OPERATIONS RESEARCH CENTER

working paper

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

GENERALIZED LINEAR PROGRAMMING

WITHOUT CONSTRAINT QUALIFICATION

by

Elie Gugenheim and Jeremy F. Shapiro

OR115-82 JUNE 1982

*This research was supported in part by the National Science Foundation under Grant No. ECS-8117105.

Generalized Linear Programming
Without Constraint Qualifications

by

Elie Gugenheim and Jeremy Shapiro

1. Introduction

In an earlier paper [2], the equivalence was established between convexification and dualization of an arbitrary mathematical programming problem. Generalized linear programming (also known as Dantzig-Wolfe decomposition) applied to such a problem was shown to be a mechanization of this result in the following sense. If the sequence of linear programming master problems generated by the method produces a convergent subsequence of optimal shadow prices, any limit point of such a subsequence is an optimal solution to the dual of the arbitrary mathematical programming problem. Moreover, the limit of the optimal master problem objective function values equal the value of the dual problem. These results hold whether or not there is a duality gap, or what is almost the same condition, whether or not generalized linear programming finds an optimal solution to the arbitrary primal problem.

A sufficient condition for generalized linear programming to converge in this sense is the existence of an interior solution. This familiar regularity condition, or constraint qualification, has the constructive effect of bounding the set of optimal shadow prices produced by the master problems thereby ensuring the existence of a convergent subsequence. Even if the primal problem does not have an interior solution, however, the optimal master problem objective function values converge to a limit in all non-trivial cases. This is because the values are monotonically decreasing and bounded from below by the finite optimal objective function value of the dual problem, which we denote by d. Clearly, the limit is at least as great as d. A loose end of the analysis in [2] is a characterization of this limit, and the purpose of this note is to give one. Our approach is to perform first a complete study of the one constraint case, then relate the results to the multiple constraint case. The problem has also been studied by Wang [4] who has developed some results complementary to ours.

2. Review

This section consists of a brief review of the necessary results and constructions. The primal problem we will consider is

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

where f is a continuous function from R^n to R, g is a continuous function from R^n to R^m with components g_i, X is a non-empty, compact set. We have made these assumptions about f, g and X to simplify the mathematical analysis, but the results are valid in more general cases.

The dual problem is

$$\begin{aligned} d = \sup\ & L(u) \\ \text{s.t.}\ & u \geq 0 \end{aligned} \qquad \text{(D)}$$

where

$$L(u) = \underset{x \in X}{\text{minimum}} \{f(x) + ug(x)\}.$$

We let $L(u,x)$ denote the quantity $f(x) + ug(x)$. It is well known and easy to show that L is a finite, concave function, $L(u) \leq v$, and $d \leq v$. The relationship of (D) to (P) which we seek but may not find is embodied in the <u>global optimality conditions</u>: For $\bar{x} \in X$, and $\bar{u} \geq 0$, these are

(i) $L(\bar{u}) = f(\bar{x}) + \bar{u}g(\bar{x})$

(ii) $\bar{u}g(\bar{x}) = 0$

(iii) $g(\bar{x}) \leq 0$.

If the global optimality conditions hold, then $\bar{x}$ is optimal in (P), $\bar{u}$ is optimal in (D), and $v = d = L(\bar{u})$ (e.g., see Shapiro [3]). We will also make use of the following result.

<u>Theorem 1 (Wang [4])</u>: Suppose for each i, either $g_i(x) \leq 0$ for all $x \in X$, or $g_i(x) \geq 0$ for all $x \in X$. Then $d = v$.

Generalized linear programming is a method for optimizing (D). Given the solutions $x^1,\ldots,x^K \in X$, the method proceeds by solving the linear programming master problem

$$\begin{aligned} d^K &= \max w \\ \text{s.t.} \quad w &\leq f(x^k) + ug(x^k) \qquad k = 1,\ldots,K \\ u &\geq 0 \end{aligned} \tag{1}$$

It is easy to see that

$$d^K = \underset{u \geq 0}{\text{maximum}} \ \underset{k=1,\ldots,K}{\text{minimum}} \ \{f(x^k) + ug(x^k)\} \geq d.$$

We can assume that a phase one procedure has been applied to the dual of (1) to select some of the x^k so that d^K is finite. Let u^K denote an optimal solution to (1). The method continues by solving

$$\begin{aligned} L(u^K) &= \underset{x \in X}{\text{minimum}} \ \{f(x) + u^K g(x)\} \\ &= f(x^{K+1}) + u^K g(x^{K+1}). \end{aligned}$$

If $L(u^K) = d^K$, then u^K is optimal in (D) and $d = d^K$. On the other hand, if $L(u^K) < d^K$, the constraint

$$w \leq f(x^{K+1}) + ug(x^{K+1})$$

is added to (1) and it is re-optimized.

Generalized linear programming optimizes (1) in the following sense.

Theorem 2 (Dantzig [1], Magnanti, Shapiro and Wagner [2]): If there is a subsequence $\{u^{K_i}\}_{i\varepsilon I}$ converging to a limit point u^*, then

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

The convergence condition hypothesized in Theorem 1 can be constructively ensured by finding an interior point; namely, an $x^0 \varepsilon X$ such that $g_i(x_0) < 0$ for $i = 1,\ldots,m$. The point is used to generate a constraint in problem (1). Employing the constraint qualification in this way bounds the set of feasible solutions u since

$$w \leq f(x^0) + ug(x^0)$$

implies

$$\sum_{i=1}^{m} u_i(-g_i(x^0)) \leq f(x^0) - w,$$

where the right hand side is bounded by $f(x^0) - L(\tilde{u})$ for any $\tilde{u} \geq 0$, and the coefficients on the left hand side $-g_i(x^0)$ are all positive. By the same reasoning, the set of feasible solutions u will be bounded if the first m constraints in (1) are written with respect to $x^i \varepsilon X$ satisfying $g(x^i) \leq 0$ and $g_i(x^i) < 0$ for $i = 1,\ldots,m$. More generally, we always have

$$\lim_K d^K = d^\infty \geq d.$$

3. Solution of the One Constraint Case

In this section, we consider problem (P) with one constraint ($m = 1$) under the assumptions that $g(x) \geq 0$ for all $x \varepsilon X$, and $g(x^1) = 0$ for some $x^1 \varepsilon X$. In addition, we assume generalized linear programming begins with the solution x^1. At iteration K, we have

$$d^K = \underset{u \geq 0}{\text{maximum}} \ \underset{k=1,\ldots,K}{\text{minimum}} \ \{f(x^k) + ug(x^k)\}.$$

A typical situation is shown in Figure 1 for $K = 3$. At iterations 2 and 3, the Lagrangean calculation produces the solutions $x \varepsilon X$ satisfying $g(x) > 0$ and $f(x) < 0$ implying $d^1 = d^2 = d^3 = f(x^1)$ and the successive dual solutions satisfy $u^1 < u^2 < u^3$.

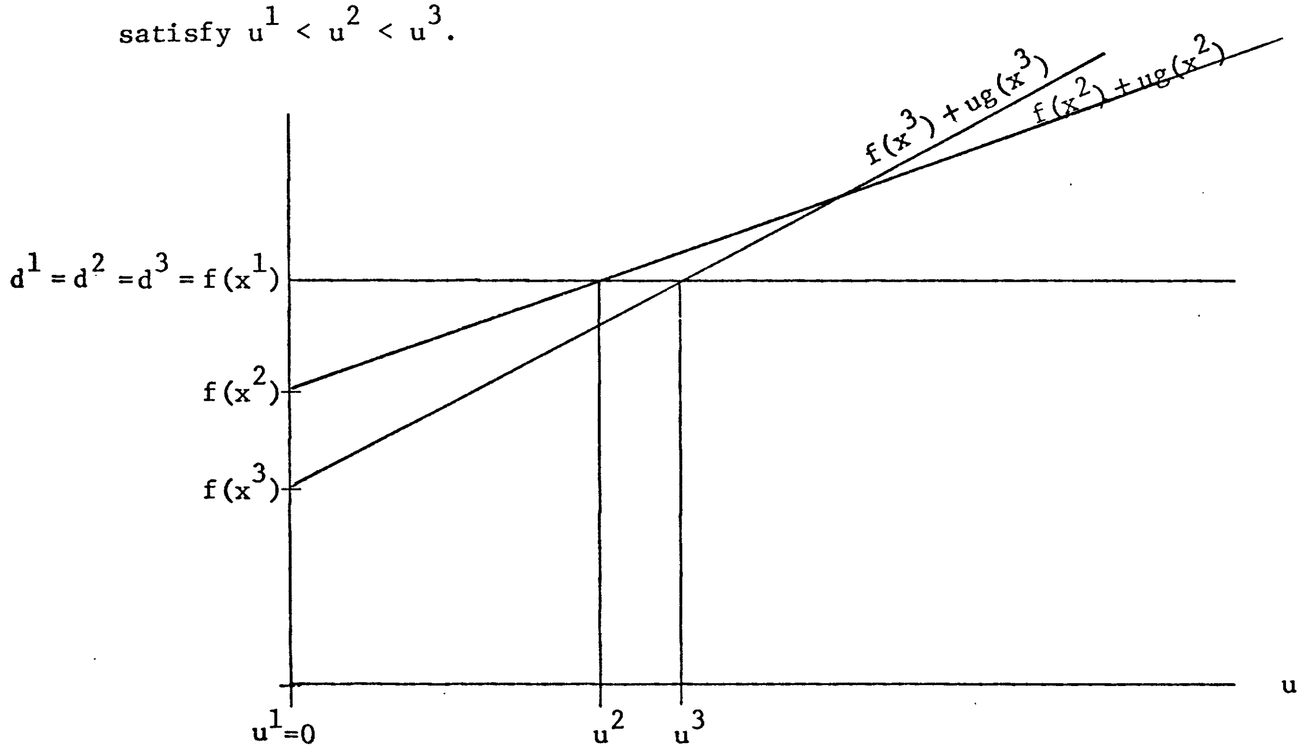

Figure 1

In general, we can see that at any iteration K,

$$d^K = \min f(x^k)$$
$$\text{s.t.} \qquad g(x^k) = 0;$$

that is, d^K equals the minimal objective function value of the feasible solutions to (P). Moreover, we can choose any u^K to be optimal in the master problem that satisfies $f(x^k) + u^K g(x^k) \geq d^K$ or $u^K \geq (d^K - f(x^k))/g(x^k)$ for all those points x^k such that $g(x^k) > 0$, and $f(x^k) < d^K$. Thus, by inspection, we take

$$u^K = \max \{0, \max \{\frac{d^K - f(x^k)}{g(x^k)} \mid g(x^k) > 0\}\} \tag{2}$$

to be the optimal solution to the master problem at iteration K.

Consider now the possibilities when we calculate

$$L(u^K) = f(x^{K+1}) + u^K g(x^{K+1}) \leq d^K.$$

On the one hand, if $g(x^{K+1}) = 0$, then we have

$$v \geq d \geq L(u^K) = f(x^{K+1}) \geq v$$

which permits us to conclude immediately that $L(u^K) = f(x^{K+1}) = v = d$ and generalized linear programming terminates. On the other hand, if $g(x^{K+1}) > 0$, then we must have $f(x^{K+1}) < d^K$ since generalized linear programming had not previously terminated implying $u^K > 0$. We can summarize the behavior of the method by the following.

Theorem 3: Suppose generalized linear programming is applied to (P) with $m = 1$ and $g(x) \geq 0$ for all $x \varepsilon X$. Suppose further that the method is initiated with the solution $x^1 \varepsilon X$ satisfying $g(x^1) = 0$ implying $d^1 = f(x^1)$. Two outcomes are possible.

(a) At some iteration $K \geq 1$, the new solution x^{K+1} generated by the method satisfies $g(x^{K+1}) = 0$ implying x^{K+1} is optimal in (P) and $\lim_K d^K = d = v$, or

(b) For all iterations $K \geq 1$, the new solution x^{K+1} satisfies $g(x^{K+1}) > 0$ implying for $K = 2,3,..$

$$d^K = d^1$$

and an optimal solution to the master problem is

$$u^K = \frac{d^1 - f(x^K)}{g(x^K)}$$

Moreover,

$$\lim_K d^K = d^1 = f(x^1) \geq v = d$$

Note that we know $v = d$ by applying theorem 3 to the one constraint problem (P). Case (b) of Theorem 2 says that $\lim_K d^K$ can equal the objective function value $f(x^1)$ of any feasible solution x^1. The next section gives a specific example of this occurrence. The remainder of this section is devoted to results characterizing when case (a) obtains. As in the case when the constraint qualification holds, the underlying idea is to bound the monotonically increasing sequence $\{u^K\}_{K=1}^{\infty}$ defined in (2) thereby eliminating case (b) as a possibility.

<u>Theorem 4</u>: For problem (P) with m = 1, the following properties are equivalent.

(1) $h(x) = \dfrac{f(x) - d}{g(x)}$ is bounded from below for all $x \in X$ satisfying $g(x) > 0$.

(2) $d = L(u)$ for some $u \geq 0$.

(3) If the generalized linear programming algorithm is initiated with any $x^1 \in X$ satisfying $g(x^1) = 0$, the sequence $\{u^K\}$ given by (2) satisfies $\lim_K u^K = u^*$ such that $L(u^*) = d$.

Proof: We show that (1) $\Rightarrow$ (2) $\Rightarrow$ (3) $\Rightarrow$ (2) $\Rightarrow$ (1):

(1) $\Rightarrow$ (2): Assume that there exists a real number A such that for all $x \in X$ such that $g(x) > 0$, we have $h(x) \geq A$. Then take any $u \geq \text{Max}\ \{0, - A\}$ and consider $x \in X$:

- If $g(x) = 0$, $L(u,x) = f(x) \geq v = d$.
- If $g(x) > 0$, $L(u,x) = f(x) + ug(x) \geq f(x) - Ag(x) \geq d$;

hence

$$L(u) \geq d, \text{ and } L(u) = d.$$

(2) $\Rightarrow$ (3): We suppose that there exists a $\bar{u} \geq 0$ such that $L(\bar{u}) = d$. Let $x^* \in X$, $g(x^*) = 0$. As explained in Theorem 3, if (a) occurs, (3) holds. If (b) occurs, then our previous remark implies $L(u^K) < d$ for all $K \geq 1$. But since L is a monotonously increasing function over $u \geq 0$, we must then have, for all $K \geq 1$, $u^K < \bar{u}$. Thus, the monotonously increasing sequence

$\{u^K\}$ is bounded by $\bar{u}$, and therefore converges to some $u^* \leq \bar{u}$ which, by Theorem 2, will satisfy: $L(u^*) = d$.

(3) $\Longrightarrow$ (2): Obvious, since there exists an $x \in X$ such that $g(x) = 0$.

(2) $\Longrightarrow$ (1): We assume that there exists a $\bar{u} \geq 0$ such that $L(u) = d$. Take any $x \in X$ with $g(x) > 0$. Then $L(\bar{u},x) = f(x) + \bar{u}g(x) \geq L(\bar{u}) = d$. Hence, $h(x) \geq -\bar{u}$, and h is bounded from below over the set $\{x \in X \mid g(x) > 0\}$ by $-\bar{u}$. ||

Alternatively, we could easily prove the following results: First, we say that "property (H)" holds if any of the (equivalent) properties (1), (2) and (3) of the theorem holds. Then, we could show:

(H) holds if the set $\{x \in X | g(x) = 0 \text{ and } f(x) > v\}$ is closed in R^n.

More generally, we have

(3) $\Longrightarrow$ (4):

$$\underset{\begin{pmatrix} x \;\in X \\ g(x) = 0 \\ f(x) = v \end{pmatrix}}{\text{Min}} \left\{ \underset{\begin{matrix} y \Longrightarrow x \\ g(y) > 0 \\ f(y) < v \end{matrix}}{\lim} h(y) \right\} > -\infty$$

Corollary 1: If (H) does not hold, then for any $x^1 \in X$ such that $g(x^1) = 0$, for all $K \geq 1$, we have $d^K = d^\infty = d^1 = f(x^1) \geq v = d$.

Theorem 4 fulfills the task we assigned to it. It shows that the occurrence of (a) or (b) in Theorem 3 is entirely independent on the choice of the starting point x^1, and it provides a convenient criterion to predict the behavior of a given problem.

Thus, our results complement the results which can be obtained when the constraint qualification holds. A counter-example ($d^\infty > d$) will be constructed in the next section as an application of our theory, showing that the sequence $\{u^K\}$ may well go to $(+\infty)$ and that d^∞ may be strictly greater than d when the primal problem does not have any interior solution.

4. A Counter Example

Since the behavior of the generalized linear programming algorithm depends entirely on the function h over $\{x \varepsilon X | g(x) > 0\}$, we can choose it and f as we wish, and obtain the desired results by the proper choice of g. In particular, we take $x = [0,1]$, $g(x) = 0$ only for $x = 0$ and $x = 1$, and $v = 0$ is attained only at $x = 0$.

$$f(x) = \begin{cases} -2\sqrt{x}, & \text{for all } x \varepsilon [0, \frac{1}{2}] \\ (-2\sqrt{2}-1) + 2(1+\sqrt{2})x, & \text{for all } x \varepsilon [\frac{1}{2}, 1] \end{cases}$$

Then, we have $h(x) = \dfrac{f(x)}{g(x)}$, for $x \varepsilon (0,1)$. Since we want

$$\left(\begin{array}{c} \lim \\ x \to 0 \\ x > 0 \\ f(x) < 0 \end{array}\right) h(x) = -\infty,$$

we choose

$$g(x) = \begin{cases} 2x, & \text{for } x \varepsilon [0, \frac{1}{2}] \\ 2 - 2x, & \text{for all } x \varepsilon [\frac{1}{2}, 1] \end{cases}$$

The functions f and y are shown in Figure 2.

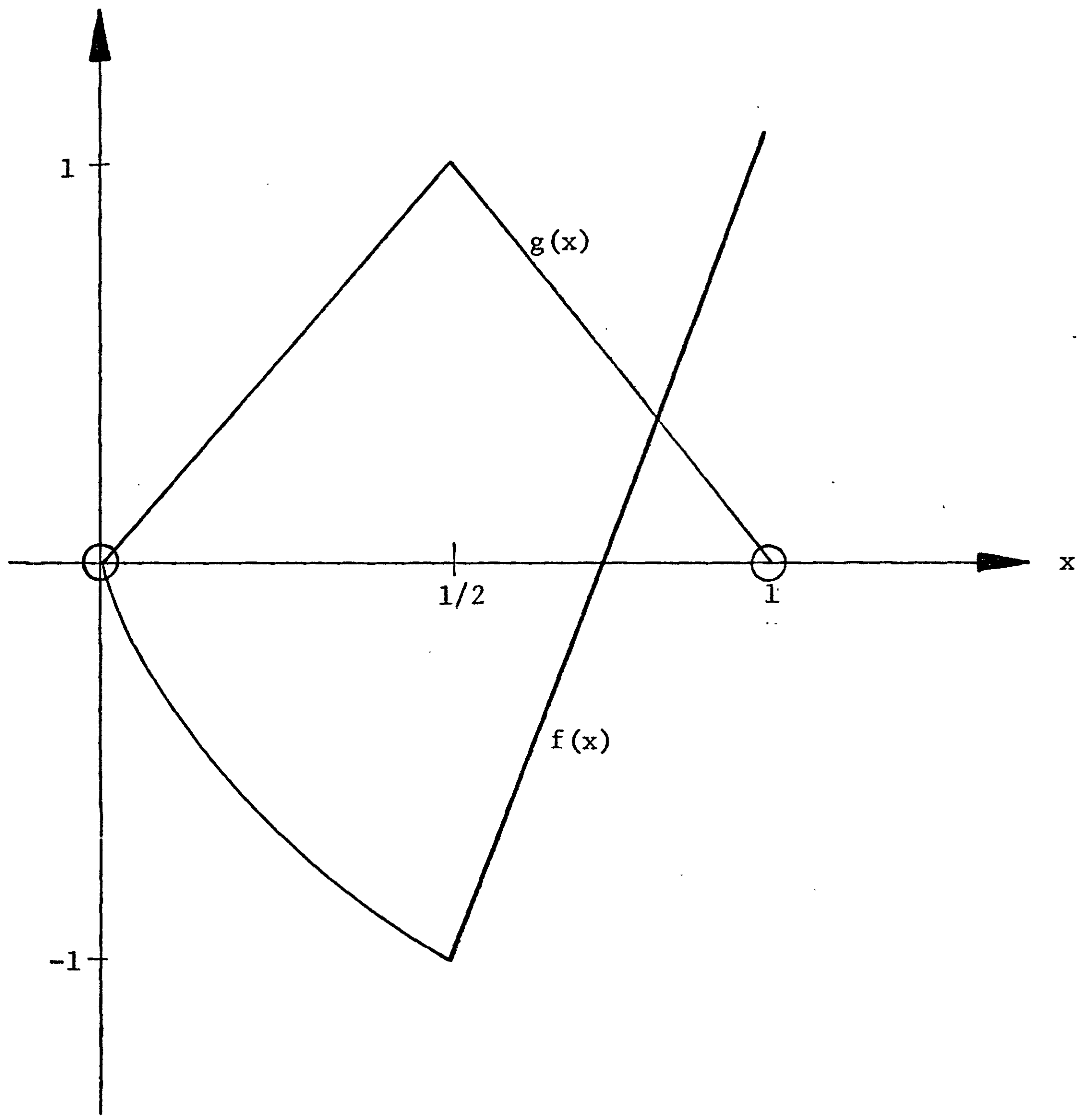

Figure 2

Thus, criterion (4) is contradicted and (H) does not hold. The reader will easily check that, in agreement with the predictions of Theorem 4, the sequence $\{u^K\}$ obtained from $x^1 = 0$ or 1 will diverge to $+\infty$, whereas the values d^K remain constant, equal to $f(x^1)$. Starting, for instance, with $x^1 = 1$, leads to $d^\infty = 1 > v = d = 0$.

Actually, one finds:

$$L(u) = \begin{cases} u - \sqrt{2}, & \text{for all } u < \frac{\sqrt{2}}{2} \\ -\frac{1}{2u}, & \text{for all } u \geq \frac{\sqrt{2}}{2} \end{cases}$$

and $L(0) = -\sqrt{2}$; thus $d = 0$ is not attained.

5. The General Case

We apply the results of section 3 to the multiple constraint case. One way to accomplish this is to make use of the function

$$G(x) = \underset{i=1,\ldots,m}{\text{maximum}}\ g_i(x)$$

Problem (P) is equivalent to the single constraint problem

$$\begin{aligned} v &= \min f(x) \\ \text{s.t.} \quad & G(x) \leq 0 \qquad\qquad (\tilde{P}) \\ & x \in X \subseteq R^n. \end{aligned}$$

Although the dual problem to $(\tilde{P})$ is not equivalent to (D), we can make use of it in analyzing (P).

Specifically, we define for $u \varepsilon R^1$ the dual problem

$$\begin{aligned} \tilde{d} &= \max \tilde{L}(u) \\ \text{s.t.} \quad & u \geq 0 \end{aligned} \qquad (\tilde{D})$$

where

$$\tilde{L}(u) = \underset{u \geq 0}{\text{minimum}} \{f(x) + uG(x)\}$$

Lemma 1: For any $u \varepsilon R^m$ and $u \geq 0$,

$$L(u) \leq L(\sum_{i=1}^{m} u_i) \text{ implying } d \leq \tilde{d}.$$

Proof: For any $x \varepsilon X$, $L(u,x) \leq L(\sum_{i=1}^{m} u_i, x)$ since $u \geq 0$ and $g_i(x) \leq G(x)$ implies $f(x) + ug(x) \leq f(x) + (\sum_{i=1}^{m} u_i) G(x)$. Thus, $L(u) \leq \tilde{L}(\sum_{i=1}^{m} u_i)$ and taking the supremum of both sides gives us $d \leq \tilde{d}$. The inequality $d \leq \tilde{d}$ may be strict but there are some properties linking (P), (D) and $(\tilde{D})$ that we can state.

<u>Theorem 5</u>:

(a) $d \leq \tilde{d} \leq v$

(b) If $v = d$ (that is, there is no duality gap between (P) and (D)) then $v = \tilde{d}$.

(c) If $d = \tilde{d}$ and d is attained at $u^* \varepsilon R^m$, then $L(\sum_{i=1}^{m} u_i^*) = \tilde{d}$.

As far as the generalized linear programming algorithm is concerned, we can use the results of section 3 to construct examples in any dimension for which the dual vectors $\{u^K\}$ do not have any convergent subsequence, no matter what starting solution x^1 is chosen. Indeed, we know that d cannot be attained if $\tilde{d} = d$ but $\tilde{d}$ is not attained. In turn, Theorem 4 provides us with several criteria guaranteeing that (H) does not hold for (P).

References

[1] Dantzig, G. B., Linear Programming and Extensions, Princeton U. Press, 1963.

[2] Magnanti, T. L., J. F. Shapiro and M. H. Wagner, "Generalized Linear Programming Solves the Dual," Management Science, 22, pp. 1195-1203, 1976.

[3] Shapiro, J. F., Mathematical Programming: Structures and Algorithms, J. Wiley and Sons, 1979.

[4] Wang, H., "On the Relationship Between a Non Convex Problem and its Convex Relaxation," Wellesley College, 1980.