OPERATIONS RESEARCH, Vol. 23, No. 3, May-June 1975

Surrogate Constraint Duality in Mathematical Programming

FRED GLOVER

University of Colorado, Boulder, Colorado

(Received original November 12, 1973; final, November 12, 1974)

This paper presents a unified development of a surrogate duality theory that is applicable to problems in which Lagrangean duality gaps limit the usefulness of standard duality approaches. A surrogate dual is created by generating a single constraint to replace the original problem constraints, rather than by absorbing these constraints into the objective function as in the Lagrangean. We give necessary and sufficient conditions for optimality both with and without the imposition of complementary slackness, and also consider a related 'overestimating' surrogate that may be used in a strategy to bracket the optimal value of the primal. The optimality conditions invite direct comparison with those for Lagrangean duality, demonstrating not only that the surrogate approach yields smaller duality gaps than the Lagrangean (as first observed by Greenberg and Pierskalla), but also giving a precise characterization of the manner and extent to which this occurs. Concepts of parametric and relative subgradients, paralleling (and generalizing) the concept of the subgradient of ordinary duality theory, lead to easily stated results that encompass both surrogate and Lagrangean duality, as well as their composite, in a single framework.

DUALITY theory in mathematical programming has customarily been based upon the use of a generalized Lagrangean function to define the dual. Very elegant results have emerged linking optimality conditions for the dual to those for the primal. Out of these results have arisen solution strategies for the primal that exploit the properties of the primal-dual interrelations. Some of these strategies have been remarkably successful, particularly for problems in which the duality gap—the amount by which optimal objective function values for the two problems differ—is nonexistent or small.

A different type of solution strategy has been proposed for solving mathematical programs in which duality gaps are likely to be large. In contrast to the Lagrangean strategy, which absorbs a set of constraints into the objective function, this strategy replaces the original constraints by a new one called a surrogate constraint.

There have been several important contributions to the literature on surrogate constraints, but there has been no body of theory, comparable to that for Lagrangean functions that provides necessary and sufficient conditions for the absence of duality gaps for a *surrogate dual*. In fact, most of the attempts to generate 'best' surrogate constraints, particularly in the setting of integer programming, have resorted to Lagrangean analysis (or its equivalent). This paper develops a surrogate duality theory that provides exact conditions under which surrogate duality gaps cannot occur. These conditions (both necessary and sufficient) are less confining than those governing the absence of Lagrangean duality gaps. Furthermore, they give a precise characterization of the difference between surrogate and Lagrangean relaxation. We develop theorems that are analogous to major results of standard duality theory, allowing direct comparisons to these earlier results.

The foundation for surrogate duality theory, as developed here, rests on the introduction of several new concepts that generalize the notions of perturbation functions, subgradients, and stability as normally employed in Lagrangean duality theory. These include *parametric perturbation functions*, *parametric* and *relative subgradients*, and *relative stability*. With the use of these new concepts, we identify an 'overestimating' function whose optimal value (together with that of the surrogate dual) brackets the optimal value of the primal. We thereby deduce inequalities involving both the parametric perturbation function and the ordinary perturbation function that constitute the fundamental building blocks of surrogate duality.

These relations do not require the complementary slackness conditions that provide a cornerstone of ordinary duality theory. Nevertheless, because complementary slackness has led to useful solution strategies in a variety of settings, we also show that it is possible to superimpose complementary slackness on surrogate duality to give alternative characterizations of optimality.

Our principal results do not rely on assumptions of convexity, since the major applications of surrogate constraints occur in situations where convexity is conspicuously missing (e.g., integer programming). However, to provide further ties to standard duality theory, we also examine the 'convex case' by introducing the notion of relative stability.

A key feature of the development is the statement of two general duality theorems for a dual defined relative to an arbitrary number of surrogate constraints together with a weighted objective. These results imply as a special case the chief consequences of Lagrangean and surrogate duality, as well as their conjunction. The nature of these results and their connections to earlier work will be discussed more fully after we introduce some fundamental definitions.

1. THE PRIMAL PROBLEM AND THE SURROGATE DUAL

The primal problem of mathematical programming will be written

$$P\colon \min_{x\epsilon X} f(x), \quad \text{subject to} \quad g(x) \leqq 0,$$

where f and each component $g_i(x)$ of the vector $g(x)$ are real-valued functions defined on X. No special characteristics of these functions or of X will be assumed unless otherwise specified. We will not bother to distinguish between row and column vectors—all vector products are dot products in the usual sense and conformable dimensious are taken for granted.

A surrogate constraint for P is a linear combination of the component constraints of $g(x) \leqq 0$ that associates a multiplier u_i with each $g_i(x)$ to produce the inequality $ug(x) \leqq 0$, where $u = (u_i)$. Clearly, this inequality is implied by $g(x) \leqq 0$ whenever $u \geqq 0$. Correspondingly, we define the surrogate problem

$$SP(u)\colon \min_{x\epsilon X} f(x), \quad \text{subject to} \quad ug(x) \leqq 0.$$

The optimal objective function value for SP(u) will be denoted by $s(u)$, or more precisely,

$$s(u) = \inf_{x\epsilon X(u)} f(x), \quad \text{where} \quad X(u) = \{x\epsilon X\colon ug(x) \leqq 0\}.$$

Since $SP(u)$ is a relaxation of P (for u nonnegative), $s(u)$ cannot exceed the optimal objective function value for P and approaches this value more closely as $ug(x) \leqq 0$ becomes a more 'faithful' representation of the constraint $g(x) \leqq 0$. Choices of the vector u that improve the proximity of $SP(u)$ to P—i.e., that provide the greatest values of $s(u)$—yield strongest surrogate constraints in a natural sense, and motivate the definition of the *surrogate dual*

$$SD\colon \max_{u \geqq 0} s(u).$$

The surrogate dual may be compared with the Lagrangean dual

$$LD\colon \max_{u \geqq 0} L(u),$$

where $L(u)$ is the function given by

$$L(u) = \inf_{x\epsilon X} \{f(x) + ug(x)\}.$$

It should be noted that $s(u)$ is defined relative to the set $X(u)$, which is more restrictive than the set X relative to which the Lagrangean $L(u)$ is defined. Also, modifying the definition of $L(u)$ by replacing X with $X(u)$, while possibly increasing $L(u)$, will nevertheless result in $L(u) \leqq s(u)$ because of the restriction $ug(x) \leqq 0$; that is, $L(u)$ may be regarded as an 'underestimating' function for both the surrogate problem and the primal problem.

Another immediate observation is that any optimal solution to the surrogate problem that is feasible for the primal is automatically optimal for the primal. (No complementary slackness conditions are required, as for the Lagrangean.) These notions have been embodied in the applications of surrogate constraints since they were first proposed and are entirely conspicuous. Taken together, they provide what may be called a 'first duality theorem' for surrogate mathematical programming. A formal statement of this theorem will be given after additional groundwork has been laid, at which time it will be possible to enlarge its content. Thereupon, the advanced surrogate duality results will be developed. We will first summarize some of the earlier contributions to surrogate constraint theory and provide a detailed interpretive guide to the results that lie ahead.

2. BACKGROUND AND PRELIMINARIES

Since their introduction,[11] surrogate constraints have been proposed by a variety of authors for use in solving nonconvex problems, especially those of integer programming. Surrogate constraints that were 'strongest' for 0–1 integer programming under certain relaxed assumptions were suggested by Balas[1] and Geoffrion.[9] The paper by Geoffrion also contained a computational study that demonstrated the practical usefulness of such proposals. Methods for generating strongest surrogate constraints according to other definitions, in particular segregating side conditions and introducing normalizations, were subsequently proposed.[12] However, all of these proposals used relaxation assumptions whose effect was to replace the original nonconvex primal problem by a linear programming problem. The structure of this LP problem is sufficiently simple that the distinction between the surrogate constraint approach and the Lagrangean approach vanishes.

The first proposal for surrogate constraints used notions that closely accord with those considered here, defining a strongest surrogate exactly as in Section 1. A theorem of reference 11 leads to a procedure for searching for optimal surrogate multipliers that can obtained stronger surrogate constraints for a variety of problems than subsequent proposals (see Note 1).

Greenberg and Pierskalla provide the first major theoretical treatment of surrogate constraints in the context of general mathematical programming. These authors show that $s(u)$ is quasiconcave, thus assuring that any local maximum for $s(u)$ is a global maximum (disregarding sequences of 'plateaus'). In addition, their paper was the first to demonstrate rigorously a smaller duality gap for the surrogate approach than for the Lagrangean approach. It also provided sufficient conditions for the nonoccurrence of surrogate duality gaps. Specifically, surrogate duality gaps are shown to be eliminated if the perturbation function (of standard duality theory) is a closed quasiconvex function. More recently, GREENBERG[14] has developed applications of 'surrogated Lagrangeans' to provide strengthened penalty functions, and GREENBERG AND PIERSKALLA[16] have unified and extended these previous developments in a nonlinear multiplier setting via a theory of quasiconjugate functions.

The developments of this paper owe a large debt to the work of Greenberg and Pierskalla, and to previous contributions to duality theory generally (e.g., the landmark work of CHARNES, COOPER, AND KORTANEK,[2] COTTLE,[3] DANTZIG,[4] EVANS AND GOULD,[5] EVERETT,[6] FENCHEL,[7] GALE,[8] LUENBERGER,[17] MANGASARIAN AND PONSTEIN,[18] ROCKAFELLAR,[19] STOER,[20] and WOLFE,[21] among others). The format of our presentation is motivated by Geoffrion's outstanding exposition of Lagrangean duality.[10]

Chiefly, the results of the next section are based upon relations between optimality conditions for the surrogate problem and four central concepts: the perturbation function (of Lagrangean duality theory), the parametric perturbation function, the parametric subgradient, and the relative subgradient. Under convexity, we also introduce a concept of relative stability that plays the same role in surrogate duality as the concept of ordinary stability in Lagrangean duality. In particular, under convexity, necessary and sufficient conditions for surrogate optimality are shown to be equivalent to the conditions of relative stability.

Optimality conditions for surrogate duality are the requirements (already discussed) that the surrogate multiplier vector u is nonnegative, x is optimal for the surrogate problem, and x is feasible for the primal problem. 'Strong' optimality conditions add the requirement of complementary slackness $[ug(x)=0]$. We first give theorems that show that the optimality conditions are essentially equivalent to the statement that the negative of the surrogate multiplier vector is a relative subgradient of the parametric perturbation function with respect to the ordinary perturbation function at the origin. This statement is less restrictive than the corresponding statement for the Lagrangean optimality conditions and affords a direct means of identifying the additional latitude of the surrogate problem that yields smaller duality gaps than provided by the Lagrangean. Strong optimality conditions for the surrogate problem are then shown to be equivalent to the statement that the negative of the surrogate multiplier vector is a parametric subgradient for the parametric perturbation function at the origin. Although the strong surrogate optimality conditions are more confining than the regular surrogate optimality conditions, they are still less restrictive than the Lagrangean conditions—i.e., an optimal solution can be available for the primal by means of the surrogate problem under complementary slackness when the Lagrangean is unable to provide such a solution (because of a Lagrangean duality gap).

It is useful to consider a special surrogate problem related to the parametric perturbation function in the same way that the surrogate problem $SP(u)$ is related to the ordinary perturbation function. Under certain circumstances, it is possible for this special surrogate problem to 'overestimate' (or even solve) the primal, while the problem $SP(u)$ lies in a duality gap. We give a theorem that provides a necessary and sufficient condition for this 'overestimation' to occur. This result demonstrates that overestimation is more frequently attained than surrogate optimality, suggesting a strategy of using both the special surrogate and the ordinary surrogate in an attempt to 'bracket' an optimal solution to the primal.

Section 4 introduces a generalized surrogate problem defined relative to an arbitrary number of surrogate constraints and a weighted objective. For this generalized problem we give two duality theorems that imply as special cases the results of the preceding section and all of the corresponding results for Lagrangean duality. In addition, these theorems give complete necessary and sufficient conditions for optimality in a combined surrogate-Lagrangean approach and have implications for surrogate duality and Lagrangean duality independent of each other. Among other things, it is shown that the restriction of earlier results to 'subgradients' defined in terms of optimal multiplier vectors is unnecessary. In both Lagrangean and surrogate duality, the concept of the relative subgradient makes it possible to extend the statement of 'optimality equivalence' to hold with any nonpositive vector in the role of the relative subgradient.

Finally, in Section 5 we consider the standard convexity assumptions usually imposed on duality theory and introduce the concept of relative stability, which provides the desired equivalences to the previous optimality results under convexity.

A simplified scheme is given for proving the central relations of Section 4 (which imply all the others for the nonconvex case). By means of this scheme it is possible to justify the main results of Section 5 by a brief commentary linking a handful of straightforward observations.

3. FUNDAMENTAL RESULTS FOR SURROGATE DUALITY

For our starting point, we define two problems $P(y)$ and $P(y, u)$ related to the primal problem P and the surrogate problem $SP(u)$:

$$P(y)\colon \min_{x\epsilon X} f(x), \quad \text{subject to} \quad g(x)\leqq y.$$

$$P(y, u)\colon \min_{x\epsilon X(u)} f(x)-ug(x), \quad \text{subject to} \quad g(x)\leqq y.$$

The standard *perturbation function* $v(y)$ is the optimal objective function value for $P(y)$:

$$v(y)=\inf_{x\epsilon X} f(x), \quad \text{subject to} \quad g(x)\leqq y.$$

The *parametric perturbation function* $v(y, u)$ is correspondingly defined to be the optimal objective function value for $P(y, u)$:

$$v(y, u)=\inf_{x\epsilon X(u)} f(x)-ug(x), \quad \text{subject to} \quad g(x)\leqq y.$$

Since $P(0)$ is the primal, $v(0)$ is the optimal objective function value for P. It should be noted that the objective function $v(y, u)$ for $P(y, u)$ may be thought of as an 'overestimating function' in the same sense that the Lagrangean is an 'underestimating function.' [The condition $ug(x)\leqq 0$ assures $f(x)-ug(x)\geqq f(x)$ for all $x\epsilon X(u)$.] More generally, given the assumption $u\geqq 0$, we can immediately write the following inequalities and equalities as a consequence of the preceding definitions.

Basic inequalities:

$v(y, u)\geqq v(y)$	and $v(y, u)\geqq s(u)$.
$v(y)\geqq s(u)$	if some x satisfying $ug(x)\leqq 0$ is optimal for $P(y)$.
$s(u)\geqq v(y)$	if some x satisfying $g(x)\leqq y$ is optimal for $SP(u)$.

The foregoing imply as special cases: $v(y)\geqq s(u)$ if $uy\leqq 0$, and $v(0, u)\geqq v(0)\geqq s(u)$.

Basic equalities:

$v(y, u)=v(y)$	if and only if some x optimal for $P(y)$ satisfies $ug(x)=0$ [in which case the set of all x optimal for $P(y, u)$ is the set of all x optimal for $P(y)$ that satisfy $ug(x)=0$].
$v(y)=s(u)$	if some x optimal for $P(y)$ satisfies $ug(x)\leqq 0$ and some x optimal for $SP(u)$ satisfies $g(x)\leqq y$ [in which case all x of the first type are optimal for $SP(u)$ and all x of the second type are optimal for $P(y)$].

The second equality is of course a consequence of the second and third inequalities. We may also conclude: if $uy\leqq 0$, then $v(y)=s(u)$ if and only if some x optimal for $SP(u)$ satisfies $g(x)\leqq y$ (in which case the set of all x optimal for $P(y)$ is the set of all x optimal for $SP(u)$ that satisfy $g(x)\leqq y$). Stated in another way:[15] an optimal solution x for $SP(u)$ is optimal for all $P(y)$ such that $uy\leqq 0$ and $g(x)\leqq y$.

These observations contain all the ingredients of a 'first duality theorem,' which we state upon formalizing the surrogate optimality conditions, as follows.

Surrogate Optimality Conditions

(i) $u\geqq 0$.

(ii) x is optimal for $SP(u)$.

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

THEOREM 1. *The surrogate optimality conditions imply x is optimal for the primal, u is optimal for the surrogate dual, and their optimal objective function values are equal. Moreover, if there is a nonnegative u such that $v(0)=s(u)$, then the set of all optimal solutions to P is precisely the set of x that together with u satisfy the surrogate optimality conditions.*

The content of this theorem, previewed in Section 2, provides the starting point for surrogate duality theory. Using the basic equalities, we can also go a step farther. We define the *strong optimality conditions* to consist of the surrogate optimality conditions plus the condition of complementary slackness:

$$\text{(iv)}\ ug(x)=0.$$

The latter is suggested by analogy with the Lagrangean and the fact that a sophisticated body of solution techniques has grown up for exploiting the complementary slackness condition. In fact, the strong optimality conditions are the same as the Lagrangean optimality conditions upon replacing (ii) by the statement that x is optimal for the Lagrangean [i.e., x minimizes $f(x)+ug(x)$ over X]. However, strong optimality conditions accommodate a somewhat broader range of possibilities than the Lagrangean optimality conditions, as will be subsequently demonstrated. By means of our earlier observations, however, we can immediately state the following addendum to Theorem 1.

COROLLARY. *The strong optimality conditions imply x is optimal for both P and $P(0, u)$, and $v(0, u)=v(0)$. Moreover, if any pair u, x satisfies the strong optimality conditions, then these conditions characterize all x that are optimal for $P(0, u)$ and all u such that $v(0, u)=v(0)$.*

The principal goal of this section will be to identify necessary and sufficient conditions for the two sets of optimality conditions to hold, so that the conclusions of Theorem 1 and its corollary will be at hand. To provide a foundation for the main results to follow, we first recall the definition of subgradient used in Lagrangean duality theory.

A vector γ^* is called a *subgradient* of a function $F(y)$ at the point y^* if $F(y)\geqq F(y^*)+\gamma^*(y-y^*)$ for all y. Not all functions have subgradients by this definition, and the direction of the inequality is sometimes reversed to enable the definition to apply—e.g., for concave functions. The subgradient, like the gradient (when the latter exists), points from y^* in the direction of those y that yield 'improved' values of $F(y)$ [i.e., $F(y)\leqq F(y^*)$ implies that y is contained in the half space $\gamma^*(y-y^*)\leqq 0$].

By extension, we will call $\gamma(\alpha)$ a *parametric subgradient* of the function $F(y; \alpha)$ at the point y^* if $F(y; \alpha)\geqq F(y^*; \alpha)+\gamma(\alpha)(y-y^*)$ for all y. The parametric subgradient is an improving direction for the function $F(y; \alpha)$ for some *given* value of the parameter α. (The fact that α is constrained to the same value throughout the foregoing inequality suggests the alternative nomenclature of 'constrained subgradient.') By means of the parametric subgradient it will be possible to identify necessary and sufficient conditions for the strong optimality conditions to hold.

Extending the notion of the parametric subgradient, we define $\gamma(\alpha)$ to be a *relative subgradient* of a function $F(y; \alpha)$ with respect to a function $G(y; \alpha)$ at the point y^* if $F(y; \alpha) \geqq G(y^*; \alpha) + \gamma(\alpha)(y - y^*)$ for all y. The relative subgradient points in a direction in which the function $F(y; \alpha)$ improves *relative to* $G(y^*; \alpha)$ for a given α. By means of this notion, relations equivalent to the surrogate optimality conditions can be expressed. In addition, the relative subgradient provides the basis for the 'composite' duality theorems of the next section.

We now connect the parametric and relative subgradients to the optimality conditions for P and SD. In the results to follow, the functions $F(y; \alpha)$ and $G(y; \alpha)$ of the preceding definitions correspond to the functions $v(y; u)$ and $v(y)$, with the multiplier vector u taking the role of the parameter α. These results, as those of Lagrangean duality, refer to 'subgradients,' taken at the origin, i.e., for $y^* = 0$ (see Note 2).

THEOREM 2. *Assume* $u \geqq 0$. *Then the surrogate optimality conditions are met and the conclusions of Theorem 1 apply if and only if* $-u$ *is a relative subgradient of* $v(y, u)$ *with respect to* $v(y)$ *at the origin; i.e.,*

$$v(y, u) \geqq v(0) - uy \quad \textit{for all } y. \tag{1}$$

It is instructive to compare Theorem 2 with the corresponding theorem for Lagrangean duality, which says that the Lagrangean optimality conditions are met if and only if $-u$ is a subgradient of $v(y)$ at the origin; i.e.,

$$v(y) \geqq v(0) - uy \quad \text{for all } y. \tag{2}$$

The additional latitude supplied by (1) over (2) is seen by reference to the inequality $v(y, u) \geqq v(y)$ (for all $u \geqq 0$). This inequality holds for two reasons. First, $v(y, u)$ is defined relative to a more restrictive constraining relation than $v(y)$ [since it must additionally accommodate $ug(x) \leqq 0$]. Second, $f(x) - ug(x)$, the objective function for $v(y, u)$, is an 'overestimating function' for the objective function of $v(y)$. Taken together, these conditions doubly ensure that the gap region for (1) is smaller than for (2)—sometimes dramatically so, as illustrated in reference 13.

It might be guessed that part of the difference in the gap regions supplied by (1) and (2) is due to the complementary slackness condition that is required for Lagrangean optimality. Indeed, by the basic equalities, $v(y, u) = v(y)$ if and only if $ug(x) = 0$ for some optimal solution x to the problem $P(y)$. Thus, the difference in restrictiveness between (1) and (2) depends precisely upon complementary slackness, not simply for the problem P, but for all problems $P(y)$. Consequently, the imposition of this condition for P accounts for only a portion of the difference between (1) and (2) and suggests that the incorporation of complementary slackness via the strong optimality conditions will still provide a latitude that is missing in the Lagrangean. This is an attractive possibility since, as already noted, complementary slackness is highly exploitable from an algorithmic standpoint. We now demonstrate the consequence of this condition as imbedded in the strong optimality conditions.

THEOREM 3. *The strong optimality conditions are met, and the conclusions of both Theorem 1 and its corollary apply, if and only if* $-u$ *is a parametric subgradient of* $v(y, u)$ *at the origin*; *i.e.*,

$$v(y, u) \geqq v(0, u) - uy \text{ for all } y. \tag{3}$$

It is clear that the gap region admitted by Theorem 3 is not as small as that admitted by Theorem 2, since $v(0, u) \geqq v(0)$; hence some of the latitude supplied by (1) is not available in (3). Nevertheless, (3) still produces a smaller gap region than (2), for the same reasons that (1) does. To see this fact, we note that $v(0, u) = v(0)$ under the assumption of complementary slackness (given $u \geqq 0$), and the imposition of this assumption in both the Lagrangean and the strong optimality conditions demonstrates that the latter are less confining than the former. By way of example, consider the problem

$$\min 10x_1 + 7x_2, \text{ subject to } 10x_1 + x_2 \geqq 1 \text{ and } x_j = 0, 1, j = 1, 2,$$

(or $-10x_1 - x_2 + 1 \leqq 0$). Since there is only one constraint, it may be taken to be the surrogate constraint; and the optimal solution is $x_1 = 0$, $x_2 = 1$, which satisfies complementary slackness. On the other hand, the optimal Lagrangean for this problem is min $0x_1 + 6x_2 + 1$, $x_j = 0, 1$, yielding solutions of $x_1 = 0$, $x_2 = 0$, and $x_1 = 1$, $x_2 = 0$, both of which are clearly nonoptimal for the original problem. The optimal objective function value of $+1$ for the Lagrangean is also substantially removed from the true optimum, demonstrating the existence of a sizeable Lagrangean optimality gap.

It is interesting to note further that the incorporation of complementary slackness into the strong optimality conditions makes it possible to suppress the assumption $u \geqq 0$ (which is required in Theorem 2) since it is a direct consequence of (3). [This is likewise accomplished for the Lagrangean via (2).]

For the final result of this section, we turn our attention to an optimal surrogate constraint for the 'overestimating function' $f(x) - ug(x)$. Since $f(x) - ug(x)$ always equals or exceeds $f(x)$ for $ug(x) \leqq 0$, it is useful to know when $\inf_{x \epsilon X(u)} f(x) - ug(x)$ provides an upper bound for $v(0)$, since then the relation $\inf_{x \epsilon X(u)} f(x) - ug(x) \geqq v(0) \geqq \inf_{x \epsilon X(u)} f(x)$ can be used to bracket the optimal objective function value for P. The conditions under which this relation holds are somewhat less stringent than the conditions of Theorems 2 and 3, as we now show.

THEOREM 4. *Assume* $u \geqq 0$. *Then the optimum for the overestimating surro-*

gate [Inf $f(x)-ug(x)$ *over* $X(u)$] *is greater than or equal to the optimum for* P *if and only if*

$$v(y, 2u) \geqq v(0) - uy \text{ for all } y. \quad (4)$$

Comparing (4) to (1), we see that the overestimating attempt will succeed in a somewhat wider variety of circumstances than those under which the surrogate optimality conditions hold, since $v(y, 2u) \geqq v(y, u)$ for all $u \geqq 0$, with equality only under complementary slackness—that is, the increased latitude of (4) over (1) is approximately comparable to the increased latitude of (1) over the Lagrangean conditions.

4. MULTIPLE SURROGATE CONSTRAINTS AND COMPOSITE DUALITY

In this section we generalize the results of Section 3 to a surrogate problem defined relative to multiple surrogate constraints and a weighted (Lagrangean) objective function. From this generalization the results of the preceding section and of Lagrangean theory emerge as a special case. In addition, characterizations of 'composite' optimality conditions unifying the two approaches are provided.

We will show that for multiple surrogate constraints, necessary and sufficient conditions for optimality can take precisely the form of (1) with u replaced by the sum of the individual surrogate multiplier vectors. Indeed, these conditions can also take a variety of other forms, all of which are subsumed by a simple inequality involving the relative subgradient.

To express our results, we first introduce notation appropriate to the general case. In place of the surrogate problem $SP(u)$ of the preceding sections we introduce the 'multi-parameter' surrogate problem $SP(u, z)$, defined by

$$SP(u, z)\colon \min_{x \epsilon X(u)} f(x) + zg(x),$$

which has the same form as $SP(u)$ except for the addition of $zg(x)$ to the objective function. Here, however, u is no longer a vector but a matrix of fixed dimensions, and the condition $ug(x) \leqq 0$ [as in the definition $X(u) = \{x \epsilon X\colon ug(x) \leqq 0\}$] therefore represents not just one but a collection of surrogate constraints. As before, $g(x) \leqq 0$ implies $ug(x) \leqq 0$ whenever $u \geqq 0$, and the Lagrangean function $f(x)+zg(x)$ is an underestimating function for $f(x)$ if $z \geqq 0$. We denote the optimal objective function value for $SP(u, z)$ by $s(u, z)$; i.e., $s(u, z) = \inf_{x \epsilon X(u)} f(x) + zg(x)$.

By the foregoing remarks, $v(0) \geqq s(u, z)$ for u and z nonnegative. In one respect, however, we will allow $f(x)+zg(x)$ to differ from the Lagrangean. After the manner of Theorem 4, we will also want to consider cases in which z may not be nonnegative in order to specify conditions under which $s(u, z)$ may provide an overestimate of $v(0)$, so that we can bracket the optimal objective function value of P. Our main result concerning this possibility will be given in an 'overestimating lemma.'

The dual of P relative to the problem $SP(u, z)$ will be defined by D: $\min_{u \geqq 0, z \geqq 0} s(u, z)$. In a solution strategy for P based on solving D [hence $SP(u, z)$], it is reasonable to impose restrictions of the form $u \epsilon U$ and $z \epsilon Z$, which, for example, limit the nonzero coefficients of u and z to different subsets of the constraints $g_i(x) \leqq 0$. [This restriction is particularly appropriate when the form of $g(x) \leqq 0$ makes it possible for the surrogate constraints of $ug(x) \leqq 0$ to form a decomposed system. Similarly, for the most part it does not seem useful to select the nonzero coefficients of z to correspond to the same constraints as one of the rows of u—unless these coefficients are allowed to be negative—since then the underestimating character of $f(x)+zg(x)$ implies that the maximum value of $s(u, z)$ always occurs for $z=0$.] More generally, $u \epsilon U$ and $z \epsilon Z$ may represent 'normalizations' of the type introduced in reference 12. In the following development we will take such restrictions to be implicit.

To achieve the requisite generality, we also replace the problem $P(y; u)$ by the problem $P(y; \alpha)$, where α is the triple (u, z, w), and $P(y; \alpha)$ is given by

$$P(y; \alpha): \min_{x \epsilon X(u)} f(x) + (z-w)g(x) \text{ subject to } g(x) \leqq y.$$

Here w is of course of the same dimension as z. The optimal objective function value $v(y; \alpha)$ for $P(y; \alpha)$ is defined in the same way that $v(y; u)$ is defined for $P(y, u)$. It may be noted that $P(y; \alpha)$ is exactly the same as $P(y; u)$ for the case in which u is a vector, $z=0$, and $w=u$.

Our chief results will be in terms of parametric and relative subgradients involving $v(y; \alpha)$, where $\gamma(\alpha)$ (in the earlier definitions of these subgradients) is $-w$.

By convention, we will assume throughout that all problems encountered have finite values. To facilitate the development, we will make two fundamental observations. Special instances of these observations underlie many of the results of standard duality theory, and it is useful to isolate and express these observations in their general form, whereupon a variety of results may be immediately justified as subcases.

As a basis for these observations, consider the two problems

$$P_R: \min_{x \epsilon R} r(x) \quad \text{and} \quad P_Q: \min_{x \epsilon Q} q(x),$$

and let r and q respectively denote their optimal objective function values. Define $t(x) = r(x) - q(x)$.

Observation 1: $r \geqq q$ if there is an optimal solution x for P_R that satisfies $x \epsilon Q$ and $t(x) \geqq 0$.

Observation 2: Assume $Q \subset R$ and $t(x) \leqq 0$ for all $x \epsilon Q$. Then $r=q$ if and only if there is an optimal solution x^* to P_R such that $x^* \epsilon Q$ and $t(x^*)=0$. Moreover, x^* is optimal for P_Q, and the set of all optimal solutions to P_Q is the set of all optimal solutions to P_R satisfying $x \epsilon Q$ and $t(x)=0$. [Observation 2 gives Theorem 1 and the corresponding theorem for Lagrangean duality as a special case, where $t(x)=0$ is the complementary slackness condition for the Lagrangean problem.]

As a direct consequence of these observations, we have the following relations.

I: *Given* $u \geqq 0$, $z \geqq 0$: $v(0) \geqq s(u, z)$; and $v(0)=s(u, z)$ if and only if there is an optimal solution to $SP(u, z)$ such that $g(x) \leqq 0$ and $zg(x)=0$.

II: *Given* $w \geqq 0$: $v(0; \alpha) \geqq s(u, z)$; and $v(0; \alpha)=s(u, z)$ if and only if there is an optimal solution to $SP(u, z)$ such that $g(x) \leqq 0$ and $wg(x)=0$.

III: *Given* $w \geqq z$: $v(0; \alpha) \geqq v(0)$; and $v(0; \alpha)=v(0)$ if and only if there is an optimal solution to P such that $ug(x) \leqq 0$ and $(z-w)g(x)=0$.

In each case, the inequality follows from observation 1 and the equality follows from observation 2.

Upon stating two lemmas, we will be able to provide the two central theorems of this section.

NONNEGATIVITY LEMMA. *$v(y; \alpha) \geqq v(y; \alpha)-wy$ for all y implies $w \geqq 0$.*

Proof. Take $y=e_j$ (the unit vector with a 1 in the jth position and 0's elsewhere). Then $wy=w_j$ and we obtain $w_j \geqq v(0; \alpha)-v(e_j; \alpha) \geqq 0$ [since $v(0; \alpha) \geqq v(y; \alpha)$ for all $y \geqq 0$], proving $w \geqq 0$.

OVERESTIMATION LEMMA. *Let $v^*(\alpha)$ be an arbitrary function of α. Then $s(u, z) \geqq v^*(\alpha)$ if and only if $v(y; \alpha) \geqq v^*(\alpha)-wy$ for all y and all $w \geqq 0$.*

Proof. For the 'if' part, let $y=g(x)$. Then $f(x)+(z-w)g(x) \geqq v^*(\alpha)-wg(x)$ for all $x \epsilon X(u)$. Hence $f(x)+zg(x) \geqq v^*(\alpha)$ for all $x \epsilon X(u)$, and taking the infimum of the left-hand side yields the desired result. For the 'only if' part, $s(u, z) \geqq v^*(\alpha)$ gives $f(x)+zg(x) \geqq v^*(\alpha)$ for all $x \epsilon X(u)$. Hence for all $w \geqq 0$ and all y such that $g(x) \leqq y$, $f(x)+zg(x)-wg(x) \geqq v^*(\alpha)-wy$. Taking the infimum of the left-hand side over $x \epsilon X(u)$ and $g(x) \leqq y$ yields the desired result (under the standard convention whereby the infimum is taken to be $+\infty$ for any y that yields an empty 'defining set' for the argument x).

These lemmas reflect key properties of the relative subgradient and can readily be generalized to functions other than $v(y; \alpha)$ and $s(u, z)$ whose 'objective functions' (over x) differ by the quantity $\gamma(\alpha)g(x)$ and whose 'constraining conditions' differ by the requirement $g(x) \leqq y$.

It is interesting to note that Theorem 4 of Section 3 is actually a special case of the overestimation lemma. In particular, if we let $v^*(\alpha)$ be the constant function $v(0)$ and $u=w=-z \geqq 0$ (when u is a vector), the inequality of the overestimation lemma in terms of the earlier notation becomes $v(y; 2u) \geqq v(0)-uy$ for all y, which is precisely (4).

Our main theorems may now be stated as follows.

DUALITY THEOREM (Optimality for P and D). *Assume $u \geqq 0$, $z \geqq 0$. Then $s(u, z)=v(0)$ if and only if there is an optimal solution x^* to $SP(u, z)$ such* *that* $g(x^*) \leqq 0$ *and* $zg(x^*) = 0$ [*whereupon the pair* (u, z) *is optimal for* D, x^* *is optimal for* P, *and all optimal solutions to* P *are solutions to* $SP(u, z)$ *of the form of* x^*]. *Moreover, a necessary and sufficient condition for* $s(u, z) = v(0)$ *is that, for all nonnegative* w, *the vector* $-w$ *is a relative subgradient of* $v(y; \alpha)$ *with respect to* $v(y)$ *at the origin; i.e.,* $v(y; \alpha) \geqq v(0) - wy$ *for all* y *and all* $w \geqq 0$.

Proof. The first part of the theorem is just the relation I, expanded by means of observation 2 to include a statement concerning the set of optimal solutions to P and noting that (u, z) must necessarily be optimal for D under the stated conditions. The second part of the theorem is an immediate consequence of the overestimation lemma with $v^*(\alpha) = v(0)$ and the inequality $v(0) \geqq s(u, z)$ from I.

The foregoing theorem gives Theorem 2 of Section 3 for the case in which u is a vector, $z = 0$, and $w = u$. The insights of Section 4 that derived from taking $w = u$ indicate that this choice is useful from a heuristic standpoint, but the duality theorem demonstrates that it is not the only one. Also, the 'natural' generalization of Theorem 2 to the case of multiple surrogate constraints, whose form was indicated at the beginning of this section, occurs when u is a matrix with $z = 0$ and $w = 1u$, where 1 is a vector of 1's (i.e., w becomes a summation of the multiplier vectors for the individual surrogate constraints). Theorem 2 is then the case in which u has a single row and 1 is a scalar.

It may be noted that complementary slackness holds for the multipliers that weight the objective function, just as for the ordinary Lagrangean. However, this complementary slackness is not 'identical' to that for the Lagrangean nor, insofar as it is independent of the vector w, does it apply to the complementary slackness that appears in the strong optimality conditions of Theorem 3. The theorem that directly specializes to both the Lagrangean case and the 'strong' surrogate case is the following.

RESTRICTED DUALITY THEOREM [Optimality for $P(0; \alpha)$, and restricted optimality for P and D]. *If* $w \geqq 0$, *then* $s(u, z) = v(0; \alpha)$ *if and only if there is an optimal solution* x^* *to* $SP\ (u, z)$ *such that* $g(x^*) \leqq 0$ *and* $wg(x^*) = 0$. *Moreover, a necessary and sufficient condition for* $w \geqq 0$ *and* $s(u, z) = v(0; \alpha)$ *is that* $-w$ *is a parametric subgradient of* $v(y; \alpha)$ *at the origin; i.e.,* $v(y; \alpha) \geqq v(0; \alpha) - wy$ *for all* y. *If* $u \geqq 0$, $z \geqq 0$, *and* $w \geqq z$, *then* $s(u, z) = v(0; \alpha) = v(0)$ *if and only if there is an* x^* *that satisfies the foregoing conditions together with* $zg(x^*) = 0$ [*whereupon the pair* (u, z) *is optimal for* D, x^* *is optimal for both* P *and* $P(0; \alpha)$, *and all solutions that are optimal for both problems have the indicated form*].

Proof. The first part of the theorem is the relation II. The second follows from II coupled with the nonnegativity lemma and the overestimation lemma, by letting $v^*(\alpha) = v(0; \alpha)$ in the latter. The last part of the theorem is a direct consequence of I, II, and III.

The restricted duality theorem gives Theorem 3 of Section 3 for $z=0$, u a vector, and $w=u$. When $u=0$ and $w=z$, in which case $v(y;\alpha)=v(y)$ for all y, the theorem reduces to the fundamental theorem of Lagrangean duality theory. In the 'multiple' surrogate constraint case, if we let $w=1u$, the theorem implies complementary slackness for every surrogate multiplier vector composing the matrix u [via the conditions $wg(x)=0$ and $g(x)\leqq 0$]. Nevertheless, as demonstrated in Section 3 for the case in which u is a vector, this type of complementary slackness is not as restrictive as that which arises for $u=0$ and $w=z$.

Thus we see that all the results of Section 3, and the corresponding results for Lagrangean duality, are special consequences of the foregoing results. The appealing feature of this development is not just its generality but its simplicity. We now identify the equivalent notions for the convex case.

5. SURROGATE DUALITY UNDER CONVEXITY

When X is a convex set, and f and the components of $g(x)$ are convex functions, then it is easy to state necessary and sufficient conditions for the existence of relative subgradients and hence provide 'convex equivalents' to the theorems of the preceding sections.

First we identify the set $Y(\alpha)$ of vectors y for which $P(y;\alpha)$ has a feasible solution: $Y(\alpha)=\{y: g(x)\leqq y \text{ for some } x\epsilon X(u)\}$. Then it is immediate that $Y(\alpha)$ is a convex set, and in addition, for a given α, $v(y;\alpha)$ is a convex function on $Y(\alpha)$, provided either $z\geqq w$ or $g(x)$ is linear.

By analogy to the concept of stability, which is equivalent under convexity to the optimality conditions for Lagrangean duality, we introduce the concept of relative stability in the definition:

P is *relatively stable* if $v(0)$ is finite and there is a scalar M such that

$$[y(0)-v(y;\alpha)]/\|y\|\leqq M \text{ for all } y\neq 0.$$

Similarly, we may define $P(0;\alpha)$ to be relatively stable by replacing $v(0)$ with $v(0;\alpha)$ in the foregoing definition. [The definition of ordinary stability is obtained by replacing $v(y;\alpha)$ with $v(y)$.] Given the form of the results of Section 5 and the relation of the relative and parametric subgradients to the ordinary subgradient, these definitions have precisely the form that one would expect to provide 'convex equivalents' to the surrogate optimality conditions, that relative stability of P and $P(0;\alpha)$ fills the intended role—specifically, of being equivalent to the existence of relative and parametric subgradients for the theorems of Section 5—is expressed in the following result.

CONVEXITY LEMMA. *Let $F(y;\alpha)$ be a convex function on a convex set $Y(\alpha)$ taking values in the extended reals. Let $\|\cdot\|$ be any norm and let $G(y^*;\alpha)$be* *finite. Then* $F(y;\alpha)$ *has a relative subgradient at* $y^*\epsilon Y(\alpha)$ *with respect to* $G(y^*;\alpha)$ *if and only if* $[G(y^*;\alpha)-F(y;\alpha)]/\|y-y^*\|\leqq M$ *for all* $y\epsilon Y(\alpha)$ *such that* $y\neq y^*$.

By making the 'natural' substitutions of terms, the proof of this lemma follows exactly the proof of Gale[8] that connects ordinary subgradients and stability, and thus we will not repeat the proof here. (See also Geoffrion[10].)

From the convexity lemma and the convexity assumptions [and $w\geqq z$ or $g(x)$ linear], it follows that relative stability of P and $P(0;\alpha)$, respectively, provide equivalents to the duality theorem and restricted duality theorem of Section 5. The relevance of this result depends ultimately on the extent to which surrogate constraints find application in convex as well as nonconvex settings.

6. CONCLUSIONS

This paper develops a surrogate duality theory that identifies necessary and sufficient conditions for optimality in a simple form that invites direct comparison with corresponding conditions for Lagrangean duality theory. These conditions are favorable to the surrogate approach whenever the size of the duality gap is an overriding consideration or the form of the original objective function lends itself usefully to optimization over a single constraint (as opposed to 'unconstrained' optimization over a modified objective). We provide characterizations of optimality under the imposition of complementary slackness that likewise yield smaller duality gaps than the Lagrangean. By means of the concepts of parametric and relative subgradients, general theorems are given for a dual containing an arbitrary number of surrogate constraints and a weighted objective. These theorems imply the results for single-constraint surrogate duality and the corresponding results for Lagrangean duality, as well as their composite. Though we focus principally on the nonconvex case, where surrogate constraints have so far had the greatest application, we also give equivalent optimality conditions under assumptions of convexity. Our results are established by means of a simplified framework that facilitates their justification and gives rise to a variety of new inferences for surrogate constraints in mathematical programming.

NOTES

1. As noted in reference 11, a feasible solution to any positive linear combination of two inequalities is always feasible for at least one member of the pair, and a 'best' linear combination in the context of surrogate constraints can be found by checking a sequence of breakpoints over which $\min f(x)$ continually increases to a maximum and thereafter continually decreases. Both of these observations hold

not only for 0–1 programming but generally. The first immediately implies quasi-concavity (as observed by Greenberg and Pierskalla[15]), while the second gives a constructive means for exploiting it in the case of $s(u)$.

2. Note that in the correspondence $G(y;\alpha)=v(y)$, which is followed whenever the relative subgradient is considered separately from the parametric subgradient, $G(y;\alpha)$ is a function of y only. The inclusion of α in $G(y;\alpha)$ in the definition permits the parametric subgradient and the relative subgradient to be encompassed within a single framework—i.e., the parametric subgradient is a relative subgradient of a function with respect to itself.

REFERENCES

1. EGON BALAS, "Discrete Programming by the Filter Method," *Opns. Res.* **19,** 915–957 (1967).
2. 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. Log. Quart.* **16,** 41–51 (1969).
3. R. W. COTTLE, "Symmetric Dual Quadratic Programs," *Quart. Appl. Math.* **21,** 237–243 (1963).
4. G. B. DANTZIG, *Linear Programming and Extensions,* Princeton University Press, Princeton, N.J., 1963.
5. J. P. EVANS AND F. J. GOULD, "Stability in Nonlinear Programming," *Opns. Res.* **18,** 107–118 (1970).
6. H. EVERETT, "Generalized Lagrange Multiplier Method for Solving Problems of Optimal Allocation of Resources," *Opns. Res.* **11,** 399–417 (1963).
7. W. FENCHEL, "Convex Cones, Sets, and Functions," Lecture Notes, Princeton University, Princeton (1953).
8. D. GALE, "A Geometric Duality Theorem with Economic Applications," *Rev. Economic Studies* **34,** 19–24 (1967).
9. A. M. GEOFFRION, "An Improved Implicit Enumeration Approach for Integer Programming," *Opns. Res.* **17,** 437–454 (1969).
10. ———, "Duality in Nonlinear Programming: A Simplified Applications-Oriented Development," in *Perspectives on Optimization,* A. M. GEOFFRION (Ed.), Addison-Wesley, Reading, Mass., 1972.
11. F. GLOVER, "A Multiphase-Dual Algorithm for the Zero-One Integer Programming Problem," *Opns. Res.* **13,** 879–919 (1965).
12. ———, "Surrogate Constraints," *Opns. Res.* **16,** 741–749 (1968).
13. ———, "Heuristics for Integer Programming Using Surrogate Constraints," M SRS 74-6, College of Business, University of Colorado, July 1974.
14. H. J. GREENBERG, "The Generalized Penalty Function Surrogate Model," *Opns. Res.* **21,** 162–178 (1973).
15. ——— AND W. P. PIERSKALLA, "Surrogate Mathematical Programs," *Opns. Res.* **18,** 924–939 (1970).
16. ——— AND WILLIAM P. PIERSKALLA, "Quasi-Conjugate Functions and Surrogate Duality," Cahier Centre d'Etude de Recherche Operationelle, **15,** 437–448 (1973).

17. D. G. LUENBERGER, "Quasi-Convex Programming," *SIAM J. Appl. Math.* **16,** 1090–1095 (1968).
18. O. L. MANGASARIAN AND J. PONSTEIN, "Minimax and Duality in Nonlinear Programming," *J. Math. Anal. Appl.* **11,** 504–518 (1965).
19. R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, N.J., 1970.
20. J. STOER, "Duality in Nonlinear Programming and the Minimax Theorem," *Numer. Math.* **5,** 371–379 (1963).
21. P. WOLFE, "A Duality Theorem for Non-Linear Programming," *Quart. Appl. Math.* **19,** 239–244 (1961).