SIAM-AMS Proceedings
Volume 9
1976

LAGRANGE MULTIPLIERS IN OPTIMIZATION

R. Tyrrell Rockafellar[1]

The theory of Lagrange multipliers began with classical studies of minimization subject to constraints. Until the late 1940's, interest was centered on deriving necessary or sufficient conditions to characterize a local minimum point in terms of differential properties at the point. The highest attainments in this direction are exemplified by results of the Chicago school, which were aimed at applications to the calculus of variations. The book of Hestenes [2] provides an illuminating exposition. Although the emphasis traditionally was on equality constraints, inequalities were also studied, and there was an increasing use and development of notions of convex analysis.

The 1950's brought a realization of the immense importance of inequality constraints in economic and industrial problems. Computational possibilities opened up and were greatly stimulated by the success of Dantzig's simplex method, an algorithm making very significant use of Lagrange multipliers. Probably the most notable development for the theory of Lagrange multipliers, however, was the connection made between the Lagrangian function and von Neumann's minimax principle for two-person games.

Kuhn and Tucker showed that for problems of convex type an optimal solution $\bar{x}$ and its associated multiplier vector $\bar{y}$ constitute a minimax saddle point $(\bar{x},\bar{y})$ of the Lagrangian function. This provided a simple <u>global</u> characterization of optimality which proved to be very fruitful, conceptually and computationally. It focused much attention on investigating the role of convexity in optimization and possible ways of exploiting duality.

Efforts have continued up to the present on refining the classical theory of local necessary and sufficient conditions. This has been despite

---

AMS(MOS) subject classifications (1970). Primary 90C25, 90C30, 49B30.

[1]Research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under AFOSR Grant number 72-2269.

the often impractical or unverifiable character of such conditions when there is no simple way of reducing the candidates for a minimum point to a few explicit nominees. Perhaps the reason is that the minimax approach to optimality, however desirable, has seemed limited to the smaller class of problems where "everything is convex".

While such is largely true for a minimax approach based on the ordinary Lagrangian, it is not necessarily valid for the infinite variety of other Lagrangian functions that now can be generated for a problem using the theory of conjugate duality. As a matter of fact, a global saddle point characterization of optimality has recently been developed for nonconvex programming problems in terms of a certain "augmented" Lagrangian (see [7], [8] and the references given there). As one might expect, this characterization is closely tied in with new methods of calculation, in particular to certain algorithms ostensibly better than standard penalty techniques.

What is especially important, however, is the possibility that now presents itself of extending the minimax approach to other general types of problems, such as in optimal control or stochastic programming. This would provide a much more satisfactory theory of optimality conditions for modern purposes.

Our aim here is to outline the elements that point towards this development. We start with an overview of the main necessary and sufficient conditions in nonlinear programming and their significance. (More details are given in the paper of McCormick in this volume.) An abstract minimization problem is then introduced. It is explained how the theory of conjugate duality in convex analysis leads to the construction of a broad class of generalized Lagrangian functions and the attempt to characterize the optimality of a point $\bar{x}$ by showing it corresponds to a minimax saddle point $(\bar{x},\bar{y})$ of such a Lagrangian, where $\bar{y}$ is a generalized multiplier vector. (A fuller exposition is contained in the recently published lecture notes [6].) Finally, we discuss the augmented Lagrangian in nonlinear programming and why it indicates the desirability of adopting this methodology even in nonconvex optimization.

1. CLASSICAL THEORY: EQUALITY CONSTRAINTS

Let us consider first a nonlinear programming problem of the form

(P_0) minimize $f_0(x)$ subject to $f_1(x) = 0,\ldots,f_m(x) = 0$,

where each $f_i : R^n \to R$ for $i = 1,\ldots,m$ is of class C^2. The set

(1) $$M = \{x \in R^n \mid f_1(x) = 0,\ldots,f_m(x) = 0\}$$

is the feasible "manifold". It may be hoped that M is in fact a differentiable manifold of dimension $n-m$ in the natural way, but of course the situation could be more complicated without further restrictions on the constraint functions.

Let $\bar{x}$ be a local solution to (P_0), i.e. a point where f_0 has a local minimum relative to M. Let

(2) $N =$ subspace of R^n generated by $\nabla f_1(\bar{x}),\ldots,\nabla f_m(\bar{x})$,

(3) $$T = N^{\perp} = \{v \in R^n \mid \nabla f_i(\bar{x})\cdot v = 0 \text{ for } i = 1,\ldots,m\},$$

where $\nabla f_i(\bar{x})$ denotes the gradient of f_i at $\bar{x}$. Heuristically, each $\nabla f_i(\bar{x})$ is normal to the hypersurface $f_i(x) = 0$ at $\bar{x}$, so that N should be the "normal space" to M at $\bar{x}$ and T the "tangent space". However, the tangent space can also be defined more directly as the set

(4) $$T' = \limsup_{t\downarrow 0} t^{-1}[M-\bar{x}]=\{v \in R^n \mid \exists t_k \downarrow 0,\ v_k \to v, \text{ with } \bar{x}+t_k v_k \in M\}.$$

Trivially $T' \subset T$. Following Hestenes [2], we say that $\bar{x}$ is a regular point of M if $T' = T$. This is easily seen to be true if

(5) $$\nabla f_1(\bar{x}),\ldots,\nabla f_m(\bar{x}) \text{ are independent.}$$

(Note that regularity is not a property of M itself, but of its representation in terms of the functions f_i.) At all events, if the optimal solution $\bar{x}$ is a regular point we must have

$$\nabla f_0(\bar{x})\cdot v = 0 \text{ for all } v \in T,$$

so that $\nabla f_0(\bar{x})$ must belong to $T^{\perp} = N$.

We thus obtain the following first-order necessary conditions for $\bar{x}$ to be an optimal solution to (P_0), assuming $\bar{x}$ is a regular point:

(6) $$f_1(\bar{x}) = 0,\ldots,f_m(\bar{x}) = 0, \text{ and } \exists\, \bar{y} = (\bar{y}_1,\ldots,\bar{y}_m) \in R^m \text{ such that } \nabla f_0(\bar{x}) + \bar{y}_1\nabla f_1(\bar{x})+\ldots+\bar{y}_m\nabla f_m(\bar{x}) = 0 .$$

These conditions have a simple representation in terms of the function

(7) $$L(x,y) = f_0(x)+y_1f_1(x)+\ldots+y_mf_m(x) \quad \text{for all } x \in R^n,\ y \in R^m ,$$

which is called the ordinary <u>Lagrangian</u> associated with problem (P_0). Namely, (6) is equivalent to

(8) $$\exists\, \bar{y} \text{ such that } \nabla_x L(\bar{x},\bar{y}) = 0 \text{ and } \nabla_y L(\bar{x},\bar{y}) = 0.$$

Observe that the same conditions can be obtained assuming merely that $\bar{x}$ was a "regular stationary point" of f_0 relative to M.

This result is interesting, because it shows that a constrained stationary point of f_0 corresponds to an unconstrained stationary point of L. Thus constraints can be eliminated at the expense of introducing additional variables y_i; these variables are called <u>Lagrange multipliers</u>. However, it must not be thought that a constrained <u>minimum</u> of f_0 typically corresponds to an unconstrained <u>minimum</u> of L. The stationary point of L may not even involve a local minimum of $L(x,\bar{y})$ with respect to x.

Second-order conditions for optimality in (P_0) can be derived using the Hessian (second derivative) matrix

(9) $$\nabla_x^2 L(\bar{x},\bar{y}) = \nabla^2 f_0(\bar{x}) + \bar{y}_1\nabla^2 f_1(\bar{x})+\ldots+\bar{y}_m\nabla^2 f_m(\bar{x}).$$

Two properties of importance are

(10) $$v\cdot\nabla_x^2 L(\bar{x},\bar{y})v \geq 0 \quad \text{for all } v \in T,$$

and

(10') $$v\cdot\nabla_x^2 L(\bar{x},\bar{y})v > 0 \quad \text{for all nonzero } v \in T.$$

The main result on necessary and sufficient conditions can be stated in these terms.

THEOREM 1. (<u>local optimality</u>)

(a) (<u>necessary conditions</u>). <u>If</u> $\bar{x}$ <u>is a local solution to</u> (P) <u>which is also a regular point of</u> M, <u>then</u> (8) <u>and</u> (10) <u>hold</u>.

(b) <u>sufficient conditions</u>). <u>If</u> (8) <u>and</u> (10') <u>hold, then</u> $\bar{x}$ <u>is an isolated local solution to</u> (P_0).

A proof of Theorem 1 can be found in the book of Hestenes [2, Ch. 1].

It deserves emphasis that the "regularity" of $\bar{x}$ is not required in the sufficient condition but only in the necessary condition. Unless every point of M is regular, this is an awkward assumption which cannot be checked without knowing the solution to the problem in advance. A more honest statement of the necessary condition would be that "if $\bar{x}$ is a local solution, then either (8) and (10) hold, or $\bar{x}$ is not a regular point of M". Even so, this is a stronger assertion than one sometimes sees. Especially in expositions inspired by the maximum principle in optimal control, it is common to include a nonnegative multiplier $\bar{y}_0$ for f_0 which, if nonzero, can be normalized as equal to 1. The necessary condition then asserted, in effect, is merely that "if $\bar{x}$ is a local solution, then either (8) and (10) hold, or (5) is false".

In the classical framework, Theorem 1 is regarded as solving (P) in principle in the sense of reducing it to the solution of a system of $m+n$ equations in $m+n$ unknowns: $\nabla L(\bar{x},\bar{y}) = 0$. Thus we should proceed by finding all possible solutions $(\bar{x},\bar{y})$ to these equations. The ones satisfying (10') furnish local solutions, although not necessarily all such solutions.

To determine the global solutions, <u>assuming the minimum in (P_0) is indeed attained</u>, we should form the collection S consisting of all the $\bar{x}$ components of pairs $(\bar{x},\bar{y})$ satisfying (8) and (10), as well as all points $\bar{x}$ of M which are not regular (or, for easier checking, all points $\bar{x}$ of M where (5) fails). For each $\bar{x} \in S$, the value $f_0(\bar{x})$ is calculated; the global solutions to (P_0) are then the elements of S for which this value is lowest. In other words, the minimization of f_0 over M is reduced to the minimization of f_0 over a much more special set S (possibly a finite set).

Unfortunately, however reasonable this procedure may seem by analogy with the elementary (and carefully concocted!) cases treated in calculus textbooks, it is hopeless from the modern point of view. The set S is likely to be too difficult to determine, or too large to work with. Anyway, solving a system of equations is generally as hard a problem, numerically speaking, as solving (P_0) itself.

If Lagrange multipliers do not lead to an effective computational procedure as just described, of what value are they? They seem nevertheless to have two main uses in the theory of computation. The first occurs in showing that some algorithm produces a sequence whose cluster points satisfy the first-order condition (8). The second lies in demonstrations that if a certain algorithm is initiated with a point close enough to an $\bar{x}$ satisfying the second-order sufficient conditions (8) and (10'), then the sequence it generates converges to $\bar{x}$, moreover at a particular rate.

The idea behind this presumably is that points satisfying (8) and (10) are "usually" local solutions to (P_0), while in "most" cases a local solution to (P_0) satisfies (8) and (10'). Theorem 1 is customarily cited as the justification of this heuristic reasoning, although it really says nothing quite along such lines.

Another use of Lagrange multipliers stems not from their role in computation but from a certain interpretation of them. This is in terms of the rate of change of the minimum in (P_0) with respect to various perturbations.

To gain an understanding, consider for each $u = (u_1,\ldots,u_m) \in R^m$ the "perturbed" problem

(P_0^u) minimize $f_0(x)$ subject to $f_i(x) - u_i = 0$ for $i = 1,\ldots,m$.

The first-order condition for (P_0^u) corresponding to (8) concerns a pair (x,y) such that

$$\nabla_x L^u(x,y) = 0 \quad \text{and} \quad \nabla_y L^u(x,y) = 0, \tag{11}$$

where

$$L^u(x,y) = f_0(x) + y_1(f_1(x) - u_1)+\ldots+y_m(f_m(x) - u_m).$$

We can express (11) as $D(u,x,y) = 0$, where D is continuously differentiable. Suppose $\bar{x}$ is a local solution to (P_0) actually satisfying (5), (8) and (10') for some $\bar{y}$. One sees that then $D(0,\bar{x},\bar{y}) = 0$, and the Jacobian $\nabla_{x,y}D(0,\bar{x},\bar{y})$ is nonsingular. Hence there exist differentiable functions $x(u)$ and $y(u)$, defined on a neighborhood of $u = 0$, such that

$$D(u,x(u),y(u)) \equiv 0, \quad x(0) = \bar{x}, \quad y(0) = \bar{y}.$$

Then not only do we have the first-order condition

$$\nabla_x L^u(x(u),y(u)) = 0 \quad \text{and} \quad \nabla_y L^u(x(u),y(u)) = 0 \tag{12}$$

satisfied for u near enough to 0, but also the second-order condition for (P_0^u) corresoponding to (10'). This can be verified using continuity and the fact that $\bar{x}$ and $\bar{y}$ satisfy (10'). Theorem 1 says that $x(u)$ is an isolated local solution to (P_0^u), the local minimum value being

$$p(u) = f_0(x(u)) = L^u(x(u),y(u)).$$

It follows from (12) that

$$\nabla p(u) = (\nabla_u L^u)(x(u),y(u)) = -y(u),$$

and in particular

$$\bar{y} = -\nabla p(0).$$

Thus if a computational method for solving (P_0^u) produces along with a solution $\bar{x}$ an associated multiplier vector $\bar{y}$, the components $\bar{y}_i$ furnish definite information about the way the minimum depends on perturbations of the constraints.

2. INEQUALITY CONSTRAINTS AND CONVEXITY

Convex analysis enters the picture as soon as inequality constraints are admitted, even if the functions in question are not convex. Although a problem may have a mixture of equality and inequality constraints, we limit ourselves for clarity to the pure inequality case:

(P_1) minimize $f_0(x)$ subject to $f_1(x) \le 0,\dots,f_m(x) \le 0.$

Again each f_i is assumed to be of class C^2 on R^n. The feasible set

(13) $$M = \{x \in R^n \mid f_1(x) \le 0,\dots,f_m(x) \le 0\}$$

may well have "edges, faces and corners", and local convexification, rather than linearization, is needed in characterizing the minimum.

Let $\bar{x}$ be a local solution to (P_1), and let I be the set of <u>active</u> constraint indices at $\bar{x}$:

(14) $$I = \{i \mid 1 \le i \le m,\ f_i(\bar{x}) = 0\}.$$

Let

(15) $$N = \text{polyhedral convex cone generated by } \{\nabla f_i(\bar{x}) \mid i \in I\}$$
$$= \{\textstyle\sum_{i=1}^m y_i \nabla f_i(\bar{x}) \mid y_i \ge 0 \text{ with } y_i = 0 \text{ if } f_i(\bar{x}) < 0\}$$

(16) $$T = N^\circ \text{ (polar cone)} = \{v \in R^n \mid \nabla f_i(\bar{x})\cdot v \le 0 \text{ for all } i \in I\}.$$

If T agrees with the tangent space T' to M at $\bar{x}$ as defined by (4), one says again that $\bar{x}$ is a *regular point*. This holds in particular if the gradients $\nabla f_i(\bar{x})$ for $i \in I$ are linearly independent.

If $\bar{x}$ is regular, local optimality implies that $\nabla f_0(\bar{x})\cdot v \geq 0$ for all $v \in T$. In other words $-\nabla f_0(\bar{x}) \in T^{oo} = N^{oo}$. But $N^{oo} = N$ by the laws of convex analysis (Lemma of Farkas). We thus get the *first-order necessary conditions*:

(17)
$$\begin{aligned} &f_i(\bar{x}) \leq 0 \quad \text{for} \quad i = 1,\dots,m \quad \text{and} \ \exists\, \bar{y} = (\bar{y}_1,\dots,\bar{y}_m) \\ &\text{with} \quad \bar{y}_i \geq 0 \quad \text{and} \quad \bar{y}_i f_i(\bar{x}) = 0 \quad \text{for} \quad i = 1,\dots,m \\ &\text{such that} \quad \nabla f_0(\bar{x}) + \bar{y}_1 \nabla f_m(\bar{x}) + \dots + \bar{y}_m \nabla f_m(\bar{x}) = 0. \end{aligned}$$

Taking the Lagrangian function to be

(18)
$$L(x,y) = \begin{cases} f_0(x) + y_1 f_1(x) + \dots + y_m f_m(x) & \text{if} \quad y_i \geq 0,\ i = 1\dots,m, \\ -\infty & \text{otherwise,} \end{cases}$$

we can express (17) equivalently as

(19)
$$\exists\, \bar{y} \quad \text{such that} \quad L(\bar{x},\bar{y}) = \max_{y \in R^m} L(\bar{x},y) \quad \text{and} \quad \nabla_x L(\bar{x},\bar{y}) = 0$$

Second-order conditions concern the Hessian matrix $\nabla_x^2 L(\bar{x},\bar{y})$ as in (9), the set $\tilde{M} = \{x \in M \mid f_i(x) = 0 \ \text{if} \ \bar{y}_i > 0\}$, and the properties

(20)
$$v \cdot \nabla_x^2 L(\bar{x},\bar{y}) v \geq 0 \quad \text{for all} \quad v \in \tilde{T}$$

and

(20')
$$v \cdot \nabla_x^2 L(\bar{x},\bar{y}) v > 0 \quad \text{for all nonzero} \quad v \in \tilde{T},$$

(21)
$$\tilde{T} = \{v \in T \mid \nabla f_i(\bar{x}) \cdot v = 0 \ \text{if} \ \bar{y}_i > 0\},$$

where $\tilde{T}$ is the tangent cone to $\tilde{M}$ at $\bar{x}$ if $\bar{x}$ is regular.

THEOREM 2. (local optimality).

(a) (*necessary conditions*). *If* $\bar{x}$ *is a local solution to* (P_1) *which is also a regular point of* M *and* $\tilde{M}$, *then* (19) *and* (20) *hold*.

(b) (*sufficient conditions*). *If* (19) *and* (20') *hold, then* $\bar{x}$ *is an isolated local solution to* (P_1).

For a proof of Theorem 2, we refer again to Hestenes [2, Ch. 1]. The assertion that the first-order condition (19) (or equivalently (17)) holds if $\bar{x}$ is a "regular" local solution to (P_1) is usually cited as a theorem of

Kuhn and Tucker [3], at least under a slightly narrower formulation of "regularity". However, historical antecedents are now known; see the article of Kuhn in this volume.

The comments made about the significance of the Lagrange multiplier vector $\bar{y}$ in Theorem 1 apply equally to the $\bar{y}$ in Theorem 2. In particular, it can be shown under mild assumptions that $\bar{y} = -\nabla p(0)$, where $p(u)$ gives a local minimum value in the perturbed problem

(P_1^u) minimize $f_0(x)$ subject to $f_1(x) \le u_1, \ldots, f_m(x) \le u_m$.

While much research has gone into improving the first-order necessary condition in Theorem 2 (or combined versions of Theorems 1 and 2) by substituting different forms of "regularity" (which make the potential nonregularity of $\bar{x}$ a more restrictive property and hence supposedly "less likely to occur"), this has had almost no impact on applications. What would be far more valuable in this area would be a family of generic theorems asserting that for "most problems" in a given class (in some well defined mathematical sense) a solution $\bar{x}$ can be characterized by certain stronger conditions, like those in Theorem 2(b). New research should be conducted in this direction.

Of course, we do not mean to say that work on necessary conditions as such is not of interest. There are many examples where such work has led to great theoretical clarification of the nature of a class of problems and how their properties can be formulated and derived in an elegant manner. As an example, we mention the recent work of Clarke [11] on necessary conditions in optimal control using new ideas of generalized gradients; this may lead to developments in other branches of optimization. However, we do feel that work on "practical" sorts of sufficient conditions has not been stressed enough in the past.

There is a dramatic change in the theory of Lagrange multipliers when we reach the case of a convex programming problem, where the functions f_i in (P_1) are all convex. Then local solutions $\bar{x}$ are global solutions, and $L(x,y)$ is convex in x for each y. Any stationary point of $L(x,y)$ with respect to x must give a global minimum with respect to x, and the first-order condition (19) can be written as

(22) $\exists\, \bar{y} \in R^m$ such that $\min_{x \in R^n} L(x,\bar{y}) = L(\bar{x},\bar{y}) = \max_{y \in R^m} L(\bar{x},y)$.

This fact was recognized by Kuhn and Tucker in their pioneering paper [3] and proved to be very fertile ground for new growth of the theory. A pair $(\bar{x},\bar{y})$ satisfying the relation in (22) is said to be a saddle point of L. As such, it can be given a game-theoretic interpretation which further enhances the meaning of the Lagrange multipliers, especially in an economic context.

It was quickly recognized that in deriving (22) as a necessary condition in convex programming no differentiability assumptions are needed. Convex analysis suffices, if an appropriate substitute for the notion of "regular point of M" is provided. A convenient assumption often invoked in this connection is the Slater condition:

(23) $\exists\, \tilde{x} \in R^n$ such that $f_i(\tilde{x}) < 0$ for $i = 1,\ldots,m$.

However, a much broader, yet also meaningful criterion can be stated in terms of the "stability" of (P_1) with respect to the perturbations introduced above.

Let $\phi(u)$ denote the global infimum in problem (P_1^u), so that $\phi(0) = \inf(P_1)$. (The infimum is taken to be $+\infty$ if the constraints in the problem cannot be satisfied.) The convexity of the functions f_i implies the convexity of ϕ. It follows then (assuming the finiteness of $\inf(P_1)$) that the one-sided directional derivatives

$$\phi'(0;w) = \lim_{t \downarrow 0} \frac{\phi(tw) - \phi(0)}{t}$$

all exist (possibly $+\infty$ or $-\infty$). One says that (P_1) is stable if there is no $\tilde{w}$ with $\phi'(0;\tilde{w}) = -\infty$. The Slater condition (23) implies that (P_1) is stable, but other verifiable criteria are also known, cf. [4], [5], [6]. A more general concept of stability for the nonconvex case will be treated in §5 (cf. [5])).

THEOREM 3. (global optimality)

(a) (necessary condition). Suppose $\bar{x}$ is a local solution to (P_1) and the functions f_i, $i = 0,1,\ldots,m$, are all convex (finite but not necessarily differentiable). Then (22) holds if and only if (P_1) is stable.

(b) (_sufficient condition_). _If_ (22) _holds, then_ $\bar{x}$ _is a global solution to_ (P_1).

A proof of Theorem 3(a) may be found in [4] or [6] (see also §4 below). The fact that (b) is true even without convexity is elementary (see §3) and was observed already by Kuhn and Tucker [3].

Parallel to the earlier characterization $-\bar{y} = \nabla p(0)$, there is the result that the Lagrange multiplier vectors $\bar{y}$ appearing in (22) are precisely those satisfying

(24) $$\phi'(0;w) \geq -w \cdot \bar{y} \quad \text{for all} \quad w \in R^m.$$

(See [4], [6]). This says that $-\bar{y}$ is a "subgradient" of ϕ at 0, written $-\bar{y} \in \partial\phi(0)$.

More generally, the perturbed problem (P_1^u) is stable if and only if the set $\partial\phi(u)$ is nonempty. Theorems of convex analysis about the existence of subgradients show that $\partial\phi(u)$ is nonempty for every "relative interior" point of the set of vectors u such that $\phi(u) < \infty$. In this sense, we can truly say that "virtually all" problems in the convex case are stable, so that (22) is both necessary and sufficient for optimality.

3. GENERALIZED LAGRANGE MULTIPLIERS

Further insight into the computational significance of the Lagrangian function L for (P_0) or (P_1) can be gained from its relationship with the essential objective function f, defined by

(25) $$f(x) = \begin{cases} f_0(x) & \text{if } x \text{ satisfies the constraints,} \\ +\infty & \text{otherwise.} \end{cases}$$

In both (P_0) and (P_1), one has

$$\sup_{y \in R^m} L(x,y) = f(x),$$

and the problem is equivalent to minimizing $f(x)$ over all $x \in R^n$.

To add flexibility to the discussion, let us now think of an abstract optimization problem having merely the form

(P) minimize $f(x)$ over all $x \in X$,

where X is a real linear space and f is an extended-real-valued function on X. Such a problem is convex is f is a convex function, but we do not make this restriction in what follows.

By a Lagrangian representation of (P), we shall mean the choice of a real linear space Y and an extended-real-valued function L on $X \times Y$ such that

(26) $$f(x) = \sup_{y \in Y} L(x,y).$$

The latter formula may be interpreted as expressing f as the pointwise supremum on X of the collection of functions $\{L(\cdot,y) \mid y \in Y\}$. The elements $y \in Y$ are regarded as generalized Lagrange multiplier vectors, and L itself is a generalized Lagrangian for (P).

For each fixed $y \in Y$, we can regard the problem of minimizing $L(\cdot,y)$ over X as a sort of "lower representative" of (P), since

(27) $$f(x) \geq L(x,y) \quad \text{for all } x.$$

Clearly

(28) $$\inf_{x \in X} f(x) \geq \inf_{x \in X} L(x,y) \quad \text{for all } y \in Y.$$

We shall say that $\bar{y}$ is a Kuhn-Tucker vector for (P) (with respect to the Lagrangian L) if

(29) $$\inf_{x \in X} f(x) = \inf_{x \in X} L(x,\bar{y}).$$

If a Kuhn-Tucker vector $\bar{y}$ were known, one could replace (P) by the possibly simpler problem of minimizing $L(x,\bar{y})$ over all $x \in X$. Indeed, in view of (29) and the inequality (27) with $y = \bar{y}$, the points furnishing the global minimum in (P) must then be precisely the points $\bar{x}$ which afford the global minimum of $L(\cdot,\bar{y})$ and also satisfy $L(\bar{x},\bar{y}) = f(\bar{x})$. In particular, if $L(\cdot,\bar{y})$ has its minimum at a unique point $\bar{x}$, then $\bar{x}$ must be the unique solution to (P).

The determination of a Kuhn-Tucker vector $\bar{y}$ corresponds to solving a certain dual problem:

(D) maximize $g(y)$ over all $y \in Y$,

where

(30) $$g(y) = \inf_{x \in X} L(x,y).$$

Observe that (28) is equivalent to

(28') $$\inf(P) \geq g(y) \quad \text{for all} \quad y \in Y,$$

while the definition (29) of a Kuhn-Tucker vector is equivalent to

(29') $$\inf(P) = \max(D) \quad \text{and} \quad \bar{y} \quad \text{solves} \quad (D).$$

Of course, computational realizations of these notions are not really envisioned in terms of first maximizing $g(y)$ to get $\bar{y}$ and then minimizing $L(x,\bar{y})$ to get $\bar{x}$, although this might be possible in some cases. More hope rests in generating a maximizing sequence $\{y^k\}$ for (D) by some method that involves calculation at each step of an approximate minimum in (30) for $y = y^k$ and thus simultaneously generates a sequence $\{x^k\}$. Perhaps this can be done in such a manner that $\{y^k\}$ converges to a Kuhn-Tucker vector $\bar{y}$, while $\{x^k\}$ or some auxilliary sequence converges to an $\bar{x}$ solving (P). In summary, the idea is to try to solve (P) by replacing it by a well chosen sequence of more favorable problems of the form

(31) $$\text{minimize} \quad L(x,y^k) \quad \text{over all} \quad x \in X.$$

Optimality in (P) can be characterized by means of the general saddle point condition:

(32) $$\exists\, \bar{y} \in Y \quad \text{such that} \quad \min_{x \in X} L(x,\bar{y}) = L(\bar{x},\bar{y}) = \max_{y \in Y} L(\bar{x},y).$$

THEOREM 4. (global optimality).

(a) (necessary condition). If $\bar{x}$ furnishes the global minimum in (P) and $\min(P) = \max(D)$, then (32) holds.

(b) (sufficient condition). If $\bar{x}$ satisfies (32), then $\bar{x}$ furnishes the global minimum in (P), and $\min(P) = \max(D)$. Moreover, the elements $\bar{y}$ occurring in (32) then furnish the global maximum in (D), and they are precisely the Kuhn-Tucker vectors for (P) with respect to the Lagrangian L.

This basic result is an elementary consequence of the relations and definitions given above.

The saddle point condition (32) is thus always sufficient for optimality, but it is necessary only to the extent that a duality relation of the type

$$\inf(P) = \max(D) \tag{33}$$

(equivalent to the existence of a Kuhn-Tucker vector $\bar{y}$) can be established. The value of the condition and the corresponding numerical approach rests therefore on our knowledge of cases where (33) is sure to be true, or "almost" sure to be true. It depends also on analyzing the various ways that workable Lagrangian representations might be constructed.

Another important question concerns the information provided by the Kuhn-Tucker vectors $\bar{y}$, if any. Can they, as in the earlier cases, be interpreted in terms of directional derivatives with respect to certain perturbations of (P)?

The theory of conjugate duality has produced some far-reaching answers that we attempt to elucidate in the next section. But the subject is by no means closed to further research.

4. PERTURBATIONS AND CONJUGATE DUALITY

General perturbations can be introduced into the abstract problem (P) by choosing a real linear space U and an extended-real-valued function F on $X \times U$ such that

$$F(x,0) = f(x) \quad \text{for all} \quad x \in X. \tag{34}$$

We shall call this a <u>perturbational representation</u> of (P). The perturbed problem corresponding to a vector $u \neq 0$ in U is

(P^u) minimize $F(x,u)$ over all $x \in X$.

For example, the perturbations already discussed for problem (P_1) correspond to choosing $U = R^m$ and

$$F(x,u) = \begin{cases} f_0(x) & \text{if} \quad f_i(x) \le u_i \quad \text{for} \quad i = 1,\dots,m \\ +\infty & \text{otherwise} \end{cases} \tag{35}$$

Given any perturbational representation, we are interested in the properties of the function

$$\phi(u) = \inf(P^u) = \inf_{x \in X} F(x,u) \tag{36}$$

around $u = 0$. If $F(x,u)$ is convex jointly in x and u (we refer to this as the <u>fully convex case</u>), ϕ is a convex function on U and its properties are readily investigated via convex analysis. Of course, one cannot have F convex in this way unless f is convex, i.e. (P) is a "convex" problem as defined earlier. However, even in the nonconvex case, we can employ various techniques of convex analysis to learn much about ϕ.

As a matter of fact, there is a very close connection between Lagrangian representations and perturbational representations of (P). This is derived from the theory of conjugate convex functions, as we now proceed to explain.

Let us suppose that the real linear spaces U and Y are <u>paired</u> by some bilinear form $\langle u,y \rangle$. This means that $\langle \cdot ,y \rangle$ is a linear function on U for each $y \in Y$, but it is the zero function only if $y = 0$; at the same time, $\langle u, \cdot \rangle$ is a linear function on Y for each $u \in U$, but it is the zero function only if $u = 0$. (For $U = R^m = Y$, one can take $\langle u,y \rangle = u \cdot y$.).

If h is any extended-real-valued function on U, the <u>closed convex hull</u> of h (with respect to the pairing of U and Y) is the function h^{**} which is the pointwise supremum of the collection of all the affine functions (i.e., functions of the form $\langle \cdot ,y \rangle$ + const.) majorized by h. If $h^{**} = h$, one says that h is a <u>closed convex function</u>. Closed convex functions on Y are defined similarly. A <u>closed concave function</u> is one whose negative is closed convex. Topological criteria for the closedness of a convex function can also be given [4], [6].

The following facts are fundamental. <u>Given any extended-real-valued function h on U, the function</u>

$$k(y) = \inf_{u \in U} \{h(u) + \langle u,y \rangle\} \tag{37}$$

<u>is closed concave on Y, and one has</u>

$$h^{**}(u) = \sup_{y \in Y} \{k(y) - \langle u,y \rangle\}. \tag{38}$$

Formula (37) thus defines a one-to-one correspondence between the collection of all closed convex functions h on U and the collection of all closed concave functions k on Y.

The theory of this correspondence (which is expounded in [4], [6]) is customarily expressed not in terms of h and k, but h and h^*, where $h^*(y) = -k(-y)$. (The closed convex function h^* is called the conjugate of h.) Many powerful facts are known. Generally speaking, there is a deep duality between local properties of h around $u = 0$ and properties of the nest of level sets $\{y \in Y \mid k(y) \geq \alpha\}$, $\alpha \in R$.

The following consequence is almost immediate.

THEOREM 5. (equivalence of representations).

For paired spaces U and Y, there is a one-to-one correspondence between all the possible Lagrangian representations L of (P) with $L(x,y)$ closed concave in y and all the possible perturbational representations F of (P) with $F(x,u)$ closed convex in u, namely:

$$L(x,y) = \inf_{u \in U} \{F(x,u) + \langle u,y \rangle\}, \tag{39}$$

$$F(x,u) = \sup_{y \in Y} \{L(x,y) - \langle u,y \rangle\}. \tag{40}$$

Moreover, $F(x,u)$ is jointly convex in x and u if and only if $L(x,y)$ is convex in x (as well as concave in y).

The only observation that needs to be made, to see the validity of Theorem 5, is that properties (26) and (34) correspond to each other under (39) and (40).

We note that the ordinary Lagrangian (18) for (P_1) is obtained from the perturbational representation (35). This is an example where $F(x,u)$ is closed convex in u but not convex in (x,u) (unless the functions f_i in (P_1) are all convex).

The main result connecting (P) and the dual problem (D) may now be stated.

THEOREM 6. (duality)

Under the correspondence in Theorem 5, the optimal value function ϕ in (36) and the dual objective function g in (30) are related by

$$g(y) = \inf_{u \in U} \{\phi(u) + \langle u,y \rangle\} \tag{41}$$

(42) $$\phi^{**}(u) = \sup_{y \in Y} \{g(y) - \langle u,y\rangle\},$$

and hence in particular one always has

(43) $$\sup(D) = \phi^{**}(0) \le \phi(0) = \inf(P).$$

Moreover, $\bar{y}$ *is a Kuhn-Tucker vector for* (P) *if and only if*

(44) $$\phi(u) \ge \phi(0) - \langle u,\bar{y}\rangle \quad \text{for all} \quad u \in U.$$

In other words, one has $\inf(P) = \max(D)$ *if and only if there exists* $\bar{y}$ *satisfying* (44).

Relation (41) is obtained at once from (39) and definitions (36) and (30), and then (42) follows, just as in general (38) follows from (37). As for (43), this is just the specialization of (36) and (42) to $n = 0$. The inequality (44) says that $g(\bar{y}) = \phi(0)$, and hence it is equivalent to (29'), an equivalent form of the definition of a Kuhn-Tucker vector $\bar{y}$.

The implication of (41) and (42), of course, is that all the theory of conjugate functions can be invoked in the study of the relationship between (P) and (D). In particular, properties of the nest of level sets $\{y \in Y \mid g(y) \ge \alpha\}$, $\alpha \in R$, are seen to be dual to local properties of $\phi(u)$ around $u = 0$.

In the "fully convex" case mentioned above, where ϕ is definitely a convex function, (44) can be expressed by directional derivatives as in (24) (with U in place of R^m). Theorems of convex analysis lead again via Theorem 4 to the conclusion that a Kuhn-Tucker vector "usually" exists, so that the saddle point condition (32) is always sufficient and "usually" necessary for the optimality of $\bar{x}$ in (P).

This is potentially a very rich result, because it furnishes for any class of convex problems a vast array of sufficient and usually necessary conditions for optimality, moreover, in a computationally suggestive form. At the same time, the theory is applicable to problems far more general than (P_1) such as in optimal control or stochastic programming.

Even for the fundamental convex programming problem (P_1) this theory yields much that is new. The ordinary Lagrangian (18) is not the only one that can be associated with (P_1). There is an infinite variety of

Lagrangians L for (P_1), each choice corresponding to a different dual problem. Of course, not all of these Lagrangians can be of practical value in computation, but the possibilities have hardly begun to be explored. Some examples based on special structure of (P_1) are treated in [5] and [9]. Another example, not involving special assumptions, is presented below.

We remark that in the fully convex case it is possible to introduce perturbations into the dual problem (D) in such a way that the dual of (D) is the primal problem (P). Solutions $\bar{x}$ to (P) can then be interpreted as Kuhn-Tucker vectors for the solutions $\bar{y}$ to (D).

For further details on conjugate duality, including some of its applications to problems in optimal control and stochastic programming, we refer to the recent lecture notes [6]. No doubt there are many applications to special classes of convex problems simply awaiting discovery.

In the nonconvex case, it might be thought that there is little hope in the existence of a $\bar{y}$ satisfying (44) and hence little promise in the saddle point condition (32) as a general criterion for optimality. This would indeed be correct, were it not for the great flexibility afforded by the theory in the choice of the Lagrangian representation. A choice described in the next section demonstrates that, for (P_0) and (P_1) at least, the saddle point approach is capable of subsuming most other aspects of optimality, such as the facts stated in §1 and §2.

5. THE AUGMENTED LAGRANGIAN IN NONCONVEX PROGRAMMING

Certain deficiencies of the ordinary Lagrangian function (18) for (P_1) can be overcome by passing to a so-called augmented Lagrangian involving an additional variable r which acts much like a penalty parameter. In this way a useful saddle point characterization of optimality, capable of taking the place of Theorems 1 and 2, can be obtained without assuming convexity of the functions f_i. We shall discuss only (P_1) for simplicity, although a combined form of (P_0) and (P_1) could easily be encompassed.

To facilitate comparisons with the facts in §2, we denote the ordinary Lagrangian by $L(x,y)$ and the augmented Lagrangian by $\hat{L}(x,\hat{y})$, where

$$\hat{y} = (y_1,\ldots,y_m,r) = (y,r) \in R^m \times R.$$

The augmented Lagrangian for (P_1) is defined by

(45) $$\hat{L}(x,\hat{y}) = f_0(x) + \sum_{i=1}^{m}\theta(f_i(x),y_i,r),$$

where

(46) $$\theta(f_i(x),y_i,r) = \begin{cases} y_i f_i(x) + r f_i(x)^2 & \text{if } f_i(x) \geq -y_i/2r,\ r > 0, \\ -y_i^2/4r & \text{if } f_i(x) \leq -y_i/2r,\ r > 0, \\ y_i f_i(x) & \text{if } y_i \geq 0,\ r = 0, \\ -\infty & \text{otherwise.} \end{cases}$$

We cannot go into the origin or computational motivation for the augmented Lagrangian here; see [1], [7], [8] and the references given there. It may be observed, however, that for $y = 0$, $r \geq 0$, one has

(47) $$\hat{L}(x,0,r) = f_0(x) + r\sum_{i=1}^{m} \max^2\{f_i(x),0\},$$

a familiar expression in a well known penalty method for solving (P_1). On the other hand, setting $r = 0$ one gets

$$\hat{L}(x,y,0) = L(x,y) \quad \text{for all } x,y.$$

The augmented Lagrangian is thus truly an extension (augmentation) of the ordinary Lagrangian, as the name suggests.

Note that $\hat{L}(x,y,r)$ is nondecreasing as a function of r and hence so is

(48) $$\hat{g}(\hat{y}) = \hat{g}(y,r) = \inf_{x\in R^n}\hat{L}(x,y,r).$$

It follows that in the augmented dual problem, namely

$(\hat{D})$ maximize $\hat{g}(y,r)$ over all $(y,r) \in R^m \times R$,

nothing is lost in restricting r to be positive, or for that matter to be as large as seems convenient in some context.

The saddle point condition for $\hat{L}$ can therefore, without real loss of generality, be posed in the form:

$$\exists\ \bar{y} \in R^m \text{ and } \bar{r} \geq 0 \text{ such that}$$

(49) $$\min_{x\in R^n} \hat{L}(x,\bar{y},\bar{r}) = \hat{L}(\bar{x},\bar{y},\bar{r}) = \max_{\substack{y\in R^m \\ r\geq 0}} \hat{L}(\bar{x},y,r).$$

It is readily checked that $\hat{L}(x,y,r)$ is closed concave in (y,r) (in fact also convex in x if the functions f_i are all convex, although we are not assuming this). Furthermore, as can be seen even from manipulating r alone,

$$\sup_{y,r} \hat{L}(x,y,r) = f(x),$$

where f is the essential objective function (25) for (P_1). We therefore do have a Lagrangian representation of (P_1) meeting the requirements of Theorem 5, and it must correspond to a certain perturbational representation: the one given by

$$\hat{F}(x,\hat{u}) = \sup_{\hat{y}}\{\hat{L}(x,\hat{y}) - \hat{u}\cdot\hat{y}\},$$

where

$$\hat{u} = (u_1,\ldots,u_m,s) = (u,s) \in R^m \times R.$$

The calculation of $\hat{F}$ is elementary to carry out, and one obtains

$$F(x,u) = \begin{cases} \hat{F}(x,\hat{u}) & \text{if } |u|^2 \le s \\ +\infty & \text{otherwise,} \end{cases}$$

where F is the ordinary perturbational representation of (P_1) in (35) and $|\cdot|$ is the Euclidean norm. The optimal value function $\hat{\phi}$ corresponding to $\hat{F}$, which is needed in applying Theorem 6, is therefore related to the ordinary optimal value function ϕ corresponding to F by

$$\hat{\phi}(\hat{u}) = \hat{\phi}(u,s) = \begin{cases} \phi(u) & \text{if } |u|^2 \le s, \\ +\infty & \text{otherwise .} \end{cases} \tag{50}$$

Condition (44) in the "augmented duality" context has the form

$$\hat{\phi}(u,s) \ge \hat{\phi}(0,0) - u\cdot\bar{y} - s\cdot\bar{r} \quad \text{for all } (u,s) \in R^m \times R, \tag{51}$$

where it may be assumed that $\bar{r} \ge 0$. We can rewrite this by means of (50) as

$$\begin{aligned} \phi(0) = \hat{\phi}(0,0) &= \inf_{u,s}\{\hat{\phi}(u,s) + u\cdot\bar{y} + s\cdot\bar{r}\} \\ &= \inf_{u\in R^m} \inf_{s\ge|u|^2} \{\phi(u) + u\cdot\bar{y} + s\bar{r}\} \\ &= \inf_{u\in R^m} \{\phi(u) + u\cdot\bar{y} + \bar{r}|u|^2\}. \end{aligned}$$

Thus (51) is equivalent (for $\bar{r} \ge 0$) to

$$\phi(u) \ge \phi(0) - u\cdot\bar{y} - \bar{r}|u|^2 \quad \text{for all } u \in R^m. \tag{52}$$

Combining portions of Theorems 4 and 6, we reach the following conclusion.

One has $\inf(P_1) = \max(\hat{D})$ if and only if the ordinary optimal value function $\phi(u) = \inf(P_1^u)$ satisfies (52) for some $\bar{y} \in R^m$ and $\bar{r} \geq 0$. In this event the saddle point condition (49) for $\hat{L}$, which is always sufficient for $\bar{x}$ to be a global solution to (P_1), is also necessary.

The value of this result lies in the fact that (52) requires only the existence of some concave quadratic "supporting function" for ϕ at $u = 0$. The corresponding condition for the ordinary Lagrangian takes the form (44) for the same ϕ (in other words limits (52) to $\bar{r} = 0$) and thus requires the existence of an affine "supporting function". Obviously, the first condition can often be fulfilled even if ϕ is not convex, but the second is rather unreasonable without convexity.

While no true generic theorems, asserting that (52) can "usually" be satisfied for some category of problems (P_1), have yet been established, a related result is known. To state this, we need some asumptions:

(i) the functions $f_0, f_1, \ldots, f_m$ are of class C^2;

(ii) for some $\alpha > \inf(P_1)$, the set

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

is bounded;

(iii) there exists $r \geq 0$ such that

$$\inf_{x \in R^n} \hat{L}(x,0,r) > -\infty$$

(cf. (47)). The following theorem is proved in a somewhat more general form in [7], [8].

THEOREM 8. (saddle points in nonconvex programming)

Assume (i), (ii) and (iii).

(a) If $\bar{x}$ satisfies the sufficient conditions (19) and (20') for a certain $\bar{y}$ and is in fact the only global solution to (P_1), then $(\bar{x},\bar{y},\bar{r})$ is a saddle point of $\hat{L}$ for all $\bar{r}$ sufficiently large.

(b) If $(\bar{x},\bar{y},\bar{r})$ is a saddle point of $\hat{L}$, then $\bar{x}$ is a global solution to (P_1) satisfying the necessary conditions (19) and (20) for this $\bar{y}$.

Assumptions (ii) and (iii) and the uniqueness in part (a) of Theorem 8 are really not the restrictions they may seem when it comes to comparing the saddle point condition to the earlier conditions for optimality, which were local in character. Indeed, if $\bar{x}$ is an isolated local solution to (P_1), these assumptions can be satisfied by replacing $f_0(x)$ by $f_0(x) + \rho(x)$, where ρ is a function which vanishes throughout some neighborhood of $\bar{x}$ but grows fast enough in the large. This modification of (P_1) has no effect on the analysis of a local solution $\bar{x}$ but allows us to regard it as the only global solution.

In this sense, we may conclude from Theorem 8(a) that the saddle point condition (49) is a sufficient condition for optimality more comprehensive in every respect than the classical differential conditions (19) and (20'). It is equally applicable without any differentiability. It corresponds directly to a perturbational interpretation of the Lagrange multipliers, namely (52), which reduces to the earlier interpretation where that was valid. It has immediate computational significance (cf. [1], [7]). Other properties that are sometimes useful in the analysis of various algorithms, such as the linear independence of the gradients $\nabla f_i(\bar{x})$ for $i \in I$ or the positivity of $\bar{y}_i$ for $i \in I$, are likewise expressible in terms of $\hat{L}$ and its second derivatives at $(\bar{x},\bar{y},\bar{r})$.

If (P_1) is a convex programming problem, the saddle point condition for $\hat{L}$ is equivalent to the one for the ordinary Lagrangian L. (Then ϕ is convex in (52), so that the term $-\bar{r}|u|^2$ is superfluous.) Thus (49) is just as necessary for optimality in the convex case as previous conditions.

In the differentiable nonconvex case, (49) appears from Theorem 8(b) to be slightly less general than (19) and (20). It is possible in some cases where $\bar{x}$ is a regular point of M satisfying (19) and (20) but not (20') that $(\bar{x},\bar{y},\bar{r})$ is not a saddle point of $\hat{L}$ for any $\bar{r}$, no matter how large. But this is a marginal situation of little genuine interest. This is quite apparent when it is remembered that the classical second-order necessary conditions are of questionable virtue anyway, except as a heuristic justification in some contexts for assuming that the second-order sufficient conditions hold. A better justification would be a generic theorem giving circumstances under which (52) can "usually" be satisfied. This would be a very worthwhile contribution.

We mentioned earlier that the first-order condition (19) has some uses by itself in the study of algorithms. That this aspect is also covered by properties of the augmented Lagrangian $\hat{L}$ is shown by the next result.

THEOREM 9. (stationary points)

Assume the functions f_i in (P_1) are of class C^1. Then $\hat{L}$ is of class C^1 on $R^n \times R^m \times (0,\infty)$, and condition (19) is equivalent to

(53) $\exists(\bar{y},\bar{r}),\ \bar{r} > 0,$ *such that* $\nabla_x L(\bar{x},\bar{y},\bar{r}) = 0$ *and* $\nabla_{y,r} L(\bar{x},\bar{y},\bar{r}) = 0.$

(If the equations hold for a particular $\bar{r} > 0$, they hold for every $\bar{r} > 0$.)

Theorem 9 is derived by direct calculation of the gradients of $\hat{L}$. Setting

$$\eta(f_i(x),y_i,r) = \max\{f_i(x),-y_i/2r\}$$

for $r > 0$, we have

$$\nabla_x \hat{L}(x,y,r) = \nabla f_0(x) + \sum_{i=1}^{m}(y_i+\eta(f_i(x),y_i,r))\nabla f_i(x),$$
$$\frac{\partial \hat{L}}{\partial y_i}(x,y,r) = \eta(f_i(x),y_i,r),$$
$$\frac{\partial \hat{L}}{\partial r}(x,y,r) = \sum_{i=1}^{m} \eta(f_i(x),y_i,r)^2,$$

from which the assertions are obvious. Recall that $\hat{L}(x,y,r)$ is concave in (y,r), so that the second of gradient conditions in (53) corresponds to a global maximum.

The conclusion to be drawn from all this is that there is virtually nothing in the previous approaches to optimality conditions for (P_1) (or more generally, combined versions of (P_0) and (P_1)) which cannot be formulated advantageously, and often more generally, in terms of the augmented Lagrangian $\hat{L}$.

A new goal for the theory of Lagrange multipliers is thereby suggested. Equally potent Lagrangian functions should be sought for other classes of not-necessarily-convex problems of optimization besides (P_1). Optimal control problems are good candidates. Some results in this direction have already been achieved by Rupp; see [10] and its references. However, only rather special kinds of control problems have been handled successfully up till now. This remains a promising area for research.

Another good objective for the future is that of devising new Lagrangians that can be used effectively in the decomposition of large-scale problems.

REFERENCES

1. D. P. Bertsekas, "On penalty and multiplier methods for constrained minimizations", S.I.A.M. Journal on Control, to appear.

2. M. R. Hestenes, Calculus of Variations and Optimal Control Theory, Wiley, 1966.

3. H. W. Kuhn and A. W. Tucker, "Nonlinear programming", Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, U. of California Press, Berkeley, 1951, 481-492.

4. R. T. Rockafellar, Convex Analysis, Princeton U. Press, 1970.

5. R. T. Rockafellar, "Some convex programs whose duals are linearly constrained", in Nonlinear Programming, J. B. Rosen, O. L. Mangasarian, and K. Ritter (editors), Academic Press, 1970, 293-322.

6. R. T. Rockafellar, Conjugate Duality and Optimization, Regional Conference Series no. 16, S.I.A.M. Publications (33 South 17th Street, Philadelphia, PA 19103), 1974.

7. R. T. Rockafellar, "Solving a nonlinear programming problem by way of a dual problem", Symposia Mathematica, to appear.

8. R. T. Rockafellar, "Augmented Lagrange multiplier functions and duality in nonconvex programming", S.I.A.M. Journal on Control 12 (1974), 268-285.

9. R. T. Rockafellar, "Convex programming and systems of elementary monotonic relations", J. Math. Analysis Appl. 19 (1967), 167-187.

10. R. D. Rupp, "A nonlinear optimal control minimization technique", Trans. A.M.S. 178 (1973), 357-381.

11. F. H. Clarke, "A maximum principle without differentiability", Bull. A.M.S. 81(1975), 219-222.