of assumption (1.5), any sequence of points $x^k \in X(u^k)$ will have a subsequence converging to some $x \in X(u)$. (The same argument was given in the proof of Theorem 1, where it was pointed out that "sequences" should really be replaced by "nets" when X is a general locally convex space.) Thus in considering a sequence of elements $y^k \in \partial^* p(u^k)$, we might just as well be considering a sequence of pairs (x^k, y^k) with $x^k \to x \in X(u)$ and both $y^k \in \partial^* p(u^k)$ and $x^k \in X(u^k)$ holding for all k. The latter conditions mean $(x^k, y^k) \in S(u^k)$, as demonstrated above. Formulas (3.10) and (3.11) therefore define the same sets as the formulas in (2.4), and the proof of Theorem 3 is complete.

Remarks. In comparing the estimate in Theorem 3 with the one in Theorem 1, we need only remember the inclusions $Y(u) \subset M(u)$ and $Y_0(u) \subset M_0(u)$ established in the proof of Theorem 1 to see that Theorem 3 is in every respect sharper. The challenge in applying Theorem 3 is to make use somehow of the properties of the augmented Lagrangian to analyze the limiting saddlepoint condition in (3.10) and (3.11) and thereby get a better grip on the nature of the multiplier vectors in $Y(u)$ and $Y_0(u)$.

In the finite-dimensional nonlinear programming case (3.2) with all functions f_i of class C^2 and no abstract constraint (i.e., $C = R^n$), we have recently used this approach in [7] to show that

$$\partial p(u) \subset \text{cl co}\{ \bigcup_{x \in X(u)} K^2(u,x) + \bigcup_{x \in X(u)} K_0^2(u,x) \} ,$$

(3.16)

$$\partial^\circ p(u) \subset \text{cl co}\{ \bigcup_{x \in X(u)} K_0^2(u,x) \} ,$$

where $K^2(u,x)$ is the set of all $y = (y_1, \ldots, y_m)$ satisfying the first and second-order conditions:

(3.17)

(a) $y_i \geq 0$ and $y_i[f_i(x) + u_i] = 0$ for $i = 1,\ldots s$,

(b) $\nabla f_0(x) + \sum_{i=1}^{m} y_i \nabla f_i(x) = 0$,

(c) $w[\nabla^2 f_0(x) + \sum_{i=1}^{m} y_i \nabla^2 f_i(x)]w \geq 0$ for all w,

such that $\nabla f_i(x) \cdot w = 0$ for all constraint indices i having $f_i(x) + u_i = 0$,

and $K_0^2(u,x)$ is the same thing but without the terms $\nabla f_0(x)$ and $\nabla^2 f_0(x)$ in (b) and (c). By utilizing (3.16) in various ways it was possible in [7] to deduce that if $x \in X(u)$ and the special constraint qualification $K_0^2(u,x) = \{0\}$ is fulfilled, then there exists some $y \in K^2(u,x)$, i.e., the conditions (a),(b),(c) are necessary for optimality.

While we do not, as yet, have other concrete examples where by means of augmented Lagrangians higher-order optimality conditions can be determined and incorporated into estimates of $\partial p(u)$, we can nevertheless sketch the pattern that might be followed in analogy with the second-order nonlinear programming results just described. In order to facilitate this, we shall assume

(3.18) $F(u,x)$ is convex in u for each x.

Let

(3.19) $C = \{x \in X \mid \exists\, u \in R^m \text{ with } F(u,x) < \infty\}$.

Obviously

(3.20) $x \notin C \Rightarrow L_{u,r}(x,y) = +\infty$ for all u,r,y,

but otherwise the definition (3.1) of $L_{u,r}(x,y)$ concerns the minimum of a coercive, strictly convex function of $u' \in R^m$ that is not identically $+\infty$. This minimum is accordingly finite and attained at a unique point, which will be denoted by

(3.21) $\quad v(u,r,x,y) = \arg\min \text{ in } (3.1)$

Then by theorems in convex analysis we have

(3.22) $$x \in C \Rightarrow \begin{cases} L_{u,r}(x,y) \text{ finite for all } u,r,y, \\ \text{with } \nabla_y L_{u,r}(x,y) = v(u,r,x,y) - u \end{cases} .$$

Let us now consider an element $y \in Y(u)$ and try to analyze it further in terms of x^k, y^k and u^k, as in (3.10). The condition $(x^k, y^k) \in S(u^k)$ in (3.10) means that for $r_k > 0$ sufficiently large one has

(3.23) $$L_{u^k,r_k}(x^k,y^k) \geq L_{u^k,r_k}(x^k,y') \quad \text{for all } x' \in X,\ y' \in R^m .$$

As we know, this entails $x^k \in X(u^k)$, so that

(3.24) $$L_{u^k,r_k}(x^k,y^k) = F(u^k,x^k) = p(u^k) < \infty .$$

Moreover the second inequality in (3.23) can be written simply as

(3.25) $$0 = \nabla_y L_{u^k,r_k}(x^k,y^k) = v(u^k,r_k,x^k,y^k) - u^k .$$

As for the first inequality in (3.23), we observe it implies for arbitrary $w \in X, w^k \to w, t_k \downarrow 0$, that

(3.26) $$L_{u^k,r_k}(x^k + t_k w^k, y^k) - L_{u^k,r_k}(x^k,y^k) \geq 0 \quad \text{for all } k .$$

For $q = 1,2,\ldots,$ and arbitrary $w \in X$ let us define the following expression which is suggested by (3.26) but independent of the particular sequence of elements u^k, x^k, y^k and r_k that might be available:

$$(3.27)\qquad \Lambda^q(u,x,y;w) = \sup_{\substack{W\in N(w)\\ \delta>0}} \Big[\limsup_{\substack{(u',x',y')\to(u,x,y)\\ F(u',x')\to F(u,x)\\ v(u',r,x',y')=u'\\ r\to\infty}} \Big[\inf_{\substack{w'\in W\\ 0<t<\delta}} \frac{L_{u',r}(x'+tw',y')-L_{u',r}(x',y')}{t^q}\Big]\Big] .$$

Then for x and y as in the definition (3.10) of Y(u) we may conclude from the foregoing that

$$(3.28)\qquad \Lambda^q(u,x,y;w) \geq 0 \quad \text{for all } w\in X .$$

Note that $\Lambda^q(u,x,y;w)$ is also positively homogeneous of degree q with respect to w:

$$(3.29)\qquad \Lambda^q(u,x,y;\lambda w) = \lambda^q\Lambda^q(u,x,y;w) \quad \text{for all } \lambda>0 .$$

We interpret (3.28) as an *abstract* q*th-order optimality condition* associated with the optimal solution x and multiplier vector y for problem (1.1). This designation is supported by the fact that the nonlinear programming result cited above is based on a demonstration that for $x\in X(u)$,

$$[\Lambda^1(u,x,y;w) \geq 0,\ \forall w\in W] \Rightarrow y\in K^1(u,x),$$

$$[\Lambda^1(u,x,y;w) \geq 0 \text{ and } \Lambda^2(u,x,y;w) \geq 0, \forall w\in W] \Rightarrow y\in K^2(u,x),$$

where $K^1(u,x)$ consists of the vectors y satisfying (3.17) (a) (b), and $K^2(u,x)$ consists as before of the ones satisfying (3.17) (a) (b) (c).

With such motivation we can introduce for $q=1,2,\ldots,$ the multiplier set

$$(3.30)\qquad Y^q(u) = \{u \mid \exists\ x\in X(u) \text{ with } \Lambda^j(u,x,y;w) \geq 0,\ \forall w,\ j=1,\ldots q\} .$$

The conclusion is then the following.

THEOREM 4.

Assume (1.5), (2.1), (3.5) *and* (3.18), *and consider any* u *with* $p(u) < \infty$. *One has*

$$(3.32)\qquad M(u) \supset Y^1(u) \supset Y^2(u) \supset \ldots \supset Y(u) \quad ,$$

$$(3.33)\qquad M_0(u) \supset Y_0^1(u) \supset Y_0^2(u) \supset \ldots \supset Y_0(u) \quad ,$$

and for $q = 1,2,\ldots,$

$$(3.34)\qquad \partial p(u) \subset \mathrm{cl\ co}\,[Y^q(u) + Y_0^q(u)], \qquad \partial^\circ p(u) \subset \mathrm{cl\ co}\, Y_0^q(u) \quad .$$

Proof.

Most of the demonstration has been built up in the prologue to the theorem, so that only the estimate in Theorem 3 needs to be applied to get (3.34). One feature has not been dealt with, however, and that is the initial inclusions in (3.32) and (3.33). Without these, it would not be possible to claim that (3.34) is any sharper an estimate than the one in Theorem 1.

The initial inclusion in (3.32) can be verified by fixing any u,x,y, such that $\Lambda^1(u,x,y;w) \geq 0$ for all $w \in X$ and showing that $(y,0) \in \partial F(u,x)$, or in other words that

$$(3.35)\qquad \langle y,h \rangle \leq F^{\uparrow}(u,x;h,w) \quad \text{for all } h \in R^m,\ w \in X \ ,$$

where

$$(3.36)\qquad F^{\uparrow}(u,x;h,w) = \sup_{\substack{U \in N(h) \\ W \in N(w)}} \Big[\limsup_{\substack{(u',x') \to (u,x) \\ F(u',x') \to F(u,x) \\ t \downarrow 0}} \Big[\inf_{\substack{h' \in U \\ w' \in W}} \frac{F(u'+th',x'+tw') - F(u',x')}{t} \Big]\Big] \ .$$

Turning to the formula (3.27) for $\Lambda^1(u,x,y;w)$, we recall the meaning (3.21) of the requirement $v(u',r,x',y') = u'$ and note that it implies

$$L_{u',r}(x',y') = F(u',x') \quad .$$

At the same time we have by definition (3.1) that

$$L_{u',r}(x'+tw',y') \le F(u'',x'+tw') - \langle y',u''-u'\rangle + (r/2)\,|u''-u'|^2, \quad \forall u'',$$

or by writing $u'' = u'+th'$,

$$(3.38) \qquad L_{u',r}(x'+tw',y') \le F(u'+th',x'+tw') - \langle y',h'\rangle + (r/2)\,|h'|^2, \quad \forall h'.$$

From (3.37) and (3.38) we obtain the following estimate for eventual application to (3.27) for $q = 1$ (here U and W denote neighborhoods of h and w, and we assume U is bounded):

$$(3.39) \qquad \limsup_{\substack{(u',x',y')\to(u,x,y)\\ F(u',x')\to F(u,x)\\ v(u',r,x',y')=u'\\ r\to\infty}} \Big[\inf_{\substack{w'\in W\\ 0<t<\delta}} \frac{L_{u',r}(x'+tw',y') - L_{u',r}(x',y')}{t}\Big]$$

$$\le \limsup_{\substack{(u',x',y')\to(u,x,y)\\ F(u',x')\to F(u,x)\\ r\to\infty}} \Big[\inf_{\substack{h'\in U\\ w'\in W\\ 0<t<\delta}} \Big[\frac{F(u'+th',x'+tw') - F(u',x')}{t} - \langle y',h'\rangle + \frac{rt|h'|^2}{2}\Big]\Big]$$

$$\le \limsup_{\substack{(u',x',y')\to(u,x,y)\\ F(u',x')\to F(u,x)\\ t\downarrow 0}} \Big[\inf_{\substack{h'\in U\\ w'\in W}} \Big[\frac{F(u'+th',x'+tw') - F(u',x')}{t} - \langle y',h'\rangle\Big]\Big]$$

where the last step is justified by the fact that no matter how $r\to\infty$ in the "lim sup", corresponding values of $t\downarrow 0$ could be chosen so that $rt|h'|^2\to 0$ uniformly in $h'\in U$ (because U is bounded).

Since (3.39) holds for any bounded $U\in N(h)$, and the infimum over $h'\in U$ increases if anything as U diminishes, we can take the supremum in U (the same as the limit as U shrinks to $\{h\}$) and see that the term $\langle y',h'\rangle$ at the end of (3.39) must wind up as $\langle y,h\rangle$. Thus to the chain of inequalities already generated we can add:

$$\leq \sup_{U \in N(h)} \Big[\limsup_{\substack{(u',x') \to (u,x) \\ F(u',x') \to F(u,x) \\ t \downarrow 0}} \Big[\inf_{\substack{h' \in U \\ w' \in W}} \frac{F(u'+th',x'+tw') - F(u',x')}{t} \Big]\Big] - \langle y,h \rangle .$$

This is then an upper bound for the first expression in (3.39). When it is invoked in the definition (3.27) of the condition $\Lambda^1(u,x,y;w) \geq 0$, we obtain by formula 3.36 the desired inequality (3.35).

The proof of the first inclusion in (3.33) is parallel, and with this the proof of Theorem 4 is complete.

REFERENCES

1. Clarke, F.H. 1975 "Generalized gradients and applications", Trans. Amer. Math. Soc. 205, 247-262.

2. Hiriart-Urruty, J.B. 1978 "Gradients généralisés de fonctions marginales", SIAM J. Cont. Opt., 16, 301-316.

3. Gauvin, J. 1979. "The generalized gradient of a marginal function in Mathematical programming", Math. of Op. res. 4,458-463.

4. Gauvin, J. and F. Dubeau, (forthcoming) "Differential properties of the marginal function in mathematical programming", Math. Prog. Studies (M. Guignard, ed.).

5. Rockafellar, R.T. (forthcoming) "Proximal subgradients, marginal values and augmented Lagrangians in nonconvex optimization", Math. of Op. Res.,

6. Rockafellar, R.T. (forthcoming) "Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming", Math. Prog. Studies (R. Wets ed.) .

7. Rockafellar, R.T. (submitted) "Marginal values and second-order necessary conditions for optimality", Math. Prog.

8. Rockafellar, R.T. 1980 "Generalized directional derivatives and subgradients of nonconvex functions", Canad. J. Math. 32, 257-280.

9. Wierzbicki, A.P., and S. Kurcyusz, 1977. "Projection on a cone: generalized penalty functionals and duality theory for problems with inequality constraints in Hilbert space", SIAM J. Control Opt. 15, 25-56.

10. Clarke, F.H. 1980 "Generalized subgradients of Lipschitz functionals", Advances in Math.

11. Rockafellar, R.T. 1974 "Augmented Lagrange multiplier functions and duality in nonconvex programming", SIAM J. Control 12, 268-285.