JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 10, No. 4, 1972

Generalized Benders Decomposition[1]

A. M. GEOFFRION[2]

Communicated by A. V. Balakrishnan

Abstract. J. F. Benders devised a clever approach for exploiting the structure of mathematical programming problems with *complicating variables* (variables which, when temporarily fixed, render the remaining optimization problem considerably more tractable). For the class of problems specifically considered by Benders, fixing the values of the complicating variables reduces the given problem to an ordinary linear program, parameterized, of course, by the value of the complicating variables vector. The algorithm he proposed for finding the optimal value of this vector employs a cutting-plane approach for building up adequate representations of (i) the extremal value of the linear program as a function of the parameterizing vector and (ii) the set of values of the parameterizing vector for which the linear program is feasible. Linear programming duality theory was employed to derive the natural families of *cuts* characterizing these representations, and the parameterized linear program itself is used to generate what are usually *deepest* cuts for building up the representations.

In this paper, Benders' approach is generalized to a broader class of programs in which the parametrized subproblem need no longer be a linear program. Nonlinear convex duality theory is employed to derive the natural families of cuts corresponding to those in Benders' case. The conditions under which such a generalization is possible and appropriate are examined in detail. An illustrative specialization is made to the variable factor programming problem introduced by R. Wilson, where it offers an especially attractive approach. Preliminary computational experience is given.

[1] Paper received April 10, 1970; in revised form, January 28, 1971. An earlier version was presented at the Nonlinear Programming Symposium at the University of Wisconsin sponsored by the Mathematics Research Center, US Army, May 4–6, 1970. This research was supported by the National Science Foundation under Grant No. GP-8740.

[2] Professor, University of California at Los Angeles, Los Angeles, California.

1. Introduction

This paper is devoted to problems of the form

$$\underset{x,y}{\text{maximize}}\, f(x, y) \text{ subject to } G(x, y) \geqslant 0, \qquad x \in X, \qquad y \in Y, \tag{1}$$

where y is a vector of *complicating* variables in the sense that (1) is a much easier optimization problem in x when y is temporarily held fixed. G is an m-vector of constraint functions defined on $X \times Y \subseteq R^{n_1} \times R^{n_2}$. We have in mind particularly situations such as the following:

(a) for fixed y, (1) separates into a number of independent optimization problems, each involving a different subvector of x;

(b) for fixed y, (1) assumes a well-known special structure, such as the classical transportation form, for which efficient solution procedures are available; and

(c) Problem (1) is not a concave program in x and y jointly, but fixing y renders it so in x.

Such situations abound in practical applications of mathematical programming and in the literature of large-scale optimization, where the central objective is to exploit special structure such as this in order to design effective solution procedures.

It is evident that there are substantial opportunities for achieving computational economies by somehow looking at (1) in y-space rather than in xy-space. We expect that, in situation (a), the computations can be largely decentralized and done in parallel for each of the smaller independent subproblems; in (b), use can be made of available efficient special-purpose algorithms; and in (c), the nonconvexities can be treated separately from the convex portion of the problem.

The key idea that enables (1) to be viewed as a problem in y-space is the concept of *projection* (Ref. 1), sometimes also known as *partitioning*. The projection of (1) onto y is

$$\underset{y}{\text{maximize}}\, v(y) \text{ subject to } y \in Y \cap V, \tag{2}$$

where

$$v(y) \equiv \underset{x}{\text{supremum}}\, f(x, y) \text{ subject to } G(x, y) \geqslant 0, \qquad x \in X, \tag{3}$$

and

$$V \equiv \{y : G(x, y) \geqslant 0 \text{ for some } x \in X\}. \tag{4}$$

Note that $v(y)$ is the optimal value of (1) for fixed y and that, by our designation of y as *complicating* variables, evaluating $v(y)$ is much easier

than solving (1) itself. Because it must be referred to so often in the sequel, the label (1–y) is introduced to refer to the optimization problem in (3)

$$\underset{x \in X}{\text{maximize}}\, f(x, y) \text{ subject to } G(x, y) \geqslant 0. \tag{1–y}$$

The set V consists of those values of y for which (1–y) is feasible; $Y \cap V$ can be thought of as the projection of the feasible region of (1) onto y-space.

It is intuitively clear that the projected problem (2) is equivalent for our purposes to (1). This will be made precise in Theorem 2.1 below. For now, it is enough to keep in mind that an optimal solution y^* of (2) readily yields an optimal solution (x^*, y^*) of (1), where x^* is any optimizing x in (1–y^*).

Benders (Ref. 2) was one of the first to appreciate the importance of (2) as a route to solving (1). The difficulty with (2), however, is that the function v and the set V are only known implicitly via their definitions (3) and (4). Benders coped with this difficulty by devising a cutting-plane method that builds up an approximation to v and V. His development treats the special case

$$X \equiv \{x : x \geqslant 0\}, \tag{5–1}$$

$$f(x, y) \equiv c^t x + \phi(y), \tag{5–2}$$

$$G(x, y) \equiv Ax + g(y) - b, \tag{5–3}$$

where ϕ is a scalar-valued function and g a vector-valued function. Under these assumptions, both v and V turn out to have exact representations using only a finite number of approximators. Linear programming duality theory yields a constructive proof of this result based on the fact that (1) is a linear program in x for each fixed y. The rationale of the computational procedure is then evident. See the original paper or Ref. 1, Section 4.1, for details.

The main result of this paper is an extension of Benders' approach to a more general class of problems with the help of nonlinear duality theory. Many new practical problems are thereby brought into range, including some difficult ones having to do with the optimal design of gaseous diffusion processes and the computation of chemical equilibria. These results are presented in the following section. In Section 3, specialization is made to the variable factor programming problem, previously treated by Wilson (Ref. 3) by another approach. In addition to its intrinsic interest, this application illustrates an instance in which (1) is not a concave program and yet can be solved optimally by concave programming techniques via the present approach. Computational

experience is given. Section 4 presents some additional discussion of the assumptions required at various stages of the development.

For the reader's convenience, pertinent but not yet universally known results from nonlinear duality theory are summarized in an appendix. The notation employed is standard. All vectors are column vectors, unless transposed.

2. Generalized Benders Decomposition

The generalization of Benders' approach to a class of problems of the form (1) is divided into five subsections. The first establishes the *master* problem and its equivalence to (1). The central idea here is to invoke the natural dual representations of v and V after passing to (2). The second subsection demonstrates how the master problem can be solved via a series of subproblems which generate dominating approximators of v and V as needed. Roughly speaking, this is accomplished by obtaining the optimal multiplier vectors for $(1\text{–}y)$ corresponding to various trial values of y. The generalized Benders decomposition procedure for (1) is then stated and discussed in Section 2.3. Theoretical convergence is taken up in Section 2.4, and Section 2.5 discusses some computational considerations.

2.1. Derivation of the Master Problem. The desired master problem is obtained from (1) by a sequence of three manipulations: (i) project (1) onto y, resulting in (2); (ii) invoke the natural dual representation of V in terms of the intersection of a collection of regions that contain it; and (iii) invoke the natural dual representation of v in terms of the pointwise infimum of a collection of functions that dominate it.

Manipulation (i) was already discussed in Section 1. The following easy theorem (Ref. 1, Section 2.1) shows that (1) and (2) are equivalent for our purposes [note that no assumptions on (1) whatever are needed].

Theorem 2.1. (*Projection*). Problem (1) is infeasible or has unbounded optimal value iff the same is true of (2). If (x^*, y^*) is optimal in (1), then y^* must be optimal in (2). If y^* is optimal in (2) and x^* achieves the supremum in (3) with $y = y^*$, then (x^*, y^*) is optimal in (1). If $\bar{y}$ is ϵ_1-optimal in (2) and $\bar{x}$ is ϵ_2-optimal in $(1\text{–}\bar{y})$, then $(\bar{x}, \bar{y})$ is $(\epsilon_1 + \epsilon_2)$-optimal in (1).

Manipulation (ii) is based on the following theorem.

Theorem 2.2. (*V-Representation*). Assume that X is a nonempty convex set and that G is concave on X for each fixed $y \in Y$. Assume

further that the set $Z_y \equiv \{z \in R^m : G(x, y) \geqslant z \text{ for some } x \in X\}$ is closed for each fixed $y \in Y$. Then, a point $\bar{y} \in Y$ is also in the set V iff $\bar{y}$ satisfies the (infinite) system

$$[\underset{x \in X}{\text{supremum}}\ \lambda^t G(x, y)] \geqslant 0, \qquad \text{all} \qquad \lambda \in \Lambda, \tag{6}$$

where

$$\Lambda \equiv \left\{\lambda \in R^m : \lambda \geqslant 0 \text{ and } \sum_{i=1}^{m} \lambda_i = 1\right\}.$$

Proof. Let $\bar{y}$ be an arbitrary point in Y. It is trivial to verify directly that $\bar{y}$ satisfies (6) if it is in V. The converse can be demonstrated with the help of nonlinear duality theory as follows. Suppose that $\bar{y}$ satisfies (6). Then,

$$\underset{\lambda \in \Lambda}{\text{infimum}}[\underset{x \in X}{\text{supremum}}\ \lambda^t G(x, \bar{y})] \geqslant 0.$$

It follows that

$$\underset{\lambda \geqslant 0}{\text{infimum}}[\underset{x \in X}{\text{supremum}}\ \lambda^t G(x, \bar{y})] = 0, \tag{7}$$

since the scaling of λ does not influence the sign of the bracketed expression and $\lambda = 0$ is allowed in (7). Now, (7) simply asserts that the dual with respect to the G-constraints of the concave program

$$\underset{x \in X}{\text{maximize}}\ 0^t x \text{ subject to } G(x, \bar{y}) \geqslant 0 \tag{8}$$

has optimal value 0. Recalling that $Z_{\bar{y}}$ is closed, we therefore have from Theorem 5.1 (see the appendix) that (8) must be feasible and, hence, that $\bar{y} \in V$.

The assumption that Z_y is closed for each $y \in Y$ is not a stringent one. Mild sufficient conditions under which it must hold are given in Section 4.1.

Manipulation (iii) is based on the following theorem.

Theorem 2.3. (*v-Representation*). Assume that X is a nonempty convex set and that f and G are concave on X for each fixed $y \in Y$. Assume further that, for each fixed $\bar{y} \in Y \cap V$, at least one of the following three conditions holds: (a) $v(\bar{y})$ is finite and $(1\text{–}\bar{y})$ possesses an optimal multiplier vector[3]; (b) $v(\bar{y})$ is finite, $G(x, \bar{y})$ and $f(x, \bar{y})$ are

[3] See the appendix for the definition of an optimal multiplier vector. Actually, it is enough to consider *generalized* optimal multiplier vectors (also defined in the appendix) in order to avoid the implicit assumption that $(1\text{–}\bar{y})$ must have an optimal solution.

continuous on X, X is closed, and the ϵ-optimal solution set of $(1\text{–}\bar{y})$ is nonempty and bounded for some $\epsilon \geqslant 0$; and (c) $v(\bar{y}) = +\infty$. Then, the optimal value of $(1\text{–}y)$ equals that of its dual on $Y \cap V$, that is,

$$v(y) = \underset{u \geqslant 0}{\text{infimum}}[\underset{x \in X}{\text{supremum}} f(x, y) + u^t G(x, y)], \qquad \text{all} \qquad y \in Y \cap V. \quad (9)$$

The proof is a direct application of Theorems 5.2 and 5.3 and weak duality. Alternative assumption (a) is one that will very often hold. Many different sufficient conditions, usually called *constraint qualifications*, are known which imply it—most of them fairly weak assumptions to preclude pathological cases. Nor is alternative (b) particularly stringent. Of course, boundedness of X or of the feasible region is enough to guarantee that the ϵ-optimal solution set is bounded for any $\epsilon > 0$, and the existence of a unique optimal solution does it for $\epsilon = 0$.

Under the assumptions of Theorems 2.2 and 2.3, then, manipulations (i)–(iii) applied to (1) yield the equivalent master problem

$$\underset{y \in Y}{\text{maximize}}[\underset{u \geqslant 0}{\text{infimum}}[\underset{x \in X}{\text{supremum}} f(x, y) + u^t G(x, y)]] \text{ subject to (6)}$$

or, using the definition of infimum as the greatest lower bound,

$$\underset{\substack{y \in Y \\ y_0}}{\text{maximize}}\, y_0 \qquad (10\text{–}1)$$

subject to

$$y_0 \leqslant \underset{x \in X}{\text{supremum}}\{f(x, y) + u^t G(x, y)\}, \qquad \text{all} \qquad u \geqslant 0, \qquad (10\text{–}2)$$

$$\underset{x \in X}{\text{supremum}}\{\lambda^t G(x, y)\} \geqslant 0, \qquad \text{all} \qquad \lambda \in \Lambda. \qquad (10\text{–}3)$$

2.2. Solving the Master Problem. The most natural strategy for solving the master problem (10), since it has a very large number of constraints, is *relaxation* (Ref. 1). Begin by solving a relaxed version of (10) that ignores all but a few of the constraints (10-2) and (10-3); if the resulting solution does not satisfy all of the ignored constraints, then generate and add to the relaxed problem one or more violated constraints and solve it again; continue in this fashion until a relaxed problem solution satisfies all of the ignored constraints [at which point an optimal solution of (10) has been found], or until a termination criterion signals that a solution of acceptable accuracy has been obtained. Details regarding the termination criterion will be supplied later. The deeper concern is the crucial issue of how a solution to a relaxed version of (10) can be tested for feasibility with respect to the ignored constraints and, in case of infeasibility, how a violated constraint can be generated.

Suppose that $(\hat{y}, \hat{y}_0)$ is optimal in a relaxed version of (10). How can this point be tested for feasibility in (10-2) and (10-3)? From Theorem 2.2 and the definition of V, we see that $\hat{y}$ satisfies (10-3) iff $(1\text{–}\hat{y})$ has a feasible solution. And if $(1\text{–}\hat{y})$ turns out to be feasible, Theorem 2.3 implies that $(\hat{y}, \hat{y}_0)$ satisfies (10-2) iff $\hat{y}_0 \leqslant v(\hat{y})$. Thus, $(1\text{–}\hat{y})$ is the natural subproblem for testing $(\hat{y}, \hat{y}_0)$ for feasibility in the master problem. This is in perfect accord with our interest in applications where y is *complicating* in the sense that $(1\text{–}y)$ is much easier than (1) itself.

Not only is $(1\text{–}\hat{y})$ the appropriate subproblem for testing the feasibility of $(\hat{y}, \hat{y}_0)$ in (10), but almost any reasonable algorithm for $(1\text{–}\hat{y})$ will yield an index of a violated constraint in the event that $(\hat{y}, \hat{y}_0)$ is infeasible. By an *index* of a violated constraint, we mean a vector $\hat{u} \geqslant 0$ such that

$$\hat{y}_0 > \operatorname*{supremum}_{x \in X} \{f(x, \hat{y}) + \hat{u}^t G(x, \hat{y})\}, \tag{11-1}$$

if (10-2) is violated, or a vector $\hat{\lambda} \in \Lambda$ such that

$$\operatorname*{supremum}_{x \in X} \{\hat{\lambda}^t G(x, \hat{y})\} < 0, \tag{11-2}$$

if (10-3) is violated. If $(1\text{–}\hat{y})$ is infeasible, it can be shown that most dual-type algorithms addressed to it yield such a $\hat{\lambda}$, as do most primal algorithms fitted with a phase-one procedure for finding an initial feasible solution if one exists [$\hat{\lambda}$ can be viewed as specifying a convex combination of constraints that has no solution in X]. If $(1\text{–}\hat{y})$ is feasible and has a finite optimal value, it follows from Theorem 5.2 that an optimal multiplier vector satisfies (11-1) if one exists and $\hat{y}_0 > v(\hat{y})$. Virtually all modern algorithms applicable to $(1\text{–}\hat{y})$ produce an optimal multiplier vector as a byproduct, if one exists, as is usually the case. Nonexistence must be associated either with an unbounded optimal value, in which case one may terminate since the same must then be true of (1), or with a finite optimal value but a pathological condition in which, by (9), $\hat{u}$ nevertheless satisfies (11-1) if it comes close enough to being optimal in the dual of $(1\text{–}\hat{y})$. Such a $\hat{u}$ will be referred to subsequently as a *near-optimal* multiplier vector.

In light of this discussion, it is reasonable to presume henceforth that $(1\text{–}\hat{y})$ will be addressed with an algorithm that is *dual-adequate* in the sense that it yields: a vector $\hat{\lambda} \in \Lambda$ satisfying (11-2) if $(1\text{–}\hat{y})$ is infeasible; an optimal multiplier vector $\hat{u}$ if one exists; or a near-optimal multiplier vector $\hat{u}$ satisfying (11-1) if no optimal multiplier vector exists, but $\hat{y}_0$ exceeds the optimal value of $(1\text{–}\hat{y})$.

Thus, we have shown how $(1\text{–}\hat{y})$ can be used to test any point $(\hat{y}, \hat{y}_0)$ for feasibility in the master problem (10) and to generate an

index ($\hat{\lambda}$ or $\hat{u}$) of a violated constraint in the event of infeasibility. This completes the discussion of how (10) can be solved by the relaxation strategy.

2.3. Statement of the Procedure. The generalized Benders decomposition procedure can now be stated formally. The hypotheses of Theorems 2.2 and 2.3 are presumed to hold, and it will be assumed for simplicity that (1) has finite optimal value.

It will be convenient in stating the procedure to define the following (see Section 2.5 for discussion):

$$L^*(y; u) \equiv \operatorname*{supremum}_{x \in X} \{f(x, y) + u^t G(x, y)\}, \qquad y \in Y, \qquad u \geqslant 0, \quad (12\text{-}1)$$

$$L_*(y; \lambda) \equiv \operatorname*{supremum}_{x \in X} \{\lambda^t G(x, y)\}, \qquad y \in Y, \qquad \lambda \geqslant 0. \quad (12\text{-}2)$$

With this understanding, the generalized Benders decomposition procedure is represented by the following steps.

Step (1). Let a point $\bar{y}$ in $Y \cap V$ be known. Solve the subproblem $(1\text{–}\bar{y})$ and obtain an optimal (or near-optimal) multiplier vector $\bar{u}$ and the function $L^*(y; \bar{u})$. Put $p = 1$, $q = 0$, $u^1 = \bar{u}$, $LBD = v(\bar{y})$. Select the convergence tolerance parameter $\epsilon > 0$.

Step (2). Solve the current relaxed master problem

$$\begin{aligned} \underset{\substack{y \in Y \\ y_0}}{\text{maximize}}\ y_0 \text{ subject to } y_0 &\leqslant L^*(y; u^j), \qquad j = 1, \ldots, p, \\ L_*(y; \lambda^j) &\geqslant 0, \qquad j = 1, \ldots, q, \end{aligned} \quad (13)$$

by any applicable algorithm. Let $(\hat{y}, \hat{y}_0)$ be an optimal solution; $\hat{y}_0$ is an *upper bound* on the optimal value of (1). If $LBD \geqslant \hat{y}_0 - \epsilon$, terminate.

Step (3). Solve the revised subproblem $(1\text{–}\hat{y})$. One of the following cases must occur.

Step (3A). *The quantity $v(\hat{y})$ is finite.* If $v(\hat{y}) \geqslant \hat{y}_0 - \epsilon$, terminate. Otherwise, determine an optimal multiplier vector $\hat{u}$ [if none exists, a near-optimal multiplier vector satisfying (11-1) will do] and the function $L^*(y; \hat{u})$. Increase p by 1 and put $u^p = \hat{u}$. If $v(\hat{y}) > LBD$, put $LBD = v(\hat{y})$. LBD is a *lower bound* on the optimal value of (1). Return to Step (2).

Step (3B). *Problem $(1\text{–}\hat{y})$ is infeasible.* Determine $\hat{\lambda}$ in Λ satisfying (11-2), and the function $L_*(y; \hat{\lambda})$. Increase q by 1 and put $\lambda^q = \hat{\lambda}$. Return to Step (2).

Remark 2.1. It is implicitly assumed in the statement of Step (2) that an optimal solution of (13) exists. This problem is necessarily feasible in view of the assumed existence of the point $\bar{y}$ in Step (1), but to preclude an unbounded optimum during the early executions of Step (2) it may be necessary to constrain (y, y_0) additionally to a region known to contain all physically significant solutions of (10).

Remark 2.2. Step (1) is set up to begin at a known $\bar{y}$ in $Y \cap V$, because this is believed to be the situation most likely to be encountered in applications. There is an advantage to utilizing experience and mathematical or physical insight in order to provide a good initial feasible solution. If such a point is unknown, however, Step (1) could be altered in the obvious way to accomodate an infeasible initial point [see Step (3B)]. An alternative would be to make the procedure as it stands function first in a phase-one mode in order to find a point $\bar{y}$ in $Y \cap V$.

Remark 2.3. The termination conditions can be understood as follows, in view of Theorem 2.1. Termination at Step (3A) implies that $\hat{y}$ in ϵ-optimal in (2), for $\hat{y}_0$ is obviously an upper bound on the optimal value of (2) [it is the optimal value of the equivalent master problem (10) with some of its constraints deleted]. Thus, any optimal solution $\hat{x}$ of $(1-\hat{y})$ yields an ϵ-optimal solution $(\hat{x}, \hat{y})$ of (1) [if only a δ-optimal solution of $(1-\hat{y})$ can be computed, the result is $(\epsilon + \delta)$-optimal in (1)]. Termination at Step (2) is similar, except that *LBD* plays the role of $v(\hat{y})$; *LBD* is always set at the greatest optimal value of $(1-y)$ found at any previous execution of Step (1) or Step (3A), and so is the best known lower bound on the optimal value of (2). When $LBD \geqslant \hat{y}_0 - \epsilon$, the subproblem corresponding to *LBD* yields an ϵ-optimal solution of (1). Note that, while the sequence of values for $\hat{y}_0$ found at successive executions of Step (2) must be monotone nonincreasing, the sequence of values for $v(\hat{y})$ found at Step (3A) need *not* be monotone nondecreasing. This is the reason for introducing *LBD*.

Remark 2.4. It is of some interest and comfort to know that the violated constraints generated at Step (3) are usually the *most* violated (or nearly so) among all violated constraints. When $\hat{u}$ is an optimal multiplier vector at Step (3A), it follows from Theorem 5.2 that it indexes a constraint among (10-2) that is most violated at $(\hat{y}, \hat{y}_0)$. When no optimal multiplier vector for $(1-\hat{y})$ exists, how near $\hat{u}$ comes to indexing a most violated constraint depends solely on how nearly it solves the dual problem of $(1-\hat{y})$. Similarly, how close $\hat{\lambda}$ comes to indexing

a most violated constraint among (10-3) at Step (3B) depends solely on how close it comes to solving the dual of (8) with $\bar{y}$ equal to $\hat{y}$ (with the dual vector normalized to Λ).

2.4. Theoretical Convergence. There are many different sets of conditions under which the generalized Benders decomposition procedure or variants of it can be proven convergent. Two convergence theorems are presented here to illustrate the possibilities. The first applies when Y is a finite discrete set and is therefore appropriate for applications to mixed integer nonlinear programming.[4] The second convergence theorem is intended for the case where Y is of infinite cardinality (e.g., convex).

Theorem 2.4. (*Finite Convergence*). Assume that Y is a finite discrete set, that the hypotheses of Theorem 2.2 hold, and that the hypotheses of Theorem 2.3 hold with condition (b) omitted. Then, the generalized Benders decomposition procedure terminates in a finite number of steps for any given $\epsilon > 0$ and even for $\epsilon = 0$.

Proof. Fix $\epsilon \geqslant 0$ arbitrarily. Finite termination is a direct consequence of the finiteness of Y and the fact that no $\hat{y}$ can ever repeat itself in a solution to (13) at Step (2). For, if $\hat{y} \notin V$, then Step (3B) generates an L_*-constraint that precludes $\hat{y}$ from ever again being feasible in (13); and, if $\hat{y} \in V$ and $v(\hat{y})$ is finite, then a constraint $y_0 \leqslant L^*(y; \hat{u})$ is generated that would imply the optimality of $\hat{y}$ under the termination criterion of Step (3A) if $\hat{y}$ were ever to occur again in a solution of (13) [if $(y_0^1, y^1 = \hat{y})$ were to solve (13) subsequently, then $y_0^1 \leqslant L^*(\hat{y}; \hat{u}) = v(\hat{y}) = v(y^1)$ would have to hold, and the termination condition would therefore be satisfied].

The reason for the slight strengthening of the hypotheses of Theorem 2.3 [i.e., requiring that (1–$\hat{y}$) has an optimal multiplier vector whenever $v(\hat{y})$ is finite] is to preclude the possibility of an infinite number of executions of Step (3A) with the same $\hat{y}$, each resulting in a slightly better near-optimal multiplier vector but always bounded strictly away from $v(\hat{y})$ in value. A slight strengthening of what is meant by a *near-optimal* multiplier vector at Step (3A) would preclude this possibility

[4] Balas (Ref. 4) has previously developed a somewhat different generalization of Benders' approach for the case where Y is a discrete set. Aside from his use of earlier (pre-Rockafellar) nonlinear duality results designed for differentiable convex functions, the main distinguishing feature of his work is the use of implicit enumeration rather than relaxation, to solve the master problem analogous to (10), his *linearized dual*.

if the existence of a true optimal multiplier vector for $(1\text{–}\bar{y})$ were in question. A similar comment could be made with regard to the second convergence theorem.

Theorem 2.5. (*Finite ϵ-convergence*). Assume that Y is a nonempty compact subset of V, that X is a nonempty compact convex set, that f and G are concave on X for each fixed $y \in Y$ and are continuous on $X \times Y$, and that the set $U(y)$ of optimal multiplier vectors for $(1\text{–}y)$ is nonempty for all y in Y and uniformly bounded in some neighborhood of each such point. Then, for any given $\epsilon > 0$, the generalized Benders decomposition procedure terminates in a finite number of steps.

Proof.[5] Fix $\epsilon > 0$ arbitrarily. Suppose that the procedure does not terminate in a finite number of steps. Let $\langle y_0^{\nu}, y^{\nu} \rangle$ be the sequence of optimal solutions to (13) at successive executions of Step (2). By taking a subsequence, if necessary, we may assume that $\langle y_0^{\nu}, y^{\nu} \rangle$ converges to a point $(\bar{y}_0, \bar{y})$ such that $\bar{y} \in Y$ [remember that $\langle y_0^{\nu} \rangle$ is a nonincreasing sequence bounded below and that $\langle y^{\nu} \rangle$ is in the compact set Y]. Similarly, we may assume by the uniform boundedness assumption that the corresponding sequence of optimal multiplier vectors $\langle u^{\nu} \rangle$ produced by Step (3A) converges to a point $\bar{u}$. Now, $y_0^{\nu+1} \leqslant L^*(y^{\nu+1}; u^{\nu})$ obviously holds by the accumulation of constraints in (13), and so by the continuity of L^* (apply Theorem 1.4 in Meyer, Ref. 5) we obtain $\bar{y}_0 \leqslant L^*(\bar{y}; \bar{u})$. To complete the proof, it remains to show that $\bar{u} \in U(\bar{y})$ and that v is lower semicontinuous at $\bar{y}$, for then $L^*(\bar{y}; \bar{u})$ must equal $v(\bar{y})$ by Theorem 5.2 and consequently $\bar{y}_0 \leqslant v(\bar{y})$; this would imply by the lower semicontinuity of v at $\bar{y}$ that $y_0^{\nu} \leqslant v(y^{\nu}) + \epsilon$ for all ν sufficiently large, which would contradict the supposition that the termination criterion at Step (3A) is never met.

To show that $\bar{u} \in U(\bar{y})$, it is enough to show that $U(y)$ is an upper-semicontinuous mapping at $\bar{y}$. To do this, we employ the characterization of $U(y)$ as the set of optimal solutions to the dual of $(1\text{–}y)$; in other words (Theorem 5.4),

$$U(y) = \{u \geqslant 0 : L^*(y; u) = \min_{u^1 \geqslant 0} L^*(y; u^1)\}.$$

Then, by the continuity of L^*, we may apply Theorem 1.5 of Ref. 5 to obtain the desired upper-semicontinuity of U at $\bar{y}$.

[5] A number of basic results involving point-to-set mappings in mathematical programming will be employed. The reader is advised to review the first five results in Meyer (Ref. 5) before attempting to follow the proof. The author is indebted to W. Hogan for his collaboration in developing this proof.

To show the lower semicontinuity of v at $\bar{y}$, it is convenient to use the characterization (9), which becomes

$$v(y) = \min_{u \geqslant 0} L^*(y; u), \qquad \text{all} \qquad y \in Y,$$

in view of (12-1) and the assumed nonemptiness of $U(y)$. Since Y is compact, the local uniform boundedness assumption implies the uniform boundedness of $U(y)$ on all of Y. Hence, there exists a compact set U^* such that $U(y) \subseteq U^*$ for all y in Y. The constraint $u \in U^*$ can therefore be introduced above without disturbing the equality, thus permitting direct application of Lemma 1.2 of Ref. 5 in order to obtain the desired lower semicontinuity of v.

The uniform boundedness assumption of Theorem 2.5 may be somewhat awkward. For this reason, we now demonstrate a condition implying it at a given point that may be more readily verifiable in practice. For example, this condition is easily verified at any point $\bar{y} \in Y$ in the case of variable factor programming (Section 3).

Lemma 2.1. Assume that X is a nonempty compact convex set and that f and G are concave on X for each fixed $y \in Y$ and continuous on $X \times Y$. If there exists a point $\bar{x} \in X$ such that $G(\bar{x}, \bar{y}) > 0$, then the set $U(y)$ of optimal multiplier vectors for $(1\text{-}y)$ is uniformly bounded in some open neighborhood of $\bar{y}$.

Proof. The existence of $\bar{x}$ implies that the set $\Omega(x, z) \equiv \{x \in X : G(x, y) \geqslant z\}$ is nonempty for all (y, z) in some open neighborhood N of $(\bar{y}, 0)$. Define the auxiliary function

$$\psi(y, z) = \sup_{x \in X} f(x, y) \text{ subject to } G(x, y) \geqslant z.$$

It can be shown that the map Ω is continuous on a neighborhood of $(\bar{y}, 0)$, say N. It follows from Theorem 1.4 of Ref. 5 that ψ must be continuous on N. Define N_1 to be an open neighborhood of $\bar{y}$ and $N_2 \equiv \{z \in R^m : 0 \leqslant z_i \leqslant \delta, i = 1, \ldots, m\}$, $\delta > 0$, such that $\bar{N}_1 \times N_2 \subset N$, where $\bar{N}_1$ is the closure of N_1. Define

$$\underline{\beta} \equiv \underset{y,z}{\text{minimum}}\, \psi(y, z) \text{ subject to } y \in \bar{N}_1,\ z \in N_2,$$

$$\bar{\beta} \equiv \underset{y,z}{\text{maximum}}\, \psi(y, z) \text{ subject to } y \in \bar{N}_1,\ z \in N_2.$$

Evidently, $-\infty < \underline{\beta} \leqslant \bar{\beta} < +\infty$. To show that the conclusion of Lemma 2.1 holds on N_1, take any point y^1 in N_1. If $U(y^1)$ is empty, there is nothing to show; otherwise, let u^1 be any point in $U(y^1)$. Then,

$$\psi(y^1, z) \leqslant \psi(y^1, 0) - u^1 z, \text{ all } z,$$

by one of the characterizations (Theorem 5.4) of $U(y^1)$. By considering each of the m choices δe_i for z, where e_i is the ith unit vector in R^m, we obtain

$$0 \leqslant u_i^1 \leqslant [\psi(y^1, 0) - \psi(y^1, \delta e_i)]/\delta \leqslant (\bar{\beta} - \underline{\beta})/\delta.$$

Thus, $U(y^1)$ is contained within the same hypercube for any y^1 in N_1.

The assumption $Y \subseteq V$ in Theorem 2.5, which precludes any possibility of Step (3B) occurring, also calls for clarification. There are many applications (such as the one discussed in Section 3) where $Y \subseteq V$ automatically holds. In other applications, prior experience or physical or mathematical insight may make it possible to enforce this assumption by redefining Y to be some subset of itself known to contain the optimal solution of (2) and on which (1–y) is known to be feasible. Without the assumption $Y \subseteq V$, however, there is a danger that Step (3B) may occur infinitely many times in succession and thereby preclude finite ϵ-convergence. One situation of importance where it is known that this danger cannot materialize is when X is given by linear constraints and G is of the form (5-3), for then it is easily shown that V has a representation in terms of a finite collection of constraints; thus, Step (3B) need occur at most a finite number of times. A finite number of occurrences of Step (3B) is not enough to spoil the finite ϵ-convergence argument of Theorem 2.5.

For applications where $Y \not\subseteq V$ and Step (3B) could possibly occur infinitely often in succession, there are two fairly obvious possibilities for preserving finite ϵ-convergence with Y of infinite cardinality. The first would be to modify the procedure to finitely truncate any excessively long sequence of successive executions of Step (3B) and go to Step (3A) with $\hat{y}$ equal to the extrapolated limit point (which, we presume, will be in $Y \cap V$ under reasonable regularity assumptions). The other possibility would be to essentially absorb Step (3B) into Step (2) by imposing the side constraint $y \in V$ on (13). One might then be able to employ a primal algorithm for (13) with a parametric algorithm for (1–y) used simultaneously to discover when the side constraint is about to be violated. Or one might solve (13) with the side constraint $y \in V$ handled by successive inside-out approximation in the spirit of the Dantzig–Wolfe decomposition (see Ref. 6, Chapter 24 for the original method and Ref. 1, Section 4.3, for the extension required here). This is the approach adopted in Ref. 7, Section 3.1, for a special structure. None of these possibilities has been worked out in complete detail, and must therefore be viewed as but tentative suggestions. It is hoped that computational studies now under way will shed light on the situation.

In closing the discussion of convergence, we make reference to the very recent results of Hogan (Ref. 8) who applies the results of Eaves and

Zangwill (Ref. 9) to demonstrate finite ϵ-convergence of the procedure even when amply satisfied constraints are dropped at Step (2). This enables the relaxed master problem (13) to have fewer constaints, thereby reducing the time required per iteration of the procedure. Whether this advantage will be offset by a greater number of required iterations is another matter for computational study.

2.5. Computational Considerations. With the theoretical development of generalized Benders decomposition complete, it is appropriate to consider the prospects for efficient computational implementation.

Since (1–y) is a concave program, the availability of a dual-adequate algorithm for Steps (1) and (3) is not a serious issue. The availability of an algorithm for solving the relaxed master problem (13) at Step (2), however, is in question unless something is known about the mathematical nature of Y, L^*, and L_*. Except for certain special cases, it is preferred that (13) be a concave program, that is, the set Y convex and the functions L^* and L_* concave on Y. Several sufficient conditions for the latter requirement in terms of the behavior of f and G are given in Section 4.2 (e.g., joint concavity of f and G on $X \times Y$).

If (13) is not a concave program, but Y includes the requirement that the components of y must be integer-valued, then implicit enumeration or some other discrete optimization technique may enable (13) to be optimized. This is certainly feasible if f and G are separable and linear in y and if Y consists of linear constraints and an integrality requirement on y, for then (13) becomes an integer *linear* program for which many solution methods are available. In fact, this is probably the single most important case arising in applications when y is a vector of 0-1 variables. The reason is that the introduction of such variables into an otherwise continous optimization problem is usually done to account for logical alternatives, which in turn typically leads to the structure in y mentioned above.

In selecting algorithms to be used for (1–y) and (13), an obvious point to bear in mind is that they should if possible have the capability of being restarted after the first iteration so as to take advantage of the most recent previous solution to their respective problems.

Attention is now turned to the as yet undiscussed question of how the functions L^* and L_* in (13) are to be obtained in explicit form. This issue must be resolved except in the unlikely circumstance that the algorithm for (13) is prepared to deal with these functions in their supremal value representations (12-1) and (12-2). It appears necessary to assume that (1) has the following property.

Property (*P*). For every $u \geqslant 0$, the supremum of $f(x, y) + u^t G(x,y)$ over X can be taken essentially independently of y, so that the function $L^*(\cdot\ ; u)$ on Y can be obtained explicitly with little or no more effort than is required to evaluate it at a single value of y.

Similarly, for every $\lambda \in \Lambda$, the supremum of $\lambda^t G(x, y)$ over X can be taken essentially independently of y, so that the function $L_*(\cdot\ ; \lambda)$ on Y can be obtained explicitly with little or no more effort than is required to evaluate it at a single value of y.

One important case in which Property (P) holds is when f and G are linearly separable in x and y, that is,

$$f(x, y) \equiv f_1(x) + f_2(y), \qquad G(x, y) \equiv G_1(x) + G_2(y). \tag{14}$$

Then, for any $u \geqslant 0$ and $\lambda \in \Lambda$,

$$L^*(y; u) = \operatorname*{supremum}_{x \in X} \{f_1(x) + u^t G_1(x)\} + f_2(y) + u^t G_2(y), \qquad y \in Y, \tag{15-1}$$

$$L_*(y; \lambda) = \operatorname*{supremum}_{x \in X} \{\lambda^t G_1(x)\} + \lambda^t G_2(y), \qquad y \in Y. \tag{15-2}$$

The variable factor programming problem discussed in Section 3 shows that Property (P) can hold even though F and G are *not* linearly separable in x and y. The same is true of certain problems having to do with the optimal design of diffusion processes[6] and the computation of chemical equilibria.[7] See Section 4.2 for further discussion.

A dual-adequate algorithm for $(1\text{–}\hat{y})$ that produces $L^*(\cdot\ ; \hat{u})$ or $L_*(\cdot\ ; \hat{\lambda})$ essentially as a byproduct as well as $\hat{u}$ or $\hat{\lambda}$ will be called L-*dual-adequate*. In the presence of Property (P), *L-dual-adequacy* usually requires very little if anything more of an algorithm for $(1\text{–}\hat{y})$ than dual-adequacy does. When (14) holds, for example, by (9) and (15-1) we have

$$L^*(y; \hat{u}) = v(\hat{y}) - f_2(\hat{y}) - \hat{u}^t G_2(\hat{y}) + f_2(y) + \hat{u}^t G_2(y), \qquad y \in Y,$$

when $\hat{u}$ is an optimal multiplier vector [for then $\hat{u}$ achieves the infimum in (9)]. This equation is only approximate when an optimal multiplier vector does not exist, but the approximation can be made arbitrarily accurate by making $\hat{u}$ nearly enough optimal in the dual of $(1\text{–}\hat{y})$ (see Theorem 2.3). Thus, $L^*(y; \hat{u})$ is obtained without even having to explicitly perform the supremum in (15-1). A similarly satisfactory situation usually exists regarding $L_*(y; \hat{\lambda})$, although a detailed justifica-

[6] The author is grateful to Prof. R. Wilson of Stanford University for pointing out this application (Ref. 10).

[7] This was pointed out to the author by R. J. Clasen and will be reported in a forthcoming paper by him.

tion of this statement would depend upon whether the algorithm produces $\hat{\lambda}$ by a phase-one or dual-type approach when $(1\text{–}\hat{y})$ is infeasible. In the context of variable factor programming, it will be seen in Section 3 that any algorithm for $(1\text{–}\hat{y})$ is L-dual-adequate so long as it produces an optimal solution $\hat{x}$ and multiplier vector $\hat{u}$ (both of which necessarily exist for any $\hat{y}$).

3. Application to Variable Factor Programming

Wilson (Ref. 3) has defined the variable factor programming problem as follows:

$$\underset{x,y}{\text{maximize}} \sum_{i=1}^{n_2} y_i f_i(x^i) \tag{16-1}$$

subject to

$$\sum_{i=1}^{n_2} x^i y_i \leqslant c, \tag{16-2}$$

$$x^i \geqslant 0, \qquad i = 1,\ldots, n_2, \tag{16-3}$$

$$Ay \leqslant b, \qquad y \geqslant 0, \tag{16-4}$$

where x^i is an m-vector [x is the n_2m-vector $(x^1,\ldots, x^{n_2})$]. One may interpret y_i as the level of activity i and f_i as the profit coefficient or yield rate for activity i when x^i is the allocation of variable factors (or stimulants) per unit of activity i. Constraint (16-2) requires the total allocation of variable factors to be within an availability vector c.

Obviously (16) is a proper generalization of the general linear programming problem. As pointed out by Wilson, it arises in petroleum refining and blending (tetraethyl lead is a stimulant), certain preprocessing industries, and agricultural planning. In the later case, for instance, y_i might represent the number of acres to be planted with crop i, and x^i might represent the amounts of various kinds of fertilizer, irrigation water, tilling labor, etc., to be used on each acre of crop i.

3.1. Theoretical Aspects. It will be assumed that each f_i is a continuous concave function on the nonnegative orthant, i.e., that the marginal productivity of the variable factors is nonincreasing. It is also assumed, without loss of generality, that each component of c is positive (if some component of c were zero, the corresponding variable factor could be dropped from the problem).

The assumptions underlying the generalized Benders decomposition procedure are easily verified as follows. Theorem 2.2 is unnecessary,

since the set V turns out to be all of R^{n_2}. This is so because $c \geqslant 0$ and $x^1 = \cdots = x^{n_2} = 0$ satisfies (16-2) for any choice of y. In Theorem 2.3, the initial assumptions are immediate. Alternative assumption (a) holds for all $\bar{y} \geqslant 0$, because the subproblem (1–$\bar{y}$), which specializes to

$$\underset{x^1 \geqslant 0, \ldots, x^{n_2} \geqslant 0}{\text{maximize}} \sum_{i=1}^{n_2} \bar{y}_i f_i(x^i) \tag{17–$\bar{y}$}$$

subject to

$$\sum_{i=1}^{n_2} x^i \bar{y}_i \leqslant c,$$

is a concave program with a finite optimal solution[8] which also satisfies Slater's interiority constraint qualification ($x^1 = \cdots = x^{n_2} = 0$ is interior since $c > 0$).

The preceding observations also show (with the help of Lemma 2.1) that the assumptions required for finite ϵ-convergence by Theorem 2.5 also hold, except possibly for x and y being unbounded. From the physical interpretation of the problem, however, it is clear that bounding constraints on x and y can be imposed without loss of generality.

Fortunately, the relevant part of Property (P), namely the first part, also holds, since for all $y \geqslant 0$ we have

$$\begin{aligned}
&\underset{x \in X}{\text{supremum}} \{f(x, y) + u^t G(x, y)\} \\
&\quad = \underset{x^1 \geqslant 0, \ldots, x^{n_2} \geqslant 0}{\text{supremum}} \left\{ \sum_{i=1}^{n_2} y_i f_i(x^i) + u^t \left(-\sum_{i=1}^{n_2} x^i y_i + c \right) \right\} \\
&\quad = u^t c + \sum_{i=1}^{n_2} y_i \underset{x^i \geqslant 0}{\text{supremum}} \{f_i(x^i) - u^t x^i\} = L^*(y; u).
\end{aligned} \tag{18}$$

Note that $L^*(y; u)$ is linear in y. Additionally, we note that any algorithm for solving (17–$\bar{y}$) at Step (1) or (3) is dual-adequate so long as it produces an optimal solution $\bar{x}$ and an optimal multiplier vector $\bar{u}$ (the existence of both for any $\bar{y}$ has been demonstrated above). Such an algorithm is also L-dual-adequate, because

$$\underset{x^i \geqslant 0}{\text{supremum}} \{f_i(x^i) - \bar{u}^t x^i\} = f_i(\bar{x}^i) - \bar{u}^t \bar{x}^i \tag{19}$$

holds for all i such that $\bar{y}_i > 0$; for i such that $\bar{y}_i = 0$, the supremum will have to be computed. Thus, $L^*(y; \bar{u})$ is at hand directly from (18)–(19) with little if any extra work once the subproblem (17–$\bar{y}$) is solved.

[8] The objective function is continuous, x^i is arbitrary for i such that $\bar{y}_i = 0$, and x^i is bounded by $0 \leqslant x^i \leqslant c/\bar{y}_i$ for i such that $\bar{y}_i > 0$.

It would seem that generalized Benders decomposition is particularly well suited to variable factor programs. Step (3B) cannot occur; the relaxed master program to be solved at Step (2)

$$\underset{\substack{y\geqslant 0\\ y_0}}{\text{maximize}}\ y_0 \text{ subject to } Ay \leqslant b, \tag{20-1}$$

$$y_0 \leqslant L^*(y; u^j), \qquad j = 1,..., p, \tag{20-2}$$

is an ordinary *linear program* (L^* is linear in y for fixed u); and the subproblem (17–y) to be solved at Steps (1) and (3) is a concave program with linear constraints and an objective function that is linearly separable in the x^i's. This subproblem has a natural and meaningful interpretation in its own right as the problem of determining the optimal use of variable factors given a tentative choice of activity levels. The violated constraints generated are always most violated constraints (see Remark 2.4).

3.2. Computational Experience. Computational experiments have been carried out for a class of variable factor programs with linear yield rate functions f_i. The generalized Benders decomposition procedure worked quite well for the test problems. The number of iterations required to find the optimal solution was typically four or five, with a range of 1–13, and seemed remarkably insensitive to problem size parameters over the range studied.

The test problems were randomly generated as follows. The A-matrix is $r \times n_2$ and 30% dense. Each nonzero element is independently chosen from a uniform distribution on the interval [1, 15]. The f_i functions are all of the form

$$f_i(x^i) = d_i + \gamma^i \cdot x^i,$$

where the scalars d_i are independently uniform on [1, 15] and

$$\gamma^i = (d_i/2)\,\bar{\gamma}^i \Big/ \left(\sum_{j=1}^{m} \bar{\gamma}_j{}^i\right),$$

where $\bar{\gamma}_j{}^i$ is zero with probability one-half or, with equal probability, an integer between 1 and 10. For each series run, the coefficients were generated according to the largest values of the size parameters, and the individual problems of the series were then obtained by deleting portions of the data. For every case, the b-vector was taken to be the row sums of the actual A matrix used, and the c-vector was always a vector with each element equal to n_2.

The structure just given can be motivated in terms of the following interpretation. A manufacturing facility has n_2 production processes,

Table 1. Number of iterations ($r = 8$).

	$n_2 = 6$	$n_2 = 9$	$n_2 = 12$	$n_2 = 15$	$n_2 = 18$
$m = 1$	$1\frac{3}{4}$ 2, 2, 2, 1	$2\frac{1}{4}$ 2, 2, 2, 3	$2\frac{1}{4}$ 2, 3, 2, 2	$2\frac{1}{4}$ 2, 3, 2, 2	$2\frac{1}{4}$ 2, 3, 2, 2
$m = 2$	$1\frac{3}{4}$ 2, 2, 2, 1	$2\frac{3}{4}$ 2, 2, 2, 5	$2\frac{3}{4}$ 2, 3, 3, 3	3 2, 3, 4, 3	3 2, 3, 2, 5
$m = 4$	3 2, 6, 2, 2	5 6, 5, 2, 7	$4\frac{3}{4}$ 5, 4, 5, 5	$4\frac{3}{4}$ 6, 4, 5, 4	$5\frac{1}{4}$ 5, 6, 5, 5
$m = 6$	3 2, 5, 2, 3	$5\frac{1}{2}$ 5, 4, 3, 10	$5\frac{3}{4}$ 5, 4, 7, 7	$5\frac{1}{2}$ 6, 5, 4, 7	$7\frac{1}{4}$ 5, 7, 11, 6
$m = 8$	$3\frac{1}{4}$ 3, 6, 3, 1	6 6, 5, 4, 9	6 5, 5, 7, 7	$7\frac{1}{4}$ 7, 6, 9, 7	9 7, 11, 13, 5

each of which consumes r different resources (materials, time on various machines, types of labor, etc.). The constraints $Ay \leqslant b$ take into account the limited availability of these resources; there are just enough resources to operate each process at unit level. The basic per unit profit is d_i for process i, but this can be increased by applying stimulants (variable factors). There is just enough of each stimulant to apply it at unit rate to every process if each process were operated at unit level, and so doing would increase each d_i by a factor of one-half.

The computational results are given in Tables 1 and 2. The four lower numbers in each cell give the number of iterations to optimality for four independent trials, and the upper number is the average. In each case, the termination criterion was that the lower and upper bounds coincide to six significant figures.

4. Discussion

The first subsection discusses an assumption of Theorem 2.2, and the second gives further insight into Property (P) and the nature of the functions L^* and L_*.

Table 2. Number of iterations ($m = 4$, $n = 12$).

$r = 4$	$r = 8$	$r = 12$	$r = 16$
$6\frac{1}{4}$ 4, 2, 11, 8	$4\frac{3}{4}$ 5, 4, 5, 5	$4\frac{3}{4}$ 5, 6, 4, 4	$4\frac{1}{4}$ 4, 4, 5, 4

4.1. Assumption of Theorem 2.2. Theorem 2.2 assumes that the set

$$Z_y \equiv \{z \in R^m : G(x, y) \geqslant z \text{ for some } x \in X\}$$

must be closed for each $y \in Y$. At first glance, it might seem that this would be so if G were continuous on X for each fixed y in Y and X were closed, but the following example shows that this is not sufficient: let $m = 1$, $X = [1, \infty)$, and $G(x, y) = -1/x$; then clearly $Z_y = (-\infty, 0)$, which is not closed. If X is bounded as well as closed and G is continuous on X for each $y \in Y$, however, it is easy to verify that Z_y must be closed for all $y \in Y$. It can also be shown (with the help of Lemma 6 of Ref. 11) that these conditions remain sufficient if the boundedness of X is replaced by the condition: X is convex and, for each fixed $y \in Y$, G is concave on X and there exists a point z_y such that the set $\{x \in X : G(x, y) \geqslant z_y\}$ is bounded and nonempty. Since X is required to be convex and G is required to be concave on X for fixed $y \in Y$, by Theorem 2.3 anyway, this condition is a useful weakening of the requirement that X be bounded. For example, the condition holds if, for each fixed y in Y, at least one of the component functions of G has a bounded set of maximizers over X (in this case, select the corresponding component of z_y to be the maximal value and let the other components be arbitrarily small).

4.2. Functions L^* and L_*. The significance of the requirement of Property (P) that the extrema in (12-1) and (12-2) *can be taken essentially independently of* y deserves further elaboration. We have already seen two classes of problems for which this requirement holds, namely linear separability in x and y [see (15)] and variable factor programming [see (18)], and have alluded to two other such problem classes. It is easy to invent still other such examples, e.g., those for which $f(x, y) + u^t G(x, y)$ can be written as $Q(h(x, u), y, u)$ for any $x \in X$, $y \in Y$, and $u \geqslant 0$, where Q is increasing in its first component and h is a scalar function of x and u. Then, clearly,

$$L^*(y; u) = Q(\operatorname*{supremum}_{x \in X}\{h(x, u)\}, y, u), \qquad y \in Y. \tag{21}$$

A similar representation would also have to be available for $\lambda^t G(x, y)$ if the second part of (P) is to hold.

When the achievement of the extrema in (12-1) and (12-2) is not in question, the following version of Property (P) is appropriate: for every $u \geqslant 0$ there exists a point $x_u \in X$ such that

$$\operatorname*{supremum}_{x \in X}\{f(x, y) + u^t G(x, y)\} = f(x_u, y) + u^t G(x_u, y), \qquad y \in Y, \tag{22-1}$$

and, for every $\lambda \in \Lambda$, there exists a point $x_\lambda \in X$ such that

$$\operatorname{supremum}_{x \in X}\{\lambda^t G(x, y)\} = \lambda^t G(x_\lambda, y), \qquad y \in Y. \tag{22-2}$$

Then, $L^*(y; u)$ is available directly from x_u, and $L_*(y; \lambda)$ directly from x_λ. Let this version be called Property (P'). It is helpful to keep in mind that the usual situation, but by no means the only possible one, is where $\hat{u}$ is an optimal multiplier vector for $(1\text{–}\hat{y})$ and $x_{\hat{u}}$ is an optimal solution.

One conjecture concerning Property (P) that has probably already occurred to the reader is that the first part implies the second part. This seems plausible since for any $\lambda \in \Lambda$ in the second part of (P) one may take $u = \theta\lambda$ in the first part, with θ so large that the influence of f becomes inconsequential by comparison with that of $\theta\lambda^t G$, with the result that $L_*(y; \lambda)$ approximately equals $(1/\theta)L^*(y; \theta\lambda)$ for large θ. This conjecture can indeed be verified when X is closed and bounded and f and G are continuous on X for each fixed $y \in Y$. The proof is most easily carried out in terms of (P'). The boundedness of X can be weakened somewhat along the lines suggested by the discussion of Section 4.1.

Finally, it is appropriate to recite some sufficient conditions for $L^*(L_*)$ to be concave as a function of y for fixed $u(\lambda)$. As noted in Section 2.5, this is usually necessary if the relaxed master problem (13) is to be solvable by an available algorithm. Suppose that Y is a convex set. We consider several possible cases. When f and G are linearly separable in x and y, it is evident from (15) that the desired conclusion holds if f_2 and G_2 are concave on Y. In variable factor programming, (18) shows that L^* is even *linear* on Y. When the Q-representation introduced above holds, L^* is concave if Q is concave in its y-argument, and L_* is concave if the analog of Q is concave in its y-argument. And if the slightly more stringent Property (P') holds, (22) reveals that the desired conclusion holds if f and G are concave on Y for each fixed $x \in X$. With regard to this last condition, it is perhaps unnecessary to say that marginal concavity of f and G in x for fixed y and in y for fixed x does *not* imply joint concavity on $X \times Y$, although of course the converse does hold. It is useful to know, however, that joint concavity does imply without any further qualification that L^* and L_* are concave on Y. The proof follows the lines of Theorem 2 in Ref. 1.

It is interesting to note that, under any of these sufficient conditions for the concavity of L^* and L_*, not only is (13) a concave program, but so is the original projected problem (2): by Theorem 2.2, $V \cap Y$ is the convex set $\{y \in Y : L_*(y; \lambda) \geqslant 0, \text{ all } \lambda \in \Lambda\}$; and by Theorem 2.3,

$v(y) = \text{infimum}_{u \geqslant 0} L^*(y; u)$ is concave on $Y \cap V$ since it is the infimum of a collection of concave functions. *Thus,* (2) *can be a concave program even though* (1) *is not a concave program.* This is certainly true in variable factor programming, as we have already pointed out. It can also be true, for example, when the sufficient condition associated with property (P') holds.

5. Appendix: Summary of Relevant Results from Nonlinear Duality Theory

All results will be stated in terms of the standard primal problem

$$\underset{x \in X}{\text{maximize}}\, f(x) \text{ subject to } g_i(x) \geqslant 0, \qquad i = 1,..., m, \tag{23}$$

where it is assumed that f and each function g_i is concave on the nonempty convex set $X \subseteq R^n$. See Ref. 11 for further details.

The dual of (23) with respect to the g_i constraints is

$$\underset{u \geqslant 0}{\text{minimize}} \left[\underset{x \in X}{\text{supremum}}\, f(x) + \sum_{i=1}^{m} u_i g_i(x) \right], \tag{24}$$

where $u = (u_1 ,..., u_m)$ is a vector of dual variables. By the weak duality theorem (Ref. 11, Theorem 2), any feasible solution of the primal must have a value no greater than the value of any feasible solution of the dual.

If $\bar{x}$ is an optimal solution of the primal problem, an optimal multiplier vector $\bar{u}$ is defined to be any nonnegative vector such that

$$\sum_{i=1}^{m} \bar{u}_i g_i(\bar{x}) = 0$$

and $\bar{x}$ maximizes

$$f(x) + \sum_{i=1}^{m} \bar{u}_i g_i(x)$$

over X. To preclude assuming that the optimal value of the primal is actually achieved by some $\bar{x}$, for many purposes it is enough to work with the concept of a generalized optimal multiplier vector: a nonnegative vector $\bar{u}$ such that, for every scalar $\epsilon > 0$, there exists a point x_ϵ feasible in the primal problem satisfying the two conditions

(i) x_ϵ is an ϵ-optimal maximizer of $f(x) + \sum_{i=1}^{m} \bar{u}_i g_i(x)$ over X,

(ii) $\sum_{i=1}^{m} \bar{u}_i g_i(x_\epsilon) \leqslant \epsilon$.

Every optimal multiplier vector is also a generalized optimal multiplier vector.

Theorem 5.1. If $\{z \in R^m : g_i(x) \geqslant z_i,\ i = 1,..., m,$ for some $x \in X\}$ is closed and the optimal value of the dual is finite, then the primal problem must be feasible.

Proof. This is an immediate corollary of Theorem 5 of Ref. 11.

Theorem 5.2. If $\bar{u}$ is a (generalized) optimal multiplier vector for the primal problem, then $\bar{u}$ is also an optimal solution of the dual and the optimal values of primal and dual are equal.

Proof. This follows immediately from Lemmas 3 and 4 of Ref. 11 and the discussion just before the latter.

It is also true that if the primal problem has no generalized optimal multiplier vector and yet the optimal values of primal and dual are equal, then the dual can have no optimal solution (by inspection of Diagram 1, Ref. 11).

Theorem 5.3. Assume X to be closed, f and each g_i to be continuous on X, the optimal value of (23) to be finite and the set

$$\{x \in X : g_i(x) \geqslant 0,\ i = 1,..., m, \text{ and } f(x) \geqslant \alpha\}$$

to be bounded and nonempty for some scalar α no greater than the optimal value of (23). Then, the optimal values of the primal and dual problems are equal.

Proof. This follows immediately from Theorems 7 and 8 of Ref. 11.

Theorem 5.4. If (23) has an optimal solution, then u is an optimal multiplier vector iff

$$\psi(z) \leqslant \psi(0) - \sum_{i=1}^{m} u_i z_i \qquad \text{for all} \qquad z,$$

where $\psi(z)$ is the optimal value of (23) with 0 replaced by z_i in the ith constraint. If, in addition, at least one optimal multiplier vector exists, then u is an optimal multiplier vector iff u is an optimal solution of the dual problem (24).

Proof. See Theorems 1 and 3 of Ref. 11.

References

1. GEOFFRION, A. M., *Elements of Large-Scale Mathematical Programming*, Management Science, Vol. 16, No. 11, 1970.
2. BENDERS, J. F., *Partitioning Procedures for Solving Mixed-Variables Programming Problems*, Numerische Mathematik, Vol. 4, 1962.
3. WILSON, R. *Programming Variable Factors*, Management Science, Vol. 13, No. 1, 1966.
4. BALAS, E., *Duality in Discrete Programming: IV. Applications*, Carnegie-Mellon University, Graduate School of Industrial Administration, Report No. 145, 1968.
5. MEYER, R., *The Validity of a Family of Optimization Methods*, SIAM Journal on Control, Vol. 8, No. 1, 1970.
6. DANTZIG, G. B., *Linear Programming and Extensions*, Princeton University Press, Princeton, New Jersey, 1963.
7. GEOFFRION, A. M., *Primal Resource-Directive Approaches for Optimizing Nonlinear Decomposable Systems*, Operations Research, Vol. 18, No. 3, 1970.
8. HOGAN, W., *Application of a General Convergence Theory for Outer Approximation Algorithms*, University of California at Los Angeles, Western Management Science Institute, Working Paper No. 174, 1971.
9. EAVES, B. C., and ZANGWILL, W. I., *Generalized Cutting Plane Algorithms*, SIAM Journal on Control, Vol. 9, No. 4, 1971.
10. GEOFFRION, A. M., *A New Global Optimization Technique for Gaseous Diffusion Plant Operation and Capital Investment*, University of California at Los Angeles, Graduate School of Business Administration, Discussion Paper, 1970.
11. GEOFFRION, A. M., *Duality in Nonlinear Programming: A Simplified Application-Oriented Development*, SIAM Review, Vol. 13, No. 1, 1971.