SIAM REVIEW
Vol. 13, No. 1, January 1971

DUALITY IN NONLINEAR PROGRAMMING: A SIMPLIFIED APPLICATIONS-ORIENTED DEVELOPMENT*

A. M. GEOFFRION†

Summary. The number of computational or theoretical applications of nonlinear duality theory is small compared to the number of theoretical papers on this subject over the last decade. This study attempts to rework and extend the fundamental results of convex duality theory so as to diminish the existing obstacles to successful application. New results are also given having to do with important but usually neglected questions concerning the computational solution of a program via its dual. Several applications are made to the general theory of convex systems.

The general approach is to exploit the powerful concept of a perturbation function, thus permitting simplified proofs (no conjugate functions or fixed-point theorems are needed) and useful geometric and mathematical insights. Consideration is limited to finite-dimensional spaces.

An extended summary is given in the Introduction.

CONTENTS

* Received by the editors September 25, 1969.

† Graduate School of Business Administration, University of California, Los Angeles, California 90024. This work was supported by the National Science Foundation under Grant GP-8740, and by the United States Air Force under Project RAND.

1. Introduction.

1.1. Objective. In this paper much of what is known about duality theory for nonlinear programming is reworked and extended so as to facilitate more readily computational and theoretical applications. We study the dual problem in what is probably its most satisfactory formulation, permit only assumptions that are likely to be verifiable, and attempt to establish a theory that is more versatile and general in applicability than any heretofore available.

Our methods rely upon the relatively elementary theory of convexity. No use is made of the differential calculus, general minimax theorems, or the conjugate function theory employed by most other studies in duality theory. This is made possible by fully exploiting the powerful concept of a certain *perturbation function*—the optimal value of a program as a function of perturbations of its "right-hand side." In addition to some pedagogical advantages, this approach affords deep geometrical and mathematical insights and permits a development which is tightly interwoven with optimality theory.

The resulting theory appears quite suitable for applications. Several illustrative theoretical applications are made; and some reasonable conditions are demonstrated under which the dual problem is numerically stable for the recovery of an optimal or near-optimal primal solution. A detailed preview of the results obtained is given after the canonical primal and dual programs are introduced.

1.2. The canonical primal and dual programs. The canonical *primal* problem is taken to be:

$$\text{(P)} \qquad \underset{x \in X}{\text{Minimize}} \quad f(x) \quad \text{subject to} \quad g(x) \leqq 0,$$

where $g(x) \triangleq (g_1(x), \cdots, g_m(x))^t$, and f and each g_i are real-valued functions defined on $X \subseteq R^n$. *It is assumed throughout that X is a nonempty convex set on which all functions are convex.*

The *dual* of (P) with respect to the g-constraints is:

$$\text{(D)} \qquad \underset{u \geqq 0}{\text{Maximize}} \quad [\underset{x \in X}{\text{infimum}}\, f(x) + u^t g(x)],$$

where u is an m-vector of dual variables. Note that the maximand of (D) is a concave function of u alone (even in the absence of the convexity assumptions), for it is the pointwise infimum of a collection (indexed by x) of functions linear in u.

Several other possible "duals" of (P) have been studied, some of which are discussed in § 6. All are closely related, but we believe (D) to be the most natural and useful choice for most purposes.

It is important to recognize that, given a convex program, one can dualize with respect to *any* subset of the constraints. That is, each constraint can be assigned to the g-constraints, in which case it will possess a dual variable of its own; or it can be assigned to X, in which case it will not possess a dual variable. In theoretical applications, the assignment will usually be dictated by the desired conclusion (cf. § 8); while in computational applications, the choice is usually made so that evaluating the maximand of (D) for fixed $u \geqq 0$ is significantly easier than solving (P) itself (cf. [17], [19], [29]).

1.3. Preview and summary of results. Section 2 presents the fundamental optimality and duality results as three theorems. Theorem 1 is the *optimality theorem.* Its first assertion is that if (P) has an optimal solution, then an optimal multiplier vector exists if and only if (P) has a property called *stability*, which means that the perturbation function mentioned above does not decrease infinitely steeply in any perturbation direction. Stability plays the role customarily assumed in Kuhn–Tucker type optimality theorems by some type of "constraint qualification." Because stability is necessary as well as sufficient for an optimal multiplier vector to exist, it is evidently implied by every known constraint qualification. It turns out to be a rather pleasant property to work with mathematically, and to interpret in many problems. The second assertion of the optimality theorem is that the optimal multiplier vectors are precisely the negatives of the subgradients of the perturbation function at the point of no perturbation. An immediate consequence is a rigorous interpretation of an optimal multiplier vector in terms of quasi "prices."

Theorem 2 is the customary *weak duality theorem*, which asserts that the infimal value of (P) cannot be smaller than the supremal value of (D). Although nearly trivial to show, it does have several uses. For instance, it implies that (P) must be infeasible if the supremal value of (D) is $+\infty$.

Theorem 3 is the powerful *strong duality theorem*: If (P) is stable, then (D) has an optimal solution; the optimal values of (P) and (D) are equal; the optimal solutions of (D) are essentially the optimal multiplier vectors for (P); and any optimal solution of (D) permits recovery of all optimal solutions of (P) (if such exist) as the minimizers over X of the corresponding Lagrangean function which also satisfy $g(x) \leqq 0$ and the usual complementary slackness condition. Note that stability plays a central role here, just as it does in the optimality theorem. It makes (D) quite inviting as a surrogate problem for (P), and precludes the possibility of a "duality gap"—inequality between the optimal values of (P) and (D)—whose presence would render (D) useless in many, if not most, potential applications.

Section 3 gives complete proofs of these key results. The main construct is the perturbation function already mentioned. By systematically exploiting its convexity, no advanced methods or results are needed to achieve a direct and unified development of the optimality and strong duality theorems. No differentiability or even continuity assumptions need be made, and X need not be closed or open.

Section 4 develops a useful geometric portrayal of the dual problem which permits construction of simple examples to illustrate the various relationships that can obtain between (P) and (D). It also yields geometric insights which suggest some of the further theoretical results developed in the next section.

Section 5 establishes six additional theorems, numbers 4 through 9, concerning (P) and (D). *Theorem* 4 asserts that the maximand of (D) has value $-\infty$ for all $u \geqq 0$ if and only if the right-hand side of (P) can be perturbed so as to yield an infimal value of $-\infty$. *Theorem* 5 asserts that if (P) is infeasible and yet the supremal value of (D) is finite, then some arbitrarily small perturbation of the right-hand side of (P) will restore it to feasibility. *Theorem* 6 amounts to the statement that (P) is stable if it satisfies Slater's qualification that there exist a point $x^0 \in X$

such that $g_i(x^0) < 0$, $i = 1, \cdots, m$. The next result, called the *continuity theorem*, gives a key necessary and sufficient condition for the optimal values of (P) and (D) to be equal: the perturbation function must be lower semicontinuous at the origin. (Stability is just a sufficient condition, in view of the strong duality theorem, unless further assumptions are made.) *Theorem* 8 provides useful sufficient conditions for the perturbation function to be lower semicontinuous at the origin; and, therefore, for a duality gap to be impossible: f and g continuous, X closed, the infimal value of (P) finite, and the set of ε-optimal solutions nonempty and bounded for some $\varepsilon > 0$. The final result of § 5, called the *converse duality theorem*, is an important companion to the strong duality theorem. It requires X to be closed and f and g to be continuous on X; and asserts that if (D) has an optimal solution and the corresponding Lagrangean function has a unique minimizer x^* over X which is also feasible in (P), then (P) is stable and x^* is the unique optimal solution. Actually, as the discussion indicates, the hypothesis concerning the uniqueness of the minimizer of the Lagrangean can be weakened.

Section 6 examines the relationships between the results of §§ 1 to 5 and previous work. After discussing the specialization to the completely linear case, a detailed comparison is made to several key papers representative of the major approaches previously applied to nonlinear duality theory. These are: Dorn [8], which exploits the special properties of quadratic programs; Wolfe [34], which applies the differential calculus and the classical results of Kuhn and Tucker; Stoer [28] and Mangasarian and Pontstein [25], which apply general minimax theorems; and Rockafellar [26], which applies conjugate convex function theory. This comparison is favorable to the methods and results of the present study.

Section 7 discusses numerical considerations of interest if (D) is to be used computationally to solve (P). Such questions have been almost totally ignored in previous studies, but must be examined if nonlinear duality theory is to be applied computationally. After first indicating some of the pitfalls stability precludes, we briefly survey the two main approaches that have been followed for optimizing (D); and, subsequently, the main topic of whether and how an optimal or near-optimal solution of a stable program can be obtained from an optimal or near-optimal solution of its dual. A result is given in *Theorem* 10 from which it follows that no particular numerical difficulties exist in solving (P), provided that an exactly optimal solution of (D) can be found. However, if only a sequence converging to an optimal solution u^* of (D) can be found, the situation appears to turn on whether the Lagrangean function corresponding to u^* has a unique minimizer over X. If it does, *Theorem* 11 shows that the situation is manageable, at least when X is compact and f and g are continuous on X. Otherwise, the situation can be quite difficult, as demonstrated by an example.

Section 8 makes the point that nonlinear duality theory can be used to prove many results in the theory of convex systems which do not appear to involve optimization at all; just as linear duality theory can be used to prove results concerning systems of linear equalities and inequalities. This provides an easy and unified approach to a substantial body of theorems. The possibilities are illustrated by presenting new proofs for three theorems. The first is a separation theorem for disjoint convex sets; the second a characterization in terms of supporting half-spaces for a certain class of convex sets generated by projection; and the third a fundamental property of a system of inconsistent convex inequalities.

Finally, in § 9 we indicate a few of the significant areas in which further work remains to be done.

1.4. Notation. The notation employed is standard. We follow the convention that all vectors are columnar unless transposed (e.g., u^t).

2. Fundamental results. After establishing some basic definitions, we state and discuss three fundamental results: the optimality theorem, weak duality theorem, and strong duality theorem. The proofs of the first and third results are deferred to § 3.

2.1. Definitions.

DEFINITION 1. The *optimal value of* (P) is the infimum of $f(x)$ subject to $x \in X$ and $g(x) \leqq 0$. The *optimal value of* (D) is the supremum of its maximand subject to $u \geqq 0$.

Problems (P) and (D) always have optimal values (possibly $\pm\infty$) whether or not they have optimal solutions—that is, whether or not there exist feasible solutions achieving these values—provided we invoke the customary convention that an infimum (supremum) taken over an empty set is $+\infty(-\infty)$.

DEFINITION 2. A vector u is said to be *essentially infeasible in* (D) if it yields a value of $-\infty$ for the maximand of (D). If every $u \geqq 0$ is essentially infeasible in (D), then (D) itself is said to be *essentially infeasible*. If (D) is not essentially infeasible, it is said to be *essentially feasible*.

The motivation for this definition is obvious: a vector $u \geqq 0$ is useless in (D) if it leads to an "infinitely bad" value of the maximand.

DEFINITION 3. A pair (x, u) is said to satisfy the *optimality conditions* for (P) if

(i) x minimizes $f + u^t g$ over X,
(ii) $u^t g(x) = 0$,
(iii) $u \geqq 0$,
(iv) $g(x) \leqq 0$.

A vector u is said to be an *optimal multiplier vector* for (P) if (x, u) satisfies the optimality conditions for some x.

An optimal multiplier vector is sometimes referred to as a "generalized Lagrange multiplier vector," or a vector of "dual variables" or "dual prices."

It is easy to verify that a pair (x, u) satisfies the optimality conditions only if x is optimal in (P). Thus the existence of an optimal multiplier vector presupposes the existence of an optimal solution of (P). The converse, of course, is not true without qualification. It also can be verified that if u is an optimal multiplier vector, then (x, u) satisfies the optimality conditions for *every* optimal solution of (P). Thus an optimal multiplier vector is truly associated with (P)—more precisely, with the optimal solution set of (P)—rather than with any particular optimal solution. On this point the traditional custom of defining an optimal multiplier vector in terms of a particular optimal solution of (P) is misleading, although it is equivalent to the definition used here.

It is perhaps worthwhile to remind the reader that the optimality conditions are equivalent to a constrained saddle point of the Lagrangean function. Specifically, one can verify that (x^*, u^*) satisfies conditions (i)–(iv) if and only if $u^* \geqq 0$, $x^* \in X$, and

$$f(x^*) + u^t g(x^*) \leqq f(x^*) + (u^*)^t g(x^*) \leqq f(x) + (u^*)^t g(x)$$

for all $u \geqq 0$ and $x \in X$. Another equivalent rendering of the optimality conditions, one which gives a glimpse of developments to come, is this: (x^*, u^*) satisfies the optimality conditions if and only if x^* is optimal in (P), u^* is optimal in (D), and the optimal values of (P) and (D) are equal.

These remarks on Definition 3 do not depend in any way on the convexity assumptions. The demonstrations are straightforward.

DEFINITION 4. The *perturbation function* $v(\cdot)$ associated with (P) is defined on R^m as

$$v(y) \triangleq \operatorname*{infimum}_{x \in X} \{f(x) \text{ subject to } g(x) \leqq y\},$$

where y is called the *perturbation vector*.

The perturbation function is convex (Lemma 1) and is the fundamental construct used to derive the relationship between (P) and (D). Evidently $v(0)$ is the optimal value of (P). Values of v at points other than the origin are also of intrinsic interest in connection with sensitivity analysis and parametric studies of (P).

DEFINITION 5. Let $\bar{y}$ be a point at which v is finite. An m-vector $\bar{\gamma}$ is said to be a *subgradient* of v at $\bar{y}$ if [1]

$$v(y) \geqq v(\bar{y}) + \bar{\gamma}^t(y - \bar{y}) \quad \text{for all } y.$$

Subgradients generalize the concept of a gradient and are a technical necessity since v is usually not everywhere differentiable. Their role is made even more important by the fact that they turn out to be the negatives of the optimal multiplier vectors (Lemma 3). An important criterion for their existence (Lemma 2) is given by the property in the next definition (cf. Gale [15]).

DEFINITION 6. (P) is said to be *stable* if $v(0)$ is finite and there exists a scalar $M > 0$ such that

$$\frac{v(0) - v(y)}{\|y\|} \leqq M \quad \text{for all } y \neq 0.$$

Stability is an easy property to understand intuitively. We shall see that it is implied by all known constraint qualifications for (P). It can be interpreted as a Lipschitz condition on the function v. If it fails to hold, then the ratio of improvement in the infimal value of (P) to the amount of perturbation can be made as large as desired (the particular norm $\|\cdot\|$ used to measure the amount of perturbation is immaterial). This is also true in the marginal sense, that is, with the

[1] The direction of this inequality would be reversed if v were concave rather than convex.

perturbations made as small as desired, as follows from the convexity of v. A consequence of this observation is that the following alternative definition of stability (used by Rockafellar) is equivalent to the one above.

DEFINITION 6′. (P) is said to be *stable* if v is finite at 0 and does not decrease infinitely steeply in any perturbation direction; that is, if

$$\lim_{\theta \to 0+} \left[\frac{v(0) - v(\theta y)}{\theta \|y\|} \right] < \infty \quad \text{for all } y \neq 0.$$

The limit defined is the negative of the directional derivative of v in the perturbation direction y ($\pm \infty$ are allowed as limits).

2.2. Optimality. Although the main focus of this study is duality theory, an inevitable by-product of the present approach is what must surely be near the ultimate of Kuhn–Tucker type optimality theorems. (Cf. Gale [15, Theorem 3] and Rockafellar [26].) The following theorem also gives a key characterization of optimal multiplier vectors.

THEOREM 1 (Optimality). *Assume that* (P) *has an optimal solution. Then an optimal multiplier vector exists if and only if* (P) *is stable; and u is an optimal multiplier vector for* (P) *if and only if* $(-u)$ *is a subgradient of* v *at* $y = 0$.

Part of the content of this theorem is the result that, if x^* is an optimal solution of (P), and (P) is stable, then there exists a vector u^* such that (x^*, u^*) satisfies the optimality conditions for (P). It is well known that some qualification of (P) is needed for this result to hold; stability plays the role of such a qualification here. It bears emphasizing, however, that the theorem reveals stability to be not only a sufficient qualification for this purpose, but also a *necessary* one. Thus, stability is implied by every "constraint qualification" ever used to prove the necessity of the optimality conditions. For example, Slater's constraint qualification that there exist a point $x^0 \in X$ such that $g_i(x^0) < 0$ for all i implies stability, as does the original Kuhn–Tucker constraint qualification. For a discussion of these and many other "classical" qualifications, see Mangasarian [24].

It is striking that none of the classical qualifications emphasizes the role of the objective function, although a careful reading reveals that each requires the objective function to be within a particular general class (e.g., defined on all of R^n, or differentiable on an open set containing the feasible region). Of course, if the objective function is sufficiently well-behaved, (P) will be stable no matter how poorly behaved the constraints are (e.g., if f is constant on X then (P) is obviously stable for any constraint set as long as it is feasible). On the other hand, it is possible for an objective function to be so poorly behaved that (P) is unstable even if all constraints are linear. For example [15], put $n = m = 1$, $f(x) = -\sqrt{x}$, $X = \{x : x \geqq 0\}$, $g_1(x) = x$; then by perturbing the right-hand side positively, the ratio

$$\frac{0 - (-\sqrt{y})}{|y|} = \frac{1}{\sqrt{y}}$$

in Definition 6 can be made as large as desired by making the perturbation amount

y sufficiently small. The obvious trouble with the objective function is that it has infinite steepness at $x = 0$.

One further useful and somewhat surprising observation on the concept of stability is in order. Namely, to verify stability it is actually necessary and sufficient to consider only a *one-dimensional* choice of y: (P) is stable if and only if $v(0)$ is finite and there exists a scalar $M > 0$ such that

$$\frac{v(0) - v(\xi, \cdots, \xi)}{\xi} \leqq M \quad \text{for all } \xi > 0.$$

The proof makes use of the fact that stability according to Definition 6 does not depend on which particular norm is used (this follows from the fact that if $\|\cdot\|_1$ and $\|\cdot\|_2$ are two different norms on R^m, then there exists a scalar $r > 0$ such that $(\|y\|_1/\|y\|_2) \geqq r$ for all nonzero $y \in R^m$), so that the Chebyshev norm

$$\|y\|_T \equiv \text{maximum } \{|y_1|, \cdots, |y_m|\}$$

can be used in Definition 6. Thus stability of (P) implies the above inequality simply by taking y of the form $(\xi, \cdots, \xi)$. Conversely, the above inequality implies that (P) is stable because

$$\frac{v(0) - v(y)}{\|y\|_T} \leqq \frac{v(0) - v(\|y\|_T, \cdots, \|y\|_T)}{\|y\|_T}, \qquad \text{all } y \neq 0,$$

by the fact that v is a nonincreasing function.

The second part of Theorem 1 gives a characterization of optimal multiplier vectors useful for sensitivity analysis and other purposes. Suppose that u^* is any optimal multiplier vector of (P) and a perturbation in the direction y' is contemplated; that is, we are interested in the perturbed problem:

$$\underset{x \in X}{\text{Minimize}} \quad f(x) \quad \text{subject to} \quad g(x) \leqq \theta y'$$

for $\theta \geqq 0$. From Theorem 1,

$$v(\theta y') \geqq v(0) - \theta(u^*)^t y' \quad \text{for all } \theta \geqq 0.$$

We thus have a lower bound on the optimal value of the perturbed problem, and by taking limits we can obtain a bound on the right-hand derivative of $v(\theta y')$ at $\theta = 0$;

$$\frac{d^+ v(\theta y')}{d\theta} \geqq -(u^*)^t y'.$$

In particular, with y' equal to the jth unit vector, we obtain the well-known result that $-u_j^*$ is a lower bound on the marginal rate of change of the optimal value of (P) with respect to an increase in the right-hand side of the jth constraint. All of this follows from the fact that any optimal multiplier vector is the negative of a subgradient of v at 0. The converse of this is also significant, as it follows that the set of all subgradients—and hence all directional derivatives of v—can be characterized in terms of the set of all optimal multiplier vectors for (P). Further details and an application to the optimization of multidivisional systems are given elsewhere [16, § 4.2].

2.3. Duality. We begin with a very easy result that has several useful consequences.

THEOREM 2 (Weak duality). *If $\bar{x}$ is feasible in* (P) *and $\bar{u}$ is feasible in* (D), *then the objective function of* (P) *evaluated at $\bar{x}$ is not less than the objective function of* (D) *evaluated at $\bar{u}$.*

To demonstrate this result, one need only write the obvious inequalities

$$\text{infimum}\,\{f(x) + \bar{u}^t g(x) | x \in X\} \leqq f(\bar{x}) + \bar{u}^t g(\bar{x}) \leqq f(\bar{x}).$$

The convexity assumptions are not needed.

One consequence is that any feasible solution of (D) provides a lower bound on the optimal value of (P); and any feasible solution of (P) provides an upper bound on the optimal value of (D). This can be useful in establishing termination or error-control criteria when devising computational algorithms addressed to (P) or (D); if at some iteration feasible solutions are available to both (P) and (D) that are "close" to one another in value, then they must be "close" to being optimal in their respective problems. In Theorem 3 we shall see that there exist feasible solutions to (P) and (D) that are as close to one another in value as desired, provided only that (P) is stable.

From Theorem 2, it also follows that (D) must be essentially infeasible if the optimal value of (P) is $-\infty$ and, similarly, (P) must be infeasible if the optimal value of (D) is $+\infty$.

THEOREM 3 (Strong duality). *If* (P) *is stable, then*

(a) (D) *has an optimal solution,*

(b) *the optimal values of* (P) *and* (D) *are equal,*

(c) *u^* is an optimal solution of* (D) *if and only if $-u^*$ is a subgradient of v at $y = 0$,*

(d) *every optimal solution u^* of* (D) *characterizes the set of all optimal solutions (if any) of* (P) *as the minimizers of $f + (u^*)^t g$ over X which also satisfy the feasibility condition $g(x) \leqq 0$ and the complementary slackness condition $(u^*)^t g(x) = 0$.*

Conclusions (a) and (d) justify taking a dual approach to the solution of (P). It is perhaps surprising that *all* optimal solutions of (P) can be found from any *single* optimal solution of (D). Another way of phrasing (d) would be to say that if u^* is optimal in (D), then x is optimal in (P) if and only if (x, u^*) satisfies optimality conditions (i), (ii) and (iv) (see Definition 3). In § 7 we shall take up at some length the matter of approaching the computational solution of (P) via its dual.

Conclusion (b) precludes the existence of what is often referred to as a *duality gap* between the optimal values of (P) and (D). Most applications of nonlinear duality theory require that there be no duality gap (e.g., see § 8 and [18]).

Conclusion (c) reveals the connection between the set of optimal solutions of (D) and the perturbation function. If (P) has an optimal solution as well as the property of stability, then, using Theorem 1, we obtain an alternative interpretation of the optimal solution set of (D): it is precisely the set of optimal multiplier vectors for (P).

It is perhaps worth noting that Lemma 4 in the next section shows that conclusion (c) holds under a slightly weaker assumption than stability, namely, when $v(0)$ is finite and the optimal values of (P) and (D) are equal.

Additional results concerning (P) and (D) are given in § 5. Relationships to known results are taken up in § 6.

3. Proof of the optimality and strong duality theorems. It is convenient to subdivide the proofs of Theorems 1 and 3 into five lemmas.

It will be necessary to refer to the set Y of all vectors y for which the perturbed problem is feasible:

$$Y \triangleq \{y \in R^m : g(x) \leqq y \text{ for some } x \in X\}.$$

Obviously, $v(y) = \infty$ if and only if $y \notin Y$.

The first lemma establishes that the perturbation function is convex on Y. This well-known result is the cornerstone of the entire development. It also is obvious that v is nonincreasing.

LEMMA 1. *Y is a convex set, and v is a convex function on Y.*

Proof. The convexity of Y follows directly from the convexity of X and the convexity of g. Since $-\infty$ is permitted as a value for v on Y, the appropriate definition of convexity for v is in terms of its epigraph [26]: $\{(y, \mu) \in R^{m+1} : y \in Y$ and $\mu \geqq v(y)\}$. Let (y^0, μ^0) and (y', μ') be arbitrary points in this set, and let θ be an arbitrary scalar between 0 and 1. Define $\bar{\theta} = 1 - \theta$. Then

$$\begin{aligned} v(\theta y^0 + \bar{\theta} y') &= \underset{x^0, x' \in X}{\text{infimum}}\, f(\theta x^0 + \bar{\theta} x') \quad \text{subject to } g(\theta x^0 + \bar{\theta} x') \leqq \theta y^0 + \bar{\theta} y' \\ &\leqq \underset{x^0, x' \in X}{\text{infimum}}\, f(\theta x^0 + \bar{\theta} x') \quad \text{subject to } g(x^0) \leqq y^0, g(x') \leqq y' \\ &\leqq \underset{x^0, x' \in X}{\text{infimum}}\, \theta f(x^0) + \bar{\theta} f(x') \quad \text{subject to } g(x^0) \leqq y^0, g(x') \leqq y' \\ &= \theta v(y^0) + \bar{\theta} v(y') \leqq \theta \mu^0 + \bar{\theta} \mu', \end{aligned}$$

where the equality or inequality relations follow, respectively, from the convexity of X, the convexity of g, the convexity of f, separability in x^0 and x', and the definitions of μ^0 and μ'. Thus the point $\theta(y^0, \mu^0) + \bar{\theta}(y', \mu')$ is in the epigraph of v, and so v must be convex on Y. This completes the proof.

Many properties of v follow directly from its convexity. For example, v must be continuous on the interior of any set on which it is finite; it must have value $-\infty$ everywhere on the interior of Y if it is $-\infty$ anywhere; its directional derivative (see Definition 6′) must exist in every direction at every point where it is finite; it must have a subgradient at every interior point of Y at which it is finite; and it must be differentiable at a given point in Y if and only if it has a unique subgradient there. These properties hold, not only for v, but for any convex function (see, e.g., [11] or [26, § 2]).

The following property is an important criterion for the existence of a subgradient of a convex function at a point where it is finite. We offer a proof to keep the development self-contained; the method of proof is due to Gale [15].

LEMMA 2. *Let $\phi(\cdot)$ be a convex function on a convex set $Y \subseteq R^m$ taking values in $R \cup \{-\infty\}$. Let $\|\cdot\|$ be any norm on R^m, and let $\bar{y}$ be a point at which ϕ is finite. Then ϕ has a subgradient at $\bar{y} \in Y$ if and only if there exists a positive scalar M such that*

$$\frac{\phi(\bar{y}) - \phi(y)}{\|y - \bar{y}\|} \leqq M \quad \textit{for all } y \in Y \quad \textit{such that } y \neq \bar{y}.$$

Proof. Suppose that ϕ has a subgradient $\bar{\gamma}$ at $\bar{y} \in Y$; i.e.,

$$\phi(y) \geqq \phi(\bar{y}) + \bar{\gamma}^t(y - \bar{y}) \quad \text{for all } y \in Y.$$

Then $\phi(y) > -\infty$ on Y, and upon rearranging and dividing by $\|y - \bar{y}\|$ (we shall use the Euclidean norm, although any norm will do) we obtain

$$\frac{\phi(\bar{y}) - \phi(y)}{\|y - \bar{y}\|} \leqq \frac{-\bar{\gamma}^t(y - \bar{y})}{\|y - \bar{y}\|} \quad \text{for all } y \in Y \quad \text{such that } y \neq \bar{y}.$$

Since the right-hand side does not exceed $\|\bar{\gamma}\|$ for any y, we obtain the desired inequality with $M = \|\bar{\gamma}\|$. (If $\|\bar{\gamma}\| = 0$, the desired inequality holds with M equal to any positive number.)

Now suppose that there exists a positive scalar M such that the stated inequality holds. We must show that there exists a subgradient of ϕ at $\bar{y}$. Since $\phi(y) > -\infty$ on Y, we may define the sets

$$\Phi \triangleq \{(y, z) \in R^{m+1} : y \in Y \text{ and } \phi(\bar{y}) - \phi(y) \geqq z\},$$
$$\Psi \triangleq \{(y, z) \in R^{m+1} : M\|y - \bar{y}\| < z\}.$$

It is easy to see that Φ and Ψ are convex sets, that $\Phi \cap \Psi$ is empty, and that Ψ is open. From elementary results on the separation of nonintersecting convex sets, it follows that Φ and Ψ can be separated by a hyperplane that does not intersect Ψ; that is, there exist an m-vector γ and scalars ρ and α such that

$$\gamma^t y + \rho z \geqq \alpha \quad \text{for } (y, z) \in \Phi, \qquad \gamma^t y + \rho z < \alpha \quad \text{for } (y, z) \in \Psi,$$

Now $(\bar{y}, 0) \in \Phi$, so $\gamma^t \bar{y} \geqq \alpha$. Actually $\gamma^t \bar{y} = \alpha$, for $\gamma^t \bar{y} \leqq \alpha$ follows from the second inequality and the fact that $(\bar{y}, \varepsilon) \in \Psi$ for all $\varepsilon > 0$. Thus the inequalities become

$$\gamma^t(y - \bar{y}) + \rho z \geqq 0 \quad \text{for } (y, z) \in \Phi,$$
$$\gamma^t(y - \bar{y}) + \rho z < 0 \quad \text{for } (y, z) \in \Psi.$$

Since $(\bar{y}, 1) \in \Psi$, we have $\rho < 0$. Put $\bar{\gamma} = -\gamma^t/\rho$. Then the first inequality becomes $\bar{\gamma}^t(y - \bar{y}) \geqq z$ whenever $(y, z) \in \Phi$, that is, whenever $y \in Y$ and $\phi(\bar{y}) - \phi(y) \geqq z$. Putting $z = \phi(\bar{y}) - \phi(y)$, we have

$$\bar{\gamma}^t(y - \bar{y}) \geqq \phi(\bar{y}) - \phi(y) \quad \text{whenever } y \in Y,$$

which says precisely that $-\bar{\gamma}$ is a subgradient of ϕ at $y = \bar{y}$. This completes the proof.

The following known result is equivalent: a convex function has a subgradient at a given point where it is finite if and only if its directional derivative is not $-\infty$ in any direction (cf. [11, p. 84], [26, p. 408]). This result would be used in place of Lemma 2 if Definition 6′ were used in place of Definition 6.

The next lemma establishes a crucially important alternative interpretation of optimal multiplier vectors.

LEMMA 3. *If* (P) *has an optimal solution, then u is an optimal multiplier vector for* (P) *if and only if $-u$ is a subgradient of v at $y = 0$.*

Proof. Suppose that u^* is an optimal multiplier vector for (P). Then there is a vector x^* such that (x^*, u^*) satisfies the optimality conditions for (P). From optimality conditions (i) and (ii) we have $f(x) \geqq f(x^*) - (u^*)^t g(x)$ for all $x \in X$.

It follows, using (iii), that for each point y in Y,

$$f(x) \geqq f(x^*) - (u^*)^t y \quad \text{for all } x \in X \quad \text{such that } g(x) \leqq y.$$

Taking the infimum of the left-hand side of this inequality over the indicated values of x yields

$$v(y) \geqq f(x^*) + (-u^*)^t y \quad \text{for all } y \in Y.$$

It is now evident that $-u^*$ satisfies the definition of a subgradient of v at 0 (since $f(x^*) = v(0)$ and $v(y) = \infty$ for $y \notin Y$). This completes the first half of the proof.

Now let $-u^*$ be any subgradient of v at 0. We shall show that (x^*, u^*) satisfies optimality conditions (i)–(iv), where x^* is any optimal solution of (P). Condition (iv) is immediate. The demonstration of the remaining conditions follows easily from the definitional inequality of $-u^*$, namely,

$$v(y) \geqq v(0) - (u^*)^t y \quad \text{for all } y.$$

To establish (iii), put $y = e_j$, the jth unit m-vector (all components of e_j are 0 except for the jth, which is 1). Then $v(e_j) \geqq v(0) - u_j^*$, or $u_j^* \geqq v(0) - v(e_j)$. Since v is obviously a nonincreasing function of y, we have $v(0) \geqq v(e_j)$ and therefore $u_j^* \geqq 0$. Condition (ii) is established in a similar manner by putting $y = g(x^*)$. This yields $(u^*)^t g(x^*) \geqq v(0) - v(g(x^*)) = 0$, where the equality follows from the fact that decreasing the right-hand side of (P) to $g(x^*)$ will not destroy the optimality of x^*. Hence $(u^*)^t g(x^*) \geqq 0$. But the reverse inequality also holds by (iii) and the feasibility of x^* in (P), and so (ii) must hold. To establish condition (i), put $y = g(x)$, where x is any point in X. Then

$$v(g(x)) \geqq v(0) - (u^*)^t g(x) \quad \text{for all } x \in X.$$

Since $f(x) \geqq v(g(x))$ for all $x \in X$ and $v(0) = f(x^*)$, we have

$$f(x) \geqq f(x^*) - (u^*)^t g(x) \quad \text{for all } x \in X.$$

In view of (ii), this is precisely condition (i). This completes the proof.

We now have all the ingredients necessary for the optimality theorem.

Proof of Theorem 1. The perturbation function is convex on Y, by Lemma 1, and is finite at $y = 0$ because (P) is assumed to possess an optimal solution. By Lemma 2, therefore, v has a subgradient at $y = 0$ if and only if (P) is stable. But subgradients of v at $y = 0$ and optimal multiplier vectors for (P) are negatives of one another by Lemma 3. The conclusions of Theorem 1 are now at hand.

It is worth digressing for a moment to note that the hypotheses of Lemma 3 and Theorem 1 could be weakened slightly if the definition of an optimal multiplier vector were generalized so that it no longer presupposes the existence of an optimal solution of (P). In particular, the conclusions of Lemma 3 and Theorem 1 hold even if (P) does not have an optimal solution, provided that $v(0)$ is finite and the concept of an optimal multiplier vector is redefined as follows. A point u is said to be a *generalized optimal multiplier* vector if for every scalar $\varepsilon > 0$ there exists a point x_ε such that (x_ε, u) satisfies the *ε-optimality conditions*: x is an ε-optimal minimizer of $f + u^t g$ over X, $u^t g(x) \geqq \varepsilon$, $u \geqq 0$, and $g(x) \leqq 0$. The necessary modification of the proof of Lemma 3 is straightforward.

The entire development of this paper could be carried out in terms of this more general concept of an optimal multiplier vector. The optimality conditions for (P), for example, would be stated in terms of a pair $(\langle x^\nu\rangle, u)$ in which the first member is a sequence rather than a single point. Such a pair would be said to satisfy the *generalized optimality conditions* for (P) if, for some nonnegative sequence $\langle \varepsilon^\nu\rangle$ converging to 0, for each ν the pair (x^ν, u) satisfies the ε^ν-optimality conditions. This generalization of the traditional optimality conditions given in Definition 3 seems to be the most natural one, when the existence of an optimal solution of (P) is in question. It can be shown that if u is a generalized optimal multiplier vector then $(\langle x^\nu\rangle, u)$ satisfies the generalized optimality conditions if and only if $\langle x^\nu\rangle$ is a sequence of feasible solutions of (P) converging in value to $v(0)$.

Although such a development might well be advantageous for some purposes, we elect not to pursue it here.[2]

The next lemma establishes (in view of the previous one) the connection between optimal multiplier vectors and solutions to the dual problem.

LEMMA 4. *Let $v(0)$ be finite. Then u is an optimal solution of* (D) *and the optimal values of* (P) *and* (D) *are equal if and only if $-u$ is a subgradient of v at $y = 0$.*

Proof. First we demonstrate the "if" part of the lemma. Let $(-u)$ be a subgradient of v at $y = 0$; that is, let u satisfy

$$v(y) \geqq v(0) - u^t(y - 0) \quad \text{for all } y.$$

The proof of Lemma 3 shows that $u \geqq 0$ follows from this inequality. Thus u is feasible in (D). Substituting $g(x)$ for y, and noting that $f(x) \geqq v(g(x))$ holds for all $x \in X$ yields

$$f(x) + u^t g(x) \geqq v(0) \quad \text{for all } x \in X.$$

Taking the infimum over $x \in X$, we obtain

$$\underset{x \in X}{\text{infimum}} \{ f(x) + u^t g(x) \} \geqq v(0).$$

It now follows from the weak duality theorem that u must actually be an optimal solution of (D), and that the optimal value of (D) equals $v(0)$. This completes the first part of the proof.

To demonstrate the "only if" part of the lemma, let u be an optimal solution of (D). By assumption,

$$[\underset{x \in X}{\text{infimum}}\ f(x) + u^t g(x)] = v(0).$$

Since $u^t g(x) \leqq u^t y$ for all $x \in X$ and y such that $g(x) \leqq y$ (remember that $u \geqq 0$), it follows that

$$f(x) + u^t y \geqq v(0)$$

for all $x \in X$ and y such that $g(x) \leqq y$. For each $y \in Y$, we may take the infimum

[2] K. O. Kortanek has pointed out in a private communication (Dec. 8, 1969) that the concept of an optimal multiplier vector can be generalized still further by considering a sequence $\langle u^\nu\rangle$ of m-vectors and appropriately defining "asymptotic" optimality conditions (cf. [21]). This would permit a kind of optimality theory for many unstable problems (such as the example in § 2.2), although interpretations in terms of subgradients of v at $y = 0$ are obviously not possible.

of this inequality over $x \in X$ such that $g(x) \leqq y$ to obtain

$$v(y) - (-u)^t y \geqq v(0) \quad \text{for } y \in Y.$$

This inequality holds outside of Y as well, since $v(y) = \infty$ there. Thus $(-u)$ satisfies the definition of a subgradient of v at 0, and the proof is complete.

The final lemma characterizes the (possibly empty) optimal solution set of (P).

LEMMA 5. *Assume that v is finite at $y = 0$ and that γ is a subgradient at this point. Then x^* is an optimal solution of* (P) *if and only if* $(x^*, -\gamma)$ *satisfies optimality conditions* (i), (ii) *and* (iv) *for* (P).

Proof. Let x^* be an optimal solution of (P). By Lemma 3, $(-\gamma)$ must be an optimal multiplier vector for (P); and so $(x^*, -\gamma)$ must satisfy the optimality conditions for (P) (see the discussion following Definition 3). This proves the "only if" part of the conclusion.

Now let $(x^*, -\gamma)$ satisfy (i), (ii) and (iv). The proof of Lemma 4 shows that $-\gamma \geqq 0$, and so (iii) is also satisfied. Hence x^* must be an optimal solution of (P). This completes the proof.

We are now able to prove the strong duality theorem.

Proof of Theorem 3. Since (P) is stable, $v(0)$ is finite and we conclude from Lemmas 1 and 2 that v has a subgradient at $y = 0$. Parts (a), (b) and (c) of the theorem now follow immediately from Lemma 4. Part (d) follows immediately from Lemma 5 with the help of part (c).

4. Geometrical interpretations and examples. It is easy to give a useful geometric portrayal of the dual problem (cf. [20], [23, p. 223], [31]). This yields insight into the content of the definitions and theorems of § 2, permits construction of various pathological examples, and even suggests a number of additional theoretical results. We need to consider in detail only the case $m = 1$.

The geometric interpretation focuses on the image of X under f and g, that is, on the image set

$$I \triangleq \{(z_1, z_2) \in R^2 : z_1 = g(x) \text{ and } z_2 = f(x) \text{ for some } x \in X\}.$$

Figure 1 illustrates a typical problem in which x is a scalar variable. The point P^* is obviously the image of the optimal solution of problem (P); that is,

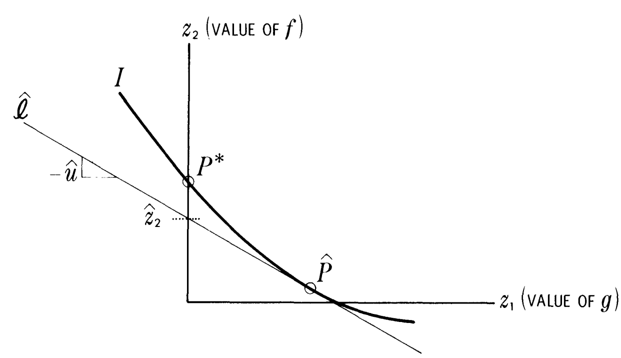


FIG. 1

$P^* = (g(x^*), f(x^*))$, where x^* minimizes f subject to $g(x) \leqq 0$ and $x \in X$. Thus the geometric interpretation of problem (P) is obvious: Find the point in I which minimizes z_2 subject to $z_1 \leqq 0$.

Consider now a particular value $\hat{u} \geqq 0$ for the scalar variable of (D). To evaluate the maximand of (D) at $\hat{u}$ one must minimize $f + \hat{u}g$ over X. This is the same as minimizing $z_2 + \hat{u}z_1$ subject to $(z_1, z_2) \in I$; as the line $z_2 + \hat{u}z_1 =$ const. has slope $-\hat{u}$ in Fig. 1, we see that evaluating the maximand of (D) at $\hat{u}$ amounts to finding the lowest line with slope $-\hat{u}$ which intersects I. This leads to the line $\hat{\ell}$ tangent to I at $\hat{P}$, pictured in Fig. 1. The point $\hat{P}$ is the image of the minimizer of $f + \hat{u}g$ over X. The minimum value of $f + \hat{u}g$ is the value of z_2 where $\hat{\ell}$ intercepts the ordinate, namely, $\hat{z}_2$ in Fig. 1 (since $(0, \hat{z}_2) \in \hat{\ell}$). The geometric interpretation of (D) is now apparent: Find that value of $\hat{u}$ which defines the slope of a line tangent to I intersecting the ordinate at the highest possible value. Or, more loosely, choose $\hat{u}$ to maximize $\hat{z}_2$. In Fig. 1, this leads to a value of u which defines a line tangent to I at P^*.

The geometric interpretation of (P) and (D) helps to clarify the content of Theorems 1, 2 and 3. The problem pictured in Fig. 1, for example, is obviously stable. In the neighborhood of $y = 0$, $v(y)$ is just the z_2-coordinate of I when z_1 equals y; and this coordinate does not decrease infinitely steeply as y deviates from 0. The subgradient of v at $y = 0$ is precisely the slope of the line tangent to I at P^*, which from Definition 3 is seen to be the negative of the optimal multiplier vector. This verifies the conclusion of Theorem 1 for this example. The geometrical verification of Theorems 2 and 3 is so easy as not to require comment here.

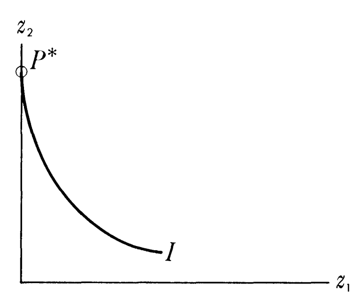


FIG. 2

An example of an unstable problem is given in Fig. 2, in which I is tangent to the ordinate at the point P^*. The value of v decreases infinitely steeply as y begins to increase above 0, and so there can be no subgradient at $y = 0$. The only line tangent to I at P^* is vertical. This checks with Theorem 1. Theorem 2 obviously holds, and Theorem 3 does not apply. The dual has optimal value equal to that of the primal, but no finite u achieves it.

This concludes the discussion of the geometrical interpretation of the definitions and theorems when $m = 1$. The generalization for $m > 1$ is conceptually straightforward, although it may be helpful to think in terms of the following convex set instead of I itself:

$$I^+ \triangleq \{(z_1, \cdots, z_{m+1}) \in R^{m+1} : z_i \geqq g_i(x), i = 1, \cdots, m, \\ \text{and } z_{m+1} \geqq f(x) \text{ for some } x \in X\}.$$

Using I^+ in place of I does not change in any way the geometrical interpretations given for (P) or (D). The line $\hat{\ell}$ now becomes, for each choice of $\hat{u}$, a supporting hyperplane (when $m > 1$) of I^+.

Now we shall put our geometrical insight to work by constructing enough pathological examples to show that the cases allowed by Theorems 1, 2 and 3 can actually occur. These examples are displayed on Diag. 1. The image set I is portrayed in each case. A dashed line indicates a missing boundary, and an arrowhead indicates that I continues indefinitely in the given direction. Although the examples are given geometrically, it is easy to write down corresponding expressions for X, f, and g. Only a single variable is needed in Examples 1, 6, 9 and 10. In Examples 1, 6 and 9, simply identify x with z_1; let $g(x) = x$; $f(x) = z_2(x)$, where $z_2(z_1)$ is the z_2-coordinate of I for a given value of z_1; and let X be the interval of z_1 values assumed by points in I. In Example 9, for instance, the corresponding problem (P) might be:

$$\underset{1 \leqq x \leqq 3}{\text{Minimize}} \quad x \quad \text{subject to} \quad x \leqq 0.$$

For Example 10, one may put $f(x) = -x$, $g(x) \equiv +1$, and $X = [-1, \infty)$. The remaining examples require two variables: identify x_1 with z_1 and x_2 with z_2, let $g(x_1, x_2) = x_1$ and $f(x_1, x_2) = x_2$; and put X equal to I. In Example 3, for instance, we obtain for (P) the problem:

$$\underset{(x_1, x_2) \in X}{\text{Minimize}} \quad x_2 \quad \text{subject to} \quad x_1 \leqq 0,$$

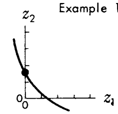
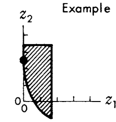
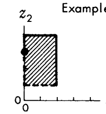
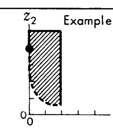
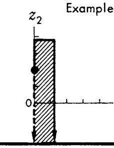
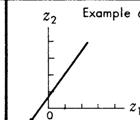
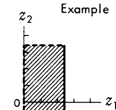
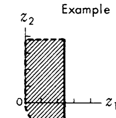
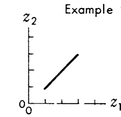
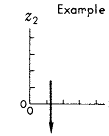
where

$$X \triangleq \{(x_1, x_2) : 0 \leqq x_1 \leqq 2, 1 < x_2 \leqq 4, \text{ and } x_2 \geqq 3 \text{ if } x_1 = 0\}.$$

In Diag. 1, we have not distinguished whether or not (P) has an optimal solution when $v(0)$ is finite. There does exist an optimum solution of (P) in each of Examples 1–5; it is denoted in each case by a heavy dot. It is easy to modify each of these examples so that no optimum solution exists: simply delete the dot from I in Examples 2–5, and delete the dot and the part of I to its right in Example 1. This shows that the cases allowed for (P) in Diag. 1 when its optimal value is finite can occur either with or without the existence of an optimal solution for (P).

5. Additional results. The geometrical insights offered in the previous section, and particularly the geometric examples of Diag. 1, suggest a number of further results concerning the relation between (P) and (D). In this section we prove several results useful in ascertaining when certain of the cases represented by Examples 2–10 cannot occur. We also present a converse duality theorem which sharpens part (d) of the strong duality theorem.

5.1. Theorems 4–6. The first result suggested by Diag. 1 is a criterion for the essential infeasibility of (D).

Possible cases for the Primal problem (**P**)		Possible cases for the Dual problem (**D**)					
		Optimal value finite				Unbounded optimum: Value $(+\infty)$	Essentially infeasible
		Optimal value $= v(0)$		Optimal value $< v(0)$			
		Has optimal solution	No optimal solution	Has optimal solution	No optimal solution		
$v(0)$ finite	Stable	Example 1 	Impossible by Theorem 3	Impossible by Theorem 3	Impossible by Theorem 3	Impossible by Theorem 2 Impossible by Theorem 3	Impossible by Theorem 3
	Unstable	Impossible by Lemma 4 and Theorem 1	Example 2 	Example 3 	Example 4 	Impossible by Theorem 2	Example 5 
Unbounded optimum: $v(0) = -\infty$		Impossible by Theorem 2	Impossible by Theorem 2	Impossible by Theorem 2	Impossible by Theorem 2	Impossible by Theorem 2	Example 6 
Infeasible $v(0) = +\infty$		Obviously impossible	Obviously impossible	Example 7 	Example 8 	Example 9 	Example 10 

DIAG. 1

THEOREM 4. (D) *is essentially infeasible if and only if* $v(y) = -\infty$ *for some* y.

Proof. We shall prove the contrapositive. Suppose that (D) is essentially feasible, so that there exists a vector $\bar{u} \geqq 0$ and a scalar M such that $f(x) + \bar{u}^t g(x) \geqq M$ for all $x \in X$. Let y be an arbitrary point in Y. Then

$$f(x) \geqq f(x) + \bar{u}^t(g(x) - y) \geqq M - \bar{u}^t y$$

for all $x \in X$ such that $g(x) \leqq y$. Taking the infimum of $f(x)$ over the indicated values of x, we obtain $v(y) \geqq M - \bar{u}^t y > -\infty$. This proves that (D) is essentially feasible only if $v(y) > -\infty$ for all $y \in Y$.

Suppose now that $v(y)$ is not $-\infty$ anywhere on Y. To show that (D) is essentially feasible, it is enough to show that v has a subgradient $\bar{\gamma}$ at some point $\bar{y}$ in Y, for then, by reasoning as in the first part of the proof of Lemma 4, we may demonstrate that $-\bar{\gamma}$ is essentially feasible in (D). Let y be any point in Y, and put $\bar{y} = y + 1$, where 1 is an m-vector with each component equal to unity. Obviously, $\bar{y}$ is in the interior of Y, and $v(\bar{y})$ is finite by supposition. Since v is convex, therefore, by a known property of convex functions (see the remark following Lemma 1) it must have a subgradient at $\bar{y}$. Thus (D) is essentially feasible if $v(y) > -\infty$ for all $y \in Y$. The proof is now complete.

Since the convexity of v implies that it has value $-\infty$ at *some* point in Y if and only if it has value $-\infty$ at *every interior* point in Y, we see that (D) will be essentially infeasible if and only if $v(y) = -\infty$ on the whole interior of Y.

The next result is suggested by the apparent impossibility of altering Examples 7 and 8 of Diag. 1 so that I is strictly separated from the ordinate (i.e., so that $y = 0$ is bounded strictly away from Y; cf. [20, Theorem 2(b)]).

THEOREM 5. *If* (P) *is infeasible and the optimal value of* (D) *is finite, then* 0 *is in the closure of* Y.

Proof. Suppose, contrary to what we wish to show, that $y = 0$ is not in the closure of Y. Then, by the convexity of Y, 0 can be strictly separated from it by a hyperplane with normal p, say: $p^t y \geqq \varepsilon > 0$ for all $y \in Y$. Certainly $p \geqq 0$, for otherwise some component of y could be taken sufficiently large to violate $p^t y > 0$. Since $g(x) \in Y$ for all $x \in X$, we obtain

$$\underset{x \in X}{\text{infimum}} \{p^t g(x)\} > 0.$$

Let $u \geqq 0$ be any vector that is essentially feasible in (D), so that

$$\underset{x \in X}{\text{infimum}} \{f(x) + u^t g(x)\} > -\infty.$$

Then $u + \theta p$ is also essentially feasible in (D) for any scalar $\theta \geqq 0$, and

$$\underset{x \in X}{\text{infimum}} \{f(x) + (u + \theta p)^t g(x)\} \geqq \underset{x \in X}{\text{infimum}} \{f(x) + u^t g(x)\} + \theta \underset{x \in X}{\text{infimum}} \{p^t g(x)\}.$$

By letting $\theta \to \infty$, we obtain the contradiction that the optimal value of (D) is $+\infty$. Hence 0 must be in the closure of Y, and the proof is complete.

This shows that the cases represented by Examples 7 and 8 (namely, (P) infeasible and a finite optimal value for (D)) can occur only if some arbitrarily small perturbation of y away from 0 will restore (P) to feasibility. Of course, if Y is closed (e.g., X compact and g continuous), then these cases cannot occur.

A similar result is suggested by the conspicuous feature which Examples 2–5 of Diag. 1 have in common: the ordinate touches I but never passes through its interior. In other words, $y = 0$ is always a boundary point of Y. This turns out to be true in general when (P) has a finite optimal value but is unstable, as follows from the next theorem.

THEOREM 6. *If $v(0)$ is finite and $y = 0$ is an interior point of Y, then* (P) *is stable.*

Proof. Since v is convex and 0 is in the interior of Y, it must have a subgradient at this point (see the remark following Lemma 1). Apply Lemma 2.

The converse of this theorem is *not* true; that is, the stability of (P) does not imply that 0 is an interior point of Y. A counterexample is provided by Example 1 with the portion of I to the left of the dot removed.

The condition that 0 is in the interior of Y can be thought of as a "constraint qualification" which implies stability (provided $v(0)$ is finite). It is equivalent to the classical qualification introduced by Slater that there exist a point $x^0 \in X$ such that $g_i(x^0) < 0$, $i = 1, \cdots, m$. In this event, it follows from Theorem 6 that only the cases represented by Examples 1 and 6 of Diag. 1 can obtain.

5.2. Duality gaps and the continuity of v. There is one more result suggested by the examples of Diag. 1 that we wish to discuss. It pertains to the possibility of a duality gap, or difference in the optimal values of (P) and (D). From Diag. 1 we can immediately observe that this can happen when (P) is feasible only if (P) is unstable, although it need not happen if (P) is unstable. Geometric considerations suggest that any difference in optimal values must be due to a lack of continuity of v at $y = 0$ (a convex function can be discontinuous at points on the boundary of its domain). And indeed this important result is essentially so, as we now show.

THEOREM 7 (Continuity). *Let $v(0)$ be finite. Then the optimal values of* (P) *and* (D) *are equal if and only if v is lower semicontinuous at $y = 0$.*

Proof. The first part of the proof makes use of the function w,

$$w(y) \triangleq \underset{u \geqq 0}{\text{supremum}} \left[\underset{x \in X}{\text{infimum}}\ f(x) + u^t(g(x) - y) \right],$$

which is to be interpreted as the optimal value of the dual problem of (P) modified to have right-hand side y rather than 0. Certainly w is a convex function on R^m, for it is the pointwise supremum of a collection of functions that are linear in y.

Suppose that v is lower semicontinuous at $y = 0$. It follows that v is finite on the interior of Y. Let $\langle y^\nu \rangle$ be the following sequence in the interior of Y converging to $0: y^\nu = 1/\nu$, where 1 is an m-vector with each component equal to unity. It follows from Theorem 6 that (P), modified to have right-hand side y^ν instead of 0, must be stable. By part (b) of Theorem 3, therefore, we conclude that $v(y^\nu) = w(y^\nu)$ for all ν. We then have

$$w(0) \geqq \liminf_{\nu \to \infty} w(y^\nu) = \liminf_{\nu \to \infty} v(y^\nu) \geqq v(0) \geqq w(0),$$

where the first inequality follows from the convexity of w, the second from the lower semicontinuity of v, and the last from the weak duality theorem. Hence $v(0)$ must equal $w(0)$, the optimal value of (D), and the "only if" part of the theorem is proved.

Now assume that $v(0)$ equals the optimal value of (D). We must show that v is lower semicontinuous at $y = 0$. Suppose to the contrary that there exists a sequence $\langle y^\nu \rangle$ of points in Y converging to 0 such that $\langle v(y^\nu)\rangle \to \bar{v} < v(0)$. We may derive the contradiction $\bar{v} \geqq$ optimal value of (D) as follows. Since $f(x) + u^t g(x) \leqq f(x) + u^t y$ holds for all $u \geqq 0$, $y \in Y$ and $x \in X$ such that $g(x) \leqq y$, we may take the infimum of both sides to obtain

$$\operatorname*{infimum}_{x \in X} \{ f(x) + u^t g(x) | g(x) \leqq y \} \leqq v(y) + u^t y \quad \text{for all } u \geqq 0 \quad \text{and} \quad y \in Y.$$

It follows that

$$\left[\operatorname*{infimum}_{x \in X} f(x) + u^t g(x) \right] \leqq v(y) + u^t y \quad \text{for all } u \geqq 0 \quad \text{and } y \in Y.$$

Hence,

$$\left[\operatorname*{infimum}_{x \in X} f(x) + u^t g(x) \right] \leqq \lim_{\nu \to \infty} (v(y^\nu) + u^t y^\nu) = \bar{v} \quad \text{for all } u \geqq 0.$$

Taking the supremum over $u \geqq 0$, we obtain the desired contradiction: optimal value of (D) $\leqq \bar{v}$. This completes the proof.

Theorem 7 motivates the need for conditions which imply the lower semicontinuity of v at 0. The following result is of fundamental interest in this regard.

THEOREM 8. *Assume that X is closed, f and g are continuous on X, the optimal value of* (P) *is finite, and $\{x \in X : g(x) \leqq 0$ and $f(x) \leqq \alpha\}$ is bounded and nonempty for some scalar $\alpha \geqq v(0)$. Then v is lower semicontinuous at $y = 0$.*

The boundedness hypothesis is obviously satisfied if, as is frequently the case in applications, the feasible region of (P) is bounded (put $\alpha = v(0) + 1$). If the boundedness of the feasible region is in question but (P) has an optimal solution, then the hypothesis holds if the optimal solution is unique or, more generally, if the set of alternative optimal solutions is bounded (put $\alpha = v(0)$). If the boundedness of the feasible region and the existence of an optimal solution of (P) are in question, then the hypothesis holds if a set of alternative ε-optimal solutions is bounded (put $\alpha = v(0) + \varepsilon$).

The proof of Theorem 8 depends on the following fundamental property of convex functions (cf. [12, p. 93]).

LEMMA 6. *Let $\phi(\cdot)$ be a convex continuous real-valued function on a closed convex set $X \subseteq E^n$. Define $\Phi_\varepsilon \triangleq \{x \in X : \phi(x) \leqq \varepsilon\}$, where ε is a scalar. If Φ_ε is bounded and nonempty for $\varepsilon = 0$, then it is bounded for all $\varepsilon > 0$.*

Proof. Let $\varepsilon > 0$ be fixed arbitrarily, and let x_0 be any point in Φ_0. The assumptions imply that Φ_ε is closed and convex, and that there exists a scalar $M > 0$ such that $\|x - x_0\| < M$ for all $x \in \Phi_0$. Suppose, contrary to the desired conclusion, that there is an unbounded sequence $\langle x^\nu \rangle$ of points in Φ_ε. The direction vectors $(x^\nu - x_0)/\|x^\nu - x_0\|$ must converge subsequentially to a limit, say Δ. Furthermore, $(x_0 + \theta\Delta) \in \Phi_\varepsilon$ for all $\theta \geqq 0$, since Φ_ε is closed and contains a sequence of points whose limit is $x_0 + \theta\Delta$ (such a sequence is $\langle x_0 + \theta(x^\nu - x_0)/\|x^\nu - x_0\| \rangle$ for ν such that $\theta/\|x^\nu - x_0\| \leqq 1$—remember that Φ_ε is convex). Define $\lambda = \phi(x_0 + M\Delta)/2\varepsilon$. Clearly $0 < \lambda < 1$, and $(x_0 + M\Delta) = \lambda(x_0 + (M/\lambda)\Delta) + (1 - \lambda)x_0$.

The convexity of ϕ therefore implies

$$\phi(x_0 + M\Delta) \leqq \lambda\phi[x_0 + (M/\lambda)\Delta] + (1 - \lambda)\phi(x_0).$$

Rearrangement of terms yields the key inequality

$$\phi\left[x_0 + \frac{M}{\lambda}\Delta\right] \geqq \frac{1}{\lambda}\phi(x_0 + M\Delta) - \frac{(1 - \lambda)}{\lambda}\phi(x_0).$$

Since the choice of λ implies that the first term on the right has value 2ε, and since the second term is obviously nonnegative, this inequality directly contradicts the known fact that $[x_0 + (M/\lambda)\Delta] \in \Phi_\varepsilon$. Hence our supposition that Φ_ε is unbounded must be erroneous, and the lemma is proved.

Proof of Theorem 8. Suppose, contrary to the conclusion, that there exists a sequence $\langle y^v \rangle$ of points in Y such that $\langle y^v \rangle \to 0$ and $\langle v(y^v) \rangle \to \bar{v} < v(0)$. For each v, there exists a point $x^v \in X$ such that $g(x^v) \leqq y^v$ and $v(y^v) \leqq f(x^v) \leqq v(y^v) + (1/v)$ (the infimal value $v(y^v)$ can be approached as closely as desired). Clearly $\langle f(x^v) \rangle \to \bar{v}$, and our assumptions imply that, for all v sufficiently large x^v must be in the set

$$\Xi \triangleq \{x \in X : g_i(x) \leqq \varepsilon, i = 1, \cdots, m \text{ and } f(x) \leqq v(0)\}$$

for any fixed $\varepsilon > 0$. But repeated application of Lemma 6 reveals that Ξ is bounded, and so $\langle x^v \rangle$ must be a bounded sequence. We may therefore assume (taking a subsequence if necessary) that $\langle x^v \rangle$ converges, say to $\bar{x}$. By the closedness of X and the continuity of g and f, we have $\bar{x} \in X$, $g(\bar{x}) \leqq 0$, and $f(\bar{x}) = \lim f(x^v) = \bar{v}$. Thus $\bar{x}$ is feasible in (P) and $\bar{v} = f(\bar{x}) \geqq v(0)$. But this contradicts the supposition $\bar{v} < v(0)$, and so v must be lower semicontinuous at $y = 0$.

5.3. A converse duality theorem. Lemma 6 is also the key to the following important partial converse to the strong duality theorem.

THEOREM 9 (Converse duality). *Assume that X is closed and that f and g are continuous on X. If* (D) *has an optimal solution u^*, $f + (u^*)^t g$ has a unique minimizer x^* over X, and x^* is feasible in* (P), *then x^* is the unique optimal solution of* (P), *u^* is an optimal multiplier vector, and* (P) *is stable.*

Proof. Since (D) has an optimal solution, we see from Diag. 1 that only the cases represented by Examples 1, 3 and 7 are possible. The last case is precluded by the assumption that x^* is feasible in (P). If we can show that v is lower semicontinuous at $y = 0$, then by Theorem 7 the first case must obtain and (P) must be stable. If we can also show that (P) has an optimal solution, then part (d) of Theorem 3 will imply that x^* must be the unique optimal solution, for it is the unique minimizer of $f + (u^*)^t g$ over X. Part (c) of Theorem 3 and Theorem 1 will also imply that u^* is an optimal multiplier vector for (P).

Thus our task is to demonstrate that v is lower semicontinuous at $y = 0$ and that (P) admits an optimal solution. To accomplish this, apply Lemma 6 with $\phi(x)$ equal to $f(x) + (u^*)^t g(x) - f(x^*) - (u^*)^t g(x^*)$. It follows that the set

$$\Phi_\varepsilon \triangleq \{x \in X : f(x) + (u^*)^t g(x) \leqq f(x^*) + (u^*)^t g(x^*) + \varepsilon\}$$

is bounded for all $\varepsilon > 0$ (Φ_0 is identical with x^*).

To see that (P) has an optimal solution, let $\langle x^v \rangle$ be a feasible sequence such that $\langle f(x^v) \rangle \to v(0)$, and put $\varepsilon = v(0) + 1 -$ optimal value (D). It is easily verified that $x^\nu \in \Phi_\varepsilon$ for all ν sufficiently large, and consequently the problem:

$$\underset{x\in\Phi_\varepsilon}{\text{Minimize}} \quad f(x) \quad \text{subject to} \quad g(x) \leqq 0,$$

whose feasible region is contained in that of (P), must have the same optimal value. Since this optimal value is actually achieved (the minimand is continuous and the feasible region is compact), this demonstrates the existence of an optimal solution to (P). The lower semicontinuity of v at $y = 0$ follows from an easy modification of Theorem 8 in which the proof uses Φ_ε (with the same choice of ε) in place of Ξ. This completes the proof.

If $f + (u^*)^t g$ is strictly convex in some neighborhood of x^*, it must have a unique minimizer over X; but the reverse implication need not hold. Of course, one would expect $f + (u^*)^t g$ to be strictly convex near x^* when (P) is a nonlinear program, as only one of the functions f and g_i such that $u_i^* > 0$ need be strictly convex near x^* for this to be so. Another way of rationalizing the uniqueness of the minimizer of $f + (u^*)^t g$ over X when (P) involves nonlinear functions is to apply Theorem 10 of § 7 with $\varepsilon = 0$; it then follows that f and each g_i with a positive multiplier would have to be *linear* over the set of alternative minimizers.

A possibly useful observation on Theorem 9 is that the uniqueness hypothesis on x^* can be weakened somewhat with the help of the concept of *g-uniqueness*.[3] If the set of all points x with some particular property is nonempty and g is constant on this set, then this set is said to be g-unique. It is g-uniqueness of x^*, rather than uniqueness, which is essential in Theorem 9. If in the hypotheses we substitute "and the set X^* of all minimizers of $f(x) + (u^*)^t g(x)$ over X is bounded and g-unique," then the conclusion still holds with the substitution "then X^* is the set of optimal solutions of (P)."

It is interesting to note how the conclusions of Theorem 9 change if x^* is not g-unique but the set X^* of all minimizers of $f + (u^*)^t g$ over X is still bounded. The set Φ_ε used in the proof remains bounded by Lemma 6, since $\Phi_0 \equiv X^*$ is nonempty and bounded. It follows that (P) is stable and has an optimal solution, u^* is an optimal multiplier vector, and the optimal solution set of (P) coincides with the points of X^* which also satisfy $g(x) \leqq 0$ and $(u^*)^t g(x) = 0$.

If X is bounded, then the hypothesis "x^* is feasible in (P)" can be omitted because its only role in the proof of Theorem 9—to ensure that (P) is feasible—can be played by Theorem 5 (Y must now be closed).

6. Relations to previous duality results. In this section we examine in more detail the relationships between our results and previous work on duality theory in linear, quadratic, and nonlinear programming. Rather than attempting an exhaustive survey or even citation of the literature, which by now is quite extensive,[4] we select several key papers as representatives for comparison. These are the well-known papers by Dorn [8], Wolfe [34], Stoer [28], Mangasarian and Ponstein [25], and Rockafellar [26]. The first paper is representative of the results that can be obtained for the special case of quadratic programming; the second of

[3] The usefulness of this concept is brought out more clearly in § 7.3.

[4] See the extensive bibliographies of [24, Chap. 8] and [26]. For a dual problem ostensibly quite different from the one considered here, see also Charnes, Cooper and Kortanek [4] (their "Farkas-Minkowski property" appears to imply stability when $v(0)$ is finite).

the results that can be obtained by means of the differential calculus and the classical results of Kuhn and Tucker; the third and fourth of the results obtainable by applying general minimax theorems; and the fifth of the more recent results obtainable by applying the theory of conjugate convex functions.

6.1. Linear programming. Let the primal linear program be:

$$\underset{x \geqq 0}{\text{Minimize}} \quad c^t x \quad \text{subject to} \quad Ax \geqq b.$$

With the identifications $f(x) \equiv c^t x$, $g(x) \equiv b - Ax$ and $X \equiv \{x \in R^n : x \geqq 0\}$, (D) becomes:

$$\underset{u \geqq 0}{\text{Maximize}} \quad [\underset{x \geqq 0}{\text{infimum}}\, c^t x + u^t(b - Ax)].$$

Observe that the maximand has value $u^t b$ if $(c^t - u^t A) \geqq 0$, and $-\infty$ otherwise. That is, $(c^t - u^t A) \geqq 0$ is a necessary and sufficient condition for essential feasibility, so that (D) may be rewritten

$$\underset{u \geqq 0}{\text{Maximize}} \quad u^t b \quad \text{subject to} \quad u^t A \leqq c^t.$$

This, of course, is the usual dual linear program.

The stability of the primal problem when it has finite optimal value is a consequence of the fact that its constraints are all linear (the perturbation function v is piecewise-linear with a finite number of "pieces"). Thus Theorems 1 and 3 apply.

The duality theorem and the usual related results of linear programming are among those now at hand, either as direct specializations or easy corollaries of the results given in previous sections. It is perhaps surprising that, even in the heavily trod domain of linear programming, the geometrical interpretation of the dual given in § 4 does not seem to be widely known.

It is interesting to examine what happens if one dualizes with respect to only a subset of the general linear constraints. Suppose, for example, that the general constraints $Ax \geqq b$ are divided into two groups, $A_1 x \geqq b_1$ and $A_2 x \geqq b_2$, and that we dualize only with respect to the second group; i.e., we make the identifications $f(x) \equiv c^t x$, $g(x) \equiv b_2 - A_2 x$ and $X \equiv \{x \in R^n : x \geqq 0 \text{ and } A_1 x \geqq b_1\}$. Then the new primal problem is still stable, and the dual problem becomes:

$$\underset{u_2 \geqq 0}{\text{Maximize}} \left[\underset{\substack{x \geqq 0 \\ A_1 x \geqq b_1}}{\text{infimum}} \; c^t x + u_2^t(b_2 - A_2 x) \right].$$

Formidable as it looks, this problem is amenable to solution by at least three approaches, all of which can be effective when applied to specially structured problems. In terminology suggested by the author in [17], the first approach is via the piecewise strategy, of which Rosen's "primal partition programming" scheme may be regarded an example [17, § 4.2]; the second approach is via outer linearization of the maximand followed by relaxation (Dantzig–Wolfe decomposition may be regarded as an example [17, § 4.3]); the third approach is via a feasible directions strategy [19]. The point to remember is that the dual of a linear program need not be taken with respect to *all* constraints, and that judicious selection in this regard allows the exploitation of special structure. This point is probably even more important in the context of structured nonlinear programs. It has been stressed previously by Falk [10] and Takahashi [29].

6.2. Quadratic programming: Dorn [8]. Let the primal quadratic program be:

$$\underset{x}{\text{Minimize}} \quad \tfrac{1}{2}x^tCx - c^tx \quad \text{subject to } Ax \leqq b,$$

where C is symmetric and positive semidefinite. The dual (D) with respect to all constraints is

$$\underset{u \geqq 0}{\text{Maximize}} \quad [\underset{x}{\text{infimum}}\, q(x;u)],$$

where

$$q(x;u) \triangleq \tfrac{1}{2}x^tCx - u^tb + (u^tA - c^t)x.$$

In what follows, we shall twice invoke the fact [14, p. 108] that a quadratic function achieves its minimum on any closed polyhedral convex set on which it is bounded below.

Consider the maximand of (D) for fixed u. The function $q(\cdot\,;u)$, being quadratic, is bounded below if and only if its infimum is achieved, which in turn can be true if and only if the gradient of q with respect to x vanishes at some point. Thus $\inf_x q(x;u)$ equals $-\infty$ if there is no x satisfying $\nabla_x q(x;u) \equiv x^tC + (u^tA - c^t) = 0$; otherwise, it equals $\frac{1}{2}x^tCx - u^tb + (-x^tC)x$ for any such x (an obvious algebraic substitution has been made). We can now rewrite the dual problem as:

$$\underset{u \geqq 0}{\text{Maximize}} \quad [-u^tb - \tfrac{1}{2}x^tCx \quad \text{for any } x \text{ satisfying } x^tC + (u^tA - c^t) = 0]$$

subject to

$$x^tC + (u^tA - c^t) = 0 \quad \text{for some } x.$$

But this is equivalent in the obvious sense to:

$$\underset{\substack{u \geqq 0 \\ x}}{\text{Maximize}} \quad -u^tb - \tfrac{1}{2}x^tCx \quad \text{subject to } x^tC + u^tA - c^t = 0.$$

In this way do the primal variables find their way back into the dual. This is precisely Dorn's dual problem.

The linearity of the constraints of the primal guarantees stability whenever the optimal value is finite, and so Theorems 1 and 3 apply. Dorn's dual theorem asserts that if x^* is optimal in the primal, then there exists u^* such that (u^*, x^*) is optimal in the dual and the two extremal values are equal. This is a direct consequence of Theorem 3. Dorn's converse duality theorem asserts that if $(\hat{u}, \hat{x})$ are optimal in the dual, then some $\bar{x}$ satisfying $C(\bar{x} - \hat{x}) = 0$ is optimal in the primal and the two extremal values are equal. To recover this result, we first note by Theorem 2 that the primal minimand must be bounded below, and hence there must be a primal optimal solution, say $\bar{x}$. By Theorem 3, the extremal values are equal, and $\bar{x}$ must be an unconstrained minimum of $q(x;\hat{u})$, i.e., $\bar{x}^tC + \hat{u}^tA - c^t = 0$. But $\hat{x}^tC + \hat{u}^tA - c^t = 0$ also holds, and so $\bar{x}^tC = \hat{x}^tC$.

In a similar manner we may obtain the symmetric dual of the special quadratic program studied by Cottle [5], and also his main results.

6.3. Differentiable nonlinear programming: Wolfe [34]. The earliest and probably still most widely quoted duality results for differentiable convex programs are those of Wolfe. He assumes $X = R^n$ and all functions to be convex and *differentiable* on R^n, and proposes the following dual for (P):

$$\text{(W)} \qquad \underset{\substack{u \geqq 0 \\ x}}{\text{Maximize}} \quad f(x) + u^t g(x)$$

subject to

$$\nabla f(x) + \sum_{i=1}^{m} u_i \nabla g_i(x) = 0.$$

Wolfe obtains three theorems *under the Kuhn–Tucker constraint qualification.* The first is a weak duality theorem. The second asserts that if x^0 is optimal in (P), then there exists u^0 such that (u^0, x^0) is optimal in (W) and the extremal values are equal. The third theorem, which requires all g_i to be linear, asserts that the optimal value of (W) is $+\infty$ if it is feasible and (P) is infeasible.

In order to compare these results with ours, we must examine the relationship between (W) and (D). Observe that, for fixed $\bar{u} \geqq 0$, $(\bar{u}, \bar{x})$ is feasible in (W) if and only if $\bar{x}$ is an unconstrained minimizer of $f + \bar{u}^t g$. In terms of the dual variables only, therefore, (W) may be written as:

$$\text{(W.1)} \qquad \underset{u \geqq 0}{\text{Maximize}} \quad [\underset{x}{\text{minimum}}\, f(x) + u^t g(x)]$$

subject to u such that the unconstrained minimum of $f + u^t g$ is achieved for some x.

Evidently (W.1) is equivalent to (W) in a very strong sense: $(\bar{u}, \bar{x})$ is feasible in (W) only if $\bar{u}$ is feasible in (W.1) and the respective objective function values are equal; and $\bar{u}$ is feasible in (W.1) only if there exists a point $\bar{x}$ such that $(\bar{u}, \bar{x})$ is feasible in (W) and the respective objective function values are equal. Problem (W.1) is, of course, identical to (D) except for the extra constraint on u. We are now in a position to recover Wolfe's results. His first theorem follows immediately from our weak duality theorem, since the extra constraint on u in (W.1) can only depress the optimal value by comparison with (D). Wolfe's second theorem follows immediately from parts (a), (b) and (d) of the strong duality theorem, since the Kuhn–Tucker constraint qualification implies stability. His third theorem would be a direct consequence of Theorem 5 were it not for the extra constraint on u in (W.1); that is, were (D) to replace (W) in his statement of the theorem. The extra constraint in (W.1) necessitates a different line of reasoning, and although one could be fashioned using only the previous results of this paper, it would not be sufficiently different from Wolfe's proof to warrant presentation here.

As Wolfe himself noted, (W) may be difficult to deal with computationally; its maximand is not concave in (u, x), and its constraint set involves nonlinear equality constraints and needn't even be convex. Another difficulty is that (W) is more prone than (D) to the misfortune of a duality gap, since it has (as revealed by (W.1)) an extra constraint on u.

6.4. Stoer [28] and Mangasarian and Ponstein [25]. The natural generalization of Wolfe's dual when some functions in (P) are not differentiable or the set X is not all of R^n is the following, which might be called the general Wolfe dual:

$$\text{(GW)} \qquad \underset{\substack{u \geqq 0 \\ x}}{\text{Maximize}} \quad f(x) + u^t g(x)$$

subject to:

$$x \text{ minimizes } f + u^t g \text{ over } X.$$

This problem bears the same relation to (D) as (W) does; the more general version of the intermediary program (W.1) that is appropriate here should be evident. Although (GW) is generally inferior to (D) for much the same reason that (W) is, it is nevertheless worthwhile to review some of the work that has been addressed to this dual.

The landmark paper treating (GW) is by Stoer [28], whose principal tool is a general minimax theorem of Kakutani. The possibility of using minimax theorems in this connection is due, of course, to the existence of an equivalent characterization of the optimality conditions for (P) as a constrained Lagrangean saddle point. Stoer's results are shown to generalize many of those obtained via the differential calculus by numerous authors in the tradition of Wolfe's paper. Because of certain technical difficulties inherent in his development, however, we shall examine Stoer's results as reworked and elaborated upon by Mangasarian and Ponstein [25]. To bring out the essential contributions of this work, we shall take considerable license to paraphrase.

Aside from some easy preliminary results which do not depend on convexity—namely, a weak duality theorem and an alternative characterization of a constrained Lagrangean saddle point (see the discussion following our Definition 3)—the Stoer–Mangasarian–Ponstein results relating (P) and (GW) can be paraphrased as three theorems, all of which require f and g to be convex and *continuous*, and X to be convex and *closed*. The first [25, Theorem 4.4a] is: Assuming (P) has an optimal solution $\bar{x}$, an optimal multiplier vector $\bar{u}$ exists if and only if $f(x) + u^t g(x)$ has the so-called "low-value property" at $(\bar{x}, \bar{u})$. We shall not quote in detail this rather technical property, but we do observe that, in view of our Theorem 1, the low-value property must be entirely equivalent (when all functions are continuous and X is closed) to the condition that (P) is stable.

The second theorem [25, Theorem 4.4b] is: Assuming (GW) has an optimal solution $(\bar{x}, \bar{u})$, there exists a minimizer x^0 of $f + \bar{u}^t g$ over X satisfying $\bar{u}^t g(x^0) = 0$ such that x^0 is optimal in (P) if and only if $f(x) + u^t g(x)$ has the so-called "high-value property" at $(\bar{x}, \bar{u})$. The high-value property is also quite technical, but its significance can be brought out by comparing the theorem with the following immediate consequence of Theorems 1 and 3 of this study: There exists an optimal solution of (D) and for any optimal $\bar{u}$ there is a minimizer x^0 of $f + \bar{u}^t g$ over X satisfying $\bar{u}^t g(x^0) = 0$ such that x^0 is optimal in (P), if and only if (P) is stable and has an optimal solution. It follows that the high-value property holds and (GW) has an optimal solution if and only if (P) is stable and has an optimal solution (when f and g are continuous and X is closed). The demonstration is straightforward, and makes use of the evident fact that if $\bar{u}$ is optimal in (D) and $\bar{x}$ minimizes $f + \bar{u}^t g$ over X, then $(\bar{x}, \bar{u})$ must be optimal in (GW).

The third result is a strict converse duality theorem, a counterpart of our Theorem 9: If $(\bar{x}, \bar{u})$ is an optimal solution of (GW) and $f(x) + \bar{u}^t g(x)$ is strictly convex in some neighborhood of $\bar{x}$, then $\bar{x}$ is an optimal solution of (P) and $(\bar{x}, \bar{u})$ satisfies the optimality conditions for (P). The difference between this and Theorem 9, besides the fact that it addresses (GW) rather than (D), is the slightly stronger hypothesis that $f + \bar{u}^t g$ be strictly convex near $\bar{x}$ (rather than simply requiring the minimizer of $f + \bar{u}^t g$ over X to be unique or just g-unique).

This discussion casts suspicion on the need for general minimax theorems as a means of obtaining strong results in duality. Such an approach may even be inadvisable, as it seems to lean toward (GW) rather than (D), and toward technical conditions less convenient than stability.

6.5. Rockafellar [26]. Finally we come to the outstanding work of Rockafellar, whose methods rely heavily upon Fenchel's theory of conjugate convex functions [11]. To make full use of this theory, it is assumed in effect that f and g_i are *lower semicontinuous* on X (so that the convex bifunction associated with (P) will be closed as well as convex, as assumed in Rockafellar's development).[5] The theory of conjugacy then yields the fundamental relationships between (P) and (D). By way of comparison, we note that one can readily deduce Lemmas 1 through 5 and Theorems 1 through 7 of this study from his results. This deduction utilizes the equivalence of Definitions 6 and 6′ for the concept of stability, and the equivalence between the optimality conditions of (P) and a constrained saddle point of the Lagrangean. The content of our Theorems 8 and 9 is not obtained, but Rockafellar does give some additional results not readily obtainable by our methods. Namely, the dual of (P) in a certain conjugacy sense is again (P); the optimal solutions of (P) are the subgradients of a certain perturbation function associated with (D); and v is lower semicontinuous at 0 if and only if the perturbation function associated with (D) is upper semicontinuous at the point of null perturbation. The significance of these additional results for applications is not clear, although they are certainly very satisfying in terms of mathematical symmetry.

Although (D) is the natural dual problem of (P) corresponding to the perturbation function v, it is interesting to note that other perturbation functions give rise to other duals to which Rockafellar's results apply immediately. It seems likely that the methods of this paper could be adapted to deal with other perturbation functions too, but as yet this has not been attempted.

7. Computational applications. This section studies a number of issues that arise if one wishes to obtain a numerical solution of a convex program via its dual—that is, if one is interested in "dual" methods for solving (P). By a dual method we mean one that generates a sequence of essentially feasible solutions that converges in value to the optimal value of (D). Such a sequence yields, by the weak duality theorem, an improving sequence of lower bounds on the optimal value of (P).

---

[5] Rockafellar has pointed out to the author in a private communication (October 13, 1969) that bifunction closedness, and hence the semicontinuity assumption on f and g_i, can be dropped except for the "additional" results referred to below. He also pointed out that results closely related to our Theorems 8 and 9 appear in his forthcoming book *Convex Analysis* (Princeton University Press).

The possible pitfalls encountered by dual methods include these: The dual may fail to be essentially feasible, even though (P) has an optimal solution; or it may be essentially feasible but fail to have an optimal solution; or it may have an optimal solution but its optimal value may be less than that of (P) (this can invalidate the obvious natural termination criterion when an upper bound on the optimal value of (P) is at hand). None of these pitfalls can occur if (P) is stable, thanks to the strong duality theorem. For this reason, and also because this property usually holds anyway, we shall *assume for the remainder of the section that* (P) *is stable*. We shall also assume that (P) has an optimal solution.

7.1. Methods for solving (D). There are several ways one may go about solving (D). Some of these are appropriate only when (P) has quite special structure, as when all functions are linear or quadratic. Others can be employed under quite general assumptions. It is perhaps fair to say that most of these methods fall into two major categories: methods of feasible directions and methods of tangential approximation. Both are based on the easily verified fact that, for any $\bar{u} \geqq 0$, if $\bar{x}$ achieves the infimum required to evaluate the maximand of (D) at $\bar{u}$ then $g(\bar{x})$ is a subgradient of the maximand at $\bar{u}$; that is,

$$\underset{x \in X}{\text{infimum}} \{ f(x) + u^t g(x) \} \leqq [f(\bar{x}) + \bar{u}^t g(\bar{x})] + g^t(\bar{x})(u - \bar{u}), \qquad \text{all } u \geqq 0.$$

The feasible directions methods typically use $g(\bar{x})$ as though it were a gradient to determine a direction in which to take a "step." Instances of such methods are found in [3], [10], [19], [22], [29], [33]. The term "Lagrangean decomposition" is sometimes applied when the maximand of (D) separates (decomposes) into the sum of several independent infima of Lagrangean functions. The ancestor of this class of algorithms is an often-overlooked contribution by Uzawa [30]. The other principal class of methods for optimizing (D) is global rather than local in nature, as it uses subgradients of the maximand of (D) to build up a tangential approximation to it. It is well known that the decomposition method of Dantzig and Wolfe for nonlinear programming [6, Chap. 24] can be viewed as such a method. See also [16, § 6] and the "global" approach in [29].

It is beyond the scope of this effort to delve into the details of methods for optimizing (D). Rather, we wish to focus on questions of common interest to almost any algorithm that may be proposed for (D) as a means of solving (P). Specifically, we shall consider the possibility of numerical error in optimizing (D) and in minimizing $f + u^t g$ over X for a given u. It is important to investigate the robustness of the resulting approximate solutions to (P) when there is numerical error of this kind. Hopefully, by making the numerical error small enough one can achieve an arbitrarily accurate approximation to an optimal solution of (P). We shall see that this is often, but not always, the case.

7.2. Optimal solution of (D) known without error. The simplest case to consider is the one in which an optimal solution u^* of (D) is known without error. Then by the strong duality theorem we know that the optimal solutions of (P) coincide with the solutions in x of the system:[6]

(i) x minimizes $f + (u^*)^t g$ over X,

(ii) $(u^*)^t g(x) = 0$,

(iv) $g(x) \leqq 0$.

If (i) is known to have a unique solution, as is very often the case when (P) is a nonlinear program (actually, g-uniqueness is enough—see the discussion following Theorem 9), then (ii) and (iv) will automatically hold at the solution. Any sequence of points converging to the solution of (i) also converges to an optimal solution of (P). Thus, there appear to be no particular numerical difficulties.

If, on the other hand, the solution set of (i) is not unique, then (ii) or (iv) or both may be relevant. The following result will be useful.

THEOREM 10. *Let $u \geqq 0$ and $\varepsilon \geqq 0$ be fixed. Let X_ε be the set of ε-optimal solutions of the problem of minimizing $f + u^t g$ over X; that is, let X_ε be the (convex) set of all points $\bar{x}$ in X such that*

$$f(\bar{x}) + u^t g(\bar{x}) \leqq \left[\operatorname*{infimum}_{x \in X} f(x) + u^t g(x) \right] + \varepsilon .$$

Then f comes within ε, and each g_i with a positive multiplier comes within (ε/u_i) of being linear over X_ε in the following sense: $x^1, x^2 \in X_\varepsilon$ and $0 \leqq \lambda \leqq 1$ implies (let $\bar{\lambda} = 1 - \lambda$)

$$\lambda f(x^1) + \bar{\lambda} f(x^2) - \varepsilon \leqq f(\lambda x^1 + \bar{\lambda} x^2) \leqq \lambda f(x^1) + \bar{\lambda} f(x^2),$$

$$\lambda g_i(x^1) + \bar{\lambda} g_i(x^2) - \frac{\varepsilon}{u_i} \leqq g_i(\lambda x^1 + \bar{\lambda} x^2) \leqq \lambda g_i(x^1) + \bar{\lambda} g_i(x^2), \qquad i : u_i > 0 .$$

Proof. The right-hand inequalities hold, of course, by convexity. Suppose that the first left-hand inequality fails for some $u \geqq 0$, $\varepsilon \geqq 0$, $x^1 \in X_\varepsilon$, $x^2 \in X_\varepsilon$, $0 \leqq \lambda \leqq 1$. Then

$$f(\lambda x^1 + \bar{\lambda} x^2) < \lambda f(x^1) + \bar{\lambda} f(x^2) - \varepsilon$$

which, when added to the other right-hand inequalities multiplied by the respective values of u_i, yields

$$\begin{aligned} f(\lambda x^1 + \bar{\lambda} x^2) &+ \sum_{i=1}^m u_i g_i(\lambda x^1 + \bar{\lambda} x^2) \\ &< \lambda \left[f(x^1) + \sum_{i=1}^m u_i g_i(x^1) \right] + \bar{\lambda} \left[f(x^2) + \sum_{i=1}^m u_i g_i(x^2) \right] - \varepsilon . \end{aligned}$$

Since X is convex, $\lambda x^1 + \bar{\lambda} x^2$ is in X and so the left-hand side has value greater than or equal to

$$\operatorname*{infimum}_{x \in X} f(x) + u^t g(x).$$

Using the fact that x^1 and x^2 are in X_ε, however, we obtain the contradiction that

[6] The rubrics (i), (ii) and (iv) are used in order to maintain correspondence with the optimality conditions as listed in Definition 3.

the right-hand side is less than or equal to this value. Hence our supposition must fail.

A similar argument shows that the other inequalities of the conclusion of the theorem must hold, completing the proof.

We must distinguish two further possibilities when (i) does not have a unique solution: either a solution of (i) can be found without error, or it cannot. Suppose that an optimal solution $\bar{x}$ of (i) can be found. Then (i) is equivalent to:

$$\text{(ia)} \qquad x \in X,$$

$$\text{(ib)} \qquad f(x) + (u^*)^t g(x) \leqq f(\bar{x}) + (u^*)^t g(\bar{x}).$$

Theorem 10 with $u = u^*$ and $\varepsilon = 0$ yields the very useful result that (ii) is a *linear* constraint so long as x satisfies (ia) and (ib) ($(u^*)^t g(x) \equiv \sum u_i^* g_i(x)$, where the sum is taken over the indices such that $u_i^* > 0$). Thus, any of a number of convex programming algorithms could be used with $\bar{x}$ as the starting point to find a feasible solution of (ia), (ib), (ii), and (iv) and thereby solve (P). Suppose, on the other hand, that only an ε-optimal solution $\bar{x}$ of (i) can be found. Then (ia) and (ib) are no longer equivalent to (i), and (ii) is no longer linear over the solution set of (ia) and (ib). However, this solution set contains the solution set of (i) (because $f(\bar{x}) + (u^*)^t g(\bar{x})$ is larger than it ought to be); and Theorem 10 implies that $(u^*)^t g(x)$ is within $\#\varepsilon$ of being linear over it, where $\#$ is the number of indices for which $u_i^* > 0$. Hence, any convex programming algorithm which solves (ia), (ib) and (iv) exactly but (ii) only to linear approximation will find a feasible solution of (P) that is within $(\# + 1)\varepsilon$ of being optimal in (P). Therefore, by taking ε sufficiently small one can find a solution of (P) that is as near optimal as desired.

In summary, we see that solving (P) once an optimal solution of (D) is known poses no special numerical difficulties.

7.3. Optimal solution of (D) not known exactly. Suppose that a particular algorithm addressed to (D) generates a sequence $\langle u^\nu \rangle$ converging to an optimal solution u^*. If the minimizers of $f + (u^*)^t g$ over X are g-unique, we can obtain a quite satisfactory result concerning the recovery of an optimal solution of (P). In the absence of this assumption, however, an example will be given to show that things can go awry.

THEOREM 11. *Assume that f and each g_i is continuous on X, X is compact,* (D) *has an optimal solution u^*, and the minimizers of $f + (u^*)^t g$ over X are g-unique. Let $\langle u^\nu \rangle$ be any nonnegative sequence converging to u^* and $\langle x^\nu \rangle$ any sequence composed, for each ν, of a minimizer of $f + (u^\nu)^t g$ over X. Then $\langle x^\nu \rangle$ has at least one convergent subsequence, and every such subsequence converges to an optimal solution of* (P).

Proof. Since $\langle x^\nu \rangle$ is in X and X is compact, there must be at least one convergent subsequence. For simplicity of notation, redefine $\langle x^\nu \rangle$ to coincide with any such subsequence. Let $X(u)$ be the set of all minimizers of $f + u^t g$ over X. Under the given assumptions it is known (e.g., [7, p. 19]) that $X(u)$ is an upper semicontinuous set-valued function of u at u^*; that is, $\langle u^\nu \rangle \to u^*$, $x^\nu \in X(u^\nu)$, $\langle x^\nu \rangle \to \bar{x}$ implies $\bar{x} \in X(u^*)$. But $X(u^*)$ is bounded and g-unique, and so by Theorem 9 (see also the ensuing discussion) $\bar{x}$ must be an optimal solution of (P). The proof is complete.

The conclusion of Theorem 11 is very comforting; but in practice one must still decide when to truncate the infinite process. We shall assume in the ensuing discussion that the hypotheses of Theorem 11 hold, except where explicitly weakened.

One natural termination criterion is based upon the easily demonstrated fact that x^ν must be an optimal solution of the following approximation to (P):

$$(\mathrm{P}^\nu) \qquad \underset{x \in X}{\text{Minimize}} \quad f(x) \quad \text{subject to } g(x) \leqq g(x^\nu).$$

The continuity of g, and the fact that $\langle x^\nu \rangle$ converges subsequentially to an optimal solution x^* of (P), implies that the right-hand side of (P^ν) converges subsequentially to $g(x^*)$ as $\nu \to \infty$; hence one may terminate when ν reaches a value for which the right-hand side of (P^ν) is "sufficiently near" to being $\leqq 0$. How near is "sufficiently" near depends upon how precisely the g constraints of (P) really must be satisfied. If a perturbation in the right-hand side of certain of the constraints cannot be tolerated, however small the change, then it may be advisable to insist that such constraints be incorporated into X. In other words, it may be advisable not to dualize with respect to such constraints in the first place.

So far we have assumed that a true minimizer x^ν of $f + (u^\nu)^t g$ over X could be found for each ν. While this may be a reasonable assumption when X is a convex polytope and f and g are linear or quadratic functions, it is desirable in the interest of generality to be able to cope with numerical inaccuracy. Results as satisfactory as those obtained above seem quite elusive unless we suppose that a minimizer of $f + (u^\nu)^t g$ over X, say x^ν, can be approached as closely as desired and, indeed, *is* approached more and more closely as ν increases. Specifically, let us suppose that a point $\bar{x}^\nu$ in X can be found satisfying $\|\bar{x}^\nu - x^\nu\| \leqq \varepsilon^\nu$, say, where $\varepsilon^\nu \geqq 0$ and $\langle \varepsilon^\nu \rangle \to 0$. It follows easily that, for each subsequence of $\langle x^\nu \rangle$ converging to x^*, the corresponding subsequence of $\langle \bar{x}^\nu \rangle$ also converges to x^*. Thus, Theorem 11 holds with $\langle \bar{x}^\nu \rangle$ in place of $\langle x^\nu \rangle$. Of course, $\bar{x}^\nu$ is not optimal in ($\bar{\mathrm{P}}^\nu$), which we define to be (P^ν) with $\bar{x}^\nu$ in place of x^ν in the right-hand side. What is true, however, is that $\bar{x}^\nu$ is within $\bar{\varepsilon}$ of being optimal in ($\bar{\mathrm{P}}^\nu$) if it comes within $\bar{\varepsilon}$ of minimizing $f + (u^\nu)^t g$ over X(see [9]). Since the right-hand side of ($\bar{\mathrm{P}}^\nu$) converges subsequentially to $g(x^*)$ as $\nu \to \infty$, we can use ($\bar{\mathrm{P}}^\nu$) in much the same way as (P^ν) to determine when to terminate, except that the magnitude of $\bar{\varepsilon}$ must also be considered.

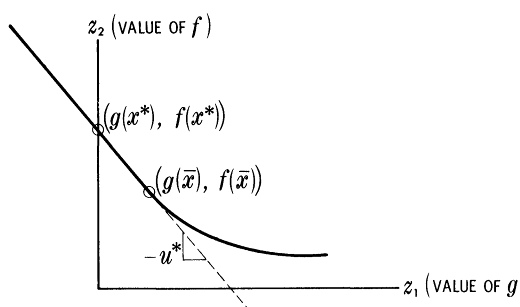


FIG. 3

Finally, we wish to emphasize the central role played by the assumption of Theorem 11 that the minimizers of $f + (u^*)^t g$ over X are *g-unique*. Failure of this assumption can lead to serious difficulties; see, for example, Fig. 3, in which any sequence $\langle u^\nu \rangle$ converging to u^* from below will lead to a sequence $\langle x^\nu \rangle$ with $g(\bar{x}) > 0$ for every subsequential limit point $\bar{x}$. Not only does $\bar{x}$ fail to solve (P), but $g(x^\nu) \geqq g(\bar{x}) > 0$ for all ν, so that the natural termination criterion based on (P^ν) will never be satisfied (unless of course the permissible tolerance in the right-hand side of (P) is sufficiently large).

8. Theoretical applications. Although the primary emphasis thus far has been on results of interest from the computational viewpoint, many of the results are also of interest from a purely theoretical point of view. Just as linear duality theory can be used to obtain many results in the theory of linear systems that do not appear to involve optimization at all (see, e.g., [6, § 6.4]), so does nonlinear duality theory readily yield many results in the theory of convex systems. In this section, we illustrate this fact by using our results to obtain new proofs of three known theorems. The first is a separation theorem for disjoint convex sets; the second is a characterization of a certain class of convex sets in terms of supporting half-spaces; and the third is a fundamental property of an inconsistent convex system. With only a modicum of ingenuity, by similar methods one may obtain many other interesting results, some of them perhaps new. Thus, nonlinear duality theory provides a unified approach to the derivation of a substantial body of theorems in the theory of convexity.

8.1. A separation theorem. Let X and $\tilde{X}$ be nonempty, closed, convex, disjoint subsets of R^k, and suppose further that X is bounded; then there exists a hyperplane in R^k that strictly separates X and $\tilde{X}$. A conventional proof of this theorem can be found, for example, in [2, p. 55], but we shall deduce it from Theorem 3. The other standard separation theorems can be obtained in a similar fashion. (The alert reader will recall that another separation theorem was used in the course of proving Theorem 3; what is being demonstrated in effect, then, is a kind of equivalence between duality theory and separation theory for convex sets.)

The hypotheses of the theorem certainly imply that the convex program

$$\underset{\substack{x \in X \\ \tilde{x} \in \tilde{X}}}{\text{Minimize}} \ \|x - \tilde{x}\|$$

has infimal value greater than 0. A convenient choice of norm is $\|x\| \triangleq$ maximum $\{|x_1|, \cdots, |x_k|\}$. Then the above problem can be rewritten

$$\underset{\substack{x \in X \\ \tilde{x} \in \tilde{X} \\ \sigma}}{\text{Minimize}} \ \sigma \quad \text{subject to} \quad \begin{aligned} \sigma &\geqq x_i - \tilde{x}_i, \qquad & i = 1, \cdots, k, \\ \sigma &\geqq -x_i + \tilde{x}_i, \qquad & i = 1, \cdots, k, \end{aligned}$$

where σ is a scalar variable. Dualizing with respect to the (linear) constraints involving σ, we obtain from Theorem 3 that the dual problem

$$\underset{u \geqq 0}{\text{Maximize}} \left[\underset{\substack{x \in X \\ \tilde{x} \in \tilde{X} \\ \sigma}}{\text{infimum}} \ \sigma + \sum_{i=1}^{k} u_i(x_i - \tilde{x}_i - \sigma) + \sum_{i=1}^{k} u_{k+i}(-x_i + \tilde{x}_i - \sigma) \right]$$

has an optimal solution u^*, and that its optimal value is greater than 0. By taking advantage of the separability of the infimand with respect to x, $\tilde{x}$ and σ, we therefore have the key inequality

$$\left[\underset{\sigma}{\text{infimum}}\ \sigma\left(1 - \sum_{i=1}^{2k} u_i^*\right)\right] + \left[\underset{x \in X}{\text{infimum}} \sum_{i=1}^{k} (u_i^* - u_{k+i}^*)x_i\right] + \left[\underset{\tilde{x} \in \tilde{X}}{\text{infimum}} \sum_{i=1}^{k} -(u_i^* - u_{k+i}^*)\tilde{x}_i\right] > 0.$$

The infimum over σ must be 0 (i.e., $\sum_{i=1}^{2k} u_i^* = 1$ must hold), for otherwise it would be $-\infty$ and the inequality could not be true. Defining $\alpha_i = u_i^* - u_{k+i}^*$ and rearranging, the inequality then becomes

$$\underset{x \in X}{\text{infimum}} \sum_{i=1}^{k} \alpha_i x_i > \underset{\tilde{x} \in \tilde{X}}{\text{supremum}} \sum_{i=1}^{k} \alpha_i \tilde{x}_i.$$

This shows that X and $\tilde{X}$ are strictly separated by the hyperplane

$$\left\{x \in R^k : \sum_{i=1}^{k} \alpha_i x_i = \alpha_0\right\},$$

where α_0 is any scalar strictly between the values of the left- and right-hand sides of the rearranged inequality.

8.2. A characterization of the set Y. The set $Y \triangleq \{y \in R^m : g(x) \leqq y$ for some $x \in X\}$ has cropped up quite often in this study. Indeed, this kind of set arises frequently in mathematical programming when it is necessary to work with the collection of perturbations for which a perturbed problem has a feasible solution. It also arises when a set must be "projected" onto a subspace; in the special case above, Y can be thought of as being obtained by projecting the set $\{(x, y) \in R^{n+m} : g(x) - y \leqq 0, x \in X\}$ onto the R^m space associated with y (see [17, § 2.1]).

It is sometimes useful to be able to characterize such sets in terms of their supporting half-spaces. In [16, p. 24], a slightly weaker form of the following theorem is demonstrated by applying a result due to Bohnenblust, Karlin and Shapley (see [2, p. 64]). Assume that $g_1, \cdots, g_m$ are convex functions on the nonempty convex set $X \subseteq R^n$ and that Y is closed. Then $y \in Y$ if and only if y satisfies the system of linear constraints

$$\lambda^t y \geqq \underset{x \in X}{\text{infimum}}\ \lambda^t g(x), \qquad \text{all } \lambda \in \Lambda,$$

where $\Lambda \triangleq \{\lambda \in R^m : \lambda \geqq 0$ and $\sum_{i=1}^m \lambda_i = 1\}$. Furthermore, every constraint in this system describes a half-space that supports Y in the obvious sense.[7]

The only part of the conclusion that cannot be proved directly and easily is the assertion that $\bar{y} \in Y$ if $\bar{y}$ satisfies the given system of constraints. To prove this using nonlinear duality theory, we observe that if $\bar{y}$ satisfies the given system of

[7] That is, for each fixed $\bar{\lambda} \in \Lambda$, either there exists a point $\bar{y}$ in Y such that $\bar{\lambda}^t\bar{y} = \text{infimum}_{x \in X}\ \bar{\lambda}^t g(x)$ or there exists a sequence $\langle y^v \rangle$ of points in Y such that $\lim_{v \to \infty} \bar{\lambda}^t y^v = \text{infimum}_{x \in X}\ \bar{\lambda}^t g(x)$.

constraints, then (as the normalization of λ is immaterial) we have

$$\underset{\lambda \geqq 0}{\text{supremum}} \left[\underset{x \in X}{\text{infimum}}\ \lambda^t(g(x) - \bar{y}) \right] = 0.$$

This is easily recognized as the assertion that the dual of

$$\underset{x \in X}{\text{Minimize}} \quad 0^t x \quad \text{subject to} \quad g(x) - \bar{y} \leqq 0$$

has optimal value 0. Applying Theorem 5 to this "primal" problem yields, since $\{y : g(x) - \bar{y} \leqq y \text{ for some } x \in X\}$ must be closed when Y is closed, that this problem must be feasible and hence that $\bar{y}$ must be in Y.

8.3. A fundamental property of an inconsistent convex system. Let $g_1, \cdots, g_m$ be convex functions on the nonempty convex set $X \subseteq R^n$. If the system $g_1(x) < 0, \cdots, g_m(x) < 0$ has no solution in X, then there exists an m-vector $u \geqq 0$ such that $\sum_{i=1}^m u_i = 1$ and

$$\left[\underset{x \in X}{\text{infimum}}\ u^t g(x) \right] \geqq 0.$$

This is essentially the Fundamental Theorem on p. 62 of [2], where a proof relying on Helly's intersection theorem can be found. To deduce it from Theorem 3, we merely observe that the system of inequalities has no solution in X only if the convex program

$$\underset{\substack{x \in X \\ \sigma}}{\text{Minimize}} \quad \sigma \quad \text{subject to} \quad g_i(x) - \sigma \leqq 0, \quad i = 1, \cdots m,$$

has infimal value $\geqq 0$. Since Slater's qualification is obviously satisfied, this program is stable and so by Theorem 3 we conclude that the dual program

$$\underset{u \geqq 0}{\text{Maximize}} \quad \left[\underset{\substack{x \in X \\ \sigma}}{\text{infimum}}\ \sigma + \sum_{i=1}^m u_i(g_i(x) - \sigma) \right]$$

has an optimal solution u^* with value $\geqq 0$. Hence

$$\left[\underset{x \in X}{\text{infimum}}\ (u^*)^t g(x) \right] + \left[\underset{\sigma}{\text{infimum}}\ \sigma \left(1 - \sum_{i=1}^m u_i^* \right) \right] \geqq 0,$$

where we have taken advantage of the separability of the infimum in the dual program. The second term must be 0—that is, $\sum_{i=1}^m u_i^* = 1$ must hold—for otherwise its value would be $-\infty$. Hence the first term is nonnegative and the theorem is proved.

A generalization of the Farkas–Minkowski theorem applicable to convex systems [2, p. 67] can be readily demonstrated by a very similar line of reasoning.

9. Opportunities for further research. It is hoped that what has been accomplished here will encourage further work in a similar vein. Much remains to be done.

In terms of importance, one could hardly do better than to work toward relaxing the convexity assumptions. We have pointed out several occasions on which these assumptions could be dispensed with entirely. For example, it is tantalizingly true that (D) is a concave program even without any assumptions at all on X, f and g. Furthermore, the astute reader may have noticed that the only role played by the assumed convexity of f, g and X in §§ 2 and 3 (and in a number of later results) is to guarantee via Lemma 2 that v and Y are convex. Thus the convexity assumptions on f, g and X can be weakened at least in Theorems 1 and 3, to the assumption that v and Y are convex. Perhaps a theory adequate for some purposes could be constructed under still weaker assumptions. Quite likely this can be done using the notions of pseudo- or quasi-convexity (e.g., [24, Chaps. 9, 10]) in place of convexity, but the real challenge would be to get along with considerably weaker assumptions. Published efforts along these lines so far leave a lot to be desired in terms of potential applicability, mainly because global results in the absence of global properties like convexity seem to require assuming global knowledge that is overly difficult to have in practice. Perhaps a local theory is all one can hope for without convexity-like assumptions.

Important opportunities are also to be found in studying questions such as those treated in § 7, relating to the robustness of computational methods addressed to the dual problem in the face of numerical error. For example, to what extent does an inability to minimize the Lagrangean function exactly for a given value of u disrupt convergence to an optimal solution of (D)? Probably such studies will have to be carried out in the context of the various specific dual methods that have been proposed. One successful study in this vein is Fox [13].

A natural extension of the theory developed here would be the construction of parallel theories for perturbation functions other than v, perhaps even for an entire class of them. Rockafellar's work strongly suggests that this is possible. Some of the alternative choices for v might prove quite useful in widening the scope of applications of duality theory.

Another direction of possible extension would be toward more general linear vector spaces. This would open up applications to optimal control, continuous programming, and other infinite-dimensional problems. In fact, a good deal has already been accomplished along these lines (e.g., Luenberger [23], Rockafellar [27], Van Slyke and Wets [31]), particularly with reference to generalizations of the results of §§ 2 and 3 of this paper. The very recent treatment by Luenberger adopts a viewpoint quite similar to the one taken here, and is especially recommended to the interested reader.

A number of questions concerning economic significance and interpretation yet remain to be explored. The paper by Gale [15], which includes a discussion of optimal multiplier vectors and the concept of stability, is a fine example of what can be done. See also Balinski and Baumol [1] and Williams [32].

Finally, we should mention that a variety of opportunities for application exist even without further extensions. Many more theoretical applications of the kind illustrated in § 8 are possible. Illustrative of the possible computational applications is a recent nonlinear generalization by this writer of Bender's partition programming method [18].

Acknowledgments. The author expresses his appreciation to B. Fox, R. Grinold, R. Marsten, R. T. Rockafellar, and C. Swoveland for constructive criticisms of the manuscript.

REFERENCES

[1] M. L. Balinski and W. J. Baumol, *The dual in nonlinear programming and its economic interpretation*, Rev. Economic Studies, XXXV, 103 (1968), pp. 237–256.

[2] C. Berge and A. Ghouila-Houri, *Programming, Games and Transportation Networks*, John Wiley, New York, 1965.

[3] S. P. Bradley, *Decomposition programming and economic planning*, Rep. 67-20, Operations Research Center, University of California, Berkeley, 1967.

[4] A. Charnes, W. W. Cooper and K. O. Kortanek, *On the theory of semi-infinite programming and a generalization of the Kuhn-Tucker saddle point theorem for arbitrary convex functions*, Naval Res. Logist. Quart., 16 (1969), pp. 41–51.

[5] R. W. Cottle, *Symmetric dual quadratic programs*, Quart. Appl. Math., 21 (1963), pp. 237–243.

[6] G. B. Dantzig, *Linear Programming and Extensions*, Princeton University Press, Princeton, 1963.

[7] G. Debreu, *Theory of Value*, John Wiley, New York, 1959.

[8] W. S. Dorn, *Duality in quadratic programming*, Quart. Appl. Math., 18 (1960), pp. 155–162.

[9] H. Everett, *Generalized Lagrange multiplier method for solving problems of optimum allocation of resources*, Operations Res., 11 (1963), pp. 399–417.

[10] J. E. Falk, *A constrained Lagrangian approach to nonlinear programming*, Doctoral thesis, University of Michigan, Ann Arbor, 1965.

[11] W. Fenchel, *Convex cones, sets, and functions*, Lecture Notes, Princeton University, Princeton, 1953.

[12] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming*, John Wiley, New York, 1968.

[13] B. L. Fox, *Stability of the dual cutting-plane algorithm for concave programming*, RM-6147-PR, The RAND Corporation, Santa Monica, Calif., 1970.

[14] M. Frank and P. Wolfe, *An algorithm for quadratic programming*, Naval Res. Logist. Quart., 3 (1956), pp. 95–110.

[15] D. Gale, *A geometric duality theorem with economic applications*, Rev. Economic Studies, 34 (1967), pp. 19–24.

[16] A. M. Geoffrion, *Primal resource-directive approaches for optimizing nonlinear decomposable systems*, Operations Res., 18 (1970), pp. 375–403.

[17] ———, *Elements of large-scale mathematical programming*, Management Sci., 16 (1970), pp. 652–691.

[18] ———, *Generalized Benders decomposition*, J. Optimization Theory Appl., to appear.

[19] E. G. Golshtein, *A general approach to the linear programming of block structures*, Soviet Physics Dokl., 11 (1966), no. 2, pp. 100–103.

[20] F. J. Gould, *Nonlinear Duality Theorems*, Rep. 6817, Center for Mathematical Studies in Business and Economics, University of Chicago, Chicago, 1968.

[21] K. O. Kortanek and J. P. Evans, *Asymptotic Lagrange regularity for pseudoconcave programming with weak constraint qualification*, Operations Res., 16 (1968), pp. 849–857.

[22] L. S. Lasdon, *Duality and decomposition in mathematical programming*, IEEE Trans. System Sci. and Cybernetics, SSC-4 (1968), pp. 86–100.

[23] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley, New York, 1969.

[24] O. L. Mangasarian, *Nonlinear Programming*, McGraw-Hill, New York, 1969.

[25] O. L. Mangasarian and J. Ponstein, *Minmax and Duality in Nonlinear Programming*, J. Math. Anal. Appl., 11 (1965), pp. 504–518.

[26] R. T. Rockafellar, *Duality in nonlinear programming*, Mathematics of the Decision Sciences, Part 1, G. B. Dantzig and A. F. Veinott, eds., American Mathematical Society, Providence, 1968, pp. 401–422.

[27] ———, *Convex functions and duality in optimization problems and dynamics*, Mathematical Systems Theory and Economics, H. W. Kuhn and G. P. Szegö, eds., Springer-Verlag, Berlin, 1969.

[28] J. Stoer, *Duality in nonlinear programming and the minimax theorem*, Numer. Math., 5 (1963), pp. 371–379.

[29] I. Takahashi, *Variable separation principle for mathematical programming*, J. Operations Res. Soc. Japan, 6 (1964), pp. 82–105.

[30] H. Uzawa, *Iterative methods for concave programming*, Studies in Linear and Nonlinear Programming, K. Arrow, L. Hurwicz and H. Uzawa, eds., Stanford University Press, Stanford, Calif., 1958.

[31] R. M. Van Slyke and R. J. B. Wets, *A duality theory for abstract mathematical programs with applications to optimal control theory*, J. Math. Anal. Appl., 22 (1968), pp. 679–706.

[32] A. C. Williams, *Nonlinear Activity Analysis and Duality*, presented at the Sixth International Symposium on Mathematical Programming, Princeton, 1967.

[33] R. Wilson, *Computation of optimal controls*, J. Math. Anal. Appl., 14 (1966), pp. 77–82.

[34] P. Wolfe, *A duality theorem for non-linear programming*, Quart. Appl. Math., 19 (1961), pp. 239–244.