OPERATIONS RESEARCH CENTER

working paper

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

Constructive Duality
in Integer Programming†

by

Marshall L. Fisher* and Jeremy F. Shapiro**

OR 008-72

April 1972

*Graduate School of Business, University of Chicago.

**Sloan School of Management, M.I.T.

†This research was supported in part by the U.S. Army Research Office (Durham) under Contract No. DA HC04-70-C-0058.

Constructive Duality in Integer Programming

by

Marshall L. Fisher* and Jeremy F. Shapiro**

1. Introduction

There are several interrelated ideas upon which this paper is based. The conceptual starting point is the cutting plane method for IP of Gomory which has long been called a "dual method" (e.g., Balinski [3; p. 254]) without specific mathematical justification. Recent research (Shapiro [27]) has led to a procedure for generating strong cuts using group theory and generalized Lagrange multipliers. We show in section 2 that this procedure is equivalent to solving a concave programming problem that is dual to a given IP problem. The strongest cuts are those written with respect to optimal dual variables.

With this perspective, several other structural and algorithmic ideas come into clearer focus. Many of these ideas are derived from mathematical programming duality theory which is applicable to any optimization problem defined over a finite dimensional vector space: that is, to any problem

$$\begin{aligned} v(b) &= \inf f(x) \\ \text{s.t.}\quad & g_i(x) \le b_i \qquad i = 1,\ldots,m \\ & x \in X \subseteq R^n, \end{aligned} \tag{1.1}$$

where the functions f and g_i and the set X are arbitrary. This duality theory has been studied by Geoffrion [11], Gould [20], Lasdon [23], Rockafellar [25], and others, for nonlinear programming which means problems where f and g_i are continuous (or at least lower semi-continuous) functions

* Graduate School of Business, University of Chicago.
** Sloan School of Management, M.I.T.

defined on a closed convex set X, and usually f and g_i are assumed to be convex. One of our purposes here is to demonstrate that this duality theory can also be used in the solution of IP problems.

It is also important, however, to recognize the differences between duality theory for convex nonlinear programming and that for integer programming. We need some constructs to discuss these differences. The first is the Lagrangian function

$$L(u) = -\sum_{i=1}^{m} u_i b_i + \inf_{x \in X}\{f(x) + \sum_{i=1}^{m} u_i g_i(x)\} \qquad (1.2)$$

defined for any $u \geq 0$. It is well known and easily shown that $L(u) \leq v(b)$ for any $u \geq 0$, and this naturally leads to the dual problem

$$w(b) = \sup_{u \geq 0} L(u). \qquad (1.3)$$

Moreover, L(u) is concave and thus (1.3) is a concave programming problem.

Clearly, $w(b) \leq v(b)$ and much of the theory of convex nonlinear programming is focused on sufficient conditions that $w(b) = v(b)$; i.e., sufficient conditions that there be no duality gap. If this is so, then any optimal solution u^* of the dual (1.3) can be used to find an optimal solution to the primal (1.1) by considering all the solutions $\tilde{x} \in X$ which satisfy $L(u^*) = -\sum_{i=1}^{m} u_i^* b_i + f(\tilde{x}) + \sum_{i=1}^{m} u_i^* g(\tilde{x})$. An optimal solution x^* to the primal also satisfies $g_i(x^*) \leq b_i$, $i = 1,\ldots,m$, and $\sum_{i=1}^{m} u_i^* g_i(x^*) = 0$ (see Geoffrion, [11; p. 9]).

In IP, duality gaps generally exist and can only be eliminated by computational effort beyond the construction and solution of a single dual problem. One means of filling in a duality gap in IP is the cutting plane method, which as we mentioned above, must use duality theory to find strong cuts. A more general algorithmic strategy, however, is branch and bound which uses cutting planes as only one tactic.

Another important duality construct which varies between different mathematical programming problems is the Lagrangean subgradient. Subgradients of the Lagrangean at a point u indicate directions of increase of the dual objective function, and they are used in ascent algorithms. In differentiable nonlinear programming, the subgradient is unique and equals the gradient of the Lagrangian. In IP, subgradients are derived from the solution of a group shortest route problem and in general there is more than one subgradient. Although they do not explicitly mention it, Held and Karp [21], [22] use an entirely analogous approach to the one here in constructing an algorithm for the traveling salesman problem. The Lagrangean subgradients for the traveling salesman dual problem are derived from the solution of minimum spanning tree problems. Fisher [6], [7] has applied this approach in an algorithm for resource-constrained network scheduling problems. The Lagrangean subgradients are derived from the solution of dynamic programming subproblems defined for each job to be scheduled.

The plan of this paper is the following. In section 2 we construct IP dual problems and discuss their properties. Section 3 contains a development of ascent methods for solving IP dual problems. Ascent methods for these problems are preferred for branch and bound because they produce monotonically increasing lower bounds. One of the main results of this paper is an adaption of the primal-dual simplex algorithm for use with ascent methods. In addition to enabling jamming to be overcome, the use of the primal-dual algorithm gives new insights into dual methods and the use of

Lagrangean subgradients. The use of IP dual problems in branch and bound is discussed briefly in section 4. It is important to emphasize that the algorithmic methods being developed here should greatly improve the group theoretic IP code discussed in [16], [17], [18]. Section 5 contains a few concluding remarks.

2. Construction of Dual Problems and Their Properties

The IP problem is written in initial form as

$$\begin{aligned} &\min \bar{c}w \\ &\text{s.t. } \bar{A}w \geq \bar{b} \\ &\qquad w \geq 0 \text{ and integer} \end{aligned} \tag{2.1}$$

where $\bar{c}$ is a (1xn) vector of integers, $\bar{A}$ is an m x n matrix of integers, $\bar{b}$ is an (mx1) vector of integers. We assume that the minimal objective function value in (2.1) is finite.

Let $\bar{B}$ by any non-singular m x m matrix made up of m columns of $(\bar{A},-I)$; assume $(\bar{A},-I)$ is partitioned as $(\bar{N},\bar{B})$. Partition $\bar{c}$ as $(\bar{c}_N,\bar{c}_B)$ and w as (x,y). An equivalent formulation of (2.1) is (e.g., see Shapiro [27]):

$$\begin{aligned} v(b) = z_B &+ \min cx \\ \text{s.t. } &Nx - b \leq 0 \\ &\sum_{j=1}^{n} \alpha_j x_j \equiv \beta (\text{mod } q) \\ &x \text{ non-negative integer,} \end{aligned} \tag{2.2}$$

where $z_B = \bar{c}_B\bar{B}^{-1}\bar{b}$, $\bar{c} = \bar{c}_N - \bar{c}_B\bar{B}^{-1}\bar{N}$, $N = \bar{B}^{-1}\bar{N}$, $b = \bar{B}^{-1}\bar{b}$. The vector

$q = (q_1,\ldots,q_r)$ where the elements q_i, $i = 1,\ldots,r$ are integers, $q_i \geq 2$, $q_i | q_{i+1}$, $\prod_{i=1}^{r} q_i = |\det \bar{B}| = D$, each $\alpha_j = (\alpha_{1j},\ldots,\alpha_{rj})$, α_{ij} integer, $0 \leq \alpha_{ij} \leq q_i - 1$, and $\beta = (\beta_1,\ldots,\beta_r)$, β_i integer, $0 \leq \beta_i \leq q_i - 1$. It is easy to see that the sums $\sum_{j=1}^{n} \alpha_j x_j$ for all possible integer values of the x_j can generate at most D distinct r-tuples of the form $\lambda_k = (\lambda_{1k},\ldots,\lambda_{rk})$, λ_{ik} integer, $0 \leq \lambda_{ik} \leq q_i - 1$. Let $G = \{\lambda_k\}_{k=0}^{D-1}$ be this collection of r-tuples with $\lambda_0 = (0,\ldots,0)$. The set G forms a finite abelian group under addition modulo q. For more discussion about G and its derivation, see Wolsey [31]. For future reference, we write out the linear programming (LP) relaxation of (2.1) relative to the basis $\bar{B}$.

$$
\begin{aligned}
\ell(b) = z_B &+ \min cx \\
\text{s.t. } Nx &- b \leq 0 \qquad\qquad (2.3) \\
x &\geq 0 .
\end{aligned}
$$

We assume $\ell(b)$ is finite.

Before we continue our main development, there are two tangential points to be mentioned. First, the perturbation function v(b) is somewhat different mathematically from the usual nonlinear programming perturbation function (e.g. Geoffrion [11; p. 6]) because of the different structure of the IP problem in the forms (2.1) and (2.2). These differences are discussed and briefly analyzed in Appendix A. The second point about the IP problem (2.1) is the absence of explicit upper bounds on the variables. We have stated the IP problem in this way for expositional convenience and the results here can be readily generalized to exploit upper bounds.

Define the set

$$X = \{x | \sum_{j=1}^{n} \alpha_j x_j \equiv \beta (\text{mod } q),\ x \text{ non-negative integer}\}.$$

We assume X is not empty; otherwise (2.1) is infeasible. For $u \geq 0$, define

$$L(u, x) = - ub + (c+uN)x,$$

and (2.4)

$$L(u) = - ub + \inf_{x \varepsilon X} (c+uN)x = \inf_{x \varepsilon X} L(u, x).$$

Problem (2.4) is a shortest route problem in the group network which has (a) nodes λ_k, $k = 0,1,\ldots,D-1$; (b) arcs $(\lambda_k - \alpha_j,\ \lambda_k)$ with arc costs $c_j + ua_j$, $j = 1,\ldots,n$, $k = 0,1,\ldots,D-1$ (see Gorry and Shapiro [16]). Let $G(\lambda_k;\ u)$ denote the cost of a shortest route from $\lambda_0 \equiv 0$ to λ_k; then $L(u) = - ub + G(\beta;\ u)$.

The cyclic nature of the group network implies that $L(u) = -\infty$ if any $c_j + ua_j < 0$, whereas $L(u)$ is attained if $c + uN \geq 0$. In the latter case, the search for x optimal in (2.4) can be limited without loss of optimality to the finite set $\{x^t\}_{t=1}^{T} \subset X$ satisfying $\prod_{j=1}^{n} (x_j^t + 1) \leq D$ (see reference [15]) in computing $L(u)$; namely

$$L(u) = - ub + \min_{t=1,\ldots,T} (c+uN)x^t. \qquad (2.5)$$

These solutions x^t are called <u>irreducible</u>.

As mentioned in the introduction, $L(u)$ is a concave function, and it is piecewise linear and continuous on the set $c + uN \geq 0$. Moreover, for $u \geq 0$, $z_B + L(u) \leq v(b)$. Thus, the IP dual problem relative to (2.2) is to find the best lower bound, or find

$$w_B(b) = z_B + \max L(u) \tag{2.6}$$

$$\text{s.t. } c + uN \geq 0$$

$$u \geq 0.$$

Note that (2.6) has a feasible solution because LP (2.3) has an optimal solution. The definition (2.5) of $L(u)$ for u in the set $c + uN \geq 0$, $u \geq 0$ allows us to immediately write the following LP formulation of the dual

$$w_B(b) = z_B + \max w$$

$$\text{s.t. } w \leq -ub + (c+uN)x^t, \; t = 1,\ldots,T \tag{2.7}$$

$$c + uN \geq 0$$

$$u \geq 0.$$

But (2.7) is precisely the dual of the master LP(8) on page 72 of [27] with columns corresponding to all the x^t included. The procedure in [27] is focused on the generation of a strong Gomory cut from (2.3), and thus that procedure is equivalent to solving the dual problem (2.6). Appendix B contains a necessary and sufficient condition that $w_B(b) < +\infty$.

For $u^0 \geq 0$ satisfying $c + u^0N \geq 0$ (e.g. u^0 optimal in (2.6) or (2.7)), a valid cut is

$$(c+u^0N)x \geq L(u^0) + u^0b \tag{2.8}$$

The inequality (2.8) is a valid cut because it holds for all $x \in X$ by the definition of $L(u^0)$, and X contains all feasible solutions to the given IP problem in the form (2.2).

It is possible, moreover, to strengthen the cut (2.8). In particular, a cut $\sum_{j=1}^{n} t_j x_j \geq t_0$ dominates the cut $\sum_{j=1}^{n} t_j' x_j \geq t_0$ if $t_j \leq t_j'$. This is because any non-negative x satisfying the former cut will satisfy the latter cut and thus the solution set admitted by the former cut is smaller than the solution set admitted by the latter cut.

Lemma 2.1. The cut (2.8) can be strengthened to the cut

$$\sum_{j=1}^{n} g_j^0 x_j \geq L(u^0) + u^0 b, \tag{2.9}$$

where

$$g_j^0 = G(\alpha_j;\ u^0).$$

Proof: The new cut is stronger because for any a_j, $G(\alpha_j;\ u^0) \leq c_j + u^0 a_j$. To see that the new cut is valid, i.e., that it does not cut off any feasible x , suppose we solve a shortest route problem from 0 to β in the group network with arc costs $G(\alpha_j;\ u^0)$ on the arcs $(\lambda_k - \alpha_j,\ \lambda_k)$. This problem will clearly yield the same value $G(\lambda_k;\ u^0)$ as the cost of the shortest paths to each λ_k. Thus $\sum_{j=1}^{n} g_j^0 x_j \geq G(\beta;\ u^0) = L(u^0) + u^0 b$ is a valid inequality for all $x \in X$, and the lemma is proven.

Consider now the augmented LP derived from (2.3) by adding the cut

$$\ell_1(b) = z_B + \min cx$$

$$\text{s.t. } Nx \leq b \qquad (2.10)$$

$$\sum_{j=1}^{n} g_j^0 x_j \geq L(u^0) + u^0 b$$

$$x \geq 0$$

Lemma 2.2. The minimal objective function value in (2.10) satisfies

$$v(b) \geq \ell_1(b) \geq z_B + L(u^0)$$

Proof: The dual of problem (2.10) is

$$d_1(b) = z_B + \max - vb + v_{m+1}(L(u^0)+u^0 b)$$

$$\text{s.t. } c_j + va_j - v_{m+1} g_j^0 \geq 0$$

$$v \geq 0,\ v_{m+1} \geq 0$$

The solution $v = u^0$, $v_{m+1} = 1$ is feasible in this problem. Thus, $d_1(b) \geq z_B + L(u^0)$, and by duality theory of LP, $\ell_1(b) \geq d_1(b)$ which establishes the right hand inequality. The left hand inequality follows directly from the property of a cut that it admits all feasible integer solutions.

It is also important to compare the lower bounds provided by the IP dual problem (2.6) for any basis B with the lower bound provided by the LP (2.3).

Lemma 2.3. Assume $w_B(b) < +\infty$. The optimal value $w_B(b)$ of the IP dual problem (2.6) satisfies

$$\ell(b) \leq w_B(b) \leq \ell(b) + \min_{t=1,\ldots,T} (c+u^*N)x^t$$

where $\ell(b)$ is the minimal cost of the LP (2.3) and u^* is any vector of optimal multipliers in (2.6).

Proof: Recall that we assume $\ell(b)$ to be finite. To prove $\ell(b) \leq w_B(b)$, note that by LP duality theory, we have

$$\ell(\mathbf{b}) = z_B + \max w$$

$$\text{s.t. } w \leq -ub$$

$$c + uN \geq 0$$

$$u \geq 0.$$

Consider any $\tilde{u}$, $\tilde{w}$ which is feasible in this last problem. Then $\tilde{u}$, $\tilde{w}$ is feasible in (2.7), the LP formulation of the IP dual problem (2.6), because $c + \tilde{u}N \geq 0$ and $x^t \geq 0$ implying $\tilde{w} \leq -\tilde{u}b + (c+\tilde{u}N)x^t$. Thus the maximal value of w in (2.7) is at least as great as the maximal value of w in the LP dual above, and the right hand inequality is established.

To prove $w_B(b) \leq \ell(b) + \min_{t=1,\ldots,T} (c+u^*N)x^t$, we use the fact that u^* optimal in (2.6) means $w_B(b) = z_B + -u^*b + \min_{t=1,\ldots,T} (c+u^*N)x^t$. The solution u^*, $w^* = -u^*b$ is feasible in the LP dual above and therefore $-u^*b \leq \ell(b) - z_B$. Adding $z_B + \min_{t=1,\ldots,T} (c+u^*N)x^t$ to both sides of this inequality gives the desired result.

Notice that for B an optimal LP basis, we have $\ell(b) = z_B$ and lemma 2.3 becomes $0 \le w_B(b) - z_B = u^*b + \min_{t=1,\dots,T} (c+u^*N)x^t \le \min_{t=1,\dots,T} (c+u^*N)x^t$.

Computational experience [16] has indicated that when B is optimal, $\frac{w_B(b)-\ell(b)}{v(b)-\ell(b)}$ is on the average a fairly high percentage. Thus, the duality theory we are proposing here gives uniformly better lower bounds than the duality theory based solely on LP (see Balas [2] and Nemhauser and Ullman [24]).

On the other hand, numerical examples have shown that it is definitely possible for $w_B(b) = \ell(b)$, particularly when B is not optimal. Lemma 2.3 says that a sufficient condition for recognizing that this has occurred after solving (2.6) and obtaining an optimal u* is $\min_{t=1,\dots,T} (c+u^*N)x^t = 0$.

Another construct used in IP is the surrogate constraint (e.g. Geoffrion [8]) and we want to demonstrate that the duality theory of this paper can be used to generate strong surrogate constraints. Let $\hat{z}$ be the cost of the best known solution to the IP problem (2.3), and let u be an arbitrary non-negative m-vector. If the feasible solution x to (2.2) satisfies $z_B + cx < \hat{z}$ then it must also satisfy

$$z_B + cx + u(Nx-b) - \hat{z} < 0 \tag{2.11}$$

since $u \ge 0$ and $Nx \le b$. The inequality (2.11) is a surrogate constraint, and on page 441 of reference [8], Geoffrion defines a strongest surrogate constraint as one for which the minimum of the left side over $x \in X$ is a maximum. In other words, we seek a u* such that

$$- u^*b + (c+u^*N)x = \max_{u \ge 0} \min_{x \in X} - ub + (c+uN)x \tag{2.12}$$

which is simply the requirement that u* be optimal in the IP dual problem.

In [8], Geoffrion uses LP duality theory to generate a strongest surrogate constraint for the zero-one integer programming problem. Although we do give the details here, it can easily be shown that the duality theory here would provide a still stronger surrogate constraint for Geoffrion's problem, although at a higher computational cost.

The next section is concerned with primal algorithmic methods for solving the IP dual problem (2.6). Before concluding this section, however, it is important to discuss briefly the many possible problem manipulations of the IP problem in the form (2.2) which give valid dual problems; i.e. problems which provide lower bounds. These manipulations can be classified as dualization outer linearization/relaxation by the taxonomy of Geoffrion [9 ; p. 656].

One manipulation of (2.2) is to omit some of the inequality constraints from Nx - b before constricting the dual. A second relaxation manipulation is to replace the system of congruences in (2.2) by a new system

$$\sum_{j=1}^{n} \{\psi_i(\alpha_{ij})\}x_j \equiv \psi_i(\beta_i)(\text{mod } q_i), \quad i=1,\ldots,r.$$

where each ψ_i is an endomorphism on Z_{q_i}, the cyclic group of order q_i. This relaxation as well as the next two are designed to control the size of the group G. The third type of relaxation of (2.2) results from an alternative reformulation of (2.1) given in section 4 of reference 18. This reformulation involves a change in the data and the number of inequality constraints and a subsequent change in the system of congruences. The reader is referred to reference 18 for more details.

The final relaxation manipulation results if we allow some of the columns of $\bar{B}$ to be activities $\bar{a}_j$ not included in the set $(\bar{A}, -I)$. In this case, we require the rows in the system $Nx - b$ corresponding to these activities to equal zero. The dual variables on these rows are then unconstrained in sign.

3. Solution of Dual IP Problems

We saw in the previous section that the dual IP problem (2.6) induced from a given IP problem (2.1) is equivalent to the linear programming problem (2.7) with many rows. One method for solving (2.6) is the generalized programming algorithm [27] applied to the LP dual of problem (2.7). This algorithm has two drawbacks.

The first drawback is that the multiplier vectors u produced by successive iterations of the generalized programming algorithm do not yield monotonically increasing lower bounds $z_B + L(u)$ to $v(b)$. Monotonically increasing lower bounds are definitely preferred for the branch and bound algorithm outlined in the next section. A second drawback of the generalized programming algorithm is its poor convergence characteristics (e.g. Held and Karp [21; p. 1146]).

Thus, our object in devising algorithm for the IP dual problem (2.6) is to construct primal algorithms for that problem, preferably ones that provide good approximate solutions quickly. Specifically, we propose an ascent algorithm to be followed by, or combined with, a primal-dual simplex algorithm. Our reasons for choosing the ascent algorithm as an opening strategy for solving (2.6) are two fold. First, it is relatively easy and efficient to use and it may produce rapid increases in the lower bounds $z_B + L(u)$.

Second, a similar ascent algorithm has been used with success in the computer code of Fisher [7] for scheduling problems and Held and Karp [22] for the traveling salesman problem. The discussion of the ascent and primal-dual simplex algorithms will be focused on the unique features of these algorithms as applied to (2.6), and the usual features will not be developed in detail.

We begin with same notation. Let

$$U_B = \{u \mid c+uN \geq 0,\ u \geq 0\}. \tag{3.1}$$

An m-vector $u \in U_B$ is said to be dual feasible. An ascent algorithm generates a sequence of points $u^k \in U_B$ given by the usual rule

$$u^{k+1} = u^k + \theta_k d^k \tag{3.2}$$

where the m-vector d^k is a direction of (possible) increase of L at u^k, and the non-negative scalar θ_k is the step length to be moved in the direction d^k. For notational simplicity, we describe one iteration of the algorithm starting at $\bar{u}$ and proceeding to the point $\bar{\bar{u}} = \bar{u} + \bar{\theta}\bar{d}$. Although L is not differentiable everywhere on U_B, directions of increase are implied by the subgradients of L.

Definition 3.1: Let $\bar{u}$ be a point satisfying $c + \bar{u}N \geq 0$, $\bar{u} \geq 0$. The vector $\bar{\gamma}$ is a subgradient of the Lagrangian L at $\bar{u}$ if

$$L(u) \leq L(\bar{u}) + (u-\bar{u})\cdot\bar{\gamma} \quad \text{for all } u.$$

The following lemma is a well known result and it is presented here for completeness.

Lemma 3.1: For $\bar{u} \in U_B$, let $\bar{x} \in X$ be a solution satisfying

$$L(\bar{u}) = -\bar{u}b + (c + \bar{u}N)\bar{x} ;$$

then

$$\bar{\gamma} = N\bar{x} - b$$

is a subgradient of L at $\bar{u}$.

Proof: For any u,

$$L(u) \leq -ub + (c + uN)\bar{x},$$

and substituting $c\bar{x} = L(\bar{u}) + \bar{u}b - \bar{u}N\bar{x}$, there results

$$L(u) \leq -ub + L(\bar{u}) + \bar{u}b - \bar{u}N\bar{x} + uN\bar{x}.$$

Rearranging this last inequality gives

$$L(u) \leq L(\bar{u}) + (u-\bar{u})\{N\bar{x} - b\},$$

which is what we wanted to show.

Clearly, if the solution to the shortest route problem (2.4) is not unique, then the subgradient is not unique. Computational experience has indicated that many subgradients at a point u are possible. As we shall see, when there are multiple subgradients, the ascent algorithm may jam because it has selected the wrong subgradient as a possible direction of increase. This difficulty is overcome by the primal-dual simplex algorithm.

Lemma 3.2: Let $\bar{\gamma}$ be any subgradient of L at the point $\bar{u} \in U_B$. The half-space $\{u|(u-\bar{u})\bar{\gamma} \geq 0\}$ in R^m contains all points u such that $L(u) \geq L(\bar{u})$.

Proof: Since $\bar{\gamma}$ is a subgradient, for any u

$$L(u) \leq L(\bar{u}) + (u - \bar{u})\cdot\bar{\gamma},$$

and if $L(u) \geq L(\bar{u})$, we have the desired result.

Thus $\bar{\gamma}$ points into the closed half containing all optimal solutions to (2.6). Lemma 3.2 is precisely lemma 1 on p. 9 of [22] in which Held and Karp describe an ascent algorithm for the traveling salesman problem. Other results from their paper are appropriate here. For example, let u be any point such that $L(u) > L(\bar{u})$. If the step size θ in the direction $\bar{\gamma} \neq 0$ at $\bar{u}$ satisfies $0 < \theta < \dfrac{2(L(u) - L(\bar{u}))}{||\bar{\gamma}||^2}$, then $||u - (\bar{u} + \theta\bar{\gamma})|| < ||u - \bar{u}||$, where $||\ \ ||$ denotes Euclidean norm. In words, a sufficiently small step size produces a point closer to a maximum point. Reference [22] also contains an investigation of a relaxation method for solving the traveling salesman dual problem.

A difference between the traveling salesman dual problem in [22] and the IP dual problem (2.6) is the presence in (2.6) of the constraints $c + uN \geq 0$, $u \geq 0$ on the dual variables. The traveling salesman dual problem is an unconstrained optimization problem while the IP dual problem here requires methods of feasible directions.

The constraints $c_j + ua_j \geq 0$ of the IP dual problem are less serious than the constraints $u_i \geq 0$ because of the tight and meaningful upper bounds available for many of the variables in the large majority of IP problems. Thus, in the ascent algorithm below, a dual solution $u \geq 0$

violating $c + uN \geq 0$ can yield a finite lower bound $z_B + L(u)$ of $v(b)$ by making the appropriate upper bound substitutions.

The ascent algorithm for the IP dual problem chooses a direction of ascent $\bar{d}$ from $\bar{u} \in U_B$ by solving the LP problem

$$\begin{aligned} &\max\ u \cdot \bar{\gamma} \\ &\text{s.t. } c_j + ua_j \geq 0,\ j \in J' \\ &\qquad u \geq 0. \end{aligned} \tag{3.3}$$

Suppose (3.3) has an optimal solution u'. It is easy to show that $w_B(b) \leq z_B + L(\bar{u}) + u' \cdot \bar{\gamma}$, and moreover that $\bar{u}$ is optimal in the IP dual problem if $u' \cdot \bar{\gamma} = \bar{u} \cdot \bar{\gamma}$. On the other hand, if $u'\bar{\gamma} > \bar{u}\bar{\gamma}$, take the direction of ascent $\bar{d} = u' - \bar{u}$. A third possibility is that (3.1) does not have an optimal solution; that is, there exists u', u'' such that $u''\bar{\gamma} > 0$ and $u' + \tau u''$ is feasible in (3.1) for all $\tau \geq 0$. In this case, take $\bar{d} = u' + \bar{\tau}u'' - \bar{u}$ for any $\bar{\tau} > 0$ such that $(u' + \bar{\tau}u'')\bar{\gamma} > \bar{u} \cdot \bar{\gamma}$.

To select the step length, we first compute

$$\theta^1_{max} = \min\{\frac{c_j + \bar{u}a_j}{-\bar{d}a_j} : \bar{d}a_j < 0\}$$

$$\theta^2_{max} = \min\{\frac{\bar{u}_i}{-\bar{d}_i} : \bar{d}_i < 0\}$$

and then

$$\theta_{max} = \min\{\theta^1_{max}, \theta^2_{max}\} > 0$$

where J' is the index set of those activities such that $c_j + ua_j$ is small. The constraints $c_j + ua_j \geq 0$, $j \notin J'$ are excluded from (3.3) because they are numerous and relatively non-binding in most applications. By construction, we have $\{u|u = \bar{u} + \theta\bar{d},\ \theta \geq 0\} \cap U_B = \{u|u = \bar{u} + \theta\bar{d},\ 0 \leq \theta \leq \theta_{max}\}$. The actual value of θ in the range $0 \leq \theta \leq \theta_{max}$ is selected by solving

$$\max_{0\leq\theta\leq\theta_{max}} L(\bar{u}+\theta\bar{d}). \tag{3.4}$$

Problem (3.4) is the problem of finding the maximum of a piecewise linear concave function of a single variable θ (see figure 4.1). This can be accomplished by a simple search procedure whose details we omit here. We simply mention that each step of the search begins with a point $\theta_{k+1} \in [0,\theta_{max}]$ and two piecewise linear segments $L(\bar{u}, x^{k-1}) + \theta\bar{d}\gamma^{k-1}$ and $L(\bar{u}, x^k) + \theta\bar{d}\gamma^k$ satisfying $L(\bar{u}, x^{k-1}) + \theta_{k+1}\bar{d}\gamma^{k-1} = L(\bar{u}, x^k) + \theta_{k+1}\bar{d}\gamma^k = L^{k+1}$. The point θ_{k+1} is tested for optimality by solving the shortest route problem (2.4) and obtaining an $x^{k+1} \in X$ satisfying $L(\bar{u}+\theta_{k+1}\bar{d}) = L(\bar{u}, x^{k+1}) + \theta_{k+1}\bar{d}\gamma^{k+1}$. If $L(\bar{u}+\theta_{k+1}\bar{d}) = L^{k+1}$, the point θ_{k+1} is optimal in (3.4). If $L(\bar{u}+\theta_{k+1}\bar{d}) < L^{k+1}$, the search procedure continues with one of the piecewise linear segments used in computing θ_{k+1} replaced by $L(\bar{u}, x^{k+1}) + \theta\bar{d}\gamma^{k+1}$. A similar search procedure is given below when we discuss the primal-dual algorithm.

Because we are working with subgradients rather than gradients, it can happen that $\bar{\theta} = 0$ and the ascent algorithm has jammed at the point $\bar{u}$. One way to try to eliminate the difficulty is to choose another subgradient for the objective function in the LP (3.3) but there is no guarantee that this will work. Appendix C gives a complete characterization of the set of subgradients at $\bar{u}$. We use this characterization in the construction of a primal-dual simplex algorithm for solving the IP dual problem which provides monotonically increasing lower bounds. This algorithm is similar to one proposed by Gomory in [13] and very similar to the algorithm for resource constrained network scheduling problems given by Fisher in [6], [7].

Suppose we are at the point $\bar{u}$ and the subgradients $\gamma^t = Nx^t - b$ for $t \in T'(\bar{u}) \subset T(\bar{u})$ have already been generated where

$$T(\bar{u}) = \{t | L(\bar{u}) = -\bar{u}b + (c+\bar{u}N)x^t\}.$$

Similarly, let

$$J(\bar{u}) = \{j | c_j + \bar{u}a_j = 0\},$$

and

$$I(\bar{u}) = \{i | \bar{u}_i = 0\}.$$

The algorithm solves the LP

$$\begin{aligned}
\sigma^* = \min\ & \sigma \\
\text{s.t.}\ & \sum_{t \in T'(\bar{u})} \lambda_t \gamma_i^t + \sum_{j \in J(\bar{u})} \mu_j a_{ij} = 0 \qquad i \notin I(\bar{u}) \\
& \sum_{t \in T'(\bar{u})} \lambda_t \gamma_i^t + \sum_{j \in J(\bar{u})} \mu_j a_{ij} \leq 0 \qquad i \in I(\bar{u}) \qquad (3.6) \\
& \sum_{t \in T'(\bar{u})} \lambda_t + \sigma = 1 \\
& \lambda_t \geq 0,\ t \in T'(\bar{u}),\ \mu_j \geq 0,\ j \in J(\bar{u}),\ \sigma \geq 0.
\end{aligned}$$

The variable σ in (3.6) is an artificial variable which we try to drive to zero. If we succeed, then as we show below, $\bar{u}$ is optimal in the IP dual (2.6). Since $\bar{u}$ is probably not optimal, the usual case is that $\sigma^* > 0$ but if this is so, the optimal solution to (3.6) provides a direction of ascent.

The following lemma is a restatement in the context of this paper of the sufficient condition for optimality of the primal-dual simplex algorithm. The proof of sufficiency we give is a new one using the property of subgradients and the structure of (3.6).

Lemma 3.3: If $\sigma^* = 0$ in (3.6), then $\bar{u}$ is optimal in the IP dual problem (2.7).

Proof: Let λ_t^*, $t \in T'(\bar{u})$, μ_j^*, $j \in J(\bar{u})$, be the optimal solution to (3.6) yielding the value $\sigma^* = 0$. Then

$$\gamma^* = \sum_{t \in T'(\bar{u})} \lambda_t^* \gamma^t + \sum_{j \in J(\bar{u})} \mu_j^* a_j$$

is a subgradient of L at $\bar{u}$ (see Appendix C) and therefore

$$L(u) \leq L(\bar{u}) + (u-\bar{u})\gamma^* \text{ for all } u.$$

But

$$(u-\bar{u})\gamma^* = \sum_{i \notin I(\bar{u})} (u_i-\bar{u}_i)\gamma_i^* + \sum_{i \in I(\bar{u})} (u_i-\bar{u}_i)\gamma_i^* = \sum_{i \in I(\bar{u})} u_i \gamma_i^*$$

because from (3.6), $\gamma_i^* = 0$, $i \notin I(\bar{u})$, $\bar{u}_i = 0$, $i \in I(\bar{u})$. Since $\gamma_i^* \leq 0$, $i \in I(\bar{u})$, we have $\sum_{i \in I(\bar{u})} u_i \gamma_i^* \leq 0$ for all $u \geq 0$, or

$$L(u) \leq L(\bar{u}) \text{ for all } u \geq 0,$$

which establishes the desired result.

In order to describe the algorithm for the usual case when $\sigma^* > 0$ in (3.6), consider the following LP that is dual to (3.6)

$$\sigma^* = \max u_{m+1}$$

$$\text{s.t. } -u\gamma^t + u_{m+1} \leq 0 \qquad t \in T'(u)$$

$$-ua_j \leq 0 \qquad j \in J(\bar{u}) \qquad (3.7)$$

$$u_{m+1} \leq 1$$

u_{m+1}, u_i unconstrained in sign, $i \notin I(\bar{u})$, $u_i \geq 0$, $i \in I(\bar{u})$.

Let u^*, $u^*_{m+1} = \sigma^*$ be an optimal solution to (3.7). We have

$$u^*\gamma^t > 0 \quad \text{for } t \in T'(\bar{u}) \qquad (3.8)$$

and

$$u^*a_j \geq 0 \quad \text{for } j \in J(\bar{u}). \qquad (3.9)$$

We choose as our direction of ascent the half line $u(\theta) = \bar{u} + \theta u^*$ for $\theta \geq 0$. Notice that condition (3.9) implies that for $j \in J(\bar{u})$, $c_j + u(\theta)a_j = \theta u^*a_j \geq 0$ for all $\theta \geq 0$. Similarly $u_i^* \geq 0$ for $i \in I(\bar{u})$ implies $u_i(\theta) \geq 0$ for $\theta \geq 0$. Thus, the range of feasible θ in the function $u(\theta)$ is determined by the $\min(\theta_1, \theta_2) = \theta_{max} > 0$ where

$$\theta_1 = \min\{\frac{c_j + \bar{u}a_j}{-u^*a_j} : j \notin J(\bar{u}) \text{ and } u^*a_j < 0\} > 0$$

$$\theta_2 = \min\{\frac{\bar{u}_i}{-u_i^*} : i \notin I(\bar{u}) \text{ and } u_i^* < 0\} > 0.$$

The step length in the direction $u(\theta)$, $\theta \geq 0$, is selected by discovering the closest point of change of slope of L to the point $\bar{u}$. Let β^0 be the subgradient satisfying

$$u^*\beta^0 = \min\{u^*\gamma^t : t \in T'(\bar{u})\}$$

and note that $u^*\beta^0 \geq \sigma^* > 0$. The amount we increase θ is determined by the constraints

$$w \leq -u(\theta)b + (c + u(\theta)N)x^t,\ t \notin T'(\bar{u})$$

of the IP dual (2.7) which we have been ignoring up to this point.

In particular, we search the line segment $[0, \theta_{max}]$ in an iterative manner generating points closer and closer to $\bar{u}$ at which the slope of $L(\bar{u} + \theta u^*)$ changes. Let $\theta_1 = \theta_{max}$ and at iteration k compute $L(\bar{u} + \theta_k u^*)$ by solving the shortest route problem (2.5). Let x^k be an optimal solution and $\beta^k = Nx^k - b$ the corresponding subgradient of L. We have

$$\begin{aligned} L(\bar{u} + \theta_k u^*) &= -\ (\bar{u} + \theta_k u^*)b + (c + (\bar{u}+\theta_k u^*)N)x^k \\ &= -\ \bar{u}b + (c + \bar{u}N)x^k + \theta_k u^*(N_x^k - b) \\ &= L(\bar{u},\ x^k) + \theta_k u^*\beta^k. \end{aligned}$$

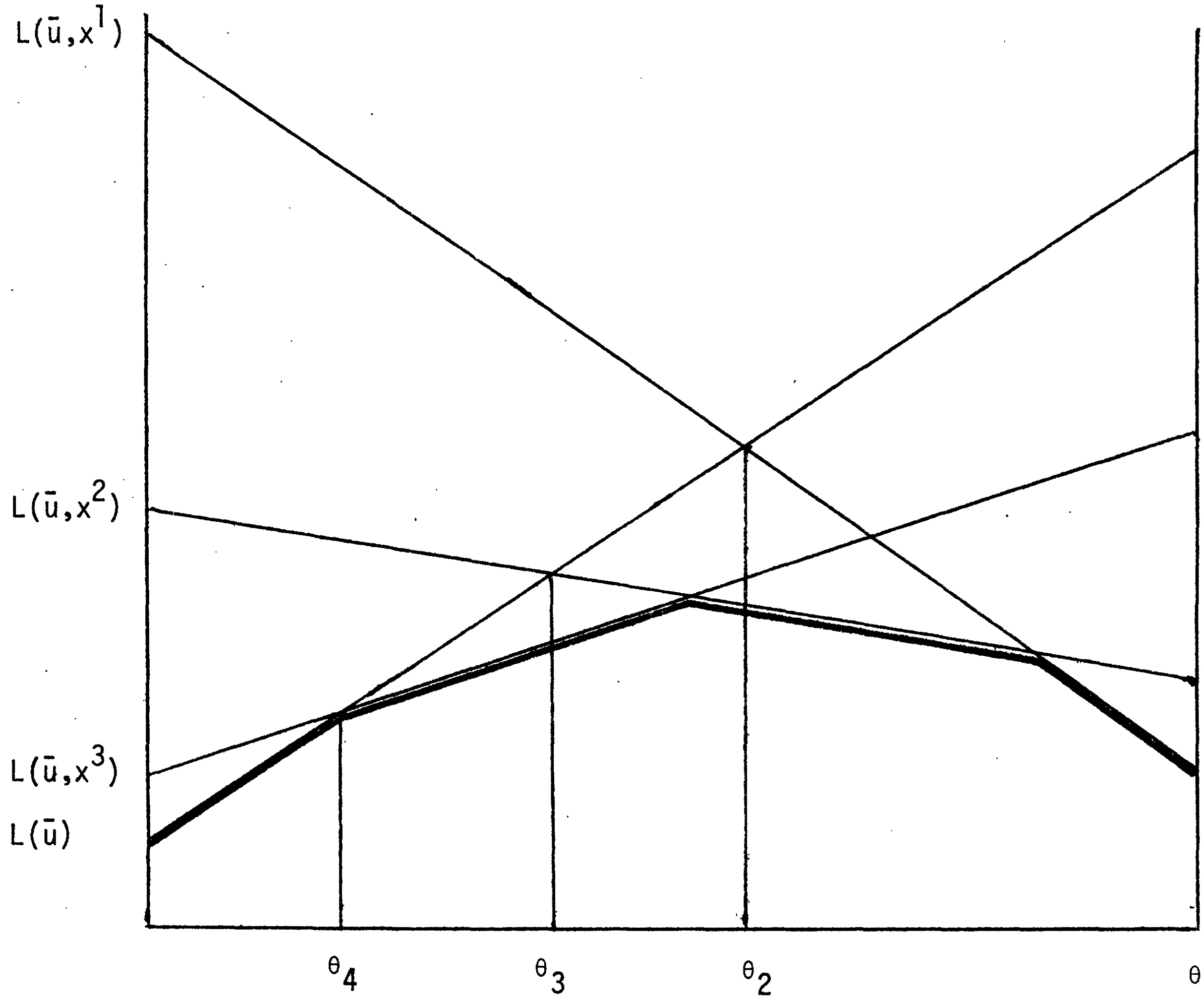
If $L(\bar{u} + \theta_k u^*) = L(\bar{u}) + \theta_k u^*\beta^0$, then we terminate the search procedure with the point $\theta^* = \theta_k$. If

$$L(\bar{u} + \theta_k u^*) < L(\bar{u}) + \theta_k u^*\beta^0, \tag{3.10}$$

then compute

$$\theta_{k+1} = \frac{L(\bar{u},\ x^k) - L(\bar{u})}{u^*\beta^0 - u^*\beta^k}\ .$$

Note that $L(\bar{u},\ x^k) \geq L(\bar{u})$ and $u^*\beta^0 > u^*\beta^k$ because of (3.10). If $\theta_{k+1} = 0$, then set $\theta^* = 0$ and the search procedure is terminated. The procedure is illustrated in figure 3.1.



$L(\bar{u},x^1)$
$L(\bar{u},x^2)$
$L(\bar{u},x^3)$
$L(\bar{u})$
θ_4
θ_3
θ_2
θ_1

Figure 3.1

Since the number of piecewise linear segments of L in the interval $[0, \theta_{max}]$ is finite, the search procedure terminates after a finite number of iterations.

Thus, there are two modes of termination of the search procedure of the line segment $[0, \theta_{max}]$. The first is when $\theta^* = 0$ in which case we have found a new solution x^k to the shortest route problem (2.5) at the point $\bar{u}$. Since when this occurs, $L(\bar{u} + \theta_k u^*, x^k) < L(\bar{u}) + \theta_k u^*\beta^0$, we have $u^*\beta^k < u^*\beta^0$, or the subgradient β^k of L at $\bar{u}$ is not one of the ones included in γ^t, $t \in T'(\bar{u})$. Moreover, $u^*\beta^k - u^*_{m+1} < 0$ so that when a column corresponding to β^k is added to problem (3.6) it can be optimized further.

On the other hand, suppose the search procedure terminates with $\theta^* > 0$. Then we create a new primal LP problem (3.6) to try to prove that the new dual vector $\bar{\bar{u}} = \bar{u} + \theta^* u^*$ is optimal in the IP dual problem. In constructing this problem, it is important to note that all of the λ_t and μ_j columns that were basic remain valid columns for problem (3.6) constructed at the point $\bar{u} + \theta^* u^*$. To see this, consider a λ_t column that is basic in the optimal solution to (3.6) at $\bar{u}$. We have $u^*\gamma^t = u^*_{m+1}$ and therefore $L(\bar{u} + \theta^* u^*) = L(\bar{u}) + \theta^* u^*\gamma^t$; i.e. $\gamma^t \in T(\bar{u} + \theta^* u^*)$. Similarly, if a μ_j column is basic in the optimal solution to (3.6) at $\bar{u}$, then $u^*a_j = 0$ which implies $c_j + (\bar{u} + \theta u^* a_j) = \theta u^* a_j = 0$, and $j \in J(\bar{u} + \theta u^* a_j)$. Thus, the initial solution to (3.6) at $\bar{u} + \theta^* u^*$ is found by retaining all the optimal basic columns from (3.6) at $\bar{u}$, add a λ_k column corresponding to β^k and resolve it. Note that the new column corresponding to β_k again prices out negatively relative to u^*, u^*_{m+1}; that is, $u^*\beta^k < u^*_{m+1}$ and the minimal value of σ can be reduced in (3.6) at $u = \bar{u} + \theta^* u^*$ from what it was at $u = \bar{u}$. Convergence of the primal-dual algorithm is thereby assured by the usual simplex criterion (see Simonnard [29; pp. 128-134]).

We conclude this section by returning to the original IP problem (2.1) (or equivalently, (2.2) and considering the relation of an optimal solution to the IP dual problem to it. In particular, suppose $\sigma^* = 0$ in problem (3.6) and therefore $\bar{u}$ is optimal in the IP dual problem. Let $x^* = \sum_{t \varepsilon T'(\bar{u})} \lambda_t^* x^t + \sum_{j \varepsilon J(\bar{u})} \mu_j^* e_j$ be the solution derived from the optimal weights in (3.6). Then it is not difficult to show that x^* is optimal in (2.2) if and only if $x^* \varepsilon X$. If $x^* \notin X$, we must proceed on the assumption that $v(b) > w(b)$ and continue the effort to find an optimal solution to (2.1) by other means. For example, we could choose to solve the augmented LP (2.10) where the added cut is written with respect to $\bar{u}$. The LP solution to this problem provides a new starting point for all of our dual analysis. It is important to mention, however, that the optimal solution to problem (3.6) for $\bar{u}$ optimal in the IP dual problem provides most of the same structural information for implementing a new dual analysis as an optimal solution to (2.10). In other words, if the solution x^* derived from (3.6) is not in X, then we would like to be able to deduce new conditions from the optimal solution to (3.6) that would render $\bar{u}$ non-optimal, although still dual feasible. The primal-dual or ascent algorithm would then proceed as before searching out still higher monotonically increasing lower bounds. Unfortunately, it has not been possible to date to construct such a procedure.

4. Use of Dual IP Problem in Branch and Bound

In this section, we discuss briefly how the dual IP methods developed in this paper can be used in conjunction with branch and bound searches. Specifically, we will illustrate the relevant ideas by showing how the branch and bound algorithm of [16] can be improved.

The branch and bound algorithm of [16] implicitly considers all non-negative integer x in (2.3). At an intermediate point of computation, the minimal cost solution found thus far is called the incumbent and denoted by $\hat{x}$ with incumbent cost $\hat{z} = z_B + c\hat{x}$. The algorithm explicitly generates IP subproblems of the following form from explicitly enumerated non-negative integer vectors $\tilde{x}$

$$
\begin{aligned}
v(b;\, \tilde{x}) &= z_B + c\tilde{x} + \min \sum_{j=1}^{j(\tilde{x})} c_j x_j \\
\text{s.t.} \quad & \sum_{j=1}^{j(\tilde{x})} a_j x_j \leq b - N\tilde{x} \qquad\qquad (4.1) \\
& \sum_{j=1}^{j(\tilde{x})} \alpha_j x_j \equiv \beta(\tilde{x}) \\
& x_j \text{ non-negative integer, } j=1,\ldots,j(\tilde{x})
\end{aligned}
$$

where $j(\tilde{x}) = \min \{j \mid \tilde{x}_j > 0\}$, $\beta(\tilde{x}) = \beta - \sum_{j=j(\tilde{x})}^{n} \alpha_j \tilde{x}_j$, and a_j is the j^{th} column of N. The summations in (4.1) are restricted to the range 1 to $j(\tilde{x})$ in order that the search be non-redundant.

If we can find an optimal solution to (4.1), then we have implicitly tested all non-negative integer $x \geq \tilde{x}$, $x_j = \tilde{x}_j$, $j=j(\tilde{x})+1,\ldots,n$, and they do not have to be explicitly enumerated. The same conclusion is true if we can ascertain that $v(b;\, \hat{x}) \geq \hat{z}$ without actually discovering the precise value of $v(b; \tilde{x})$. If either of these two cases obtain, then we say that $\tilde{x}$ has been fathomed. If $\tilde{x}$ is not fathomed, then we abandon (4.1) and create new subproblems of the same form as (4.1) from the solutions $\tilde{x} + e_j$, $j=1,\ldots,j(\tilde{x})$.

We attempt to fathom (4.1) by solution of the following IP dual problem using the methods of section 3

$$\begin{aligned} w(b; \tilde{x}) &= z_B + c\tilde{x} + \max w \\ \text{s.t. } w &\le -u(b-N\tilde{x}) + (c+uN)x^t, \; t \in \tilde{T} \\ & c_j + ua_j \ge 0, \; j=1,\ldots,j(\tilde{x}) \\ & u \ge 0, \end{aligned} \tag{4.2}$$

where $\tilde{T}$ is the index set of ineducible $x^t \in X(\beta(\tilde{x}))$ such that $x_j = 0$, $j = j(\tilde{x})+1,\ldots,n$. The use of (4.2) in analyzing a given IP subproblem (4.1) is illustrated in figure 4.1 which we will now discuss step by step. The number of algorithmic options available to us is enormous, but space does not permit an extensive discussion of them.

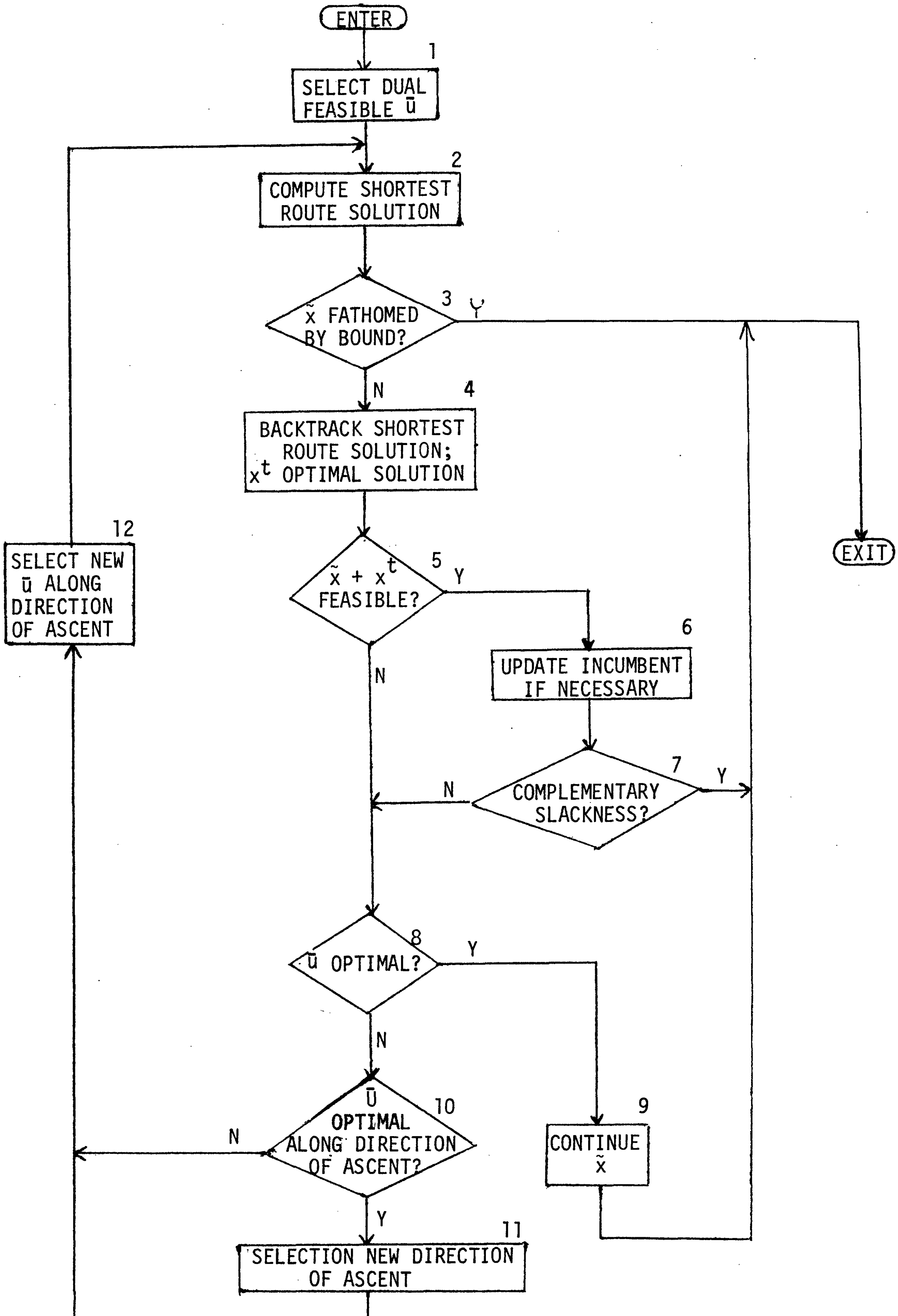
STEP 1: An initial vector $\bar{u} \ge 0$ satisfying $c_j + \bar{u}a_j \ge 0$, $j=1,\ldots,j(\tilde{x})$, is required. Such a vector should be available from previous computation.

STEP 2: Compute

$$\begin{aligned} G(\beta(\tilde{x}); \bar{u}) &= \min \sum_{j=1}^{j(\tilde{x})} (c_j+\bar{u}a_j)x_j \\ \text{s.t. } & \sum_{j=1}^{j(\tilde{x})} \alpha_j x_j \equiv \beta(\tilde{x}) (\text{mod } q) \\ & x_j \text{ non-negative integer}, \; j=1,\ldots,j(\tilde{x}). \end{aligned} \tag{4.3}$$

A good feasible solution to (4.3) is probably available from previous computation.

Figure 4.1


ENTER
1
SELECT DUAL FEASIBLE ū
2
COMPUTE SHORTEST ROUTE SOLUTION
3
x̃ FATHOMED BY BOUND?
Y
N
4
BACKTRACK SHORTEST ROUTE SOLUTION; x^t OPTIMAL SOLUTION
EXIT
12
SELECT NEW ū ALONG DIRECTION OF ASCENT
5
x̃ + x^t FEASIBLE?
Y
N
6
UPDATE INCUMBENT IF NECESSARY
7
COMPLEMENTARY SLACKNESS?
N
Y
8
ū OPTIMAL?
Y
N
10
Ū OPTIMAL ALONG DIRECTION OF ASCENT?
N
Y
9
CONTINUE x̃
11
SELECTION NEW DIRECTION OF ASCENT

STEP 3: The enumerated solution $\tilde{x}$ is fathomed by bound if

$$z_B + c\tilde{x} - \bar{u}(b-N\tilde{x}) + G(\beta(\tilde{x}); \bar{u}) \geq \hat{z} \tag{4.4}$$

because the left hand side of (4.4) is a lower bound on $v(b; \tilde{x})$.

STEPS 4 and 5: If $\tilde{x}$ is not fathomed by found, then we may want to see if the optimal solution x^t to (4.3) satisfies $Nx^t \leq b - N\tilde{x}$. The determination of x^t requires extra computation (back tracking) which may not be deemed worth the investment.

STEP 6: Notice that the implication of a no branch from step 3 is that

$$z_B + c(\tilde{x}+x^t) + \bar{u}(N(\tilde{x}+x^t) - b) < \hat{z} \ . \tag{4.5}$$

The third term on the left is non-positive since $\bar{u} \geq 0$ and $\tilde{x} + x^t$ is feasible. Thus it may or may not be true that $z_B + c(\tilde{x}+x^t) < \hat{z}$, but if it is, then $\hat{x}$ is replaced by $\tilde{x} + x^t$.

STEP 7: If $\tilde{x} + x^t$ is feasible and the term $\bar{u}(N(\tilde{x}+x^t) - b) = 0$, then we are in the fortunate case of having discovered that x^t is optimal in (4.1) and $\bar{u}$ optimal in (4.2) and $v(b; \tilde{x}) = w(b; \tilde{x})$. In other words, there is no duality gap for (4.1) and moreover, we have found an optimal solution x^t to (4.1). In this case, $\tilde{x}$ is fathomed and we exit. The chances for a yes branch from step 7 are slight but the test becomes important if one is willing to relax the test to: Is $\bar{u}(N(\tilde{x}+x^t) - b) > - \varepsilon$ for suitable $\varepsilon > 0$. This type of heuristic is employed by some production mixed integer programming codes.

STEPS 8 and 9: If $\bar{u}$ is optimal, then (4.3) has failed to bring about a fathoming of $\tilde{x}$ and $\tilde{x}$ is continued. An alternative to abandoning (4.1) and continuing $\tilde{x}$ is to add cuts to (4.1).

STEPS 10, 11, 12: These steps are concerned with the ascent and primal-dual methods of section 3.

5. Conclusions

In this paper we have developed a special application of mathematical programming duality theory to IP. We have been able to derive some new results and new algorithm methods; specifically, we have constructed in section 3 the ascent and primal-dual simplex algorithms for the IP dual problem (2.6). An area of future research is the extension of these dual methods and algorithms to other combinatorial optimization problems. We also mention that the primal-dual algorithm developed here is directly applicable to the traveling salesman dual problem of [21], [22].

As we mentioned at the end of section 3, another area of future research is a more direct interpretation and use of the cutting plane method of IP for filling in duality gaps. Conversely, the methods outlined here are designed primarily to make IP problems easier to solve, and thus there is some justification in waiting for computational experience with the IP duality theory before developing extensions. Another related research area to be investigated is the application of Gould's multiplier function theory to try to fill in the IP duality gap. It is important to mention in this regard that the use of functions rather than scalars in the shortest route problem (2.4) can make that problem much more difficult to solve.

Finally, it appears that the constructive IP duality theory can be combined with the constructive duality theory inherent in the papers of Falk [4], Falk and Soland [5], and Geoffrion [12].

Appendix A

IP Primal and Dual Perturbation Functions

Our concern here is a brief study of the primal and dual perturbation functions of the right hand side $\bar{b}$ in (2.1). These perturbations functions are somewhat different than the usual perturbation functions used in the study of nonlinear programming problems (e.g., see Geoffrion [11; p. 6]). Since the duality theory here is based on the presence of integer data in (2.1), the perturbation functions are defined only for integer vectors $\bar{b}$. Second, the algebraic nature of the transformation of (2.1) into the equivalent form (2.2) from which the duals are constructed necessitates the study of a family of perturbation functions rather than a single one for the primal and one for the dual.

We must take into account some algebraic structure of the group $G = \{\lambda_k\}_{k=0}^{D-1}$ induced by the basis $\bar{B}$ (see section 2) in order to define consistent perturbation functions. Our discussion of this structure will be brief and the reader is referred to Shapiro [28], Wolsey [31] for more details. Let Z^m denote the set of integer points in R^m, and let Φ denote the mapping of Z^m onto G. In particular, Φ is constructed from a diagonalization procedure applied to $\bar{B}$ which yields $\Phi(e_i) = \varepsilon_i \in G$, $i=1,\ldots,m$, where e_i is the i^{th} unit vector in R^m. The group identity of an arbitrary vector $a \in Z^m$ is computed by $\sum_{i=1}^{m} a_i\varepsilon_i$. The mapping Φ naturally partitions Z^m into equivalence classes Λ by the equivalence relation $\bar{b}^1, \bar{b}^2 \in \Lambda$ if and only if $\sum_{i=1}^{m} \bar{b}_i^1\varepsilon_i \equiv \sum_{i=1}^{m} \bar{b}_i^2\varepsilon_i \equiv \lambda$. It is easy to show that this equivalence relation is equivalent to $\bar{b}^1, \bar{b}^2 \in \Lambda$ if and only if $\bar{b}^1 = \bar{b}^2 + \bar{B}v$, v integer. For notational convenience, we let $b \in \Lambda$ denote $\bar{B}b \in \Lambda$.

With this background, we can define perturbation functions for the IP problem (2.1) as a function of its right hand side $\bar{b}$. For arbitrary $\lambda \varepsilon G$, define

$$X(\lambda) = \{x | \sum_{j=1}^{n} \alpha_j x_j \equiv \lambda \pmod q, \text{ x non-negative integer}\}$$

and let $\{x^t\}_{t \varepsilon T_\lambda}$ denote the finite set of irreducible points of $X(\lambda)$. For each class Λ, we define the perturbation function v_λ for elements $b \varepsilon \Lambda$ as

$$\begin{aligned} v_\lambda(b) = \bar{c}_B b &+ \min cx \\ &\text{s.t. } Nx \leq b \\ &\qquad x \varepsilon X(\lambda). \end{aligned} \tag{A.1}$$

Recall that $\bar{c}_B$ is the vector of cost coefficients of the basic variables in (2.1). As before, $x \varepsilon X(\lambda)$ ensures that the basic variables $y = b - Nx$ are integer, and $Nx \leq b$ ensures that they are non-negative. Problem (A.1) clearly makes sense only if the requirement $x \varepsilon X(\lambda)$ is coupled with the requirement $b \varepsilon \Lambda$.

The dual to problem (A.1) is

$$\begin{aligned} w_\lambda(b) = \bar{c}_B b &+ \max w \\ &\text{s.t. } w \leq -ub + (c+uN)x^t, \ t \varepsilon T_\lambda \\ &\qquad c + uN \geq 0 \\ &\qquad u \geq 0 \end{aligned} \tag{A.2}$$

The function $v_\lambda(b)$ has some asymptotic properties that are studied in [28]. More generally, however, its behavior is erractic and hard to describe. On the other hand, the dual perturbation function $w_\lambda(b)$ has a more regular behavior. First of all, since $w_\lambda(b)$ is the optimal value of a linear program, it is easy to show that it is a convex function of b. The structure of the shortest route problem (2.5) gives us some structure indicating how the D perturbation functions w_λ are related.

Lemma A.1: Suppose $\lambda \not\equiv 0$ and $w_\lambda(b) < +\infty$. Then

$$w_\lambda(b) \leq \bar{c}_j + w_{\lambda-\alpha_j}(b-a_j),\ j=1,\ldots,n$$

where $\bar{c}_j$ is the cost coefficient of activity $\bar{a}_j$ in (2.1).

Proof: Let u* be optimal in (A.2); then

$$w_\lambda(b) = \bar{c}_B b + G(\lambda;\ u^*),$$

wnere

$$G(\lambda;\ u^*) = \min_{x \in X(\lambda)} (c+u^*N)x.$$

Now, (2.5) is a shortest route problem and therefore (see [16])

$$G(\lambda;\ u^*) = \min\{c_j + G(\lambda-\alpha_j;\ u^*);\ j=1,\ldots,n\}.$$

This gives us for any j

$$w_\lambda(b) \leq \bar{c}_B b + c_j + G(\lambda-\alpha_j;\ u^*). \tag{A.3}$$

The vector u* may not be optimal in (A.2) with b replaced by b - a_j, and therefore,

$$w_{\lambda-\alpha_j}(b-a_j) \geq \bar{c}_B(b-a_j) + G(\lambda-\alpha_j;\ u^*). \tag{A.4}$$

Combining (A.3) and (A.4) yields

$$w_\lambda(b) \leq c_j + \bar{c}_B a_j + w_{\lambda-\alpha_j}(b-a_j),$$

and the desired result is obtained by noting that the LP reduced cost $c_j = \bar{c}_j - \bar{c}_B a_j$.

Appendix B

There is an additional geometric insight about IP dual problems that merits discussion and analysis. The dual of the dual IP problem (2.7) can be viewed as the linear programming problem

$$\begin{aligned} &z_B + \min cx \\ &\text{s.t. } x \in [X] \cap \{x \mid Nx \le b\} \end{aligned} \qquad (B.1)$$

where [X] denotes the convex hull of X. Problem (B.1) is problem (8) of [27] in a different form.

Lemma A.1. A necessary and sufficient condition that the dual objective function $w_B(b) < +\infty$ is that

$$[X] \cap \{x \mid Nx \le b\} \neq \phi.$$

Proof: Suppose $[X] \cap \{x \mid Nx \le b\} = \phi$; we will establish necessity by showing that $w_B(b) = +\infty$. Consider the following LP problem for trying to find an $\tilde{x} \in [X] \cap \{x \mid Nx \le b\}$. The problem is

$$\begin{aligned} &\min \sigma \\ &\text{s.t. } \sum_{t=1}^{T} \lambda_t(-Nx^t) + \sum_{j=1}^{n} \mu_j(-a_j) \ge -b \\ &\qquad \sum_{t=1}^{T} \lambda_t + \sigma = 1 \end{aligned} \qquad (B.2)$$

$$\lambda_t \geq 0, \ t = 1,\ldots,T$$

$$\mu_j \geq 0, \ j = 1,\ldots,n$$

$$\sigma \geq 0.$$

Implicit in this formulation is the fact that the set $\{x^t\}_{t=1}^{T}$ contains all the extreme points of [X] and e_j, $j = 1,\ldots,n$, are the extreme rays of [X] (see reference [15]). Let σ^* denote an optimal solution to problem (B.2), and let u^* denote an m-vector of optimal dual variables, with u^*_{m+1} the optimal dual variable on the convexity row. By LP duality theory, we have

$$\sigma^* = -u^*b + u^*_{m+1} > 0, \tag{B.3}$$

$$-u^*Nx^t + u^*_{m+1} \leq 0, \ t = 1,\ldots,T, \tag{B.4}$$

and

$$u^*N \geq 0, \tag{B.5}$$

$$u^* \geq 0. \tag{B.6}$$

Combining (B.3) and (B.4) gives

$$u^*(Nx^t - b) > 0, \ t = 1,\ldots,T.$$

Let $\tilde{u}$ be any point satisfying $c + \tilde{u}N \geq 0$, $\tilde{u} \geq 0$ (in section 2, we have assumed such a point exists). For any $\theta \geq 0$, $c + (\tilde{u}+\theta u^*)N \geq 0$, $\tilde{u} + \theta u^* \geq 0$ by (B.5) and (B.6). Thus, $L(\tilde{u}+\theta u^*)$ is finite for all $\theta \geq 0$. In particular,

$$L(\tilde{u}+\theta u^*) = \min_{t=1,\ldots,T} - (\tilde{u}+\theta u^*)b + (c+(\tilde{u}+\theta u^*)N)x^t$$

$$= \min_{t=1,\ldots,T} - \tilde{u}b + (c+\tilde{u}N)x^t + \theta u^*(Nx^t-b)$$

$$\geq L(\tilde{u}) + \theta \cdot \min_{t=1,\ldots,T} u^*(Nx^t-b)$$

where the inequality follows from the fact that the minimum of a sum is greater than or equal to the sum of the minimums. Since $\min_{t=1,\ldots,T} u^*(Nx^t-b) > 0$, the term on the right goes to $+\infty$ as θ goes to $+\infty$ implying $w_B(b) = +\infty$.

In order to prove sufficiency, it is only necessary to recognize that (B.2) with the objective function $z_B + \min \sum_{t=1}^{T} \lambda_t(cx^t) + \sum_{j=1}^{n} \mu_j(c_j)$ is the dual to the IP dual problem (2.7) with maximal value $w_B(0)$. Thus, if $[X] \cap \{x|Nx \leq b\} \neq \phi$, phase I will product a feasible solution and this solution provides a finite upper bound on $w_B(b)$. This completes the proof.

As a final point, we mention that it is possible for $\{x|Nx \leq b\} \subset [X]$ in which case the dual IP problem is nothing but the LP relaxation (2.3) of the given IP (2.1). This is the case which may be detected by the result of lemma 2.3.

Appendix C

Characterization of Lagrangian Subgradients

A characterization of Lagrangian subgradients for a class of dual problems similar to (2.6) is given by Grinold in [19]. This analysis would be applicable here except for the fact that the set of solutions to the Lagrangian problem (2.4) can be unbounded. Our development will follow [19] with modifications for overcoming this difficulty.

We begin with same notation. It is assumed throughout this appendix that we are considering $\bar{u}$ which satisfies $c + \bar{u}N \geq 0$, $\bar{u} \geq 0$. Let

$$\partial L(\bar{u}) = \{\gamma \mid L(u) \leq L(\bar{u}) + (u-\bar{u})\gamma \text{ for all } u\} \tag{C.1}$$

denote the set of Lagrangian subgradients at the point $\bar{u}$. Let

$$T(\bar{u}) = \{t \mid L(\bar{u}) = -\bar{u}b + (c+\bar{u}N)x^t\} \tag{C.2}$$

denote a subset of T, the index set of the irreducible points, and for $t \in T(\bar{u})$, define

$$\gamma^t = Nx^t - b. \tag{C.3}$$

Finally, let

$$J(\bar{u}) = \{j \mid c_j + \bar{u}a_j = 0\}$$

denote the subset of $\{1,\ldots,n\}$ corresponding to those activities with zero reduced cost at $\bar{u}$. The characterization of $\partial L(\bar{u})$ we seek is given by the following theorem.

Theorem C.1: The set of subgradients $\partial L(\bar{u}) = P(\bar{u})$ where

$$P(\bar{u}) = \{p | p = \sum_{t \varepsilon T(\bar{u})} \lambda_t \gamma^t + \sum_{j \varepsilon J(\bar{u})} \mu_j a_j, \sum_{t \varepsilon T(\bar{u})} \lambda_t = 1,$$

$$\lambda_t \geq 0,\ t \ \varepsilon\ T(\bar{u});\ \mu_j \geq 0,\ j\ \varepsilon\ J(\bar{u})\}.$$

Proof: To show $P(u) \subset \partial L(\bar{u})$, we begin with the inequality for any $t \ \varepsilon\ T(\bar{u})$

$$L(u) \leq L(\bar{u}) + (u-\bar{u})\gamma^t \text{ for all } u \tag{C.4}$$

established by lemma 3.1. We show that

$$L(u) \leq L(\bar{u}) + (u-\bar{u})(\gamma^t + \sum_{j \varepsilon J(\bar{u})} \mu_j a_j) \text{ for all } u \tag{C.5}$$

where the weights μ_j are arbitrary non-negative numbers. Consider the inner product

$$(u-\bar{u}) \sum_{j \varepsilon J(\bar{u})} \mu_j a_j = \sum_{j \varepsilon J(\bar{u})} \mu_j u a_j - \sum_{j \varepsilon J(\bar{u})} \mu_j \bar{u} a_j = \sum_{j \varepsilon J(\bar{u})} \mu_j (c_j + u a_j)$$

where the last expression follows from the definition of $J(\bar{u})$. Thus, for u satisfying $c_j + ua_j \geq 0$, $j \ \varepsilon\ J(\bar{u})$ we have $(u-\bar{u}) \sum_{j \varepsilon J(\bar{u})} \mu_j a_j \geq 0$ and (C.5) follows from (C.4). For u not satisfying $c_j + ua_j \geq 0$, $j \ \varepsilon\ J(\bar{u})$, we have $L(u) = -\infty$ and (C.5) holds trivially. The result $P(\bar{u}) \subset \partial L(\bar{u})$ follows by weighting each inequality (C.5) by $\lambda_t \geq 0$, $\sum_{t \varepsilon T(\bar{u})} \lambda_t = 1$, and summing them.

In order to establish $L(\bar{u}) \subset P(\bar{u})$, we need some intermediate results. These results give a characterization of the directional derivative of L at the point $\bar{u}$ in the direction v, denoted by $\nabla L(\bar{u};\, v)$. When the limit exists, this quantity is given by

$$\nabla L(\bar{u};\, v) = \lim_{\alpha \to 0^+} \frac{L(\bar{u} + v) - L(\bar{u})}{\alpha}\,. \tag{C.6}$$

The limit does not exist if $va_j < 0$ for any $j \in J(\bar{u})$ because then $c_j + (\bar{u}+\alpha v)a_j < 0$ for any $\alpha > 0$ which implies $L(\bar{u}+\alpha v) = -\infty$ for any $\alpha > 0$.

Lemma C.1: For any m-vector v satisfying $va_j \geq 0$, $j \in J(\bar{u})$, we have

$$\inf_{p \in P(\bar{u})} vp = \min_{t \in T(\bar{u})} v\gamma^t = \min_{p \in P(\bar{u})} vp.$$

Proof: The proof is straightforward and is omitted.

Lemma C.2: For directions v satisfying $va_j \geq 0$, $j \in J(\bar{u})$,

$$\nabla L(\bar{u};\, v) = \min_{t \in T(\bar{u})} v\gamma^t.$$

Proof: For any $t \in T(\bar{u})$ and $\alpha > 0$,

$$L(\bar{u}+\alpha v) \leq -\ (\bar{u}+\alpha v)b + (c+(\bar{u}+\alpha v)N)x^t$$

which implies

$$\frac{L(\bar{u}+\alpha v) - L(\bar{u})}{\alpha} \leq v\gamma^t.$$

Thus

$$\limsup_{\alpha \to 0} \frac{L(\bar{u}+\alpha v) - L(\bar{u})}{\alpha} \leq \min_{t \in T(\bar{u})} v\gamma^t \tag{C.7}$$

To prove the reverse inequality, consider a sequence $\{\alpha_s\}$ converging to zero from above. Note that $c_j + (\bar{u}+\alpha_s v)a_j \geq 0$, $j=1,\ldots,n$, for α_s sufficiently small. For each α_s sufficiently small, there is an x^t for some $t \in \{1,\ldots,T\}$, say x^s, satisfying

$$L(\bar{u}+\alpha_s v) = -\ (\bar{u}+\alpha_s v)b + (c+(\bar{u}+\alpha_s v)N)x^s. \tag{C.8}$$

Since the set of such points is finite, there is a distinguished one, say x^{s^*}, which occurs infinitely often. Without loss of generality, we can assume α_s converging to zero from above such that x^{s^*} is optimal in (C.8). By the continuity of L on the set satisfying $c + uN \geq 0$, $u \geq 0$, we have

$$L(\bar{u}) = -\bar{u}b + (c+\bar{u}N)x^{s^*},$$

and $s^* \in T(\bar{u})$.

Thus,

$$\lim_{\alpha_s \to 0^+} \frac{L(\bar{u}+\alpha_s v) - L(\bar{u})}{\alpha_s} = v\gamma^{s^*} \geq \min_{t \in T(\bar{u})} v\gamma^t. \tag{C.9}$$

Comparing (C.8) and (C.9), we have

$$\nabla L(\bar{u}; v) = \min_{t \in T(\bar{u})} v\gamma^t,$$

which is what we wanted to show.

We complete the proof of Theorem C.1 by proving

Lemma C.3: The set $\partial L(\bar{u}) \subset P(\bar{u})$.

Proof: Suppose $\bar{\gamma} \in \partial L(\bar{u})$ but $\bar{\gamma} \notin P(\bar{u})$; we will show a contradiction. Since $P(\bar{u})$ is a closed convex set, there exists a hyperplane v^* which strictly separates $\bar{\gamma}$ from $P(\bar{u})$. Hence for some $\varepsilon > 0$,

$$v^*\bar{\gamma} + \varepsilon \leq v^*p \text{ for all } p \in P(\bar{u}).$$

There are two possibilities for v^* and we will show a contradiction in each case. First, suppose $v^*a_j < 0$ for some $j \in J(\bar{u})$. By hypothesis, for any $t \in T(\bar{u})$, we have $v^*\bar{\gamma} < v^*\gamma^t$. But $\gamma^t + \mu_{j_1} a_{j_1} \in P(\bar{u})$ for any

$\mu_{j_1} \geq 0$ and it is clear that $v^*\gamma^t + \mu_{j_1} v^*a_{j_1} < v^*\bar{\gamma}$ for μ_{j_1} sufficiently large. This gives us a contradiction in the first case.

Suppose on the other hand that $v^*a_j \geq 0$, $j \in J(\bar{u})$. Then

$$\nabla L(\bar{u};\, v^*) \leq v^*\bar{\gamma} < \min_{p \in P(\bar{u})} v^*p = \nabla L(\bar{u};\, v^*)$$

where the left inequality follows the fact that $\bar{\gamma}$ is a subgradient, the strict inequality is the hypothesis on v^*, and the right equality follows from lemmas C.1 and C.2.

References

1. R. E. Alcaly and A. K. Klevorick, "A Note on the Dual Price of Integer Programs", Econometrica, XXXIV (1966), 206-214.

2. E. Balas, "Minimax and Duality for Linear and Nonlinear Mixed Integer Programming", Chapter 19 in Integer and Nonlinear Programming, edited by J. Abadie, North Holland (1970).

3. M. L. Balinski, "Integer Programming: Methods, Uses, Computation", Management Science, XII (1961), 253-254.

4. J. E. Falk, "Lagrange Multipliers and Nonconvex Programming", SIAM Journal on Control, VII, No. 4 (1969), 534-545.

5. J. E. Falk and R. M. Soland, "An Algorithm for Separable Nonconvex Programming Problems", Management Science, XV, No. 9 (1969), 550-569.

6. M. L. Fisher, "Optimal Solution of Scheduling Problems Using Generalized Lagrange Multipliers and Relaxation", Report 720, Center for Mathematical Studies in Business and Economics, University of Chicago, February, 1971.

7. M. L. Fisher, "Optimal Solution of Scheduling Problems Using Lagrange Multipliers: Part II", presented at the Scheduling Symposium, North Carolina State University, Raleigh, North Carolina, May 15-17, 1972.

8. A. M. Geoffrion, "An Improved Implicit Enumeration Approach for Integer Programming", Operations Research, 17, No. 3 (1969), 437-454.

9. A. M. Geoffrion, "Elements of Large-Scale Mathematical Programming I", Management Science, XVI (1970), 652-675.

10. A. M. Geoffrion, "Elements of Large-Scale Mathematical Programming II", Management Science, XVI (1970), 676-691.

11. A. M. Geoffrion, "Duality on Nonlinear Programming: A Simplified Applications-Oriented Development", SIAM Review, XIII, No. 1 (1971), 1-37.

12. A. M. Geoffrion, "Generalized Benders Decomposition", J. Optimization Theory Appl., to appear.

13. R. E. Gomory, "Large and Nonconvex Problems in Linear Programming", Proc. Symp. Interactions between Math. Res. and High-Speed Computing of the A.M.S., XV (1963), 125-239.

14. R. E. Gomory, "On the Relation Between Integer and Non-Integer Solutions to Linear Programs", Proceedings of the National Academy of Sciences, LIII (1965), 260-265.

15. R. E. Gomory, "Some Polyhedra Related to Combinatorial Problems", J. Linear Alg. Appl., II (1969), 451-558.

16. G. A. Gorry and J. Shapiro, "An Adaptive Group Theoretic Algorithm for Integer Programming Problems", Management Science, XVII, No. 5 (1971), 285-306.

17. G. A. Gorry, W. D. Northup, J. F. Shapiro and L. A. Wolsey, "A Group Theoretic Integer Programming Algorithm: System Design and Computational Experience", Proceedings of XXXVI Share Meeting, Los Angeles, March, 1971.

18. G. A. Gorry, J. F. Shapiro and L. A. Wolsey, "Relaxation Methods for Pure and Mixed Integer Programming Problems", Management Science, 18, No. 5 (Jan., Part I, 1972), 229-239.

19. R. Grinold, "Lagrangian Subgradients", Management Science, XVII, No. 3 (1970), 185-188.

20. F. J. Gould, "Extensions of Lagrange Multipliers in Nonlinear Programming", SIAM J. on Appl. Math., XVII, No. 6 (1969), 1280-1297.

21. M. Held and R. M. Karp, "The Traveling Salesman Problem and Minimum Spanning Trees", Operations Research, XVIII, No. 6 (1970), 1138-1162.

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

23. L. S. Lasdon, "Duality and Decomposition in Mathematical Programming", IEEE Trans. System Science and Cybernetics, SSC-4 (1968), 86-100.

24. G. L. Nemhauser, and A. Ullman, "A Note on the Generalized Lagrange Multiplier Solution to an Integer Programming Problem", Operations Research, XVI, No. 2 (1968), 450-453.

25. R. T. Rockafellar, "Duality in Nonlinear Programming", Mathematics of the Decision Sciences, Part I, G. B. Dantzig and A. F. Veinott (eds.), American Mathematical Society, Providence (1968), 401-422.

26. R. T. Rockafellar, Convex Analysis, Princeton, New Jersey: Princeton University Press (1970).

27. J. F. Shapiro, "Generalized Lagrange Multipliers in Integer Programming", Operations Research, XIX, No. 1 (1971), 68-76.

28. J. F. Shapiro, "Turnpike Theorems for Integer Programming Problems", Operations Research, XVIII (1970), 432-440.

29. M. Simonnard, Linear Programming, Englewood Cliffs, New Jersey: Prentice-Hall, Inc. (1966).

30. J. A. Tomlin, "An Improved Branch-and-Bound Method for Integer Programming", Operations Research, 14 (1971), 1070-1074.

31. L. A. Wolsey, "Mixed Integer Programming: Discretization and the Group Theoretic Approach", Technical Report No. 42, MIT Operations Research Center, June, 1969.