OPERATIONS RESEARCH CENTER

working paper

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

GENERALIZED LINEAR PROGRAMMING SOLVES THE DUAL

by

T.L. Magnanti, J.F. Shapiro, and M.H. Wagner

Sloan School of Management
M.I.T.

OR 019-73 September 1973

Supported in part by the U.S. Army Research Office
(Durham) under contract DAHC04-73-C-0032.

ABSTRACT

The generalized linear programming algorithm allows an arbitrary mathematical programming minimization problem to be analyzed as a sequence of linear programming approximations. Under fairly general assumptions, it is demonstrated that any limit point of the sequence of optimal linear programming dual prices produced by the algorithm is optimal in a concave maximization problem that is dual to the arbitrary primal problem. This result holds even if the generalized linear programming problem does not solve the primal problem. The result is a consequence of the equivalence that exists between the operations of convexification and dualization of a primal problem. The exact mathematical nature of this equivalence is given.

1. Introduction

Around 1960, Dantzig and Wolfe [3, 4] showed how linear programming can be used to solve mathematical programming problems that are not linear and to decompose large scale linear programming problems with special structure. Their approach has been variously called generalized linear programming, column generation, or Dantzig-Wolfe decomposition. Subsequently, the generalized linear programming algorithm has been widely interpreted as a method for decentralizing decision making by the calculation of internal prices on shared resources so that local decision makers can account for external economies or diseconomies (Baumol and Fabian [1]; Dantzig [2; Chapter 23]).

The object of this paper is to state and demonstrate a fundamental property of generalized linear programming. Suppose the algorithm is applied to an arbitrary minimization problem called the primal problem. Then, for nearly all problems of practical importance, any limit point of the sequence of dual prices produced by the algorithm is optimal in a concave maximization problem that is dual to the primal problem. This result holds even if the generalized linear programming algorithm does not solve the primal problem.

This property of generalized linear programming is a consequence of a mathematical equivalence that exists between convexification and dualization of a mathematical programming problem; namely, that the optimal objective function values of the convexified problem and the dual problem are equal. Moreover, given an arbitrary primal problem, direct methods for finding a globally optimal solution exist only if the primal problem has convex structure. A common algorithmic approach regardless of the structure of the primal problem is to replace it by its convexified relaxation. The equivalence between convexification and dualization indicates that when this is done, it is the dual problem that is actually being solved and therefore the specific algorithm used should take into account the structure of the dual problem and its relation to the primal problem.

For example, if the primal problem is a convex programming problem, then the Kuhn-Tucker optimality conditions are necessary and sufficient. In this case, the primal and dual problems are perfectly symmetric in that their optimal objective function values are equal and any pair of primal and dual solutions satisfying the Kuhn-Tucker conditions are optimal in their respective problems (Rockafellar [29; Chapter 28]). If an exact optimal solution is desired to a convex programming problem, then an algorithm may be chosen according to its rate of convergence to a primal-dual pair satisfying the Kuhn-Tucker conditions. Wolfe [37] analyzes the rates of convergence of generalized linear programming and compares it to other algorithms for convex programming problems.

On the other hand, a large number of management science applications of generalized linear programming are non-convex;

e.g., multi-item production control (Dzielinski and Gomory [6]), resource constrained network scheduling (Fisher [8]), cutting stock (Gilmore and Gomory [17]), network synthesis (Gomory and Hu [18]), traveling salesman (Held and Karp [20]), integer programming (Shapiro [30]), and multi-commodity flow (Tomlin [32]). For these problems, the same symmetry between the primal and dual problems cannot be guaranteed or even expected. In particular, there can be a so-called duality gap between the optimal values of the primal and dual problems, or equivalently, there can be a relaxation gap between the primal problem and its convexified relaxation. Thus, there may be no practical dual pricing mechanism for finding an optimal primal solution.

Nevertheless, construction and solution of dual problems is useful in analyzing and solving non-convex problems. For example, Shapiro [31] gives a dual method for aggregating, ordering and eliminating integer programming activities. Held and Karp [20, 21] use duality theory in conjunction with branch-and-bound in the construction of an efficient algorithm for solving the traveling salesman problem. The traveling salesman dual problem they use gives tight lower bounds for branch-and-bound and indicates effective branching strategies. The use of dual problems in conjunction with branch-and-bound to solve a variety of discrete optimization problems is given in [10]. Although dual solutions for non-convex problems admit some economic interpretation (e.g., Shapiro [30; p. 70]), their use is primarily algorithmic.

For non-convex problems, dual ascent methods are preferred to generalized linear programming because they provide monotonically increasing lower bounds for use in branch-and-bound. Specifically, there is the dual ascent approach of Lasdon [22], adaptations of the primal-dual simplex algorithm (Fisher and Shapiro [9, 10], Grinold [19]) approximation algorithms based on simplicial approximation [10], and subgradient relaxation (Held and Karp [21]). Approximation in the dual is attractive because the dual is itself an approximation to the primal problem and therefore it is more important to obtain quickly good dual solutions than to converge in a limiting sense to an optimal dual solution.

Generalized linear programming has also been extensively proposed as a method for decomposing large scale linear programming problems with special structure; e.g., block diagonal or staircase structures (Dantzig [2; Chapter 23] Lasdon [23]). Unfortunately, computational experience with the algorithm on these problems has been somewhat disappointing (Orchard-Hays [28; p. 240]). Thus, it appears that the primary importance of the algorithm may be in identifying the dual problem and its potential usefulness. As was the case for non-convex problems, it may be preferable to use approximation methods on the dual to obtain quickly good solutions to the primal rather than an exact method which requires an exorbitant computional investment to obtain optimal or even good solutions.

The plan of this paper is the following. Section 2 contains a statement of general primal and dual problems and a demonstration of the mathematical equivalence between convexification and dualization of the primal problem. The following section contains a brief review of the generalized linear programming algorithm and a proof that any limit point of the sequence of linear programming dual prices is an optimal dual solution. Section 4 applies the previous theory to an analysis of a phase one procedure for finding an initial feasible solution for the generalized linear programming algorithm. Mathematical generalizations of some of the results in previous sections are given in section 5. Some concluding remarks are given in section 6 and there is one appendix.

A number of authors have given results related to the ones being presented here; e.g., Falk [7], Geoffrion [14, 16], Lasdon [22], H. Wagner [33], Wolfe [37]. Nevertheless, to the best of our knowledge, the fact that generalized linear programming solves the dual has never been explicitly demonstrated in any generality in a published paper and remains unknown to most of the economics and management science community for whom generalized linear programming is a familiar idea. In his thesis, M. Wagner [34] established the property for a restricted dual problem using different mathematical arguments.

Our proof follows Dantzig's proof in [2; Chapter 24] that generalized linear programming solves the convex programming problem, but we omit the convexity assumptions on the primal problem. In this sense, generalized linear programming anticipated some of the later research into mathematical programming duality (Gale [13], Geoffrion [15], Rockafellar [29], and others). Whittle [35] gives results similar to ours on the equivalence of the operations of convexification and dualization of mathematical programming problems; see also Luenberger [25]. Our subsequent analysis of the generalized linear programming algorithm can be viewed as a mechanization of this fundamental property.

2. Statement of the Primal and Dual Problems and a Lemma

The primal problem we wish to consider is

$$
\begin{aligned}
v = \min\ & f(x) \\
\text{s.t. } & g(x) \le 0 \\
& x \in X \subseteq R^n
\end{aligned}
\qquad (2.1)
$$

where f is a continuous real-valued function, g is a continuous function from R^n to R^m and X is a non-empty compact set. If (2.1) is infeasible, then we take the primal value $v = +\infty$. Problem (2.1) is not a completely arbitrary mathematical programming problem defined on R^n. We have chosen f, g continuous and X compact to simplify the mathematical analysis below thereby permitting an uncluttered demonstration of our main results. A number of mathematical generalizations are discussed in Section 5. A final point here is that the vast majority of real-life applications of mathematical programming can be put into the form (2.1) without difficulty.

Let u be an m-vector of prices, and define

$$L(x,u) = f(x) + ug(x)$$

and

$$L(u) = \min_{x \in X} \{f(x) + ug(x)\} = \min_{x \in X} \{L(x,u)\}. \qquad (2.2)$$

The usual assumption is that (2.2) is much easier to solve than (2.1). It is well known that L(u) is a continuous concave function and that $L(u) \leq v$ for any $u \geq 0$ (see Rockafellar [29]). The dual problem is to find the best lower bound to v; namely, find

$$d = \sup L(u) \quad \text{s.t. } u \geq 0. \tag{2.3}$$

Clearly, $d \leq v$ and without additional assumptions we cannot expect $d = v$. A characterization of when equality holds is as follows.

By its definition, $L(u) \leq f(x) + ug(x)$ for every $x \varepsilon X$. Thus if we plot (see Figure 1) the values $(f(x), g(x))$ in R^{m+1} the hyperplane $L(u) = y_o + uy$ lies below the resulting set, i.e. substituting $y = g(x)$ gives $y_o = L(u) - ug(x) \leq f(x)$. Also, L(u) is the intercept of this hyperplane with $y = 0$. Letting

$$[f, g] = \bigcup_{x \varepsilon X} \{(\eta, \xi) : \eta \geq f(x), \xi \geq g(x)\}$$

and $[f, g]^c$ be the convex hull of $[f, g]$, we easily see that the $L(u) = y_o + uy$ must be a supporting hyperplane for $[f, g]^c$ as well. We formally record this result as the following well known lemma [25].

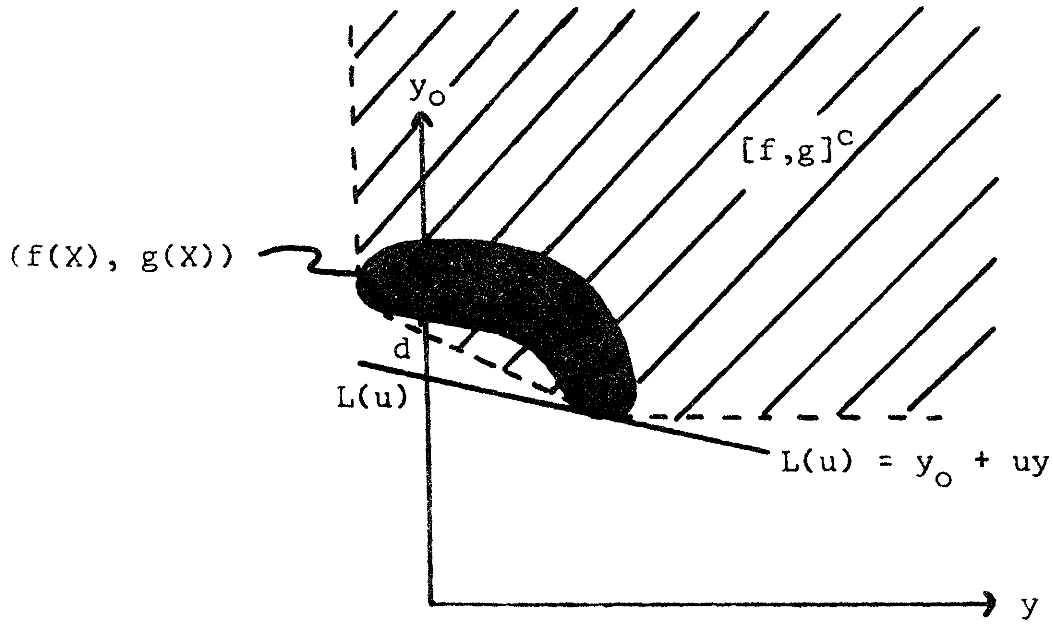

Figure 1

Lemma 2.1: For any $u \geq 0$, the hyperplane $L(u) = y_o + uy$ supports the set $[f, g]^c$.

Proof: If $(\eta, \xi) \in [f, g]^c$, than there exist $(\eta^1, \xi^1), \ldots, (\eta^{m+2}, \xi^{m+2}) \in [f, g]$ and non-negative weights $\lambda_1, \ldots, \lambda_{m+2}$ satisfying $\sum_{k=1}^{m+2} \lambda_k = 1$, $\sum_{k=1}^{m+2} \eta^k \lambda_k = \eta$, $\sum_{k=1}^{m+2} \xi^k \lambda_k = \xi$ (this is Caratheodory's Theorem [29; p.155]). By the definition of $[f, g]$ there must exist $x^k \in X$ satisfying $\eta^k \geq f(x^k)$, $\xi^k \geq g(x^k)$, $k = 1, \ldots, m+2$. These inequalities imply that for any $u \geq 0$, $\sum_{k=1}^{m+2} f(x^k) \lambda_k + u \sum_{k=1}^{m+2} g(x^k) \lambda_k \leq \eta + u\xi$. But then, since $L(u) \leq f(x^k) + ug(x^k)$, $L(u) = \sum_{k=1}^{m+2} L(u) \leq \eta + u\xi$, i.e., $y_o = L(u) - u\xi \leq \eta$. Finally, if $L(u) = f(\bar{x}) + ug(\bar{x})$ then the hyperplane $L(u) = y_o + uy$ supports $[f, g]^c$ at $(f(\bar{x}), g(\bar{x})) \in [f, g]$. ||

Next we define

$$v^c(\xi) = \inf \{\eta : (\eta, \xi) \in [f, g]^c\}$$

which is taken to be $+\infty$ if there is no $(\eta, \xi) \in [f, g]^c$; $v^c = v^c(0)$ is the convexified value of the primal problem. We are now in a position to prove the basic result establishing the equivalence of convexification and dualization of the primal problem (2.1).

Lemma 2.2: The optimal dual objective function value equals the optimal objective function value of the convexified primal; namely

$$v^c = d$$

Proof: ($d \leq v^c$): If $v^c = +\infty$, there is nothing to prove; otherwise, select an arbitrary $(\eta, 0) \in [f, g]^c$. Then from lemma 2.1, for any $u \geq 0$, $L(u) \leq \eta + u \cdot 0 = \eta$. Thus,

$$d = \sup_{\text{s.t. } u \geq 0} L(u) \leq \eta$$

and since $(\eta, 0)$ was chosen arbitrarily from $[f, g]^c$, we can conclude $d \leq v^c$.

($v^c \leq d$): If $v^c = -\infty$, there is nothing to prove; otherwise, let $r < v^c$ be an arbitrary real number. Then $(r, 0) \notin [f, g]^c$. Since $[f, g]^c$ is a closed convex set (see Appendix 1), there is a hyperplane $u_o y_o + uy = \beta$ strictly separating $(r, 0)$ and $[f, g]^c$; namely, the non-zero vector $(u_o, u) \in R^{m+1}$ and real number β satisfy:

$$u_o r + u \cdot 0 < \beta \leq u_o \eta + u \xi \text{ for all } (\eta, \xi) \in [f, g]^c . \quad (2.4)$$

Since η and each component ξ_j of ξ are unbounded from above over $[f, g]^c$ it can easily be shown using the right most inequality in (2.4) that $u_i \geq 0$, $i = 0, 1, \ldots, m$.

To complete the proof that $v^c \leq d$, we distinguish two cases.

(i) There exists a point $(\eta, 0) \in [f, g]^c$ for some $\eta \in R$. Then it follows from (2.4) that $u_o \neq 0$; otherwise $0 < \beta \leq 0$. By scaling u_o, u and β we may assume that $u_o = 1$. But then since $(f(x), g(x)) \in [f, g]$ for every $x \in X$, (2.4) implies that

$$L(u) = \min_{x \in X} \{f(x) + ug(x)\} \geq \beta > r$$

and therefore since $r < v^c$ was arbitrary

$$d = \sup_{\text{s.t. } u \geq 0} L(u) \geq v^c$$

(ii) There does not exist a point $(\eta, 0) \in [f, g]^c$ implying $v^c = +\infty$. Then the sets $\{(r, 0) : r \in R\}$ and $[f, g]^c$ are disjoint closed convex sets implying the existence[1] of $(u_o, u) \in R^{m+1}$ and a scalar β such that

$$u_o\, r + u \cdot 0 < \beta \leq u_o\, \eta + u\, \xi$$

for all $r \in R$ and $(\eta, \xi) \in [f, g]$. As in part (i), we have $(u_o, u) \geq 0$. Letting $r \to +\infty$, the left most inequality implies that $u_o = 0$ and thus that $\beta > 0$. Thus, $u\, \xi \geq \beta$ for any $(\eta, \xi) \in [f, g]$ and for any $K > 0$, we have

$$(Ku)\, \xi \geq K\beta \quad .$$

Letting $\xi = g(x)$, this inequality implies that

$$\min_{x \in X} [(Ku)\, g(x)] \geq K\beta \quad .$$

This, if $\ell = \min_{x \in X} f(x)$, we have

$$L(Ku) = \min_{x \in X} [f(x) + (Ku)\, g(x)] \geq \ell + K\beta \quad .$$

Letting K go to $+\infty$ demonstrates that

$$d = \sup_{\text{s.t. } u \geq 0} L(u) = +\infty = v^c. \;||$$

1. Since the sets are not compact, the existence of a strictly separating hyperplane needs additional proof. Such a proof uses the properties f, g continuous, X compact. Details are omitted.

3. Review of the Algorithm and Convergence Properties

At iteration K, the generalized linear programming algorithm solves the master LP

$$d^K = \min \sum_{k=1}^{K} f(x^k)\lambda_k$$

$$\text{s.t.} \quad \sum_{k=1}^{K} g(x^k)\lambda_k \leq 0 \qquad (3.1)$$

$$\sum_{k=1}^{K} \lambda_k = 1$$

$$\lambda_k \geq 0, \; k=1,\ldots,K,$$

where the points $x^k \in X$, $k=1,\ldots,K$ have been previously generated. We assume (3.1) has a feasible solution; a phase one procedure for finding a feasible solution is given in Section 4. The LP dual to (3.1) is

$$d^K = \max w$$

$$\text{s.t.} \quad w \leq f(x^k) + ug(x^k), \; k=1,\ldots,K \qquad (3.2)$$

$$u \geq 0 \; .$$

Let λ_k^K, $k=1,\ldots,K$, and the m-vector u^K, denote optimal solutions to the LP primal problem (3.1) and the LP dual problem (3.2), respectively. The generalized linear programming algorithm proceeds by solving the Lagrangean

$$L(u^K) = \min_{x \in X}\{f(x) + u^K g(x)\}$$
$$= f(x^{K+1}) + u^K g(x^{K+1}). \tag{3.3}$$

Note that by its definition in problem (2.3)

$$d = \sup w$$
$$\text{s.t. } w \leq f(x) + ug(x) \qquad \text{for all } x \in X$$
$$u \geq 0$$

Since this problem has at least as many constraints as (3.2), $d^K \geq d$; also by definition, $L(u^K) \leq d$. These inequalities give us immediately

Lemma 3.1: At iteration K of the generalized linear programming algorithm

(i) $d^K \geq d$

(ii) If $L(u^K) \geq d^K$, then $L(u^K) = d = d^K$; that is, u^K is optimal in the dual (2.3)

Thus, the generalized linear programming algorithm terminates with an optimal solution u^K to the dual problem (2.3) if $L(u^K) \geq d^K$. If $L(u^K) < d^K$, the algorithm proceeds by adding a column corresponding to x^{K+1} to (3.1) or equivalently, a row corresponding to x^{K+1} to (3.2).

We consider now the convergence properties of this algorithm when $L(u^K) < d^K$ for all K. The proof of lemma 3.2 is similar to the proof given by Dantzig in [2] for convergence of the generalized linear programming algorithm in the case when (2.1) is a convex problem. We have simply dropped all reference to convexity properties of the primal problem (2.1).

Lemma 3.2: If there exists an index set $\mathcal{K} \subseteq \{1, 2, \ldots\}$ such that the subsequence $\{u^K\}_{K \varepsilon \mathcal{K}}$ is convergent, say to the limit point u^*, then

(i) u^* is optimal in the dual problem (2.3)

and

(ii) $\lim_K d^K = d = L(u^*)$

Proof: By the definition of problem (3.2), we have for all $k = 1, 2, \ldots, K$

$$f(x^k) + u^K g(x^k) \geq d^K \geq d \tag{3.4}$$

where the right inequality is from lemma 3.1. Let $d^\infty = \lim d^K$; this limit exists because the d^K are monotonically decreasing and bounded from below by any $L(u)$ for any $u \geq 0$. Taking the limit in (3.4) for $K \varepsilon \mathcal{K}$, we obtain

$$f(x^k) + u^* g(x^k) \geq d^\infty \geq d \text{ for each } k = 1, 2, \ldots \tag{3.5}$$

Since $g(\cdot)$ is continuous and X is compact, there is a real number B such that $|g_i(x)| \leq B$ for all $x \varepsilon X$ and $i = 1, 2, \ldots, m$. Then

$$|L(x^{K+1}, u^K) - L(x^{K+1}, u^*)| = |(u^K - u^*)\, g(x^{K+1})|$$

$$\leq B \sum_{i=1}^{m} |u_i^K - u_i^*|$$

Consequently, given $\varepsilon > 0$ there is a $K_1 \in \mathcal{K}$ such that for all $K \in \mathcal{K}$, $K \geq K_1$, the right hand side is bounded by ε and therefore

$$L(u^K) = L(x^{K+1}, u^K) \geq L(x^{K+1}, u^*) - \varepsilon$$

$$= f(x^{K+1}) + u^*\, g(x^{K+1}) - \varepsilon$$

Thus, from (3.5) and the definition of d,

$$d \leq d^\infty \leq f(x^{K+1}) + u^*\, g(x^{K+1}) \leq L(u^K) + \varepsilon \leq d + \varepsilon \;.$$

Since $\varepsilon > 0$ was arbitrary, we can conclude that $d^\infty = \lim d^K = d$. Also, the last line implies that

$$\lim_{K \in \mathcal{K}} L(u^K) = L(\lim_{K \in \mathcal{K}} u^K) = L(u^*) = d,$$

where the first equality follows from the continuity of L. ||

In order to establish our main result that generalized linear programming solves the dual, we required a converging subsequence of the linear programming dual prices u^K produced by problem (3.1). It can be shown that a sufficient condition that there exists a converging subsequence of the dual prices u^K is that there exist an $x^o \in X$ such that $g_i(x^o) < 0$, $i = 1, \ldots, m$, (Fisher and Shapiro [10]).+ This is a sufficient condition that $\lim d^K = d$; it may be that $\lim d^K = d$ in all cases, but we have been unable to prove it.

Note that under the hypothesis of lemma 3.2, the generalized linear programming algorithm mechanizes the duality result of lemma 2.2 as announced in Section 1. This follows at once from the algorithm and the weak duality half ($d \leq v^c$) of lemma 2.2. At each step of the algorithm, $d^K = \sum_{k=1}^{K} f(x^k)\ \lambda_k^K$ and $\sum_{k=1}^{K} g(x^k)\ \lambda_k^K \leq 0$; thus, $(d^K,0) \in [f, g]^c$ implying $v^c \leq d^K$ and since $\lim d^K = d$, we can conclude $v^c \leq d$ and therefore $v^c = d$. There is, however, a subtle distinction between the results of lemma 2.2 and 3.2. When the generalized linear programming algorithm converges, we not only have $d = v^c$, but we have found a $u^* \geq 0$ such that $L(u^*) = d$; namely, we have attainment of the dual supremum objective function value in (2.3).

4. Phase One of the Generalized Linear Programming Algorithm

Our discussion here will be brief because the phase one procedure for finding a feasible solution and its properties is closely analogous to the phase two procedure discussed in the previous section for finding an optimal solution to the dual problem (2.3). At iteration K of phase one, the generalized linear programming algorithm solves the master LP (cf problem (3.1)).

$$\begin{aligned}
\sigma^K = \min\ & \sigma \\
\text{s.t.}\ & \sum_{k=1}^{K} g_i(x^k)\lambda_k - \sigma \leq 0,\ i=1,\ldots,m, \\
& \sum_{k=1}^{K} \lambda_k = 1 \\
& \sigma \geq 0,\ \lambda_k \geq 0,\ k=1,\ldots,k
\end{aligned} \tag{4.1}$$

The LP dual to (4.1) is (cf. problem (3.2))

$$\begin{aligned}
\sigma^K = \max\ & w \\
\text{s.t.}\ & w \leq ug(x^k),\ k=1,\ldots,K \\
& \sum_{i=1}^{m} u_i \leq 1 \\
& u \geq 0.
\end{aligned} \tag{4.2}$$

Let λ_k^K, k=1,...,K, and the m-vector u^K, denote the optimal solutions to (4.1) and (4.2). The generalized linear programming algorithm proceeds by solving the problem

$$\begin{aligned} P(u^K) &= \min_{x \varepsilon X} u^K g(x) \\ &= u^K g(x^{K+1}) \end{aligned} \tag{4.3}$$

The dual problem implicitly approximated by (4.1) and (4.2) is

$$\begin{aligned} \sigma = \max\ & P(u) \\ \text{s.t.}\ & \sum_{i=1}^{m} u_i \leq 1 \\ & u \geq 0 \end{aligned} \tag{4.4}$$

The properties of this phase one procedure are summarized below. Note that since the u^K are generated from the compact set $\sum_{i=1}^{m} u_i \leq 1$, $u \geq 0$. They have at least one limit point.

<u>Lemma 4.1</u>:

(i) $\sigma^K \geq \sigma$

(ii) If $P(u^K) \geq \sigma^K$, then $P(u^K) = \sigma = \sigma^K$

(iii) Any limit point u^* of the sequence $\{u^K\}_{K=1}^{\infty}$ of optimal solutions in (4.2) are optimal in the dual problem (4.4).

(iv) $\lim_{K} \sigma^K = \sigma = P(u^*)$, where u^* is any limit point from (iii).

Just as the phase two generalized linear programming problem provides nonlinear duality results, the phase one procedure very easily provides feasibility results in the form of a general theorem of the alternative (Mangasarian [26]). Note that since $\lim \sigma^K = \sigma$, we have $\lim \sigma^K > 0$ if and only if $\sigma > 0$; that is, if and only if there is a $u \geq 0$ such that $\min_{x \in X} ug(x) > 0$.

Lemma 4.2: Exactly one of the following alternatives is valid

(i) There is a $u \geq 0$ such that $\min_{x \in X} [ug(x)] > 0$

(ii) $0 \in G^c$, the convex hull of G where

$$G = \bigcup_{x \in X} \{\eta \in R^m : \eta \geq g(x)\} .$$

Proof: First, note that both alternatives cannot hold; for suppose that $\alpha = \min_{x \in X} [ug(x)] > 0$ and that $0 \in G^c$.

Then (by Caratheodory's theorem again) there are $x^1, x^2, \ldots, x^{m+1} \in X$ with $\eta^k \geq g(x^k)$ and non-negative weights $\lambda_1, \ldots, \lambda_{m+1}$ satisfying $\sum_{j=1}^{m+1} \lambda_j = 1$ and $0 = \sum_{j=1}^{m+1} \lambda_j \eta^j \geq \sum_{j=1}^{m+1} \lambda_j g(x^j)$. Multiplying both sides of this inequality by u leads to the contradiction

$$0 = u \cdot 0 \geq \sum_{j=1}^{m+1} \lambda_j [ug(x^j)] \geq \left(\sum_{j=1}^{m+1} \lambda_j\right) \alpha > 0.$$

On the other hand, if $\min_{x \in X} [ug(x)] \leq 0$ for every $u \geq 0$, then by its definition $\sigma = 0$. By lemma 4.1, $\lim \sigma^K = \sigma = 0$ with $\sigma^K \geq \sum_{k=1}^{m+1} \lambda_k^K g(x^k)$ $i = 1, \ldots, m$; that is $\sigma^K e \in G^c$ where e is a column vector of m ones. But since G^c is closed (see the Appendix) and $\lim \sigma^K e = \sigma e = 0$, $0 \in G^c$. ||

The following results establish that convergence to a starting feasible solution for phase two is finite if there exists an interior point solution to the original primal problem (2.1)

<u>Lemma 4.3</u>: The phase one generalized linear programming algorithm either converges finitely to a feasible solution for the phase two LP problem (3.1) or $\sum_{i=1}^{m} u_i^K = 1$ for all K.

Proof: If $\sum_{i=1}^{m} u_i^K < 1$ for any K, then $\sigma^K = 0$ by complementary slackness between the linear programs (4.1) and (4.2). ||

<u>Corollary 4.1</u>: If there exists an $x^o \in X$ satisfying $g_i(x^o) < 0$, $i=1,\ldots,m$, then the phase one generalized linear programming algorithm converges finitely to a feasible solution for problem (3.1).

Proof: Suppose the phase one generalized linear programming algorithm does not converge finitely to a feasible solution for problem (3.1). Then $\sum_{i=1}^{m} u_i^K = 1$ for all K and there is a limit point u^* of the sequence $\{u^K\}$ such that $\sum_{i=1}^{m} u_i^* = 1$. The existence of the interior point x^o implies $P(u^*) \leq u^* g(x^o) < 0$ which is a contradiction since $P(u^*) \geq \lim_K \sigma^K = \sigma > 0$ by lemma 4.1 (iv).

5. Mathematical Generalizations

In this section, we relax the assumptions on f, g and X used in the construction, analysis and solution by generalized linear programming of problem (2.1). The equivalence of convexification and dualization given in lemma 2.2 remains essentially valid, but the specification and convergence properties of the generalized linear programming algorithm is put in question.

To study problem (2.1) with no assumptions on f, g and X, we need to extend the definitions of $L(u)$ and $v^c(\xi)$ to

$$L(u) = \inf_{x \in X}\{f(x) + ug(x)\},$$

and

$$v^c(\xi) = \inf\{\eta : (\eta, \xi) \in cl([f,g]^c)\}$$

where $cl(\cdot)$ denotes closure. Again, we set $v^c = v^c(0)$. It is no longer true that $d = v^c$ (see lemma 2.2) in all cases. For example, if the primal and dual problems are both infeasible linear programs, then $d = -\infty$, $v^c = +\infty$. Below we show that this is the exceptional case and otherwise $d = v^c$. The following lemma will be useful in the demonstration. It is based in part upon the observation that if $L(u)$ is finite, Lemma 2.2 remains valid when the assumptions on f, g and X are omitted.

<u>Lemma 5.1</u>: Suppose that for every real number r^*, $(r^*,0) \notin cl([f,g]^c)$ and there is a $u \geq 0$ with $L(u)$ finite. Then given any real numbers $r' > r$,

$$(r,0) \notin cl(S)$$

where

$$S = (\{(r',0)\} \cup [f,g])^c.$$

<u>Proof</u>: Suppose to the contrary that $(r,0) \in cl(S)$. Then there exists

$(r^k, y^k) \in S$ with $\lim r^k = r$, $\lim y^k = 0 \in R^m$. By Caratheodory's theorem again, there exist $(\eta^{j,k}, \xi^{j,k}) \in [f,g]$ and non-negative numbers $\lambda_0^k, \lambda_1^k, \ldots, \lambda_{m+2}^k$, $\sum_{j=0}^{m+2} \lambda_j^k = 1$ satisfying

$$r^k = \lambda_0^k r' + \sum_{j=1}^{m+2} \lambda_j^k \eta^{j,k}$$
$$y^k = \sum_{j=1}^{m+2} \lambda_j^k \xi^{j,k} \qquad (5.1)$$

By considering subsequences if necessary, we may assume that $\lim \lambda_0^k = \lambda_0$ with $0 \leq \lambda_0 \leq 1$. We distinguish two cases.

Case 1) $\lambda_0 = 1$. From lemma 2.1, $\eta^{j,k} \geq L(u) - u\xi^{j,k}$ and thus

$$\sum_{j=1}^{m+2} \lambda_j^k \eta^{j,k} \geq \left(\sum_{j=1}^{m+2} \lambda_j^k\right) L(u) - u \sum_{j=1}^{m+2} \lambda_j^k \xi^{j,k} = (1-\lambda_0^k) L(u) - uy^k.$$

But then from (5.1), $r^k \geq \lambda_0^k r' + (1-\lambda_0^k)L(u) - uy^k$. Letting $k \to +\infty$ in this expression leads to the contradiction that $r = \lim r^k \geq r'$.

Case 2) $\lambda_0 < 1$. By considering subsequences if necessary, we may assume that $\lambda_0^k < 1$, $k=1,2,\ldots$

Thus $\mu_j^k = \dfrac{\lambda_j^k}{1-\lambda_0^k} \geq 0$, $\sum_{j=1}^{m+2} \mu_j^k = 1$. Rearranging (5.1),

$$\left(\frac{r^k - \lambda_0^k r'}{1-\lambda_0^k}\right) = \sum_{j=1}^{m+2} \mu_j^k \eta^{j,k} \text{ approaches } \frac{r-\lambda_0 r'}{1-\lambda_0}$$

and

$$\left(\frac{y^k}{1-\lambda_0^k}\right) = \sum_{j=1}^{m+2} \mu_j^k \xi^{j,k} \text{ approaches } 0.$$

But this states that $r^* = \dfrac{r-\lambda_0 r'}{1-\lambda_0}$ satisfies $(r^*, 0) \in cl([f,g]^c)$. This contradiction shows that $(r,0) \in cl(S)$ is impossible.||

We next give the main duality result. For the most part the proof is the same as that of lemma 2.2 with $[f,g]^C$ replaced by $cl([f,g]^C)$.

Theorem 5.2 (Fundamental duality theorem):

If $v^C < +\infty$ or $d > -\infty$, then $v^C = d$.

Proof: ($d \leq v^C$): If $v^C = +\infty$ or if $d = -\infty$, there is nothing to prove. Thus, suppose $d > -\infty$ and X is not empty. For any $u \geq 0$, by lemma 2.1 $L(u) \leq \eta + u\xi$ for every $(\eta,\xi) \in [f,g]^C$. Thus the inequality also holds for every $(\eta,\xi) \in cl([f,g]^C)$ and the argument of lemma 2.2 applies.

($v^C \leq d$): If $v^C = -\infty$ there is nothing to prove; if $v^C < +\infty$, i.e. there is a $(r,0) \in cl([f,g]^C)$, then case (i) of lemma 2.2 applies with $[f,g]$ replaced by $cl([f,g]^C)$.

The only remaining case has $v^C = +\infty$, i.e. there is no $(r^*,0) \in cl([f,g]^C)$. By hypothesis, there is a $\bar{u} \geq 0$ with $L(\bar{u}) > -\infty$. If $L(\bar{u}) = +\infty$ (i.e. X is empty) then $d = v^C$. Otherwise take r, r' with $r' > r$ in lemma 5.1. Then since $(r,0) \notin cl(S)$ in that lemma, there is a hyperplane $\{(y_0,y): u_0y_0+uy=\beta\}$ strictly separating $(r,0)$ and S, thus

$$u_0r + u\cdot 0 < \beta \leq u_0y_0 + uy \tag{5.2}$$

for every $(y_0,y) \in cl[f,g]^C$ and for $(y_0,y) = (r',0)$. The proof now proceeds as in case (i) of lemma 2.1, i.e. $u_j \geq 0$ $(j=1,\ldots,m)$, taking $(y_0,y) = (r',0)$ implies that $u_0 \neq 0$ and thus by scaling $u_0 = 1$ and then by taking $(y_0,y) = (f(x), g(x))$, (5.2) implies that

$$L(u) = \inf_{x\in X}\{f(x) + ug(x)\} \geq \beta > r.$$

Since r was arbitrary $d = \sup_{u\geq 0} L(u) = +\infty = v^C$. ||

A consequence of this result is a theorem of the alternative that somewhat extends lemma 4.2. If G is closed as well as convex, the theorem is similar to theorem 2.2 of Geoffrion [16] and theorem 21.3 of Rockafellar [29].

Corollary 5.2 (theorem of the alternative):

Exactly one of the following two alternatives is valid:

(i) there is a $u \geq 0$ such that $\inf_{x \in X}[ug(x)] > 0$.

(ii) $0 \in cl(G^C)$ where G^C is the convex hull of the set $\bigcup_{x \in X}\{\xi \in R^m : \xi \geq g(x)\}$.

Proof: Let f be identically zero in the optimization problem

$$\inf_{x \in X} f(x)$$

$$\text{subject to } g(x) \leq 0.$$

Then $cl[f,g]^C = \{(0,\xi):\xi \in cl(G^C)\}$ and v^C is either zero or $+\infty$ corresponding to $0 \in cl(G^C)$ and $0 \notin cl(G^C)$ respectively. Also, $L(u) = \inf_{x \in X}[ug(x)]$ here and if $L(u) > 0$ for some u then $L(Ku) \to +\infty$ as $K \to +\infty$. Thus $d = +\infty$ if and only if condition (i) applies. In summary, condition (i) $\Leftrightarrow d = +\infty \Rightarrow v^C = +\infty \Leftrightarrow 0 \notin cl(G^C)$, that is exactly one of conditions (i) and (ii) is valid.||

We consider now possible extensions of the generalized linear programming algorithm to the problem (2.1) with the relaxed assumptions given at the beginning of this section. The generalized linear programming algorithm must be modified for the cases when L(u) is finite but not attained or when $L(u) = -\infty$. Modifications to the algorithm in the former case are possible and will not be discussed here (see Fox [13]).

Modifications may be more difficult when there are $u \geq 0$ such that $L(u) = -\infty$. Since such points are of no interest, we can rewrite the dual problem as

$$w = \sup_{u \varepsilon U} L(u), \tag{5.3}$$

where

$$U = \{u : u \geq 0 \text{ and } L(u) > -\infty\}.$$

It can easily be shown that U is a convex set, but it can be empty although v is finite. A sufficient condition for U to be empty is that for some $y \varepsilon R^m$, $v(y) = -\infty$ for problem (2.1) with right hand side y. This is the contrapositive of lemma 2.1. This condition is given by Geoffrion [14; p. 18] who also establishes necessity when (2.1) is a convex programming problem.

If U is not empty, but strictly contained in the non-negative orthant, the generalized linear programming algorithm may not work. This is because the algorithm can generate a non-negative $\bar{u} \notin U$ leaving us without a rule for generating a meaningful constraint to add to (3.2) as a result of computing $L(\bar{u})$. An exception is when U can be explicitly represented and the constraints violated by $\bar{u}$ can be added to (3.2). This is the case, for example, in the integer programming dual problem of Fisher and Shapiro [10].

In spite of these difficulties, let us suppose that the generalized linear programming algorithm can be applied unambiguously to problem (2.1) with our relaxed assumptions. Lemma 3.1 which characterizes finite convergence to an optimal solution to the dual is still valid. However, infinite convergence to the optimal value for the dual in lemma 3.2 requires that the sequence $u^k g(x^{k+1})$ is bounded.

6. Concluding Remarks

We have demonstrated in this paper how linear programming applied to some fairly arbitrary mathematical programming problems produces optimal solutions to the duals to these problems. In another paper, we will demonstrate how linear programming can be used in a direct fashion in the construction of dual ascent methods.

Our main motivation in writing this paper has been more than pedagogical. Computational experience with generalized linear programming on large scale problems has been disappointing because the dual prices generated at each iteration can fluctuate greatly and bear little relation to the optimal dual prices. Moreover, convergence can be quite slow. In our opinion, a more flexible approach to solving dual problems is required, including approximation methods which do not use the simplex method in any of its forms.

Appendix A

The sets $G = \{\zeta \in R^m : \zeta \geq g(x) \text{ for some } x\varepsilon X\}$ and $[f, g] = \{(\eta, \xi) : \eta \geq f(x) \text{ and } \xi \geq g(x) \text{ for some } x\varepsilon X\}$ play a central role in the duality theory studied here. These sets may be expressed as $S^+ = \{\zeta \in R^m : \zeta \geq y$ for some $y \in S\}$ where $S = \{g(x) : x\varepsilon X\}$, the image of X under g, for G and $S = \{(f(x), g(x)) : x\varepsilon X\}$ for [f, g]. In this appendix, we consider briefly some relationships between S and S^+ and in the process, establish that G^c and $[f, g]^c$ are closed.

Lemma A.1:

$$(S^+)^c = (S^c)^+$$

proof: (i) $(S^c)^+ \subseteq (S^+)^c$

If $y \in (S^c)^+$, then $y \geq \sum_{j=1}^{m+1} \lambda_j s_j$ for some $s_j \in S$ and $\lambda_j \geq 0$, $\sum_{j=1}^{m+1} \lambda_j = 1$. Let $\delta = y - \sum_{j=1}^{m+1} \lambda_j s_j \geq 0$; then $y = \sum_{j=1}^{m+1} \lambda_j (s_j + \delta) = \sum_{j=1}^{m+1} \lambda_j s_j^+$ where $s_j^+ = s_j + \delta \in S^+$. Thus $y \in (S^+)^c$ and (i) is established.

(ii) $(S^+)^c \subseteq (S^c)^+$

If $y \in (S^+)^c$, then $y = \sum_{j=1}^{m+1} \lambda_j s_j^+ \geq \sum_{j=1}^{m+1} \lambda_j s_j$ for some $s_j^+ \in S^+$, $s_j \in S$ and $\lambda_j \geq 0$ satisfying $\sum_{j=1}^{m+1} \lambda_j = 1$. But the point $\sum_{j=1}^{m+1} \lambda_j s_j \in S^c$ and this enables us to conclude that $y \in (S^c)^+$ establishing (ii).||

Lemma A.2: If S is compact, then both S^+ and $(S^c)^+$ are closed.

Proof: If S is compact, then so is S^c and consequently, the lemma need only be exhibited for S^+. Let s^+ be a point of closure of S^+, and let $\{s_j^+\}$ be a sequence of points in S^+ converging to it. By definition of S^+, there exists a sequence $\{s_j\}$ of points in S satisfying $s_j^+ \geq s_j$. Since S is compact, there is a subsequence $\{s_{k_j}\}$ converging to a point $s \varepsilon S$. Thus, the subsequence $\{s_{k_j}^+\}$ converges to the point $s^+ \geq s$ which establishes the fact that $s^+ \varepsilon S^+$.

Corollary A.1

If X is compact and g is continuous, then the set G^c is closed. If in addition f is continuous $[f, g]^c$ is closed.

Proof: The set $S = \{g(x) : x \varepsilon X\}$ is compact and by lemma A.2, $(S^c)^+$ is closed. By lemma A.1, then so is $G^c = (S^+)$. The same argument applies to $[f, g]^c$.

References

1. Baumol, W. J. and T. Fabian, "Decomposition, Pricing for Decentralization and External Economies," Management Science, 11, pp. 1-32 (1964).

2. Dantzig, G. B., Linear Programming and Extensions, Princeton University Press, Princeton, N.J. (1963).

3. Dantzig, G. B. and P. Wolfe, "Decomposition Principles for Linear Programs, Operations Research, 8, pp. 101-111 (1960).

4. Dantzig, G. B. and P. Wolfe, "The Decomposition Algorithm for Linear Programming," Econometrica, 29, pp. 767-778 (1961).

5. Dantzig, G. B. and R. M. Van Slyke, "Generalized Linear Programming," Chapter 3 in Optimization Methods for Large Scale Systems, D. A. Wisner, editor, McGraw-Hill (1970).

6. Dzielinski, B. P. and R. E. Gomory, "Optimal Programming of Lot Size Inventory and Labor Allocations," Management Science, 11, pp. 874-890 (1965).

7. Falk, J. E., "Lagrange Multipliers and Nonconvex Programming," SIAM J. on Control, 7, pp. 534-545 (1969).

8. Fisher, M. L., "Optimal Solution of Scheduling Problems Using Generalized Lagrange Multipliers and Relaxation" (to appear in Operations Research).

9. Fisher, M. L. and J. F. Shapiro, "Constructive Duality in Integer Programming," Working Paper No. OR008-72, M.I.T. Operations Research Center, April 1972 (to appear in SIAM. J. on Appl. Math.).

10. Fisher, M. L.annd J. F. Shapiro, "Constructive Duality in Discrete Optimization," CMSBE Report 7321, Graduate School of Business, University of Chicago, April 1973.

11. Ford, L. R. and D. R. Fulkerson, "A Suggested Computation for Maximal Multi-Commodity Network Flows," Management Science, 5, pp. 97-101 (1958).

12. Fox, B. L., "Stability of the Dual Cutting-Plane Algorithm for Concave Programming," RM-6147-PR, The Rand Corporation (1970).

13. Gale, D., "A Geometric Duality Theorem with Economic Applications," Rev. Economic Studies, 34, pp. 19-24 (1967).

14. Geoffrion, A. M., "Primal Resource-Directive Approaches for Optimizing Nonlinear Decomposable Systems," Operations Research, 18, pp. 375-403 (1970).

15. Geoffrion, A. M., "Duality in Nonlinear Programming: A Simplified Applications-Oriented Development," SIAM Review, 13, pp. 1-37 (1971).

16. Geoffrion, A. M., "Generalized Benders Decomposition," J. of Optimization Theory and Applications, 10, pp. 237-260 (1972).

17. Gilmore, P. C. and R. E. Gomory, "A Linear Programming Approach to the Cutting-Stock Problem, Part II," Operations Research, 11, pp. 863-888 (1963).

18. Gomory, R. E. and T. C. Hu, "An Application of Generalized Linear Programming to Network Flows," J. SIAM, 10, pp. 260-283 (1962).

19. Grinold, Richard C., "Steepest Ascent for Large Scale Linear Programs," SIAM Review, 14, pp. 447-464 (1972).

20. Held, M. and R. M. Karp, "The Traveling Salesman Problem and Minimum Spanning Trees," Operations Research, 18, pp. 1138-1162 (1970).

21. Held, M. and R. M. Karp, "The Traveling Salesman Problem and Minimum Spanning Trees: Part II," Mathematical Programming, 1, pp. 6-25 (1971).

22. Lasdon, L. S., "Duality and Decomposition in Mathematical Programming," IEEE Trans. System Sci. and Cybernetics, 4, pp. 86-100 (1968).

23. Lasdon, L. S., Optimization Theory for Large Systems, New York: The Macmillan Company, 1970.

24. Leventhal, T, G. Nemhauser and L. Trotter, Jr., "A Column Generator Algorithm for Optimal Traffic Assignment," Department of Operations Research Report, Cornell University, 1972.

25. Luenberger, D. G., Optimization by Vector Space Methods, New York: John Wiley and Sons, 1969.

26. Mangasarian, O. L., Nonlinear Programming, New York: McGraw-Hill, 1969.

27. Magnanti, T. L., "A Linear Approximation Approach to Duality in Nonlinear Programming," Technical Report OR016-73, Operations Research Center, M.I.T., April 1973.

28. Orchard-Hays, W., Advanced Linear Programming Computing Techniques, New York: McGraw-Hill, 1968.

29. Rockafellar, R. T., Convex Analysis, Princeton, N.J.: Princeton University Press, 1970.

30. Shapiro, J. F., "Generalized Lagrange Multipliers in Integer Programming," Operations Research, 19, pp. 68-76 (1971).

31. Shapiro, J. F., "A Decomposition Algorithm for Integer Programming Problems with Many Columns," Proceedings of 25th Meeting of A.C.M., Boston, pp. 528-533 (1972).

32. Tomlin, J. A., "Minimum-Cost Multicommodity Network Flows," Operations Research, 14, pp. 45-51 (1966).

33. Wagner, H. M., Principles of Operations Research, Englewood Cliffs, N.J.: Prentice-Hall, 1969.

34. Wagner, M. H., "Constructive Fixed Point Theory and Duality in Nonlinear Programming," Technical Report No. 67, Operations Research Center, M.I.T. (1971).

35. Whittle, P., Optimization Under Constraints, New York: John Wiley and Sons, 1971.

36. Williams, A. C., "A Treatment of Transportation Problems by Decomposition," J. SIAM, 11, pp. 855-871 (1963).

37. Wolfe, P., "Convergence Theory in Nonlinear Programming," pp. 1-36 in Integer and Nonlinear Programming, J. Abadie, editor, North-Holland (1970).

38. Zangwill, W. I., Nonlinear Programming: A Unified Approach, Englewood Cliffs, N.J.: Prentice-Hall, 1969.