A Convergent Duality Theory for Integer Programming

DAVID E. BELL
Cambridge University, Cambridge, England

JEREMY F. SHAPIRO
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received original June 1975; final, June 1976)

We present a constructive procedure for generating a finite sequence of increasingly stronger dual problems to a given integer programming problem. The last dual problem in the sequence yields an optimal solution to the integer programming problem. We show that this dual problem approximates the convex hull of the feasible integer solutions in a neighborhood of the optimal solution it finds. The theory is applicable to any bounded integer programming problem with rational data.

IN THIS PAPER we develop a complete duality theory for any bounded integer programming problem. For expositional convenience, we present the theory for the zero-one integer programming problem:

\[ v = \min c x, \quad Ax = b, \quad x_i = 0 \text{ or } 1, \quad \text{(IP)} \]

where \(A\) is an \(m \times n\) matrix and the coefficients of \(A\), \(b\) and \(c\) are integer. We let \(a_j\) denote a column of \(A\), \(c_j\) denote the cost coefficient of this column, and \(b_i\) denote the \(i\)th component of \(b\). Any bounded IP problem with rational data and inequality as well as equality constraints can be solved by the procedure we present here with only minor modifications. None of our results relies on the zero-one nature of the variables. Since the essence of integer programming is the selection of small integer values for decision variables, the assumption of boundedness is not a serious one.

Fisher and Shapiro [6] give a mathematical programming problem that is dual to (IP) and that may solve it. In all cases their IP dual problem provides lower bounds on \(v\), the minimal objective function value in (IP). The greatest lower bound is shown to be at least as great as the lower bound attainable by solving the linear programming relaxation of (IP) \((0 \leq x_j \leq 1)\), adding all the Gomory cuts, and then solving the restricted linear programming relaxation. Fisher and Shapiro also give a primal-dual ascent algorithm for solving their IP dual problem. Computational experience with it on a variety of IP problems is reported in Fisher, Northup, and Shapiro [5].
If the greatest lower bound given by this dual does not equal $v$, a duality gap exists. Bell [1] and Bell and Fisher [2] have given methods that extend the dual approach to improve the lower bound when a duality gap exists. Bell’s method involves the construction of a sequence of groups of increasing size, called supergroups, which are used to construct new IP dual problems. In this paper we give a more direct method for constructing a sequence of supergroups and IP dual problems terminating with one that yields an optimal solution to (IP).

The main contribution of this paper is a complete and direct integration of constructive abelian group theory and constructive duality theory in the analysis of IP problems. If an optimal solution to a given IP dual problem fails to yield an optimal solution to (IP), then we construct a new and stronger IP dual problem. Details of this approach are given in Section 1, where we contrast it with the alternate method of Bell in [1] and with branch-and-bound, which can be viewed as a dual perturbation technique.

There are some additional novel and advantageous features to the IP dual approach given in this paper. The group-theoretic methods use only the original integer data of (IP), rather than these data transformed by one or more linear programming basis inverses, as has previously been the case for group-theoretic integer programming methods. This feature not only makes the methods easier to understand and use, it also makes them more numerically stable.

Another novel and advantageous feature of our approach is that the Lagrangian calculations, which are part of the dual optimization, produce zero-one solutions that are optimal in (IP) if the right-hand side is suitably modified. Some of these solutions may be attractive if the violated constraints are somewhat soft. Alternatively, it may be possible for some IP problems to heuristically adjust near-feasible and low cost zero-one solutions to good feasible solutions. A final point is that constraining the variables to be zero-one facilitates the use of subgradient optimization to approximately solve the IP dual problems. Subgradient optimization has shown considerable promise as a fast approximate method for discrete optimization problems.

For future reference, we define the finite set $F = \{x | Ax = b, x_j = 0 \text{ or } 1\}$. We allow the possibility that $F$ is empty. It is well known that (IP) can be solved in theory by solving the linear programming problem

$$\min \ cx, \quad x \in [F]$$

(1)

where $[ \ ]$ denotes convex hull. This is because the simplex method applied to (1) will yield an optimal extreme point $x^*$ of $[F]$, which implies $x^* \in F$ and, therefore that $x^*$ is an optimal solution to (IP). Solving (1), however, is not usually a practical approach since $[F]$ is generally very
difficult to state explicitly as a system of inequalities. One of the interpretations of the duality theory presented here is that it approximates \([F]\) in the neighborhood of an optimal solution to (IP).

The following section contains the iterative supergroup and IP dual procedure and the proof of convergence. Section 2 contains a short discussion about the relation of the IP dual problems to the set \([F]\). A numerical example is presented in Section 3. Section 4 gives a procedure that eliminates redundant congruences in the IP dual problem construction. The methods to be presented here can be extended to mixed IP, and a paper on this topic is being prepared. The methods also can be used to perform IP sensitivity analysis; some results are given in [24]. An important area of future research is an integration of these IP dual methods with the convex analysis results of Burdet and Johnson [4].

1. SUPERGROUP CONSTRUCTION AND CONVERGENCE PROOF

The overall strategy presented here for solving (IP) is to construct a sequence of IP dual problems, each one strictly stronger than the previous one, until an optimal solution to (IP) is found or it is proven infeasible. We show how this is done by: (1) constructing an IP dual problem from a given Abelian group; (2) stating the algorithmic principles for solving it; (3) discussing the conditions under which it finds an optimal solution to (IP); and if an optimal solution to (IP) is not found, (4) constructing a supergroup and a stronger IP dual problem. Our development is self-contained and treats only briefly a number of considerations relating the IP dual methods to other IP methods and its practical use in solving IP problems.

Consider the Abelian group \(G = \mathbb{Z}_{q_1} \oplus \mathbb{Z}_{q_2} \oplus \cdots \oplus \mathbb{Z}_{q_r}\) where the \(q_i\) are positive integers and \(\mathbb{Z}_{q_i}\) is the cyclic group of order \(q_i\). Except for the initial group \(G^0 = \mathbb{Z}_{q_1}\), the \(q_i\) will be greater than one. The order of \(G\), denoted by \(|G|\), is \(\prod_{i=1}^{r} q_i\). We also are given a homomorphism \(g\) from the Abelian group \(\mathbb{Z}^m\) consisting of all integer \(m\)-vectors under ordinary addition into \(G\). The homomorphism \(g\) is specified by knowing \(e_i \equiv g(e_i)\), \(i = 1, \ldots, m\), where \(e_i\) is the \(i\)th unit vector in \(\mathbb{Z}^m\). Then for any integer \(m\)-vector \(z\), \(g(z)\) is given by \(g(z) = \sum_{i=1}^{m} z_i e_i\). The homomorphism \(g\) is used to construct the mathematical programming problem

\[
\begin{align*}
 v' &= \min cx, \quad Ax = b, \quad x \in X \\
\end{align*}
\]

where \(X = \{x \mid \sum_{j=1}^{n} x_j = \beta, \quad x_j = 0 \text{ or } 1\}\) and \(\alpha_i \equiv g(a_i), \quad \beta \equiv g(b)\). We use "\(\equiv\)" to denote group equality; for the groups in this paper the equations are systems of congruences. For future reference, let \(X = \{x_1^T\}_{i=1}^{T}\) where \(T\) is at most \(2^n\).

**Lemma 1.** Problem (2) and (IP) are equivalent in the sense that
\( F = \{ x | Ax = b \} \cap X \); that is, they have the same solution sets and therefore the same minimal objective function values.

Proof. The proof that \( F \supseteq \{ x | Ax = b \} \cap X \) is trivial given the zero-one constraints in \( X \). Thus, since \( F \subseteq \{ x | Ax = b \} \) it suffices to show \( F \subseteq X \). To this end, let \( x \) be any solution in \( F \). We have \( x_j = 0 \) or \( 1 \) for all \( j \) and since \( g \) is a homomorphism from \( Z^n \) into \( G \), \( \beta = g(b) = g(Ax) = \sum_{j=1}^{n} g(a_j)x_j = \sum_{j=1}^{n} \alpha_j x_j \), hence \( x \in X \).

Although (2) and (IP) are equivalent problems, the reformulation (2) permits greater resolution when the following IP dual is constructed. For any \( u \in R^n \), define the Lagrangian function

\[
L(u) = ub + \min_{x \in X} (c - uA)x.
\]  

If \( X \) is empty, then (IP) is infeasible and we can take \( L(u) = + \infty \). Otherwise, it is well known and easily shown that \( L(u) \leq v \) and \( L \) is a concave continuous function (Rockafellar [22]). The IP dual problem is constructed by finding the best lower bound

\[
d = \max L(u), \quad u \in R^n. \tag{D}
\]

Although \( L(u) \) may be finite for all finite \( u \), it may be that the maximum in (D) is \( + \infty \), in which case (IP) is once again infeasible.

The relationship between the IP dual problem (D) and (IP) that we seek, but may not find, is summarized by the following.

Optimality Conditions. The pair \( (x^*, u^*) \) with \( x^* \in X \) is said to satisfy the optimality conditions if

\[
\begin{align*}
(i) \quad L(u^*) &= u^*b + (c - u^*A)x^*; \\
(ii) \quad Ax^* &= b.
\end{align*}
\]

It is easily shown that if \( x^* \) and \( u^* \) satisfy the optimality conditions then they are optimal in (IP) and (D), respectively, and \( d = v \). Thus, our strategy in trying to solve (IP) is to compute an optimal solution \( u^* \) to the concave unconstrained maximization problem (D) in the hope that we will find an \( x^* \in X \) satisfying (i) and (ii) to complement it.

A heuristic for obtaining good solutions, which might be useful if some of the constraints in \( Ax = b \) are soft, is available each time \( L \) is computed at a specific \( \bar{u} \). Letting \( \bar{x} \) be an optimal solution in (3), it is easy to see by direct appeal to the optimality conditions that \( \bar{x} \) is optimal in (IP) if its right-hand side is changed to \( A\bar{x} \).

To see how (D) can be optimized we observe that it is equivalent to the large-scale linear programming problem

\[
d = \max \nu, \quad \nu \leq ub + (c - uA)x^t, \quad t = 1, \ldots, T \\
\nu \in R^t, \quad u \in R^m. \tag{4}
\]
For any \( u \in \mathbb{R}^m \), the optimal objective function value in (4) is \( \nu(u) = ub + \min_{t=1, \ldots, T} (c-uA)x^t \) which is simply \( L(u) \). Thus (D) and (4) are equivalent problems. The linear programming dual of (4) is

\[
d = \min \sum_{t=1}^T (cx^t) \lambda_t, \quad \sum_{t=1}^T (Ax^t) \lambda_t = b \\
\sum_{t=1}^T \lambda_t = 1, \quad \lambda_t \geq 0, \quad t = 1, \ldots, T. \tag{5}
\]

Problem (5) has an enormous number of columns, but it could be solved by generalized linear programming. It was found [5, 6] that generalized linear programming has some undesirable features for our purposes, such as (1) it does not provide monotonically increasing lower bounds for use in branch and bound; (2) it has exhibited inconsistent performance; (3) it may not combine well with subgradient optimization. The primal-dual simplex algorithm given in [5, 6] overcomes these difficulties and provides us with insight into optimizing the IP dual problem (D). In particular, the primal-dual simplex algorithm can be used to solve the pair of linear programming problems (4) and (5) and implicitly gives us the necessary and sufficient conditions for the optimality of (D), which we now state.

Consider any \( u \in \mathbb{R}^m \) and let \( T(u) = \{ t | L(u) = ub + (c-uA)x^t \} \). Then the complementary slackness conditions of LP tell us that \( u^* \) is optimal in (D), and equivalently (4), if and only if there is a solution to the system

\[
\sum_{i \in T(u^*)} (Ax^i) \lambda_i = b \\
\sum_{i \in T(u^*)} \lambda_i = 1, \quad \lambda_i \geq 0, \quad t \in T(u^*). \tag{6}
\]

Algorithmically, the primal-dual ascent algorithm given in [5] proceeds by testing a point \( u \in \mathbb{R}^m \) for optimality by trying to establish the conditions (6). If these conditions fail to hold, then a direction of increase of \( L \) is found and the algorithm proceeds to a new point \( u' \) such that \( L(u') > L(u) \).

The set \( T(u) \) may be large, however, and the algorithm in [6] begins with a subset \( T'(u) \) consisting of one or two elements and builds up \( T(u) \) until \( u \) is proven optimal in (D) or \( u' \) is found such that \( L(u') > L(u) \). Computational experience is given in Fisher, Northup, and Shapiro [5].

Suppose the point \( u^* \) is optimal in (D) and thus the following phase-one problem has minimal objective function value equal to zero.

\[
\min \sum_{i=1}^m (s_i^+ + s_i^-), \quad \sum_{i \in T(u^*)} (Ax^i) \lambda_i + Is^+ - Is^- = b \\
\sum_{i \in T(u^*)} \lambda_i = 1, \quad \lambda_i \geq 0 \quad t \in T(u^*), \quad s^+ \geq 0, \quad s^- \geq 0. \tag{7}
\]

Let \( \lambda^*_t, t \in T(u^*) \), denote an optimal solution to (7). The solution \( \bar{x} = \sum_{i \in T(u^*)} \lambda^*_t x^i \) satisfies conditions (i) and (ii) of the optimality conditions, and it is an optimal solution to (IP) if it is integer. Our concern is what to do if \( \bar{x} \) is not integer because more than one \( \lambda^*_t \) in (7) is positive.

**Remark.** The computational effort required to do the Lagrangian minimization (3) is dependent on \(|G|\). The time required is no more than a
few seconds for \(|G|\) up to 3000 (see [12]). Computational experience with
the primal-dual ascent algorithm on a single IP dual problem is given in
[5]. This algorithm can spend too much time picking a direction of ascent
that is an interpretation of the function of (7). An alternative approach
that can be integrated with the primal-dual is subgradient optimization,
which generates a sequence of dual solutions \(\{u^k\}_{k=1}^\infty\) to (D) as follows.
Evaluate \(L(u^k) = u^kB + (c - u^kA)x^j\) for some \(j \in T(u^k)\), and use \(\gamma^k = b - Ax^j\)
as a direction of ascent. The new dual solution in this direction is \(u^{k+1} = u^k + \theta_k \gamma^k\) and if \(\theta_k\) obeys \(\theta_k \to 0^+\) and \(\sum \theta_k \to +\infty\), then it can be shown that
\(L(u^k) \to d\) (Poljak [20]). Finite convergence to any value \(d < d\) can be
achieved if \(\theta_k = \alpha_k (d - L(u^k)) / ||\gamma^k||^2\), where \(||\gamma^k||\) denotes Euclidean norm
and \(\varepsilon_1 < \alpha_k < 2 - \varepsilon_2\) for \(\varepsilon_1 > 0, \varepsilon_2 > 0\). Subgradient optimization has worked
very well on large-scale linear programming problems similar to (4) for
approximating discrete optimization problems (Held and Karp [15], Held,
Wolfe, and Crowder [16], Marsten, Northup, and Shapiro [18]). A hybrid
computational approach is indicated using subgradient optimization as an
opening strategy followed by the primal-dual or some other exact algo-

The following lemma is central to our development, as it identifies when
the IP dual problem (D) provides an optimal solution to (IP).

**Lemma 2.** If only one \(\lambda_k\) is positive in an optimal basic solution to (7) when
the minimal objective function value is zero, then the corresponding solution
\(x^k\) is optimal in (IP). On the other hand, if more than one \(\lambda_k\) is positive, then
all the \(x^k\) corresponding to basic \(\lambda_k\) are infeasible in (IP).

**Proof.** If only one \(\lambda_k\) is positive, then \(\lambda_k = 1\) and we have \(Ax^k = b\). By
construction, we have \(x^k \in X\) and \(L(u^k) = u^kb + (c - u^kA)x^k\). Thus, \((x^k, u^k)\)
satisfy the optimality conditions and \(x^k\) is optimal in (IP).

In the latter case, suppose \(x^k, k = 1, \ldots, K\), corresponds to basic \(\lambda_k\) in
(7), and suppose to the contrary that \(x^k\) satisfies \(Ax^k = b\) but \(0 < \lambda_1 < 1\). Then

\[
\begin{pmatrix}
  b \\
  1
\end{pmatrix}
\lambda_1 + \sum_{k=2}^K \begin{pmatrix}
  Ax^k \\
  1
\end{pmatrix} \lambda_k = \begin{pmatrix}
  b \\
  1
\end{pmatrix}
\text{ or } \sum_{k=2}^K \begin{pmatrix}
  Ax^k \\
  1
\end{pmatrix} \lambda_k = \begin{pmatrix}
  b \\
  1
\end{pmatrix} (1 - \lambda_1).
\]

This in turn implies (since \(\lambda_1 < 1\))

\[
\sum_{k=2}^K \begin{pmatrix}
  Ax^k \\
  1
\end{pmatrix} \lambda_k / (1 - \lambda_1) = \begin{pmatrix}
  b \\
  1
\end{pmatrix} = \begin{pmatrix}
  Ax^j \\
  1
\end{pmatrix};
\]

that is, \((Ax^j, 1)^T\) can be written as a linear combination of the other
columns in the optimal basis in (7), which is impossible.

Using the result of Lemma 2, we will show how to derive a new group
equation (i.e., system of congruences) that eliminates from the resulting
Dual Problems for Integer Programming

dual those infeasible $x^k$ that were taken in convex combination to provide an optimal solution to (D). There are two cases to distinguish—when the maximal dual objective function value $d$ is fractional and when it is integer.

**Case 1** ($d$ fractional). This is the easier case to resolve, as shown by

**Lemma 3.** Suppose the maximal dual objective function value $d$ in the IP dual(D) is fractional. Then the congruence

\[(w^* A)x \equiv w^* b \pmod{W}\]  \hspace{1cm} (8)

admits all feasible IP solutions but excludes all $x^t$ for $t \in T(u^*)$, where $d = L(u^*)$ and $u^* = w^* / W$ for $w^*$ integer.

**Proof.** Note first that $d = L(u^*)$ fractional implies $u^*$ is fractional since $L(u^*) = cx^t + u^*(b - Ax^t)$ for all $t \in T(u^*)$ and $cx^t$ and $b - Ax^t$ are integer. Without loss of generality, we can assume $u^*$ is rational, say $u^* = w^* / W$ for $w^*$ integer, because there is always a rational $u^*$ that is optimal in (D) and the primal-dual ascent algorithm from [5 and 6] generates only rational $u$ if the initial dual solution is rational.

Any feasible IP solution $x$ satisfies $cx + u^*(b - Ax) \equiv 0 \pmod{1}$ since $cx$ is integer and $b - Ax = 0$. This implies $w^*/W(b - Ax) \equiv 0 \pmod{1}$ or $(w^* A)x \equiv w^* b \pmod{W}$. On the other hand, for all $t \in T(u^*)$, $d = L(u^*) = cx^t + u^*(b - Ax^t)$ equals a fraction although $cx^t$ is integer, which implies $u^*(b - Ax^t) \equiv 0 \pmod{1}$ or $(w^* A)x^t \not\equiv w^* b \pmod{W}$.

The implication of Lemma 3 is that when $d$ is fractional, we add the congruence (8) to the set $X$ defined in the reformulation (2) of (IP). This has the effect of extending the group underlying the dual construction to $G \oplus Z_w$, and the vector $w^*$ provides the homomorphic mapping of $Z^m$ onto $Z_w$.

**Case 2** ($d$ integer). This case is somewhat more difficult because when $d$ is integer, there may be some $x^t$ for $t \in T(u^*)$, as yet undiscovered, which is an optimal solution in (IP). The method about to be presented is also applicable when $d$ is fractional, but the congruence (8) is easier to obtain and no less powerful in effect.

The crucial observation for modifying (D) when it fails to solve (IP) is that (IP) is equivalent to another integer programming problem:

\[
\min \sum_{t=1}^T (cx^t) \lambda_t, \quad \sum_{t=1}^T (Ax^t) \lambda_t = b \\
\sum_{t=1}^T \lambda_t = 1, \quad \lambda_t \geq 0 \quad \text{and integer.}
\]  \hspace{1cm} (9)

The linear programming problem (5) is the linear programming relaxation of (9). The optimal solution to (7) is also optimal in the LP relaxation of (9). Hence if more than one $\lambda_k$ is positive in (7) when the minimal objective function value is zero, we may continue our IP dual analysis by
applying the Smith reduction procedure to this optimal basis (see Wolsey [25] or Garfinkel and Nemhauser [8]). The result of the Smith reduction is a homomorphism $h$ from the Abelian group $Z^{m+1}$ into the Abelian group $H = Z_{p_1} \oplus \cdots \oplus Z_{p_t}$, where $\prod_{i=1}^{t} p_i$ equals the absolute value of the determinant of the optimal basis in (7). The supergroup we will use in this case is $G \oplus H$.

The properties of the homomorphism $h$ that we need to identify for future use are

(i) For $z \in Z^{m+1}$, $h(z) = \sum_{i=1}^{m+1} z_i \rho_i$ where $\rho_i \equiv h(e_i)$;
(ii) For $\lambda_i$ basic in (7) $h(Ax^i, 1) = \sum_{i=1}^{m+1} \sum_{j=1}^{n} a_{ij} x^i_j \rho_i + \rho_{m+1} \equiv 0$;
(iii) $h(b) = \sum_{i=1}^{n} b_i \rho_i + \rho_{m+1} \neq 0$;
(iv) $\rho_i \equiv 0$ if $s_i^+$ or $s_i^-$ is basic in (7).

Property (i) is true because $h$ is a homomorphism. Actually, the Smith reduction procedure produces the $\rho_i$, which are then used to compute $h(z)$ for all $z \in Z^{m+1}$. From (iv) we see that many of the $\rho_i$ will be 0 because many of the $s_i^+$ and $s_i^-$ will be basic at a zero level. Computational experience thus far with the IP dual methods indicates that few $\lambda_i$ columns will be present when (7) is found to have objective function value equal to zero. Property (iii) holds because the optimal solution in (7) is assumed to be fractional.

The group $H$ and the homomorphism $h$ are used to define the set $X' = \{ x | \sum_{j=1}^{n} \alpha'_j x_j = \beta', x_j = 0 \text{ or } 1 \}$, where $\alpha'_j = \sum_{i=1}^{m} a_{ij} \rho_i$ and $\beta' = \sum_{i=1}^{m} b_i \rho_i$. As with $X$, we have $F \subseteq X'$. The critical property of this construction is given by the following lemma.

**Lemma 4.** If more than one $\lambda_k$ is positive in an optimal basic solution to (7) with objective function value equal to zero, then the infeasible solutions $x^1, \cdots, x^K \in X$ corresponding to the basic $\lambda_k$ are not contained in $X'$.

**Proof.** Since the columns $(Ax^k, 1)^\tau$ are basic, by property (ii) of the homomorphism $h$, we have $\sum_{j=1}^{n} \left( \sum_{i=1}^{m} a_{ij} \rho_i \right) x^k_j = -\rho_{m+1}$ or, by the definition of $\alpha_j$

$$\sum_{j=1}^{n} \alpha'_j x^k_j \equiv -\rho_{m+1}. \quad (10)$$

On the other hand, by property (iii) of $h$ we have $\sum_{i=1}^{m} b_i \rho_i \neq -\rho_{m+1}$ or, by the definition of $\beta'$

$$\beta' \neq -\rho_{m+1}. \quad (11)$$

The lemma is established by comparing (10) and (11).

**Remark.** Note that the construction in Case 2 does not require $d$ to be integer and could be used in Case 1 as well. The construction of the supergroup can clearly be done with respect to any basis in (7) with more than one $\lambda_k > 0$ and not just from a basis giving an objective function value of zero. Alternatively, the congruence (8) can be derived from any fractional
dual solution. This is important since convergence to the exact optimal IP
dual solution may be slow (see [5]).

The implication of Lemma 4 is that we want to redefine the IP dual
problem so that we now require \( x \in X \cap X' \). The Lagrangian minimization
(3) remains qualitatively the same calculation, but now over a larger
group \( G \oplus H \). It may be that the group constraints in \( X' \) imply some of
those defining \( X \). Since it is important to keep down the order of the group,
a check should be made to eliminate such redundant equations. In Section
4 we show how a simple modification of problem (7) guarantees that
\( X' \subseteq X \) so that all of \( G \) may be discarded. This has the additional advantage
that the Smith reduction procedure gives a minimal representation for
\( G \oplus H \). Since this is not an essential part of the development, we retain
for now the current form of (7).

**Iterative Dual Method**

*Step 0 (Initialization).* Start with \( G^0 = Z_1 \) and construct the IP dual
problem to (IP)

\[
d^0 = \max L^0(u), \quad u \in R^m \tag{D^0}
\]

where \( L^0(u) = ub + \min_{x \in X^0} (c - uA)x \), and \( X^0 = \{ x | x_j = 0 \text{ or } 1 \} \). Go to
Step 1 with \( l = 0 \).

*Step 1.* Solve the IP dual problem (D\(^l\)); if \( d^l = + \infty \), (IP) is infeasible
and the method is terminated. Otherwise, let \( u^l \) denote an optimal solution
to (D\(^l\)) and let \( x^l \) be the convex combination of points in \( X^l \) satisfying
the IP dual optimality conditions (6). If \( x^l \) is integer, then it is optimal
and the method terminates. If \( x^l \) is not integer, go to Step 2.

*Step 2.* If \( d^l \) is fractional, add (8) to the equations defining \( X^l \) and thus
define a proper subset \( X^{l+1} \) of \( X^l \). If \( d^l \) is integer, use the Smith reduction
procedure to construct the group \( H \) from the optimal basis to (7) and
add the constraints

\[
h \begin{pmatrix} Ax \\ 1 \end{pmatrix} = h \begin{pmatrix} b \\ 1 \end{pmatrix}
\]

to those defining \( X^l \) to define a proper subset \( X^{l+1} \) of \( X^l \). Let \( G^{l+1} = G^l \oplus H \)
and \( L^{l+1} \) the Lagrangian defined over \( X^{l+1} \). The new dual problem is
\( d^{l+1} = \max L^{l+1}(u), \quad u \in R^m \). Return to Step 1.

*Remark.* The initial IP dual problem (D\(^0\)) is simply the linear program-
ming relaxation of (IP) when \( x_j = 0 \) or 1 is replaced by \( 0 \leq x_j \leq 1 \) (see
Nemhauser and Ullman [19]).

**Theorem 1.** The iterative dual procedure converges finitely to an optimal
solution to (IP) or proves (IP) is infeasible.

*Proof.* The solution of each IP dual problem (D\(^l\)) defined over \( X^l \subseteq X^0 \)
is finite because \(X^t\) is finite, implying \((D^t)\) is a linear programming problem. If the solution \(x^t = \sum_{i \in T(w)} \lambda_i x_i^t\) from (7) is not integer, then the new set \(X^{t+1}\) satisfies \(|X^{t+1}| \leq |X^t| - 2\). This is true because nonintegrality of \(x^t\) implies by Lemma 2 that at least two \(\lambda_i\) are positive in the optimal solution to (7) with objective function value equal to zero, which in turn implies by Lemmas 3 and 4 that at least two solutions from \(X^t\) are eliminated in the construction of \(X^{t+1}\). Clearly, since \(X^0\) is finite, this reduction process must terminate finitely with an optimal solution to (IP) in the case \(F \neq \phi\) or with \(d^l = +\infty\) for some \(l\) in the case \(F = \phi\).

Remark. Total enumeration of the set \(X^0 = \{x|x_j = 0\ or\ 1\}\) would be another finitely convergent procedure for (IP). The strength of the IP dual approach is that the sets \(X^t\) are considerably smaller than \(X^0\), and the computation of \(L^t(u)\) for various values of \(u\) does not explicitly use more than a small fraction of \(X^t\). The practical imperfection in the IP dual theory is that \(|G^t|\) may grow too large, although some measures are possible to effectively reduce \(|G^t|\). Another positive feature of the IP dual approach is the monotonically increasing lower bounds \(L^t(u)\) for use in branch-and-bound (see [6]).

Remark. Alternative methods are available for strengthening the IP dual problem when more than one \(\lambda_k\) is positive in (7) and the minimal objective function value is zero. For example, since \(Ax^k = b\) for \(\lambda_k > 0\) by Lemma 2, select a row \(i\) such that \(\sum_{j=1}^n a_{ij} x_j^k = b_i \neq 0\). Select an integer \(p_i\) such that \(p_i\) does not divide \(d_{ij}\), and add the congruence \(\sum_{j=1}^n a_{ij} x_j^k = b_i \text{ (mod } p_i\text{)}\) to the constraints of \(X\). It can be easily verified that \(x^k\) does not satisfy this congruence. This is the approach of [1].

A second method for continuing the dual analysis is branch-and-bound. Again suppose \(\lambda_1 > 0, \ldots, \lambda_K > 0\) in (7) with \(K \geq 2\). Since the columns \((Ax^1, 1)^T, \ldots, (Ax^K, 1)^T\) are linearly independent, there must be some component \(j\) and indices \(k_1\) and \(k_2\) such that \(x_j^{k_1} = 1\) and \(x_j^{k_2} = 0\). We branch on this variable creating two subproblems, one with \(x_j\) fixed at zero and the other with \(x_j\) fixed at one. The result is that the necessary and sufficient optimality conditions (6) would be destroyed for each subproblem. In one subproblem \(x^{k_1}\) would be excluded from these conditions, and in the other \(x^{k_2}\) would be excluded.

It should be clear that the sequential IP dual problem construction given in the iterative dual method is stronger than either of these alternative methods because it simultaneously eliminates all the \(x^t\) for \(t \in T(u^*)\) when \(d\) is fractional and all the infeasible \(x^k\) for \(k \in T(u^*)\) with \(\lambda_k > 0\) in (7) when \(d\) is integer.

2. RELATION TO THE CONVEX HULL OF FEASIBLE INTEGER SOLUTIONS

The equivalence between dualization and convexification of mathematical programming problems (Magnanti, Shapiro, and Wagner [17]) per-
mits us to give a convex analysis interpretation of the results of the previous section.

**Lemma 5.** \( \{ x | x = \sum_{i=1}^{n} x^i \lambda_i, \lambda_i \text{ feasible in the IP dual problem} \} = \{ x | Ax = b, 0 \leq x_j \leq 1 \} \cap [X] \).

**Proof.** The proof is straightforward and is therefore omitted.

**Theorem 2.** The IP dual problem is equivalent to

\[
d = \min cx, \quad x \in \{ x | Ax = b, 0 \leq x_j \leq 1 \} \cap [X].
\]

**Proof.** The IP dual problem can be written as the linear programming problem (5). In view of Lemma 5, the result follows immediately.

Thus, a given IP dual problem approximates (IP) by minimizing the objective function over the intersection of the linear programming feasible region with the convex polyhedron \([X]\). When the IP dual problem solves (IP) in the sense that the optimality conditions are found to hold, then \([X]\) has cut off enough of the LP feasible region for an optimal solution to (IP) to be discovered. Thus, in this case the IP dual problem has found a local approximation to \([F]\), the convex hull of feasible integer solutions, in the neighborhood of an optimal solution.

**3. NUMERICAL EXAMPLE**

Consider the zero-one IP problem

\[
v = \min -4x_1 - 7x_2 + 2x_3 + 4x_4 - x_5 + 8x_6 + 2x_7 \\
2x_1 - x_2 + x_3 + 3x_4 - 2x_5 + 2x_6 - 2x_7 = 1 \\
-4x_1 - 3x_2 + 2x_3 + x_4 + x_5 + 2x_6 - 2x_7 = -4 \\
x_j = 0 \text{ or } 1, \quad j = 1, \ldots, 7.
\]

The first IP dual problem \((D^0)\) with \(G^0 = Z_1\) is the dual to the linear programming relaxation. The optimal solution to \((D^0)\) is \(u^0 = (1, 1)\) with dual objective function value \(L(u^0) = d^0 = -9\). Since this is integral, we examine the optimal basis for (7), which has an objective function value equal to zero and on which the Smith reduction is performed. The matrix corresponding to this basis is

\[
\begin{pmatrix}
0 & 5 & 0 \\
1 & -4 & -4 \\
0 & 1 & 1
\end{pmatrix}
\]

It is derived from the optimal solution \(x_1 = x_2 = x_3 = 1, \ x_4 = 1/5, \ x_5 = 4/5, \ x_6 = x_7 = 0\) to \((D^0)\), which is a convex combination of the solutions \(x_1 = x_2 = x_3 = x_4 = 1\) and \(x_1 = x_2 = x_3 = x_5 = 1\) (ignoring variables at zero).
This basis induces the group $Z_5$, and the homomorphism $g$ from $Z^3$ onto $Z_5$ is given by $g(e_1) = 1$, $g(e_2) = g(e_3) = 0$. We use this homomorphism to derive the new congruence

$$g \left( \begin{array}{c} Ax \\ 1 \end{array} \right) = g \left( \begin{array}{c} b \\ 1 \end{array} \right) \pmod{5},$$

which is

$$2x_1 + 4x_2 + x_3 + 3x_4 + 3x_5 + 2x_6 + 3x_7 = 1 \pmod{5}.$$  \hspace{1cm} (17)

The IP dual problem $(D^1)$ derived from (16) with this new congruence defining the subset of integer solutions $X^1$ has the optimal solution $u^1 = (0, 1, 1, \frac{1}{2})$ with dual objective function value $d^1 = L(u^1) = -7.5\frac{5}{8}$. The optimal solutions in $X^1$ are $x_2 = x_3 = x_4 = x_5 = 1$ and $x_1 = x_2 = 1$. The fractional value of the objective function allows us to add the congruence $u^t A x = u^t b \pmod{1}$ or $4x_1 + 5x_2 + 2x_3 + x_4 + x_5 + 6x_7 = 4 \pmod{8}$ in the definition of the set $X^2$.

The IP dual problem $(D^2)$ has the optimal solution $u^2 = (0, 0)$ with one optimal solution from $X^2$, $x_1 = x_3 = x_7 = 1$. This is the optimal solution to the IP since the optimality conditions are satisfied.

4. THE SUPERGROUP REPRESENTATION

After the first iteration of the algorithm we are able to represent the group easily as a direct sum of cyclic subgroups $G = Z_{q_1} \oplus \cdots \oplus Z_{q_r}$, where each $q_{i+1}|q_i$, by performing the Smith reduction on the optimal basis to (7). Subsequent groups are not represented so efficiently because each is defined as a sum, not necessarily a direct sum, of the previous group and a new set of cyclic groups derived from the new optimal basis of (7).

By making a slight modification in our procedure, we can derive iteratively an integer matrix from which the whole subgroup $G^K$ may be, by a Smith reduction, represented as a direct sum of cyclic subgroups. Gorry, Northup, and Shapiro [12] have observed empirically that the number of such subgroups is small and apparently independent of the group size, which in this case would mean a compact representation for $G^K$. As shown by the following lemma, the modification also removes some of the redundancies among the derived group constraints introduced by the artificial convexity row in (7).

**Lemma 6.** Suppose that $g$ is the homomorphism from $Z^{m+1}$ onto the group $G$ derived from a Smith reduction on an $(m+1)$ by $(m+1)$ integer matrix

$$M = \begin{pmatrix} Q & c \\ r & 1 \end{pmatrix},$$

and that $g'$ is the homomorphism from $Z^m$ onto the group $G'$ derived from a Smith reduction on the $m \times m$ integer matrix $M' = Q - rc$, where the outer
product is used to form the matrix \( rc \). Then, the two sets of group constraints

\[
g\left(\begin{array}{c} Ax \\ 1 \end{array}\right) = g\left(\begin{array}{c} b \\ 1 \end{array}\right) \quad (18)
\]

\[
g'(Ax) = g'(b) \quad (19)
\]

have the same set of integer solutions.

**Proof.** The constraints (18) are equivalent, by the nature of the Smith reduction, to

\[
M^{-1}\left(\begin{array}{c} Ax \\ 1 \end{array}\right) \equiv M^{-1}\left(\begin{array}{c} b \\ 1 \end{array}\right) \pmod{1}, \quad (20)
\]

which by subtracting \( M^{-1}(a) \) from each side is equivalent to

\[
M^{-1}\left(\begin{array}{c} Ax \\ 0 \end{array}\right) \equiv M^{-1}\left(\begin{array}{c} b \\ 0 \end{array}\right) \pmod{1}. \quad (21)
\]

Let

\[
M^{-1} = \left(\begin{array}{c|c} \tilde{Q} & \tilde{r} \\ \hline \tilde{s} & 1 \end{array}\right).
\]

Then (21) is equivalent to

\[
\tilde{Q}Ax \equiv \tilde{Q}b \pmod{1} \quad (22a)
\]

and

\[
\tilde{r}Ax \equiv \tilde{r}b \pmod{1}. \quad (22b)
\]

The relation \( MM^{-1} = I \) yields two relevant equations, \( r\tilde{Q} + \tilde{r} = 0 \) and \( \tilde{Q} = (Q - rc)^{-1} \). The first shows that (22b) is an integer combination of the constraints in (22a) and is thus redundant. The second shows that (22a) is equivalent to

\[
(Q - rc)^{-1}(Ax - b) \equiv 0 \pmod{1}. \quad (23)
\]

By definition of \( g' \) this is equivalent to

\[
g'(Ax) = g'(b). \quad (24)
\]

The chain of equivalent relations (18), (20), (21), (23), (24), (19), proves the lemma.

For the moment let us assume that new group constraints are generated at each iteration from an optimal basis to problem (7). We will use induction and assume that at iteration \( K \) there exists an \( m \times m \) integer matrix \( M_K \) from which a homomorphism \( \bar{g}^K \) is defined such that \( \bar{g}^K(Ax) \equiv g^K(b) \) is equivalent to

\[
g^K\left(\begin{array}{c} Ax \\ 1 \end{array}\right) = g^K\left(\begin{array}{c} b \\ 1 \end{array}\right),
\]

the set of group equations defining \( X^K \).
Replace problem (7) by the following nearly equivalent one,

\[
\min \sum_{i=1}^{n} (s_i^+ + s_i^-),
\]

\[
\sum_{i=1}^{T} \lambda_i \begin{pmatrix} Ax_i - b \\ 1 \end{pmatrix} + \begin{pmatrix} M_K s^+ \\ 0 \end{pmatrix} - \begin{pmatrix} M_K s^- \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}
\]

(7')

\[s^+, s^- \geq 0, \quad \lambda_i \geq 0 \quad t = 1, \cdots, T.\]

Suppose that its optimal basis is \( M \) when the objective value is zero. Now each \( x^t \in X_K \) satisfies \( M_K^{-1}(Ax^t - b) \equiv 0 \pmod{1} \) or, equivalently, \( Ax^t - b = M_K y^t \) for some integer vector \( y^t \). Thus it is possible to express \( M \) as

\[
\begin{pmatrix} M_K & 0 \\ 0 & 1 \end{pmatrix} Y \quad \text{where} \quad Y = \begin{pmatrix} I' \vert Y' \end{pmatrix}
\]

and \( I' \) is a matrix having columns of the form \( \pm e_i \) and \( Y' = (y^1, \cdots, y^t) \), corresponding to the basic \( x \) solutions.

The new homomorphism derived from the basis \( M \), \( h^{K+1} \) is thus equivalent to the set of group constraints

\[
Y^{-1} \begin{pmatrix} M_K^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Ax - b \\ 0 \end{pmatrix} \equiv \begin{pmatrix} 0 \\ 0 \end{pmatrix} \pmod{1}.
\]

(25)

Now \( Y \) is an integer matrix, so that these constraints imply that \( M_K^{-1}(Ax - b) \equiv 0 \pmod{1} \) for all solutions that satisfy (25).

Stated more directly, the constraints

\[
h^{K+1} \begin{pmatrix} Ax \\ 1 \end{pmatrix} \equiv h^{K+1} \begin{pmatrix} b \\ 1 \end{pmatrix} \quad \text{imply those of} \quad g^K \begin{pmatrix} Ax \\ 1 \end{pmatrix} \equiv g^K \begin{pmatrix} b \\ 1 \end{pmatrix}.
\]

Hence, the matrix \( M \) is sufficient information from which to derive the next homomorphism \( g^{K+1} \). Moreover, since \( M \) is of the form used in Lemma 6, an \( m \times m \) integer matrix \( M_{K+1} \) may be defined from which an equivalent set of group constraints may be derived. Thus the existence of \( M_K \) and the use of (7') ensure the existence of \( M_{K+1} \) equal to \( Q_K - r_K e_K \). It remains only to remark that at iteration 0 we have \( M_0 = I_m \), the identity matrix, since the constraints \( I_M^{-1}(Ax - b) \equiv 0 \pmod{1} \) are vacuous.

This neat representation is spoiled if new group constraints are derived by means other than from the optimal basis to (7'), e.g., from the Lagrangian constraint (8). It is, nevertheless, possible to derive an integer matrix \( N_{K+1} \) from which \( g^{K+1} \) is defined directly and non-redundantly. The matrices \( M_K \) are kept as before except that \( M_K \) is not updated if the group constraints from the optimal basis to (7') are not used. Suppose that the set of additional group constraints to be used at iteration \( K+1 \) is \( C_{K+1}(Ax - b) \equiv 0 \pmod{1} \) where, for example, one row of \( C_{K+1} \) will be \( -u^K \) if the constraint \( L(u_K, x) \equiv 0 \pmod{1} \) is used. Thus, the group
constraints defining $X^{K+1}$ are
$$
\begin{pmatrix}
M_{K+1} & 0 \\
C_{K+1} & I
\end{pmatrix}
\begin{pmatrix}
Ax - b \\
0
\end{pmatrix} \equiv \begin{pmatrix}
0 \\
0
\end{pmatrix} \pmod{1}.
$$

The inverse of the matrix on the left is
$$
\begin{pmatrix}
M_{K+1} & 0 \\
-C_{K+1}M_{K+1} & I
\end{pmatrix};
$$
hence if we define
$$
N_{K+1} = \begin{pmatrix}
M_{K+1} & 0 \\
-D_{K+1}C_{K+1}M_{K+1} & D_{K+1}
\end{pmatrix},
$$
where $D_{K+1}$ is a diagonal matrix that clears the fractions in $C_{K+1}$, then the homomorphism $g^{K+1}$ may be obtained directly from $N_{K+1}$ via the Smith reduction. Note that it is $M_{K+1}$, not $N_{K+1}$, that is still used in (7').

The details of this section may have disguised the simple underlying procedure. What we have by this modification is a capability to retain the power of the Smith representation of a group by maintaining a matrix that completely defines the current group. Simultaneously we have removed some of the redundancies among the constraints.

ACKNOWLEDGMENT

This research was supported in part by the U. S. Army Research Office (Durham) under Contract Nos. DAHC04-63-C-0032 and DAAG29-76-C-0064.

REFERENCES