90

Mathematical Methods in Operations Research.Sofia,1981,p.103-110

GENERALIZED SUBGRADIENTS IN NONCONVEX PROGRAMMING

R.T. Rockafellar*

Optimization problems typically involve a number of parameters besides the variables over which the optimization takes place. The role of these parameters can be important not only in applications where the parameters can take on different values, but also in theoretical analysis and computation.

To take a central case which is actually of far greater generality than it might seem, let us consider for each $u=(u_1,\ldots,u_m)\in R^m$ the problem

$$(P_u)\qquad \text{minimize } f_0(x) \text{ subject to } f_i(x)+u_i\le 0 \text{ for } i=1,\ldots,m.$$

Assume for simplicity that the functions f_i are all of class C^2 on R^n, and that the set

$$(1)\qquad \{x\in R^n \mid f_i(x)\le\alpha_i \text{ for } i=0,1,\ldots,m\}$$

is bounded for every choice of constants α_i. Let

$$(2)\qquad \begin{array}{ll} p(u)=\inf(P_u) & \text{(optimal value)} \\ X(u)=\operatorname{argmin}(P_u) & \text{(optimal solutions).} \end{array}$$

What can be said about the way $p(u)$ and $X(u)$ depend on the parameter vector u?

This question turns out to be fundamental in understanding the nature of optimality conditions in (P_u), among other things. The awkward fact is that despite the smoothness of the functions f_i, the function p can well fail to be differentiable in the ordinary sense, or even continuous. About all that can be concluded immediately from our assumptions is that p is lower semicontinuous from R^m to $R\cup\{\infty\}$, and $X(u)$ is nonempty and compact for every u such that $p(u)<\infty$. (We interpret the infimum in (P_u) as ∞ when there are no feasible solutions.)

Nevertheless, p cannot be totally lacking in differentiability properties of some sort or another, because these are closely tied in with the nature of Lagrange multipliers for (P_u), at least in special cases. For each $x\in X(u)$ let

$$(3)\qquad I(u,x)=\{i>0 \mid f_i(x)+u_i=0\} \qquad \text{(active constraint indices),}$$

---

*Research supported in part by AFOSR grant number 77-3204 at the University of Washington, Seattle.

(4) $K^1(u,x)$ = set of $y=(y_1,\ldots,y_m)\in R^m$ satisfying

$$y_i \begin{cases} \geq 0 & \text{for } i\in I(u,x), \\ =0 & \text{for } i\notin I(u,x), \end{cases}$$

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

(5) $K_0^1(u,x)$ = same but with the term $\nabla f_0(x)$ omitted.

The elements of $K^1(u,x)$ are the first-order Lagrange multiplier vectors for (P_u) at x. The condition $K_0^1(u,x)=\{0\}$ is (an equivalent form of) the Mangasarian-Fromovitz constraint qualification [1], and as is well known, it is equivalent to $K^1(u,x)$ being both nonempty and compact. It is satisfied in particular, therefore, when $K^1(u,x)$ consists of a unique vector y.

The following kind of result in this context is classical (cf. [2]): if x is the unique element of $X(u)$, y is the unique element of $K^1(u,x)$, and certain second-order conditions of a strong kind are satisfied, then p is differentiable on a neighborhood of u with

(6) $$\nabla p(u) = y.$$

In particular, the directional derivatives

(7) $$p'(u;h) = \lim_{t\downarrow 0} \frac{p(u+th)-p(u)}{t}$$

exist and satisfy $p'(u;h) = y\cdot h$ for all $h\in R^m$.

In the convex programming case of (P_u), where every f_i is a convex function, there are other strong indications of a connection between Lagrange multiplier vectors y and differentiability properties of p. In this case p is a convex function; its epigraph

(8) $$E = \{(u,\alpha)\in R^m\times R \mid \alpha \geq p(u)\}$$

is a convex set (closed, because p is lower semicontinuous). The sets of subgradients and singular subgradients of p at u, a point where p is finite, are then given by

(9)
$$\partial p(u) = \{y\in R^m \mid (y,-1)\in N_E(u,p(u))\},$$
$$\partial^0 p(u) = \{y\in R^m \mid (y,0)\in N_E(u,p(u))\},$$

where $N_E(u,p(u))$ is the normal cone to E at the point $(u,p(u))$ in the sense of convex analysis [3].

THEOREM 1. *In the convex programming case, one has for arbitrary* $x\in X(u)$ *that*

$$\partial p(u) = K^1(u,x), \qquad \partial^0 p(u) = K_0(u,x).$$

This is immediate from standard results (cf. [3]). It shows that at least in the convex programming case (generalizations will come), one has $\partial p(u)$ nonempty and compact if and only if $\partial^0 p(u) = \{0\}$. Under such circumstances one has furthermore that

$$p'(u;h) = \max_{y \in \partial p(u)} y \cdot h = \max_{y \in K^1(u,x)} y \cdot h \tag{10}$$

by formulas of convex analysis [3,§23]. Thus here again is a case where Lagrange multipliers associated with optimal solutions to (P_u) tell us something about rates of change of p, and vice versa.

Can such results be extended in some form to nonconvex programming? This seems essential, if we are to understand and take full advantage of the meaning of Lagrange multipliers in the general case.

Clarke [4] has provided a concept of generalized subgradient that makes such an extension possible. He has defined the concept of the normal cone to an arbitrary closed set in Euclidean space in a robust way which agrees with previous definitions of normality for convex sets and for smooth manifolds. Taking $N_E(u,p(u))$ in (9) to be the normal cone in Clarke's sense, we get his subgradient set $\partial p(u)$ and the corresponding singular subgradient set $\partial^0 p(u)$. These are well defined closed convex sets, whether or not p happens to be a convex function; only the lower semicontinuity of p is utilized. They agree with the subgradient sets of convex analysis when p is convex.

An alternative but equivalent description of the generalized subgradient sets of Clarke is the following (cf. [5, 4S and 4T], [6]). Call y a <u>lower semigradient</u> of p at u if

$$p(u') \geq p(u) + y \cdot (u'-u) + o(|u'-u|). \tag{11}$$

Then, in an extended sense, $\partial p(u)$ is the closed convex hull of all the limit points y of sequences $\{y^k\}$ such that y^k is a lower semigradient at some point u^k, and both $u^k \to u$ and $p(u^k) \to p(u)$. The "extended sense" refers to the need for allowing also certain "direction points" of R^m as possible limits of sequences $\{y^k\}$ that are unbounded, and to take these "direction points" into account when forming the convex hull. (See [3, §8 and §17] for more about such matters.) The cone $\partial^0 p(u)$ is the closed convex hull of the rays in R^m corresponding to the "direction points" realized as limits.

THEOREM 2 (Rocakfellar [7]). <u>The following conditions are equivalent at any</u> u <u>where</u> $p(u) < \infty$:

(a) $\partial^\circ p(u) = \{0\}$;

(b) $\partial p(u)$ <u>is nonempty and compact</u>;

(c) p <u>is Lipschitzian on a neighborhood of</u> u.

In the case described in the theorem, there is a simple relationship between $\partial p(u)$ and certain nonclassical derivatives of p introduced by Clarke [4]:

$$(11) \qquad \limsup_{\substack{u' \to u \\ t \downarrow 0}} \frac{p(u'+th) - p(u')}{t} = \max_{y \in \partial p(u)} y \cdot h.$$

In particular, $\partial p(u)$ consists of a unique y if and only if

$$(12) \qquad \lim_{\substack{u' \to u \\ t \downarrow 0}} \frac{p(u'+th) - p(u')}{t} = y \cdot h \quad \text{for all } h \in R^m.$$

The latter means that p is *strongly differentiable* at u and $\nabla p(u) = y$; it holds certainly if p is of class C^1 on a neighborhood of u.

Even when p is not Lipschitzian as in Theorem 2, there is a close relationship between $\partial p(u)$ and certain special *subderivatives* of p, defined by

$$(13) \qquad p^{\uparrow}(u;h) = \lim_{\varepsilon \downarrow 0} \Big[\limsup_{\substack{u' \to u \\ p(u') \to p(u) \\ t \downarrow 0}} \Big[\inf_{|h'-h| \le \varepsilon} \frac{p(u'+th') - p(u')}{t} \Big] \Big].$$

Difficult as this formula may seem, it has some surprising properties which we demonstrated in [8]: $p^{\uparrow}(u;h)$ is always convex as a function of h; one has $\partial p(u) = \emptyset$ if and only if $p^{\uparrow}(u;0) = -\infty$, while otherwise

$$(14) \qquad p^{\uparrow}(u;h) = \sup_{y \in \partial p(u)} y \cdot h \quad \text{for all } h \in R^m,$$

$$\partial p(u) = \{ y \in R^m \mid y \cdot h \le p^{\uparrow}(u;h) \quad \text{for all } h \}.$$

The subderivatives (13) reduce to the ones in (11) in the Lipschitzian case in Theorem 2.

Armed with this new concept, let us return now to the discussion of Lagrange multiplier vectors for (P_u) and their relationship to p. The first person to get results on this topic was Clarke [9], who showed that $K^1(u,x) \neq \emptyset$ if (P_u) is <u>globally calm</u> in the sense that

$$\liminf_{\substack{h' \to h \\ t \downarrow 0}} \frac{p(u+th') - p(u)}{t} > -\infty \text{ for all } h. \tag{15}$$

Generalized subgradients of p in the nonconvex case were first estimated by Gauvin [10].

THEOREM 3 (Gauvin [10]). *Suppose u is such that $p(u) < \infty$ and the constraint qualification $K_0^1(u,x) = \{0\}$ is satisfied for every $x \in X(u)$. Then p is Lipschitz continuous on a neighborhood of u, and*

$$\partial p(u) \subset \mathrm{cl\,co}\{ \bigcup_{x \in X(u)} K^1(u,x) \}. \tag{16}$$

Recently we have extended this result as follows.

THEOREM 4 (Rockafellar [11]). *Suppose u is such that $p(u) < \infty$. Then*

$$\partial p(u) \subset \mathrm{cl\,co}\{ \bigcup_{x \in X(u)} K^1(u,x) + \bigcup_{x \in X(u)} K_0^1(u,x) \}, \tag{17}$$

$$\partial^\circ p(u) \subset \mathrm{cl\,co}\{ \bigcup_{x \in X(u)} K_0^1(u,x) \}.$$

This yields Gauvin's theorem by way of Theorem 2. The cited result of [11] is actually much more powerful and general than the formulation given here would indicate. It deals with multipliers for (P_u) even when the functions f_i are not of class C^2, just locally Lipschitzian. An abstract constraint $x \in D(u)$ can also be present. (The multiplier theorems of Clarke and Gauvin mentioned above also apply to some of these broader situations.)

Rather than discuss such generalizations here, we shall describe next some results in the opposite direction. These take positive advantage of the limitations we have placed here on (P_u). For $x \in X(u)$, define

$$W(u,x) = \{ w \in R^n \mid \nabla f_i(x) \cdot w = 0 \text{ for all } i \in I(u,x) \},$$

$K^2(u,x)$ = set of all $y \in K^1(u,x)$ satisfying

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

$K_0^2(u,x)$ = set of all $y \in K_0^1(u,x)$ satisfying

$$w \cdot [\sum_{i=1}^m y_i \nabla^2 f_i(x)] w \geq 0 \text{ for all } w \in W(u,x).$$

The set $K^2(u,x)$ gives what may be called second-order Lagrange multiplier vectors for (P_u), although other conditions have also been proposed in this connection.

THEOREM 5 (Rockafellar [12]). Suppose u is such that $p(u) < \infty$. Then

$$(18) \qquad \partial p(u) \subset \operatorname{cl\,co}\{ \bigcup_{x \in X(u)} K^2(u,x) + \bigcup_{x \in X(u)} K_0^2(u,x)\},$$

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

The difference between this and Theorem 4, of course, is that the sets $K^2(u,x)$ and $K_0^2(u,x)$ are likely to be smaller than $K^1(u,x)$ and $K_0^1(u,x)$, so the estimates in (18) are sharper than the ones in (17). As a corollary of Theorem 5 via Theorem 2, we get the following improvement over Gauvin's theorem.

THEOREM 6 (Rockafellar [12]). Suppose u is such that $p(u) < \infty$ and the constraint qualification $K_0^2(u,x) = \{0\}$ is satisfied for every $x \in X(u)$. Then p is Lipschitz continuous on a neighborhood of u, and

$$(19) \qquad \partial p(u) \subset \operatorname{cl\,co}\{ \bigcup_{x \in X(u)} K^2(u,x)\}.$$

Another result which can be derived from Theorem 5, although by a trickier method, is a new multiplier rule. We state it in terms of yet another kind of constraint qualification, which localizes the calmness condition of Clarke already described: we define (P_u) to be locally calm at x if there do not exist sequences $x^k \to x$, $u^k \to u$, such that x^k is feasible for (P_{u^k}) and

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

This is true for all $x \in X(u)$ when (P_u) is globally calm in Clarke's sense (see [11, Prop. 12]).

THEOREM 7 (Rockafellar [12]). Let x be a locally optimal solution to (P_u), and suppose either that (P_u) is locally calm at x or $K_0^2(u,x) = \{0\}$. Then $K^2(u,x) \neq \emptyset$.

This result differs in character from previous theorems on second-order necessary conditions, and of course it has a totally different sort of proof, since the theory of generalized subgradients of nonconvex, nondifferentiable, functions is applied to p. The corresponding "standard" result would involve a constraint qualification in terms of the existence of certain kinds of arcs leading from x into the feasible set for (P_u). Putting this qualification in the form of a condition which ensures that the implicit function theorem can be used to obtain the existence of such arcs, we would have the following. _Let_ x _be a locally optimal solution to_ (P_u), _and suppose that_ $K_0^1(u,x) = \{0\}$ _and the set_ $K^1(u,x)$ _consists of a unique_ y. _Then_ $y \in K^2(u,x)$, _and_ y _has some other second-order properties too._

But in fact the assumptions in this "standard" result imply that $K_0^2(u,x) = \{0\}$ and $K^2(u,x) = K^1(u,x) = \{y\}$. Therefore, we have merely the description of a special case of Theorem 7 where the properties defining $K^2(u,x)$ automatically entail certain other properties. Viewed in this light, Theorem 7 is seen to extend the theory of second-order necessary conditions quite substantially.

All of the results discussed in this article apply also to problems having equality constraints, with the obvious modifications. We have avoided equality constraints only in order to keep the notation simpler.

References

1. O.L. Mangasarian and S. Fromovitz, "The Fritz John necessary optimality conditions in the presence of equality and inequality constraints," J. Math. Anal. Appl. 17(1967), 37-47.

2. S.M. Robinson, "Stability theory for inequalities, part II: differentiable nonlinear systems," SIAM j. Numer. Anal. 13(1976), 497-513.

3. R.T. Rockafellar, Convex Analysis, Princeton Univ. Press, Princeton, N.J., 1970.

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

5. R. T. Rockafellar, The Theory of Subgradients and Its Applications to Problems of Optimization: Convex and Nonconvex Functions, Heldermann Verlag (Herderstr. 6-7, D-100 Berlin 41), 1981.

6. R. T. Rockafellar, "Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization," Math. of Op. Res. 6(1981).

7. R. T. Rockafellar, "Clarke's tangent cones and the boundaries of closed sets in R^n, Nonlin. Analysis. Th. Math. Appl. 3(1979), 145-154.

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

9. F.H. Clarke, "A new approach to Lagrange multipliers," Math. of Op. Res. 1(1976), 165-174.

10. J. Gauvin, "The generalized gradient of a marginal function in mathematical programming," Math. of Op. Res. 4(1979), 458-463.

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

12. R. T. Rockafellar, "Marginal values and second-order nedessary conditions for optimality," Math. Prog., to appear.