Mathematical Programming 17 (1979) 320–334.
North-Holland Publishing Company

SOME RELATIONSHIPS BETWEEN LAGRANGIAN AND SURROGATE DUALITY IN INTEGER PROGRAMMING

Mark H. KARWAN

Industrial Engineering, State University of New York at Buffalo, Amherst, NY, U.S.A.

Ronald L. RARDIN

Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, U.S.A.

Received 21 April 1976
Revised manuscript received 19 March 1979

Lagrangian dual approaches have been employed successfully in a number of integer programming situations to provide bounds for branch-and-bound procedures. This paper investigates some relationship between bounds obtained from lagrangian duals and those derived from the lesser known, but theoretically more powerful surrogate duals. A generalization of Geoffrion's integrality property, some complementary slackness relationships between optimal solutions, and some empirical results are presented and used to argue for the relative value of surrogate duals in integer programming. These and other results are then shown to lead naturally to a two-phase algorithm which optimizes first the computationally easier lagrangian dual and then the surrogate dual.

Key words: Integer Programming, Surrogate Duality, Lagrangian Relaxation, Lagrangian Duality, Surrogate Constraint.

1. Introduction

The general integer linear programming problem[1] can be stated as:

$$(P)\qquad \text{minimize}\quad cx,\qquad \text{subject to}\quad Ax \le b,\quad x \in S$$

where

$$S = \{x \ge 0\colon\ Gx \le h, x \text{ satisfies some discrete constraints}\}.$$

Here, as usual, A is a $m \times n$ matrix, x, b, c are vectors of the appropriate dimension, and G is a $k \times n$ matrix with h a $k \times 1$ vector. To avoid many pathological cases assume S is bounded and (P) is feasible.

The surrogate constraint relaxation of the problem (P) associated with any $v \ge 0$ is

[1] Many of the results of this investigation can be shown to hold for much more general versions of (P) with nonlinear constraints and objective functions. However, we will consider only the above formulation of (P) because if conforms with much of the theoretical and applied literature in integer programming.

$$(P^v) \qquad \text{minimize} \quad cx,$$
$$\text{subject to} \quad v(Ax-b) \le 0, \quad x \in S.$$

Define the function $\nu(\cdot)$ = the value of an optimal solution to problem $(\cdot)$. Then the surrogate dual is:

$$(D_S) \qquad \max_{v \ge 0} \{\nu(P^v)\}.$$

The more widely known lagrangian relaxation is:

$$(P_u) \qquad \text{minimize} \quad cx + u(Ax-b),$$
$$\text{subject to} \quad x \in S$$

with the corresponding lagrangian dual

$$(D_L) \qquad \max_{u \ge 0} \{\nu(P_u)\}.$$

The first major theoretical treatment of surrogate duality in mathematical programming is that of Greenberg and Pierskalla [10]. More recently, Glover [7], summarized these results and presented a unified development of a surrogate duality theory and Greenberg and Pierskalla [11] reported a complementary theoretical approach. Glover noted that the approach of Geoffrion [5], which advocated discarding the surrogate constraint viewpoint due to "being subsumed by the simplest lagrangian relaxation technique" utilized relaxation assumptions whose effect was to replace the original nonconvex problem by one whose structure was such that the "distinction between the surrogate constraint approach and the lagrangian approach vanishes". A very important result [10] is that any true surrogate duality gap (i.e., $\nu(P)-\nu(D_S)$) is at least as small and often smaller than the lagrangian duality gap ($\nu(P)-\nu(D_L)$). Coupled with the success of many lagrangian dual formulations and procedures (notably Held and Karp's treatment of the traveling salesman problem [12], Geoffrion's work on the plant location problem [6], and Fisher's algorithm for scheduling in the presence of resource constraints [3]) this leads one to hope for some similar success in using surrogate constraint formulations and their corresponding duals in solving integer programs. Only in rare cases would the dual directly produce a solution to (P), but the dual solution values may yield excellent bounds in a branch-and-bound procedure. The bounds would probably be more difficult to obtain because of the "knapsack" nature of the problems (P^v); but they might still be a powerful aid if they improve enough on lagrangian values.

In [5] Geoffrion gives an excellent treatment of many of the issues in attempting to apply lagrangian duality to integer programs. However, the only similar work for surrogate duality is the search algorithm presented in Banerjee [1] (to which we return to Section 5). In this paper we attempt to present a more complete parallel to Geoffrion's work by investigating a number of key computational issues affecting the relation between surrogate and lagrangian duality in integer programs.

The plan of this paper is as follows: a generalization of Geoffrion's integrality property is developed in Section 2 and shown to have implications for the relationships between the surrogate and lagrangian duals. In Section 3 we further compare the surrogate and lagrangian duals by studying the gaps between the two duals rather than the primal-to-dual gaps which are investigated elsewhere (see for example [8] and [9]). Empirical results on gaps between the duals are presented in Section 4. Section 5 deals with the problem of finding optimal surrogate multipliers. The subgradient search methods used in many lagrangian schemes are shown to be inappropriate in the surrogate case. A Benders' type procedure like Banerjee's is possible, however, and is shown to parallel the lagrangian case. Section 6 concerns placing the objective function in the set of surrogate constraints. It is shown that this is not a useful formulation of the surrogate dual.

2. Integrality properties

A key element of Geoffrion's [5] development of lagrangian duality in integer programming is the idea of an integrality property in the lagrangian relaxation. In order to conveniently present Geoffrion's ideas we shall extend the notation of Section 1. Let $P(T)$ be the version of P with the constraint $x \in S$ replaced by $x \in T$. The corresponding relaxation and dual problems are $P_u(T)$, $P^v(T)$, $D_L(T)$, and $D_S(T)$. Let $[T]$ represent the convex hull of the set T. Also, let a bar over the name of any constraint set represent the same set with all integrality requirements relaxed. Thus $P(\bar{S})$ is the linear programming relaxation of P. Finally, define the function $\Omega(\cdot)$ to represent the set of optimal solutions to the problem $(\cdot)$.

Integrality properties are relevant when taken in the context of a branch-and-bound algorithm in which we use a dual to bound the restricted version of P obtained when a number of variables have been fixed at specific values. In this context Geoffrion [5] shows that some lagrangian formulations of P possessing this property can just as well be replaced by $P(\bar{S})$, i.e., the additional complexity of obtaining a bound from the lagrangian dual leads to no improvement over the standard bound from the linear programming relaxation of (P). Formally:

Definition. The problem (P_u) has the *integrality property* if $\nu(P_u(S)) = \nu(P_u(\bar{S}))$ for all $u \geq 0$, i.e., if the lagrangian relaxation can be solved as a linear program for all u.

Letting $\bar{u}$ denote an optimal dual multiplier vector for the ordinary linear program $P(\bar{S})$, we can express Geoffrion's key results as follows:

Theorem 2.1. (i) $\nu(P(\bar{S})) = \nu(D_L(\bar{S})) \le \nu(P_{\bar{u}}(S)) \le \nu(D_L(S)) = \nu(D_L([S])) = \nu(P([S]))$.

(ii) *If the problem* (P_u) *has the integrality property, then* $\nu(P(\bar{S})) = \nu(P_{\bar{u}}(S)) = \nu(D_L(S))$.

Proof. See Geoffrion [4].

Theorem 2.1 has a number of important implications for developing lagrangian strategies for integer programming problems:

(a) $\nu(P(\bar{S})) = \nu(D_L(\bar{S}))$ implies that relaxing the constraint set S when solving the lagrangian dual precludes any potential bound improvement over the bound from the linear programming relaxation $(P(\bar{S}))$.

(b) $\nu(D_L(\bar{S})) \le \nu(P_{\bar{u}}(S)) \le \nu(D_L(S))$ implies that the optimal dual multiplier vector for $P(\bar{S})$, although not necessarily an optimal lagrange multiplier for $P_u(S)$, provides at least as good and possibly a better bound than $\nu(P(\bar{S}))$.

(c) The set S can be replaced by the convex hull of S, $[S]$, and no change in lagrangian dual solution value will occur.

(d) $\nu(D_L(S)) = \nu(P([S]))$, i.e., the value of the lagrangian dual may be obtained by solving the linear program formed when S is replaced in (P) by $[S]$.

Most important, part (ii) of Theorem 2.1 indicates that the lagrangian dual does not provide increased bounding power over the linear programming relaxation when the lagrangian relaxation has the integrality property. Of course, successful applications of lagrangian duality do occur even when the integrality property is present (see [5] and [8]). In these cases the appropriate linear programming problem is too large to solve directly. However, the integrality property for the lagrangian relaxation arises in many natural formulations of primal problems for which direct solution of $P(\bar{S})$ is common. We present only two here and refer the reader to Geoffrion [5] and Karwan [13] for others.

Example 1 (Geoffrion [5]). Consider the constraints of S to consist of only a restriction that some variables be integer and all satisfy integral lower and upper bounds, L_j and U_j. Clearly the integrality property is present in this case since $\nu(P_u(S)) = \min_{x \in S} cx + u(Ax - b)$ has a solution obtained by choosing $x_j = L_j$ if $(c + uA)_j > 0$ and $x_j = U_j$ if $(c + uA)_j \le 0$. The same solution is optimal when S is relaxed to constraints $L_j \le x_j \le U_j$, x_j continuous.

Example 2. Consider the following formulation of the integer multi-commodity minimum cost flow problem $(MCMC)$:

$$(MCMC) \quad \text{minimize} \quad \sum_{k=1}^{r} \sum_{(i,j) \in A} c_{ij}^k f_{ij}^k,$$

$$\text{subject to} \quad \sum_{j \in N} f_{ij}^k - \sum_{j \in N} f_{ji}^k = \begin{cases} v^k & i = s^k, \\ 0 & i \ne s^k, t^k, \\ -v^k & i = t^k, \end{cases} \quad \text{for all } i \in N \text{ and all } k,$$

$$\sum_{k=1}^{r} f_{ij}^k \le u_{ij} \qquad \text{for all } (i,j) \in A,$$

$$f_{ij}^k \ge 0 \qquad \text{for all } (i,j) \in A \text{ and all } k,$$

$$f_{ij}^k \text{ integer} \quad \text{for all } (i,j) \in A \text{ and all } k$$

where

N is the set of nodes of the network,
A is the set of arcs of the network,
f_{ij}^k is the flow of commodity k from node i to node j,
v_k is the total required flow of commodity k from s^k to t^k,
u_{ij} is the capacity of arc (i,j),
s^k and t^k are the source and sink, respectively, for commodity k.

Placing the capacity constraints, $\sum_{k=1}^r f_{ij}^k \le u_{ij}$ for all $(i,j) \in A$, in the objective function via a lagrange multiplier produces a lagrangian relaxation which separates into r single commodity flows. It is well known (see Ford and Fulkerson [4]) that the single commodity flow problem is unimodular. Thus, the integrality constraint is superfluous in the lagrangian relaxations, and the integrality property holds.

In the surrogate case we can also define an integrality property; $P^v(S)$ has the surrogate integrality property if $\nu(P^v(S)) = \nu(P^v(\bar{S}))$ for all $v \ge 0$. But we shall see that the resulting bounding weaknesses do not generally hold for surrogate formulations. The role of the integrality property in the surrogate case revolves around the relationship between $\nu(D_L(S))$ and $\nu(D_S(S))$. The following theorem summarizes our results.

Theorem 2.2. (i) $\nu(D_L(S)) = \nu(D_S([S]))$.
(ii) $\nu(P(\bar{S})) = \nu(D_L(\bar{S})) = \nu(D_S(\bar{S}))$.
(iii) *If $P^v(S)$ has the integrality property, then*

$$\nu(P(\bar{S})) = \nu(P_{\bar{u}}(S)) = \nu(D_L(S)) = \nu(D_S(S)).$$

Proof. (i)

$$\nu(D_S([S]) = \nu\left[\max_{v\ge 0} \min\{cx\colon v(Ax-b) \le 0, \quad x \in [S]\}\right].$$

The inner minimization problem is convex so it equals its lagrangian dual. Thus,

$$\nu(D_S([S])) = \nu\left[\max_{v\ge 0} \max_{\alpha \ge 0} \min\{cx + \alpha v(Ax-b)\colon \quad x \in [S]\}\right].$$

Since α and v appear only as a product inside the maximization, we can reduce the problem to one in $u = \alpha v$ giving us

$$\nu(D_S([S])) = \nu\left[\max_{u\ge 0} \min_{x \in [S]} \quad cx + u(Ax-b)\right] \triangleq \nu(D_L([S])).$$

By Theorem 2.1 the last quantity is equivalent to $\nu(D_L(S))$.

(ii) The proof is identical to part (i) with $[S]$ replaced by $(\bar{S})$.

(iii) By the integrality property of $P^v(S)$, $\nu(P^v(S)) = \nu(P^v(\bar{S}))$ for all $v \geq 0$ including any optimal v. Thus, $\nu(D_S(S)) = \nu(D_S(\bar{S}))$. Using this result and the definitions of various duals we have $\nu(P(\bar{S})) = \nu(D_L(\bar{S})) \leq \nu(P_{\bar{u}}(S)) \leq \nu(D_S(S)) = \nu(D_S(\bar{S}))$.

It then follows from part (ii) that $\nu(P(\bar{S})) = \nu(P_{\bar{u}}(S)) = \nu(D_L(S)) = \nu(D_S(S))$.

There are several implications of Theorem 2.2. First, note that if $P^v(S)$ has the integrality property, neither $D_L(S)$ nor $D_S(S)$ can improve on the bound from $P(\bar{S})$ (though, of course, they might provide a more efficient way to calculate $\nu(P(\bar{S}))$. Also, if we relax $P^v(S)$ and solve it as a linear program when it does not have the integrality property, we still only get $\nu(P(\bar{S}))$ as a dual value.

It seems obvious that it would be harder to find an example for which $P^v(S)$ has the integrality property than one where $P_u(S)$ has the property. Our next theorem will demonstrate that this observation is very much the case. First we note that all examples in the literature of the integrality property holding for a problem $P_u(S)$ are based on the structure of the constraint set $Ax \leq b$ and not on the cost vector c.

Theorem 2.3. *If S is finite and $\nu(P(\bar{S})) = \nu(D_S(S))$ for all cost vectors c, then all of the extreme points of $P(\bar{S})$ are contained in S.*

Proof. Choose an arbitrary extreme point, x^{LP}, of $P(\bar{S})$ and a cost vector c such that x^{LP} is the unique optimal solution to $P(\bar{S})$. Since x^{LP} is unique, we can perturb the cost function in any arbitrary direction d, and still keep x^{LP} unique, for a sufficiently small step α in the direction d.

By assumption $\nu(D_S(S)) = \nu(D_S(\bar{S})) = \nu(P(\bar{S}))$. So there must exist $x \in \Omega(D_S(S))$ such that $cx = cx^{LP}$, i.e., $x \in S$ which lie on the plane L defined by $cx = \nu(P(\bar{S}))$.

Since S is a finite set, there exists an $\alpha > 0$ such that $|cx - cx^{LP}| > \alpha$ for all $x \in S$ with $x \notin L$. That is the $x \in S$ which do not lie in the plane L, must be at least the positive distance α from L. Therefore, for a sufficiently small change αd in c, all such x will not be contained in the plane L' defined by $(c + \alpha d)x = (c + \alpha d)x^{LP}$. But by assumption $\nu(D_S(S)) = \nu(P(\bar{S}))$ for all c. Thus, there must exist $x \in S$ such that $x \in L$ *and* $x \in L'$.

Choose $d = e_1$, the first unit vector. Then for α_1 sufficiently small, there must exist $x \in S$ such that $x \in L$ and $x \in L_1$ where L_1 is the plane defined by $(c + \alpha_1 e_1)x = (c + \alpha_1 e_1)x^{LP}$. But $cx = cx^{LP}$ since $x \in L$, so we must have $\alpha_1 e_1 x = \alpha_1 e_1 x^{LP}$ which implies $x_1 = x_1^{LP}$. That is, the first component of x^{LP} is the same as the first component of an $x \in S$.

Now all $x \in S$ such that $x \notin L$ or $x \notin L_1$ are at some strictly positive distance from the plane L_1. Choose another direction $d = e_2$. Again, for $\alpha_2 > 0$, sufficiently small, there must exist $x \in S$ with $x \in L$ and $x \in L_1$ such that

$$(c + \alpha_1 e_1 + \alpha_2 e_2)x = (c + \alpha_1 e_1 + \alpha_2 e_2)x^{\mathrm{LP}}.$$

But $x \in L_1$ implies $x_2 = x_2^{\mathrm{LP}}$.

Now continue to iteratively construct new costs $c + \alpha_1 e_1 + \alpha_2 e_2 + \cdots + \alpha_j e_j$ such that there must exist an $x \in S$ with $x_i = x_i^{\mathrm{LP}}$, $i = 1, 2, \dots, j$. Letting $j = n$ implies that $x^{\mathrm{LP}} \in S$. But x^{LP} was chosen arbitrarily. So all of the extreme points of $P(\bar{S})$ are contained in S.

Theorem 2.3 says that if $\nu(P(\bar{S})) = \nu(D_{\mathrm{S}}(S))$ for all cost vectors c, i.e., for a class of problems defined by the structure of its constraints rather than its costs, then the linear programming relaxation $P(\bar{X})$ has all of its extreme points feasible to $P(S)$. Thus the original integer problem can be solved as a linear program. Since the integrality property is a sufficient condition for $\nu(P(\bar{S})) = \nu(D_{\mathrm{S}})$, we have the following corollary:

Corollary 2.4. *If $(P^v(S))$ has the integrality property for all c, then $\nu(P(S)) = \nu(P(\bar{S}))$.*

Proof. By the integrality property for all c and Theorem 2.2, $\nu(D_{\mathrm{S}}(S)) = \nu(P(\bar{S}))$ for all c. The conclusion now follows via Theorem 2.3.

Corollary 2.4 says that we need not worry much about the integrality property pertaining to the surrogate relaxation for most problems. If S is finite, then $\nu(D_{\mathrm{S}}(S)) = \nu(P(\bar{S}))$ as a consequence of constraint set properties only if $P(S)$ can be solved as a linear program. In general, this would happen only for very specially structured problems under a condition such as unimodularity. Neither the two examples presented above nor any of the examples of the lagrangian integrality property in [5] and [13] have the surrogate integrality property.

To further see the bounding advantage of $D_{\mathrm{S}}(S)$ note that one can generally improve on the $\nu(P(\bar{S}))$ bound for $\nu(P(S))$ by solving the one subproblem $P^{\bar{u}}(S)$ where $\bar{u}$ is obtained from the linear program $P(\bar{S})$. This case can occur even when the lagrangian relaxation has the integrality property. Fig. 1 presents a graphical example of just such a problem (Example 1 of the previous section).

Let

$$S = \{x : 0 \le x \le 1, x \text{ integer}\}.$$

The optimal lagrangian solution takes on the same value as the solution to $P(\bar{S})$, because $P_u(S)$ has the lagrangian integrality property, i.e., $\nu(D_{\mathrm{L}}(S)) = \nu(P(\bar{S})) = cx^0$. The plane $\bar{u}(Ax - b) = 0$ will pass through x^0 and be parallel to the cost contours of c. An optimal surrogate solution must occur at points x^1, x^2, or x^3, all lying in higher level sets of the objective function cx. Thus, $\nu(P^{\bar{u}}(S)) > \nu(P_{\bar{u}}(S)) = \nu(P(\bar{S}))$.

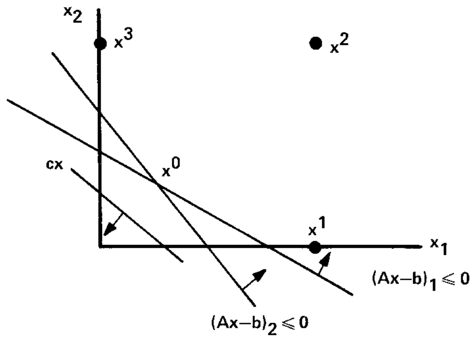

Fig. 1. Example of integrality property for (P_u), but not (P^u).

3. Gaps between lagrangian and surrogate duals

Greenberg and Pierskalla [10] show that

$$\nu(D_L) \leq \nu(D_S) \leq \nu(P).$$

Furthermore, they develop conditions for no gap to exist between D_S and P. We have shown in the previous section that one condition for no gap to exist between the surrogate and lagrangian duals is that P^v have the integrality property for all cost vectors c. However, Theorem 2.3 and Corollary 2.4 demonstrated that the integrality property will not occur in many problems of interest.

We concentrate here on further characterizations of when such a gap does or does not exist, i.e., when $\nu(D_L)$ = or $< \nu(D_S)$. Recall that the function $\Omega(\cdot)$ represents the set of optimal solutions to the problem $(\cdot)$. Consider first the following lemma.

Lemma 3.1. *If $\nu(D_L) = \nu(D_S)$ and $u \in \Omega(D_L)$, then $u \in \Omega(D_S)$ and $\Omega(P^u) = \{x \in \Omega(P_u)\colon u(Ax-b) = 0\}$.*

Proof. To show $u \in \Omega(D_S)$ we note that for all $x \in S$, $cx < \nu(D_L)$ implies $u(Ax-b) > 0$. Thus, all $x \in S$ with $cx < \nu(D_L)$ are infeasible in (P^u), and $\nu(P^u) \geq \nu(D_L)$. But then $\nu(D_L) \leq \nu(P^u) \leq \nu(D_S) = \nu(D_L)$ and u solves (D_S).

Now if $x \in \Omega(P^u)$, then $u(Ax-b) \leq 0$. But by $x \in S$ and $u \in \Omega(D_L)$ we have $\nu(D_L) \leq cx + u(Ax-b) \leq cx \leq \nu(D_S) = \nu(D_L)$. Thus, $x \in \Omega(P_u)$ and $u(Ax-b) = 0$. Conversely, if $x \in \Omega(P_u)$ and $u(Ax-b) = 0$, then $\nu(D_L) = cx = \nu(D_S)$ and $u(Ax-b) \leq 0$ so that $x \in \Omega(P^u)$.

The observations in Lemma 3.1 lead immediately to our next theorem.

Theorem 3.2. *Either $\nu(D_S) > \nu(D_L)$ or for every $u \in \Omega(D_L)$, there exist $x \in \Omega(P_u)$ such that $u(Ax-b) = 0$.*

Proof. Lemma 3.1 shows that if $\nu(D_S) = \nu(D_L)$, then the set $\Omega(P^u)$ is identical to $\{x \in \Omega(P_u): u(Ax-b)=0\}$. The theorem follows directly from the fact that $\Omega(P^u)$ is nonempty.

Theorem 3.2 states that the surrogate and lagrangian dual values can be equal only if there exist complementary x's for every optimal lagrange multiplier u, i.e., x's with $u(Ax-b)=0$. This result has some value in arguing the merits of the surrogate approach, since it seems unlikely in most integer programs that an exactly complementary x could be found for every optimal lagrangian multiplier u.

Also, recall that complementarity plays a key role in the theory of gaps between $\nu(D_L)$, $\nu(D_S)$ and $\nu(P)$. That is, if an optimal solution to (P_u) is feasible in (P) and satisfies $u(Ax-b)=0$, then it is optimal in (P). In the surrogate case the corresponding result does not require complementarity. Thus, it is not surprising that complementarity arises here as the key issue in gaps between $\nu(D_L)$ and $\nu(D_S)$. We now proceed to apply Theorem 3.2 in a constructive test to detect a gap between $\nu(D_L)$ and $\nu(D_S)$.

Theorem 3.3. *Let u be an optimal multiplier for (D_L) and define the (possibly empty) set $\Gamma = \{x \in \Omega(P_u): u(Ax-b)=0\}$. If we solve*

$$\begin{aligned} &\text{maximize} && \alpha, \\ &\text{subject to} && d(Ax-b) \geq \alpha \quad \text{for all } x \in \Gamma, \qquad d \geq 0 \end{aligned}$$

and the optimal $\alpha = 0$, then $\nu(D_L) = \nu(D_S)$. Moreover, if S is a finite set and $\alpha \neq 0$, $\nu(D_S) > \nu(D_L)$.

Proof. Clearly, $\alpha \geq 0$, and when Γ is empty, α is arbitrarily large which corresponds to the result of Theorem 3.2 that $\nu(D_S) > \nu(D_L)$. If Γ is non-empty, then from Lemma 3.1 we know that $\Omega(P^u) = \Gamma$. If u is not the optimal surrogate multiplier, then there must be $w \geq 0$, such that $w(Ax-b) > 0$ for every $x \in \Omega(P^u)$, i.e., such that all $x \in \Omega(P^u)$ are made infeasible in (P^w). If $\alpha = 0$ there is no such w and $\nu(D_L) = \nu(D_S)$. If $\alpha > 0$, then there exists a direction d such that $(u + \beta d)(Ax-b) > 0$ for all $x \in \Gamma$, for all $\beta > 0$. Moreover, by the optimality of u in (D_L), we know that $u(Ax-b) > 0$ for all $x \in S$ such that $cx < \nu(D_L)$. Thus, if S is finite, for the direction d chosen above, there must exist some $\beta > 0$ such that $(u+\beta d)(Ax-b) > 0$ for all $x \in S$ with $cx \leq \nu(D_L)$. This implies that we can improve on $\nu(D_L)$ in (D_S) by using $u + \beta d$ as a surrogate multiplier. That is to say, $\nu(D_L) = \nu(P^u) < \nu(P^{u+\beta d}) \leq \nu(D_S)$, and the theorem follows.

4. Some empirical results on gaps

Sections 3 and 4 have presented some theoretical characteristics leading to the supposition that $\nu(D_S)$ will often exceed $\nu(D_L)$, i.e., that there will be a gap between the duals. However, such a supposition can probably only be confirmed by empirical testing.

A limited set of such tests were part of the more complete computational study reported in [13] and [14]. The problems considered were randomly-generated 0–1 capital budgeting-type problems having $S = \{\text{binary } x\}$. Various combinations of the number of rows (m), the number of columns (n) and the density of nonzero entries in the constraint matrix (μ) were considered. All problems had nonnegative constraint matrices and possessed the lagrangian integrality property as in Example 1 above.

Gap results from those empirical tests are summarized in Table 1. Percents shown can be calculated as

$$100(\nu(D_S(S)) - (D_L(S)))/(\nu(P(S)) - \nu(D_L(S))).$$

Results were quite variable, but all 30 test problems had some gap between $\nu(D_L)$ and $\nu(D_S)$. Moreover, in many cases a substantial part of the $\nu(D_L)$ to $\nu(P)$ gap was closed by the surrogate dual, $\nu(D_S)$.

Table 1
Percent of lagrangian dual to primal gap closed by surrogate dual in empirical tests

	Mean percent and [range] by problem size		
Density	$m = 5, n = 10$	$m = 10, n = 20$	$m = 15, n = 30$
$\mu = 0.10$	74.5% [8.3% – 100%]	15.2% [3.7% – 45.5%]	5.9% [1.7% – 14.0%]
$\mu = 0.25$	51.7% [14.4% – 100%]	17.4% [10.5% – 28.1%]	6.6% [3.3% – 8.7%]

5 Problems solved per cell.

5. Finding optimal multipliers

With the exception of the Dantzig–Wolfe or Benders' type procedures, the dual multiplier search schemes in the lagrangian case largely depend on some subgradient-oriented direct search method (see for example the survey in [2]). Bazaraa and Goode [2] present necessary and sufficient conditions for ascent and steepest ascent directions based on the subgradients, $A\bar{x} - b$, at all $\bar{x}$ optimal solutions to the lagrangian relaxation P_u for the present set of multipliers u. We will develop similar necessary conditions for an ascent direction in the surrogate case but will also present a counterexample to any sort of subgradient search scheme. These necessary conditions were derived independently by Banerjee [1].

Theorem 5.1. *Let $\bar{u} \geq 0$ be a current surrogate multiplier. Any direction d for which there exists $\alpha \geq 0$ satisfying $\nu(P^{\bar{u}+\alpha d}) > \nu(P^{\bar{u}})$ (i.e., which is an ascent direction for the surrogate problem) must satisfy*

$$d(Ax - b) > 0 \quad \textit{for all } x \in \Omega(P^{\bar{u}}). \tag{1}$$

Proof. Let x^* be any element of $\Omega(P^{\bar{u}})$. By feasibility of $(P^{\bar{u}})$, $\bar{u}(Ax^* - b) \leq 0$. If $d(Ax^* - b) \leq 0$, then $(\bar{u} + \alpha d)(Ax^* - b) \leq 0$ for all $\alpha \geq 0$. It follows that x^* is feasible for $P(^{\bar{u}+\alpha d})$ and $\nu(P^{\bar{u}+\alpha d}) \leq cx^* = \nu(P^{\bar{u}})$. Thus d can be an ascent direction at $\bar{u}$ only if $d(Ax - b) > 0$ for all $x \in \Omega(P^{\bar{u}})$.

Corollary 5.2. *Any direction d for which there exists $\alpha \geq 0$ satisfying $\nu(P^{\bar{u}+\alpha d}) > \nu(P^{\bar{u}})$ is an ascent direction for the lagrangian problem*

$$(P_0^{\bar{u}}) \qquad \begin{array}{ll} \text{minimize} & cx + 0(Ax - b), \\ \text{subject to} & x \in S, \quad \bar{u}(Ax - b) \leq 0. \end{array}$$

Proof. $(P_0^{\bar{u}})$ is identical to $(P^{\bar{u}})$. Thus, $\Omega(P^{\bar{u}}) = \Omega(P_0^{\bar{u}})$ and the ascent directions for $(P_0^{\bar{u}})$ are exactly the ones satisfying (1).

Corollary 5.2 implies that (1) is not only a necessary condition but is also a sufficient condition for d to be an ascent direction in problem $(P_0^{\bar{u}})$. The steepest ascent direction for $(P_0^{\bar{u}})$ can be shown to be the shortest subgradient under an appropriate norm. We might hope that there is an ascent direction in $(P^{\bar{u}})$ corresponding to the shortest subgradient in $(P_0^{\bar{u}})$ with respect to some norm. However, in the following example, the subgradient is unique and is not an ascent direction for the surrogate dual, even though one exists. Consider the following problem:

$$\begin{array}{lrl} \text{minimize} & x_1 + x_2 - x_3, & \\ \text{subject to} & -x_1 + x_2 - 1 \leq 0, & \\ & -x_2 + 1 \leq 0, & \\ & x_1 + x_2 + 11x_3 - 10 \leq 0, & \\ & x \in S & \end{array}$$

where

$$S = \{x\colon 0 \leq x_i \leq 2, \text{ integer}, i = 1, 2, 3\}.$$

Let $\bar{u} = (5, 1, 20)$, $\bar{u}(Ax - b) = 15x_1 + 24x_2 + 220x_3 - 204$. Then $(P^{\bar{u}})$ implies $x^* = (0, 0, 0)$ with $Ax^* - b = (-1, 1, -10)$ unique. A superoptimal point $\hat{x} = (0, 0, 1)$ is cut off by $\bar{u}$ since $\bar{u}(A\hat{x} - b) = 16 > 0$. $Ax - b = (-1, 1, 1)$. Since x^* is unique, choose $d = Ax^* - b$ as our search direction.

(i) In order to cut off x^* and increase the value of the surrogate dual we must have $(\bar{u} + \alpha d)(Ax^* - b) > 0$. This implies we must have $\bar{u}(Ax^* - b) + \alpha d(Ax^* - b) > 0$ or $-204 + 102\alpha > 2$.

(ii) We must at the same time, however, keep $\hat{x}$ infeasible. Thus, we must have $(\bar{u}+\alpha d)(A\hat{x}-b)>0$ or $16-8\alpha>0$ which implies $\alpha<2$. It follows that Ax^*-b is not an ascent direction. But, $d=(-5,11,-19)$ is an ascent direction. With $\alpha=1$, we get our new $u=(5,1,20)+(-5,11,-19)=(0,12,1)$. $u(Ax-b)=x_1-11x_2+11x_3+2\le 0$. The optimal solution to (P^u) has a solution value of one and is given by $x=(0,1,0)$ and $(0,2,1)$.

The above example shows that the subgradient does not necessarily provide a direction of ascent in surrogate multiplier space. We can, however, construct an algorithm, derived independently by Banerjee [1], for the surrogate dual which parallels the Benders' style procedure in [2] for the lagrangian case. Benders' procedure for the lagrangian dual may be summarized as follows:

Step 0. Let $k=1$, $u_k=0$.

Step 1. Solve P_{u_k} with optimal solution x_k and $y_k=Ax_k-b$. If $y_k\le 0$ and $u_k y_k=0$, then stop, x_k is optimal to the primal problem. Otherwise, go to Step 2.

Step 2. Solve the linear program

$$\begin{aligned} &\text{maximize} && z, \\ &\text{subject to} && cx_j+uy_j\ge z, \quad j=1,2,\dots,k, \\ & && u\ge 0. \end{aligned}$$

If $z=\nu(P_{u_k})$, then (u_k,x_k) is an optimal solution to D_L. Otherwise, denote the optimal by u_{k+1} and replace k by $k+1$. Go to Step 1.

The necessary conditions given in Theorem 4.2 to improve the surrogate ralaxation (P^u) lead to a similar algorithm for the surrogate dual as follows:

Step 0. Set $k=1$, $u_k=u_k^*=0$, $\alpha^*=$ value of an incumbent solution $=-\infty$.

Step 1. Solve P^{u_k} with optimal solution

$$x_k \quad \text{and} \quad y_k=Ax_k-b.$$

If $y_k\le 0$, then stop, x_k is optimal to the primal problem. If $cx_k>\alpha^*$, let $\alpha^*=cx_k$, $u_k^*=u_k$.

Step 2. Solve the linear program

$$\begin{aligned} &\text{maximize} && z, \\ &\text{subject to} && uy_j\ge z \quad j=1,2,\dots,k, \\ &1u\le 1, \quad u\ge 0 \end{aligned}$$

where $1u\le 1$ simply normalizes the u's since $(P^u)=(P^{\beta u})$ for all $\beta>0$. If the maximal $z\le 0$, stop. (u_k^*,x_k^*,α^*) is optimal for D_S. Otherwise, denote the optimal u by u_{k+1} and replace k by $k+1$. Go to Step 1.

Clearly, if the set S is a finite discrete set then we have finite convergence since every time Step 2 is successful ($z>0$) we have cut off another member of S from the set of feasible solutions to P^{u_k}. If $z\le 0$, we stop and an incumbent solution is optimal since there exists no multipliers u_{k+1} which will cut off all points $x_1,x_2,\dots,x_k$ previously generated. We need to keep an incumbent solution since the multiplier u_k does not necessarily increase $\nu(P^{u_k})$ monotonically. In other words, x_k satisfying $u_k(Ax_k - b) \leq 0$ may be infeasible in the problem $(P^{u_{k-1}})$. However, if $cx_k < \nu(P^{u_{k-1}})$, then we will have $\nu(P^{u_k}) < \nu(P^{u_{k-1}})$. Keeping an incumbent solution is not necessary in the lagrangian case unless we plan to stop before reaching optimality by terminating at the iteration k where $z - \nu(P_{u_k}) < \epsilon$ for some specified $\epsilon > 0$.

Refinements and empirical results for this algorithm are reported in [13] and [14] for a set of randomly generated 0–1 integer programming test problems. Our intent here is to note the potential advantages due to the similarity of the two algorithms. The similarity of these two algorithms can be exploited in conjunction with the lagrangian-surrogate gap detection test in Theorem 3.3. These results suggest a two phase, lagrangian/surrogate algorithm which proceeds as follows:

Phase *I*: (a) solve (D_L) for the optimal multiplier u.

(b) Identity all complementary solutions to (P_u), i.e., $x \in \Omega(P_u)$ with $u(Ax - b) = 0$.

Phase *II*: (a) Solve the linear program in Theorem 3.3 to determine if we should undertake a surrogate dual optimization, beginning with u or $u + \beta d$ as a starting surrogate multiplier.

(b) If the optimal α in Theorem 3.3 is positive, attempt to improve further on the surrogate multiplier.

At the termination of the lagrangian Benders' procedure we have available to us the optimal lagrangian multiplier u^* which we can use in Step 1 of our surrogate algorithm. Since $\nu(P^u) \geq \nu(P_u)$ for all $u \geq 0$ we might go right into Phase II of our two phase procedure. We could also begin Step 2 of our surrogate algorithm with the set of x_j's such that $cx_j \leq \nu(D_L)$ generated in the lagrangian procedure. This would insure immediate improvement on $\nu(D_L)$ if any gap between $\nu(D_L)$ and $\nu(D_S)$ exists. Further exploitation of the obvious similarities between the two algorithms should lead to a significant savings in storage and computation time for any computer program for implementing the two phase approach. Even if the full procedure were not used, then in the context of a branch-and-bound procedure in which a lagrangian dual has been used successfully it may prove beneficial to improve the bound on any branch by continuing with the surrogate algorithm for at least a few iterations.

Finally, note that if (P^u) has the integrality property, then (P_u) can be solved as a linear program and complementary $x \in \Omega(P_u)$ which cannot be "cut off" in (P^u) are assured. Our previous discussion concluded that while integer programs of interest which have the *surrogate* integrality problem are probably uncommon, several of the most natural *lagrangian* formulations of integer programs do have the integrality property. The two phase approach above seems well suited to such cases. If the lagrangian relaxation can be seen to have the integrality property, then to find the optimal u we need only to solve the linear programming problem $P(\bar{S})$. We can proceed immediately to Phase II in the above procedure.

6. Including cost in the surrogate

Geoffrion [5] shows the value of including $\hat{z}$, the objective function value associated with the best known feasible solution to (P), in a lagrangian formulation. Including the cost in the surrogate formulations is done by adding $cx - \hat{z}$ into the surrogate linear combination of the constraints $Ax \leq b$. The following theorem will show that the optimal surrogate multiplier on $(cx - \hat{z})$ is zero for any given value of $\hat{z}$ which is an upper bound on $\nu(P)$. Hence, there is no need to include the cost in the surrogate constraints.

Theorem 6.1. *Consider the surrogate relaxation*

$$(P^{u,w}) \qquad \text{minimize} \quad cx,$$
$$\text{subject to} \quad u(Ax - b) + w(cx - \hat{z}) \leq 0, \quad x \in S$$

where $z \geq \nu(P)$. *Then* $w = 0$ *in some optimal solution to*

$$(D_S) \qquad \max_{u \geq 0, w \geq 0} \nu(P^{u,w}).$$

Proof. For any $u \geq 0$ and $w \geq 0$, $x \in \Omega(P^{u,w})$ implies $cx \leq \nu(P)$ and thus $cx \leq \hat{z}$. Thus, for such x, $w(cx - \hat{z}) \leq 0$. It follows that any $x \in \Omega(P^{u,0})$ is feasible in $(P^{u,w})$ and so $\nu(P^{u,0}) \geq \nu(P^{u,w})$. Thus, an optimal solution to (D_S) must exist which has $w = 0$.

References

[1] K. Banerjee, "Generalized lagrange multipliers in dynamic programming", Research Report ORC 71-12, Operations Research Center, University of California, Berkeley (1971).

[2] M.S. Bazaraa and J.J. Goode, "A survey of various tactics for generating lagrangian multipliers in the context of lagrangian duality", unpublished research report in the School of Industrial and Systems Engineering, Georgia Institute of Technology (1974).

[3] M.L. Fisher, "Optimal solution of scheduling problems using generalized lagrange multipliers: Part I", *Operations Research* 21 (1973) 1114–1127.

[4] L.R. Ford and D.R. Fulkerson, *Flows in networks* (Princeton University Press, Princeton, NJ, 1962).

[5] A.M. Geoffrion, "Lagrangian relaxation and its uses in integer programming", *Mathematical Programming Study* 2 (1974) 82–114.

[6] A.M. Geoffrion, "The capacitated plant location problem with additional constraints", Paper presented to the Joint National Meeting of AIIE, ORSA, and TIMS, Atlantic City, NY, November (1972).

[7] F. Glover, "Surrogate constraint duality in mathematical programming", *Operations Research* 23 (3) (1975) 434–451.

[8] H.J. Greenberg, "Bounding nonconvex programs with conjugates", *Operations Research* 21 (1973) 346–347.

[9] H.J. Greenberg, "The generalized penalty function surrogate model", *Operations Research* 21 (1973) 162–178.

[10] H.J. Greenberg and W.P. Pierskalla, "Surrogate mathematical programming", *Operations Research* 18 (1970) 1138–1162.

[11] H.J. Greenberg and W.P. Pierskalla, "Quasi-conjugate functions and surrogate duality", *Cahiers du Centre d'Études de Recherche Opérationnelle* 15 (1973) 437–448.

[12] M. Held and R.M. Karp, "The traveling salesman problem and minimum spanning trees", *Operations Research* 18 (1970) 1138–1162.

[13] M.H. Karwan, "Surrogate constraint duality and extensions in integer programming", Dissertation, Georgia Institute of Technology (Atlanta 1976).

[14] M.H. Karwan and R.L. Rardin, "Surrogate dual multiplier search procedures in integer programming", School of Industrial and Systems Engineering, Report Series No. J-77-13, Georgia Institute of Technology (1976).