CUTTING PLANES FROM CONDITIONAL BOUNDS: A NEW APPROACH TO SET COVERING

by

Egon Balas

DRC-70-3-79

September 1979

Graduate School of Industrial Administration
Carnegie-Mellon University
Pittsburgh, PA 15213

This report was prepared as part of the activities of the Management Sciences Research Group, Carnegie-Mellon University, under Grant MCS76-12026 A02 of the National Science Foundation and Contract N0014-75-C-0621 NR 047-048 with the U.S. Office of Naval Research.

Abstract

A conditional lower bound on the minimand of an integer program is a number which would be a valid lower bound if the constraint set were amended by certain inequalities, also called conditional. If such a conditional lower bound exceeds some known upper bound, then every solution better than the one corresponding to the upper bound violates at least one of the conditional inequalities. This yields a valid disjunction, which can be used to partition the feasible set, or to derive a family of valid cutting planes. In the case of a set covering problem, these cutting planes are themselves of the set covering type. The family of valid inequalities derived from conditional bounds subsumes as a special case the Bellmore-Ratliff inequalities generated via involutory bases, but is richer than the latter class and contains considerably stronger members, where strength is measured by the number of positive coefficients. We discuss the properties of the family of cuts from conditional bounds, and give a procedure for generating strong members of the family. Finally, we outline a class of algorithm based on these cuts. Our approach was implemented and extensively tested in a computational study whose results are reported in a companion paper [2]. The algorithm that emerged from the testing seems capable of solving considerably larger set covering problems than earlier methods.

CUTTING PLANES FROM CONDITIONAL BOUNDS:

A NEW APPROACH TO SET COVERING

by

Egon Balas

1. Introduction

We consider the set covering problem

$$\text{(SC)} \qquad \min \{cx \mid Ax \geq e, \; x_j = 0 \text{ or } 1, \; j \in N\}$$

where $A = (a_{ij})$ is $m \times n$, $e \in R^m$, $e = (1,\ldots,1)$, $c \in R^n$, and $a_{ij} \in \{0,1\}$, $i \in M = \{1,\ldots,m\}$, $j \in N = \{1,\ldots,n\}$. We will denote by a^i and a_j the i-th row and j-th column of A, respectively. Without loss of generality, we assume that $c_j > 0$, $\forall j \in N$. Using established terminology, we call a vector x satisfying the constraints of (SC) a _cover_, and the set of indices j such that $x_j = 1$, the _support_ of the cover. A cover is called _prime_ if no proper subset of its support defines a cover.

This problem, and its equality-constrained counterpart, the set partitioning problem, are useful mathematical models for a great variety of scheduling and other important real world problems, like crew scheduling, truck delivery, tanker routing, information retrieval, fault detection, stock cutting, offshore drilling platform location, etc., and a literature of considerable size exists on solution methods for these models (see [9] for a survey of set covering and set partitioning; [7] for a computational study and comparison of several solution techniques; and [4] for a more recent survey of set partitioning, which also contains a bibliography of applications of both models).

In this 'paper we propose a new approach to set covering, based on the idea of conditional bounds. In section 2 we introduce this concept for arbitrary mixed integer programs, and show how it can be used to derive valid disjunctions. The latter in turn can be used either to partition the feasible set in the framework of a branch and bound approach, or to derive a family of valid cutting planes. In case of a set covering problem, the cutting planes derived from conditional bounds are themselves of the set covering type. These cuts are discussed in section 3, where the Bellmore-Ratliff inequalities generated from involutory bases are shown to be a special case of the larger family of inequalities defined in this paper. In section 4 we examine some basic properties of our cutting planes. The family of cuts from conditional bounds is rather large, and in section 5 we discuss a procedure for generating "strong" members of the family. Section 6 outlines a class of algorithms based on the cutting planes introduced in this paper, and using heuristics as well as subgradient optimization rather than the simplex method. Several versions of this approach were implemented and tested computationally in a joint study of Andrew Ho and the author, that is summarized in a companion paper [2]. The algorithm that emerged from this testing seems capable of solving larger problems in less time and more reliably than earlier methods.

The approach discussed here was first circulated under [1].

2. Disjunctions from Conditional Bounds

The central idea of our approach is to derive valid inequalities for the set covering problem from conditional bounds. Since this concept is valid and useful for arbitrary mixed integer programs, we will introduce it in this more general context.

In solving pure or mixed integer programs by branch and bound, if the feasible set is tightly constrained, it is sometimes possible to derive disjunctions stronger than the usual dichotomy on a single variable. On the other hand, the feasible set of any integer program becomes more or less tightly constrained after the discovery of a "good" solution (in particular, of an optimal solution), provided that one restricts it to those solutions better than the current best. Such a "tightly constrained" state of the feasible set can often be expressed in the form of an inequality $\pi x \leq \pi_0$, with $\pi \geq 0$ and $\pi_0 > 0$, as will be discussed later on. The smaller π_0 relative to the other coefficients π_j, the tighter the inequality. Whenever such an inequality is at hand, the following result can be used to generate a valid disjunction.

Here we denote disjunction by the symbol V, and the meaning of

$$\bigvee_{i=1}^{k} A_i = A_1 \vee A_2 \vee \dots \vee A_k$$

is that at least one of the conditions $A_1, \dots, A_k$ must hold.

<u>Theorem 1</u>. Let $\pi \in R^n$, $\pi_0 \in R_+$, $N = \{1,\dots,n\}$, and $Q_i \subseteq N$, $i = 1,\dots,p$, $1 \leq p \leq n$. There exists $v \in R^p_+$ such that

(1) $$\sum_{i \mid j \in Q_i} v_i \leq \pi_j, \quad j \in N$$

and

(2) $$\sum_{i=1}^{p} v_i > \pi_0,$$

if and only if every integer $x \in R^n_+$ that satisfies $\pi x \leq \pi_0$ also satisfies the disjunction

(3) $$\bigvee_{i=1}^{p} (x_j = 0, \ j \in Q_i).$$

<u>Proof</u>. Let $G = (g_{ij})$ be the $p \times n$ matrix defined by

(4) $$g_{ij} = \begin{cases} 1 & j \in Q_i \\ 0 & j \in N \setminus Q_i \end{cases} \quad i = 1,\dots,p,$$

and let $e = (1,\dots,1)$ have p components. From (1) and (2), G contains as a submatrix the identity matrix of order p, whose columns are $j(i)$, $i = 1,\dots,p$.

From Farkas' Theorem of the Alternative (nonhomogeneous version, see Duffin [8]), one and only one of the following two systems has a solution (here T denotes transpose):

$$\text{I} \begin{cases} \pi x \le \pi_0 \\ Gx \ge e \\ x \ge 0 \end{cases} \qquad \text{II} \begin{cases} ev > \pi_0 \\ G^T v \le \pi \\ v \ge 0 \end{cases}$$

System II is the same as (1), (2), and $v \in R^m_+$. Thus there exists $v \in R^m_+$ satisfying (1) and (2) if and only if system I has no solution, i.e., if and only if every $x \in R^n_+$ such that $\pi x \le \pi_0$, violates at least one inequality of $Gx \ge e$. But an integer $x \in R^n_+$ violates the i-th inequality of $Gx \ge e$, i.e., the inequality

$$\sum_{j \in Q_i} x_j \ge 1,$$

if and only if it satisfies $x_j = 0$, $j \in Q_i$; hence it violates at least one inequality of $Gx \ge e$ if and only if it satisfies the disjunction (3).||

<u>Example 1</u>. The inequality

$$9x_1+8x_2+8x_3+7x_4+7x_5+6x_6+6x_7+5x_8+5x_9+5x_{10}+4x_{11}+4x_{12}+3x_{13}+3x_{14}+3x_{15}+2x_{16}+2x_{17} \le 10,$$

together with the condition $x \ge 0$, x_j integer, $\forall j$, implies the disjunction

$$(x_j = 0,\ j = 1,2,3,4,5,6,7) \vee (x_j = 0,\ j = 1,8,9,10,11,12,13,14) \vee$$

$$\vee (x_j = 0,\ j = 2,3,8,9,10,15,16,17).$$

Indeed, setting $v_1 = 6$, $v_2 = 3$ and $v_3 = 2$, we have $6 + 3 + 2 > 10$, i.e., (2) holds; and defining the sets Q_i, $i = 1, 2, 3$, to be those used in the above disjunction, condition (1) is satisfied. This can easily be seen from Table 1, whose rows are the incidence vectors of the sets Q_i, while the numbers on top are the π_j and those to the right are the v_i. The columns of the table correspond to the inequalities (1), which for the

9	8	8	7	7	6	6	5	5	5	4	4	3	3	3	2	2	
1	1	1	1	1	1	1											6
1							1	1	1	1	1	1	1				3
	1	1					1	1	1					1	1	1	2

Table 1.

vector v = (6, 3, 2) are $6 + 3 \leq 9$, $6 + 2 \leq 8, \ldots, 2 \leq 2$, all satisfied.

Remark 1.1 Theorem 1 remains true if $\pi x \leq \pi_0$ is replaced by $\pi x < \pi_0$ and (2) is replaced by

$$(2') \qquad \sum_{i=1}^{p} v_i \geq \pi_0 .$$

Proof. If the indicated changes are made in systems I and II, the Theorem of the Altenative still holds. ||

One way of obtaining a "tight" inequality $\pi x \leq \pi_0$ (or $\pi x < \pi_0$) in order to derive from it a conveniently strong disjunction, is as follows. Consider the mixed integer program

$$(P) \qquad \min \{cx \mid Ax \geq b,\ x \geq 0,\ x_j \text{ integer},\ j \in N_1 \subseteq N\},$$

let z_U be a known upper bound on the value of (P), and let the vectors u and s satisfy

$$(5) \qquad u \geq 0, \quad s = c - uA \geq 0.$$

Then multiplying $Ax \geq b$ by $-u$ and adding the resulting inequality, $-uAx \leq -ub$, to $cx < z_U$, yields the inequality $sx < z_U - ub$, satisfied by every feasible solution x to (P) such that $cx < z_U$. Using this, and setting, for $i=1,\ldots,p$, $v_i = s_{j(i)}$ for some $j(i) \in N$, $\pi = s$ and $\pi_0 = z_U - ub$, then applying the "only if" part of Theorem 1, as modified by Remark 1.1, we obtain the following.

Corollary 1.2. Let z_U be an upper bound on the value of (P), and let u, s satisfy (5). If there exists $S \subseteq N_1$, $S=\{j(1),\ldots,j(p)\}$, $1 \le p \le |N_1|$, such that

(6) $$\sum_{j\in S} s_j \geq z_U - ub,$$

then for any collection of sets $Q_i \subseteq M_i$, $i = 1,\ldots, p$, such that

(7) $$\sum_i \sum_{j\in Q_i} s_{ij} \le s_j, \quad j\in N,$$

every feasible solution x to (P) for which $cx < z_U$, satisfies the disjunction

(3) $$\bigvee_{i=1}^{p} (x_j = 0,\ j\in Q_i).$$

Note that if p=1, i.e., (3) has a single term, then (3) converts to the condition $x_j = 0$, $j \in Q_1$. Somewhat more generally, we have

Remark 1.3. Let z_U, u and s be as in Corollary 1.2, and define

$$Q_o = \{j\in N_1 | s_j \geq z_U - ub\}.$$

Then every feasible solution x to (P) such that $cx < z_U$ satisfies $x_j = 0$, $j\in Q_o$.

Corollary 1.2 has an interpretation (and alternative proof) in terms of conditional bounds, which yields some insight and is appealing to intuition. Consider the pair of dual linear programs

(L) $$\min \{cx|Ax \geq b,\ x \geq 0\}$$

and

(D) $$\max \{ub|uA \leq c,\ u \geq 0\},$$

associated with (P).

Clearly, for any u feasible to (D), ub is a lower bound on the value of (L), hence of (P). Now suppose the constraint set of (P) (and (L)) is amended by the system $Gx \geq e$ defined by (4). Then (L) and (D) become

(L_G) $\min \{cx | Ax \geq b,\ Gx \geq e,\ x \geq 0\}$

and

(D_G) $\max \{ub + ve | uA + vG \leq c,\ u \geq 0,\ v \geq 0\}$

respectively, and ub + ve is a lower bound on the value of (L_G), hence of (P_G), the problem obtained from (P) by adding to its constraints $Gx \geq e$. Now if a vector v can be found that together with G satisfies the constraints of (D_G) and $ub + ve \geq z_U$, then, since $cx \geq ub + ve$, every feasible solution to (L_G), hence to (P_G), satisfies $cx \geq z_U$. It follows that every feasible solution x to (P) such that $cx < z_U$, must violate the constraint set $Gx \geq e$, hence (as x_j is integer-constrained for $j \in N_1$) must satisfy the disjunction (3). If we set $v_i = s_{j(i)}$, i=1,...,p, with s defined as in (5), then the above conditions on v are a paraphrase of (6), (7), and we obtain Corollary 1.2.

The inequalities $Gx \geq e$ are not part of the problem (P), and the sole purpose of introducing them is to conclude that if they were to hold, that would imply a lower bound at least equal to the upper bound z_U, hence any solution x better than the one that produced z_U, must violate at least one of them. We therefore call these inequalities, as well as the lower bound obtained from them, conditional.

In a broader context, the idea of deriving a valid ("unconditional") constraint from one or several conditional constraints may have many other applications. One of them appears in a recent paper by Kovács and Dienes [10], where a properly chosen inequality is used to derive a bound from the fact that either the inequality or its complement must be satisfied by any feasible solution.

From Corollary 1.2, a valid disjunction (3) can be derived for the problem (P) if an upper bound z_U^* is known, a feasible solution u to the dual linear program (D) is at hand, and a subset S of N_1 can be found for which (6) holds. This latter condition is usually easy to satisfy, and we will have more to say about this later on. Given such a set S, however, every collection of subsets Q_i of N_i that satisfies (7) gives rise to a valid disjunction (3), and the question arises of choosing one that yields a disjunction as "strong" as possible, i.e., one with p as small as possible, and the sets Q_i, $i=1,\dots,p$, as large as possible. Next we state a simple heuristic that generates a disjunction (3) with that objective in mind.

1. Choose a minimal subset $S \subseteq N_1$ such that

$$\sum_{j \in S} s_j \ge z_U - ue,$$

and order $S = \{j(1),\dots,j(p)\}$ according to decreasing values of $s_{j(i)}$.

2. Set $Q_1 = \{j \in N \mid s_j \ge s_{j(1)}\}$ and define recursively

$$Q_i = \{j \in N \mid s_j \ge s_{j(i)} + \sum_{k=1}^{i-1} (s_{j(k)} - s_j) g_{kj}\}, \quad i=2,\dots,p,$$

where $g_{kj} = 1$ if $j \in Q_k$, and $g_{kj} = 0$ if $j \notin Q_k$.

The sets Q_i, $i=1,\dots,p$, then define a valid disjunction (3).

A disjunction (3) can be used either for branching, or for generating cuts. If used for branching, this disjunction can be strengthened so as to define a partition of the feasible set; namely, (3) can be replaced by

$$(3') \qquad \bigvee_{i=1}^{p} \left(x_j = 0,\ j \in Q_i;\ \sum_{j \in Q_k} x_j \ge 1,\ k = 1,\dots,i-1 \right).$$

Note that, by construction of the sets Q_i, $s_j \geq s_{j(i)} > 0$ for $j \in Q_i$, $i = 1,\ldots, p$, and thus on all branches except the one corresponding to $i = 1$, the lower bound ub given by the dual solution associated with the reduced cost vector s, can be strengthened immediately after branching, by associating with each inequality

$$\sum_{j \in Q_k} x_j \geq 1$$

the positive multiplier $x_{j(k)}$. In other words, on the i-th branch ($i > 1$) the lower bound ub can be replaced right after branching by ub + $s_{j(1)}+\ldots+s_{j(i-1)}$.

The above described branching rule, while often considerably stronger than the traditional one, can occasionally be a lot weaker. Therefore, the best way of using it is to judiciously combine it with other branching rules, according to criteria that make sure it is only used at such nodes of the search tree where it can be expected to perform relatively well. It is in this fashion that disjunctions of the type (3) are being successfully used for branching in our set covering algorithm that also uses them to generate cutting planes (see the companion paper [2]), and in a restricted Lagrangean algorithm for the traveling salesman problem [7].

Next we turn to the other use of disjunctions of type (3), namely for generating cutting planes. In the case of the set covering problem, these cutting planes turn out to be of the same type as the original constraints.

3. The Cutting Planes

From now on, we address ourselves to the set covering problem

(SC) $\quad \min \{cx \mid Ax \geq e,\ x_j = 0 \text{ or } 1,\ j \in N\}$

introduced in section 1. (Here A is $m \times n$). We will denote

$$N_i = \{j\in N \mid a_{ij} = 1\}, \quad i\in M.$$

Consider the i-th term of a disjunction (3), i.e., $x_j = 0$, $j\in Q_i$. Clearly, every cover x that satisfies the i-th term of (3), also satisfies the inequalities

$$\sum_{j\in N_h\setminus Q_i} x_j \geq 1, \quad h\in M$$

and hence, for any choice of indices $h(i)\in M$, $i=1,\ldots,p$, every cover that satisfies (3), also satisfies the disjunction

$$\bigvee_{i=1}^{p} \Big(\sum_{j\in N_{h(i)}\setminus Q_i} x_j \geq 1 \Big),$$

which is easily seen to imply (for integer x) the inequality $\sum x_j \geq 1$, with the summation taken over the union of the sets $N_{h(i)}\setminus Q_i$, $i=1,\ldots,p$.

Combining this reasoning with Corollary 1.2 yields the following.

<u>Theorem 2</u>. Let z_U be an upper bound on the value of (SC), and let u, s satisfy (5). If there exists a set of column indices $S = \{j(1),\ldots,j(p)\}$, $S \subseteq N$, such that

(8) $$\sum_{j\in S} s_j \geq z_U - ue,$$

then for any set of p row indices $h(i)\in M$, $i=1,\ldots,p$, and any collection of p subsets $Q_i \subseteq N$, $i=1,\ldots,p$, satisfying

(7) $$\sum_{i\mid j\in Q_i} s_{j(i)} \leq s_j, \quad j\in N,$$

every cover x such that $cx < z_U$ satisfies the inequality

(9) $$\sum_{j\in W} x_j \geq 1,$$

where

$$W = \bigcup_{i=1}^{p} (N_{h(i)} \setminus Q_i). \tag{10}$$

<u>Remark 2.1</u> The family of cuts (9) remains the same if the condition $Q_i \subseteq N$ in Theorem 2 is replaced by $Q_i \subseteq N_{h(i)}$, $i=1,\ldots,p$.

<u>Proof</u>. From (10), the change does not affect the set W which defines inequality (9).||

The inequalities (9) are valid cutting planes in the sense of being satisfied by every cover better than a given one. Further, they are of the set covering type. Since these properties are the same as those of the Bellmore-Ratliff cuts [5] obtained by the use of involutory bases, we next examine the relationship between the latter and our inequalities from conditional bounds. First, we show that the Bellmore-Ratliff inequalities are a subclass of the class of inequalities (9). Then we show by way of example that the subclass in question is a proper one.

<u>Theorem 3</u>. The Bellmore-Ratliff inequalities [5] are a subclass of the class defined in Theorem 2.

<u>Proof</u>. Let $\bar{x}$ be a prime cover, B an involutory basis associated with $\bar{x}$, and $c_j - c_B a_j$ the j-th reduced cost, where c_B is the m-vector whose i-th component is $c_{j(i)}$, if the basic variable associated with row i is (the structural variable) $x_{j(i)}$, and 0 if the basic variable associated with row i is slack. (When B is an involutory basis, the reduced costs are known to be of this form). Let the columns of B be indexed by I, and denote $F = \{j \in N \mid c_j - c_B a_j < 0\}$. The Bellmore-Ratliff cut associated with $\bar{x}$ and B is then

$$\sum_{j \in F} x_j \geq 1. \tag{11}$$

To obtain this cut via our procedure, set S = IPN, S = (j(1),...,j(p)}, i.e., let S be the index set of the basic structural variables, and set u = 0, s = c. Then u and s satisfy (5), and S satisfies (8) (with equality) for $z \geq c\bar{x}$.

Next let h(i) be the row index associated with basic variable $x_{j(i)}$, and set $Q_i = N_{h(i)} \backslash F$, i=l,...,p. It is easy to see that these sets Q_i satisfy (7). Substituting for C^ in (10) then yields

$$W = \bigcup N_{h(i)} \cap F.$$

On the other hand, from the definition of F it follows that j $F implies jeN. /#N for some ie{l,...,p}, hence

$$F \subseteq \left(\bigcup_{i=1}^{p} N_{h(i)} \right)$$

and therefore W = F. Thus (11) is a special case of (9).J|

Note that the cutting planes derived by Bowman and Starr [6] **via** a vector partial ordering are a special case of the Bellmore-Ratliff inequalities, hence they can also be obtained by our procedure.

Next we illustrate by an example the fact that the Bellmore-Ratliff inequalities are a proper subclass of the class of inequalities (9), and in some cases those inequalities (9) that cannot be derived by the Bellmore-Ratliff procedure are considerably stronger than the ones that can.

Example 2. Consider the set covering problem whose costs c_j and coefficient matrix A are shown in Table 2.

The 0-1 vector $\bar{x}$ whose support is {2,3,5,12,13,17} is a cover, satisfying with equality all the inequalities except for 1 and 8, which are oversatisfied. The Bellmore-Ratliff procedure generates cuts from the involutory bases that can be associated with $\bar{x}$, and it can obtain one cut from every such basis. The variables x_3, x_4 and x_5 can be basic only in rows

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
"j	3	1	1	3	1	2	2	3	3	3	3	3	3	4	4	4	5	6	8	9
1						1							1				1	1		1
2	1								1						1	1	1	1		1
3		1						1		1									1	
4			1											1						1
5				1					1		1						1		1	
6					1						1			1		1				1
7							1								1		1		1	1
8												1	1						1	
9								1						1			1	1		1
10										1			1			1			1	
11						1						1						1	1	1

Table 2.

3, 4 and 6 respectively. Since rows 1 and 8 are slack, x_{12} and x_{13} can be basic only in rows 11 and 10 respectively. Finally, x_{17} can be basic in any of the 4 rows 2, 5, 7, 9; and accordingly there are 4 involutory bases that can be associated with $\bar{x}$. We will denote them by B_2, B_5, B_7 and B_9, according as x_{17} is basic in row 2, 5, 7 or 9 respectively. The basis B_2 (after row permutations) is shown in Table 3. All variables whose index exceeds 20 are slacks.

	2	3	5	12	13	17		25	27	29	21	28
3	1						1					
4		1					1					
6			1				1					
11				1								
10					1		1					
2						1						
5						1		-1				
7						1			-1			
9						1				-1		
1					1	1					-1	
8				1	1		1					-1

Table 3.

The 4 cutting planes that can be obtained by the Bellmore-Ratliff procedure, depending on which basis is used, are

$$x_1 + x_6 + x_9 + x_{10} + x_{15} + x_{16} + x_{18} + x_{20} \geq 1 \quad , \quad \text{from } \wedge$$

$$x_4 + x_6 + x_9 + x_{10} + x_{11} + x_{19} \geq 1 \quad , \quad \text{from } B_5$$

$$x_6 + x_7 + x_{10} + x_{15} + x_{19} + x_{20} \geq 1 \quad , \quad \text{from } B_7$$

$$x_6 + x_8 + x_{10} + x_{14} + x_{18} + x_{20} \geq 1 \quad , \quad \text{from } V$$

On the other hand, using the conditional bound approach, we construct (by inspection or a heuristic) the dual vector u = (0,1,1,1,1,1,2,0,1,2,2) which, together with the associated reduced cost vector

$$s = (2,0,0,2,0,0,0,1,1,0,1,1,1,1,1,0,0,2,0,1),$$

satisfies the condition (5).

The cover x whose support is {2,3,5,12,13,17} yields $z_y = e\bar{x} = 14$; and the dual vector u yields the lower bound ue = 12.

Since $z^\wedge - ue = 2$, $Q_0 = \{j \in N \mid s_j \geq 2\} = \{1,4,18\}$, and thus (Remark 1.3) every cover better than $\bar{x}$ satisfies $x_1 = x_4 = x_{18} = 0$. Hence we replace N by N\{1,4,18}. Further, to apply Corollary 1.4, we pick the column indices j(1) = 12, j(2) = 13; for which (8) holds, since $s_{12} + s_{13} = s \geq z^\wedge - ue$. Next we pick the row indices h(1) = 8, h(2) = 5, and choose the sets $Q_1 = \{12,13\}$, $Q_2 = \{9,11\}$ to obtain $N_{h(1)} \setminus Q_1 = \{6,19\}$ and $N_{h(2)} \setminus Q_2 = \{10,16,19\}$, hence W = {6,10,16,19}. In choosing the sets Q_i we make sure that (7) is satisfied, and apart from that try to make each successive $N_{h(i)} \setminus Q_i$ add to W as few new elements as possible. We have thus obtained the cut

$$x_6 + x_{10} + x_{16} + x_{19} \geq 1$$

which has only 4 positive coefficients, whereas each of the involutory basis cuts has at least 6.

The above inequality cuts off $\bar{x}$. This is due to the way we chose the column indices j(i) and the row indices h(i), i=l,,..,p, as will be shown in the next section. If we do not care about cutting off a specified cover, we can obtain inequalities which are "stronger" in the sense of having fewer positive coefficients. Thus, for instance, if we choose j(i) = 13, j(2) = 9, and h(l) = 8, h(2) = 5, we can generate the cut

$$x_{17} + x_{19} * *$$

(by setting (^ = [12,13], Q_2 = {9,11}); and for j(l) = 13, j (2) = 14, h(l) = 8, h(2) = 4, we obtain the cut

$$x_3 + x_{19} \geq 1$$

(by choosing Q_1 = {12,13} and Q_2 = {14,20}).

4. Some Properties of Cuts from Conditional Bounds

The family of cuts defined by Theorem 2 is vast, and one is interested of course in computationally cheap procedures for generating "strong" members of this class. In this section we investigate some properties of the cuts (9) that will be helpful toward that goal.

The first practical question that arises is whether condition (8) can always be met, and how. Since s depends on u, it should not be surprising that one answer to this question comes in terms of additional conditions on u.

Theorem 4. Let the vectors $\bar{u}$ and ¥ satisfy (5), and let $\bar{x}$ be a cover with support $S(\bar{x})$. If

$$\bar{u}(A\bar{x}-e) = 0,$$

then (8) holds for $S = S(\bar{x})$.

~~Proof~~. Consider the pair of dual linear programs.

$$(L_1) \qquad \min \{cx \mid Ax \geq e,\ x_j \geq 1,\ j \in S(\bar{x}),\ x_j \geq 0,\ j \in N \setminus S(\bar{x})\}$$

and

$$(D_1) \qquad \max \{ue + \sum_{j \in S(\bar{x})} s_j \mid ua_j + s_j = c_j,\ j \in N;\ u \geq 0,\ s \geq 0\}\ .$$

Clearly, $\bar{x}$ is a feasible solution to (L_1), and $(\bar{u},\bar{s})$ is a feasible solution to (D_1). Further, $\bar{x}$ and $(\bar{u},\bar{s})$ satisfy the complementary slackness conditions $\bar{u}(A\bar{x}-e)$ and $(\bar{x}_j-1)\bar{s}_j = 0$, $j \in S(\bar{x})$, $\bar{x}_j\bar{s}_j = 0$, $j \in N \setminus S(\bar{x})$; hence they are optimal solutions to (L_1) and (D_1) respectively. Therefore

$$\bar{u}e + \sum_{j \in S(\bar{x})} \bar{s}_j = c\bar{x}$$

which together with $z_u \leq c\bar{x}$ proves the statement. ||

For any cover x, denote

$$T(x) = \{i \in M \mid a^i x = 1\}.$$

Then as an immediate consequence of Theorem 4, we have

Remark 4.1. Let $\bar{x}$ be a cover, and let $(\bar{u},\bar{s})$ satisfy (5). If $\bar{u}$ also satisfies

$$\bar{u}_i = 0,\ \forall\ i \in M \setminus T(\bar{x})$$

then (8) holds for $S = S(\bar{x})$.

Thus, if an upper bound z_U and vectors u, s satisfying (5) are at hand, but condition (8) does not hold, it can be made to hold by successively setting to 0 components u_i of u such that $i \in M \setminus T(\bar{x})$. At worst all such components may have to be set to 0; then (8) will hold.

Before turning to other characteristics of the cuts (9), we now state a basic property of the set covering problem. Let the set covering polytope P be the convex hull of all integer n-vectors satisfying $Ax \geq e$, $x \geq 0$, i.e.,

$$P = \text{conv}\,\{x \in R^n \mid Ax \geq e,\ x \geq 0,\ x_j \text{ integer},\ j \in N\}.$$

We then have the following

Theorem 5. The inequality

$$\sum_{j \in N_i} x_j \geq 1 \tag{12}$$

where $i \in M$, defines a facet of P if and only if there exists no $k \in M$ such that $N_k \subset N_i$, $N_k \neq N_i$.

Proof. The "only if" part is obvious. To prove the "if" part, we assume there is no $k \in M$ such that $N_k \subset N_i$, $N_k \neq N_i$, and we exhibit n linearly independent integer n-vectors that satisfy $Ax \geq e$, $x \geq 0$ and for which (1) holds with equality.

Let $|N_i| = p$, and assume w.l.o.g. that N_i is the set of the first p indices in N. Let $y = (1,\ldots,1)$, $y \in R^{n-p}$, and let e_i and f_i be the unit vector in R^p and R^{n-p} respectively, whose i-th component is 1. Now consider the p n-vectors (e_i, y), $i=1,\ldots,p$, and the n-p n-vectors $(e_1, y+f_{i-p})$, $i=p+1,\ldots,n$. Since there is no $k \in M$ such that $N_k \subset N_i$, $N_k \neq N_i$, each of these nonnegative integer vectors satisfies $Ax \geq e$; and since each one of them has a single 1 among its first p components, they all satisfy (12) with equality. Further, the $n \times n$ matrix whose rows are these vectors is

$$X = \begin{pmatrix} I_p & Y_p \\ E & I_{n-p} + Y_{n-p} \end{pmatrix}$$

where for k = p and k = n - p, I_k is the identity of order k, while Y_k is the $k \times (n - p)$ matrix whose entries are all equal to 1; and E is the $(n - p) \times p$ matrix whose first column consists of 1's, and whose remaining columns consist of 0's. Now define the matrix

$$Z = \begin{pmatrix} I_p + Y_p & -Y_p \\ -E & I_{n-p} \end{pmatrix}$$

Using the fact that $EY_p = Y_{n-p}$, it is easy to see that $XZ = I_n$, i.e., $Z = X^{-1}$ and hence X is nonsingular. This proves that the n vectors introduced above are linearly independent.||

In a cut-generating procedure it is important to make sure that no cut is repeated. Next we give a necessary and sufficient condition for a cut to be "new." Let (SC) stand for the set covering problem amended with all the cutting planes generated up to some point, and let

$$(9) \qquad \sum_{j \in W} x_j \geq 1$$

be the next cut generated. We then say that the inequality (9) is *new*, if there is no $i \in M$ such that $N^i \subseteq W$.

Remark 5.1. The inequality (9) is new if and only if $N \setminus W$ is the support of a cover for (SC).

Proof. The cut (9) is new if and only if $N^i \not\subseteq W$, $\forall i \in M$; hence if and only if $N^i \cap \bar{W} \neq \emptyset$, $\forall i \in M$. But this condition holds if and only if $N \setminus W$ is the support of some cover.||

While the condition of Remark 5.1 is straightforward, it is easier to embed in a cut generating procedure when paraphrased as follows.

Remark 5.1.a. The inequality (9) is new if and only if it cuts off (is violated by) some cover of (SC).

The next Theorem gives conditions on the column indices $j(i)$ and row indices $h(i)$ used in generating inequality (9), to guarantee that the inequality obtained cuts off a specified cover. We will denote

$$M_j = \{i \in M \mid a_{ij} = 1\}, \quad j \in N.$$

Theorem 6. Let $\bar{x}$, u, v, S and Q_i, $i=1,\ldots,p$, be as in Theorem 2, and let $j(i) \in Q_i$, $i=1,\ldots,p$. If $\bar{x}$ is a cover such that $S \subseteq S(\bar{x})$ and

(13) $\quad h(i) \in T(\bar{x}) \cap M_{j(i)}, \quad i=1,\ldots,p,$

then the inequality (9) cuts off (is violated by) $\bar{x}$.

Proof. Assume $S \subseteq S(\bar{x})$ and (13) holds. From $h(i) \in M_{j(i)}$ we have $j(i) \in N_{h(i)}$, $i=1,\ldots,p$; and since $j(i) \in S \subseteq S(\bar{x})$ implies $\bar{x}_{j(i)} = 1$, while $h(i) \in T(\bar{x})$ implies $|S(\bar{x}) \cap N_{h(i)}| = 1$, $i=1,\ldots,p$, it follows that

$$S(\bar{x}) \cap N_{h(i)} = \{j(i)\}, \quad i=1,\ldots,p.$$

Further, since $j(i) \in Q_i$, $i=1,\ldots,p$, we have

$$S(\bar{x}) \cap (N_{h(i)} \setminus Q_i) = \emptyset, \quad i=1,\ldots,p,$$

and hence $S(\bar{x}) \cap W = \emptyset$, i.e., the inequality (9) cuts off $\bar{x}$.||

Remark 6.1. Every inequality (9) for which the conditions of Theorem 6 are satisfied, defines a facet of

$$P^* = \text{conv}\,\{x \in R^n \mid Ax \geq e,\ \sum_{j} x_j \geq 1,\ x_j \text{ integer},\ j \in N\}.$$

Proof. Follows from Remark 5.1 and Theorem 5.||

Theorems 2 and 6 provide rules for generating a sequence of valid cutting planes that are all distinct, and furthermore, are all facets of the current polytope P*. This latter property, however, does not imply that all inequalities generated this way are equally strong. Since all the inequalities in question have coefficients equal to 0 or 1 and a right hand side equal to 1, we will use the number of coefficients equal to 1 as a measure of their strength (the fewer the 1's, the stronger the inequality). Note that some facets of the set covering polytope may be much weaker than others, according to this criterion. Thus, for instance, all 5 inequalities represented by the rows of the matrix A in Table 4 define facets of the set covering polytope corresponding to A, yet inequality 4, with only two 1's, is much stronger than inequality 5, which has ten 1's.

	1	1		1		1		1			1	
		1	1				1		1	1		
A =	1				1			1		1		1
		1			1							
			1	1	1	1	1	1	1	1	1	1

Table 4.

Thus, although they all define facets of the current polytope P*, the cutting planes obtainable via the rules of Theorem 6 are not all equally desirable. The next section discusses a procedure for generating conveniently strong members of the family.

5. Generating Cuts

The strength of an inequality (9), i.e., the size of the set W, depends on the integer p and the size of the sets $N_j(Q_i)$, $i = 1,\dots,p$, of Theorem 6. To have p conveniently small, the procedure chooses the set s = tj(i)>•••tj(p))> corresponding to the p largest reduced costs s_j, jeS($\bar{x}$), where p is the smallest integer for which (8) is satisfied. Each row index h(i), i = l,...,p, is of course chosen from T($\bar{x}$)n$M_{j(i)}$, as prescribed by Theorem 6. Further, in order to have W as small as possible, the sequence of row indices h(i) is chosen so as to make as small as possible at each step ke(l,...,p} the set $^W A^W_\kappa$, $_{\kappa-'1}$, where W_0 = 0 and

$$W_k = \bigcup_{i=1}^{k} (N_{h(i)} \setminus V * k=1 \_\_ p\_$$

Since for any S satisfying (8), |s| = 1 implies (Remark 1.3) that the variable x_j such that S = {j} can be permanently set to 0, we assume this has already been done for all such singleton sets, and hence |S| $\geq$ 2 for all S satisfying (8).

Let M and N be the row and column index sets of the current problem (SC), let $\bar{x}$ be a prime cover for (SC), and denote, as before,

$$S(\bar{x}) = [j \in N | \bar{x}_j = l], \quad T(X) = [ieMJa^1 \wedge !\}.$$

Further, let u and s satisfy (5), and assume (8) holds for S = S^+A [j eS (x) |s $_j$X)} .

<u>Cut-Generating Procedure</u>

<u>Step 0</u>. Initialize W = 0, S = S^+, $z = ue_L$, i = 1 and go to 1.

<u>Step 1</u>. Define

$$s_{j(i)} = \max_{j \in S} s_j, \qquad Q = \{j \in N | s_j \geq s_{j(i)}\}.$$

Choose h(i) such that

$$|N_{h(l)} , \setminus QLW| = \underset{h \in T \; 0 \; . M_{j(i)}}{\mathrm{rain}} \; |N_h, \setminus QUW|$$

(breaking ties arbitrarily), and set

W «- W II CN \0^ 7 *- 7 4- $s_{j(i)}$.

If $7_L \geq$ ZJJ, go to step 2. Otherwise, let

^{S}i " I /^{S}j • ^{S}j(D $^{J\in QnN}$b(i)
J s J otherwise,

$S \leftarrow S \setminus \{j(i)\}$,

set i - i+1, and go to 1.

<u>Step 2</u>. Add to the constraint set of (SC) the inequality

$$\sum_{j \in W} x_j \geq 1.$$

In at most $|S^+|$ iterations, this procedure generates an inequality (9) satisfied by every cover x such that ex < z_u and violated by $\bar{x}$. Indeed, S (initialized as S^+) is diminished at every iteration by one element, hence there are at most $|S^+|$ iterations. Further, since (8) holds for $S = S^+$, after $p \leq |s^+|$ iterations, (8) holds for the current set S (i.e., $^{Z}{}_{T}$ >^{Z}T$_{T}$) » an(* we S° to Step 2 to generate a cut (9). For the sets Q_i = Q^{nN}h/$_{(i)}$\» i=l,...,P, j(i)cQ_i, and (7) is satisfied (by the definition of Q and s_j at every iteration). Finally, the choice of h(i) guarantees (13). Thus all requirements of Theorem 6 are met.

A couple of minor improvements are at hand. Choosing the largest s_j to define j(i) at every iteration has the purpose of minimizing the size of the set S in (8). But at the last iteration choosing the largest s_j may not be indicated, if a smaller s_j suffices to satisfy (8). Thus a better rule for choosing j(i) in Step 1 is to set

$$v_i = \min\left\{\max_{j \in S} s_j,\ \min_{j \in S} [s_j \mid s_j \geq z_U - z_L]\right\}$$

and then choose as j (i) one of the indices $j \in S$ for which $s_j = v_i$. Furthermore, whenever this index is not unique, i.e., $|J| > 1$, where $J = \{j \in S \mid s_j \geq v_i\}$, the choice of j(i) and h(i) can often be improved by first setting $Q = \{j \in N \mid s_j \geq v_i\}$, next choosing h(i) so as to minimize $|N_h \backslash Q \cup W|$ over all $h \in T(\bar{x}) \cap M_J$, where

$$M_J = \bigcup_{j \in J} M_j;$$

and then choosing the unique index $\{j\} = J \cap N_{h(i)}$ as j(i).

Example 3. Consider the set covering problem of example 2 (Table 2), with $c_4 = 3$ replaced by $c_4 = 1$. Then the cover $\hat{x}$ whose support is $S(\hat{x}) = \{2,4,13,20\}$ gives $z_U = c\hat{x} = 14$, and $T(\bar{x}) = M\backslash\{1\}$. The vector u of example 2 yields the same reduced costs s_j as in that example, except for s_4, which now is 0. The lower bound ue is 12, and since $s_j \geq 14 - 12 = 2$ for j = 1, 18, we set $x_1 = x_{18} = 0$, and replace N by $N\backslash\{1,18\}$. Condition (8) holds for $S = S^+ = [13,20]$, since $s_{13} + s_{20} \geq 2$.

Step 0. W = 0, S = {13,20}, $z_L = ue = 12$.

Step 1. $v_1 = \min\{1,2\} = 1$, J = {13,20}, Q = {8,9,11,12,13,14,15,20}, $M_J = M_{13} \cup M_{20} = M\backslash\{3,5\}$. To choose h(1), we minimize $|N_h\backslash Q|$ over $h \in T(\bar{x}) \cap M_J = M\backslash\{1,3,5\}$, and find that the minimum is 1, attained for k = 4,8,9. We arbitrarily choose h(1) = 4, and set $W = N_4\backslash Q = \{3\}$, $z_L = 12 + 1 = 13$. The s_j remain unchanged except for j = 14,20, the new values for the latter being $s_{14} = s_{2Q} = 0$. We set S = {13}, i=2, and go to

Step 1. $v_2 = \min [l,l) = 1$, J = {13}, Q = {8,9,11,12,13,15},

$M_J = M_{13} = \{1,8,10\}$. To choose h(2), we minimize $|N_h \backslash Q|$ over

heTWOMj = {8,10}, and find h(2) = 8. We set $W = \{3\} U (N_g \backslash Q) = \{3,19\}$,

$z_{XJ} = 13 + 1 = 14$, and since $z_{LI} \geq z_U$, we go to

Step 2. We add to (SC) the inequality

$$x_3 + x_{19} \geq 1.$$

6. A Class of Algorithms

The cutting planes discussed in this paper can best be used in a framework that takes maximum advantage of their properties. To obtain a cutting plane from conditional bounds, one needs a feasible solution (u,s) to the dual of the linear program associated with (SC). Such a solution also provides a lower bound ue on the value of (SC). On the other hand, the easiest way to make sure that the cuts that one generates are all distinct, is to have each inequality cut off some cover satisfying all the previously produced inequalities. Thus to obtain a sequence of distinct cutting planes, one also needs a sequence of covers. Each cover x in turn, provides an upper bound ex on the value of (SC).

The best approach then seems one that alternates between (a)generating a cover x for the current problem, and (P) generating a dual solution (u,s) and using it to derive an inequality that cuts off x. In such a procedure, the value of (SC) is bounded from above by the sequence of covers obtained under *(a)*; and bounded from below by the sequence of dual solutions produced under (β). The rate of convergence of the algorithm is the rate at which the gap $z_u - z_L$ between the two bounds decreases.

Since every inequality generated in the procedure cuts off at least one new cover, and since the number of distinct covers is finite, the procedure outlined above is finite, irrespective of the methods used to generate the sequence of covers x and dual solutions (u,s). Its efficiency on the other hand depends crucially on the efficiency of those methods.

Several versions of the approach outlined above were implemented and thoroughly tested in a computational study summarized in the companion paper [2]. The algorithm that emerged as a result of the testing uses several different heuristics intermittently to generate prime covers, and produces dual solutions (u,s) both by heuristics and by subgradient optimization. When the decrease in the gap $z_u - z_L$ slows down, the algorithm branches, using either a disjunction of the type discussed in this paper, or a dichotomy derived from other considerations, according to some measure of comparative strength. The algorithm is particularly well suited for low density problems, and its performance on set covering problems taken from the literature compares favorably with earlier methods. Randomly generated test problems with up to 200 constraints and 2000 variables have been successively run.

For details of the algorithm and results of the computational tests the reader is referred to [2].

References

[1] E. Balas, "Set Covering with Cutting Planes from Conditional Bounds." MSRR # 399, Carnegie-Mellon University, July 1976.

[2] E. Balas and A. Ho, "Set Covering Algorithms Using Cutting Planes, Heuristics and Subgradient Optimization: A Computational Study." MSRR # 438, Carnegie-Mellon University, July 1979.

[3] E. Balas and N. Christofides, "A Restricted Lagrangean Approach to the Traveling Salesman Problem." MSRR # 439, Carnegie-Mellon University, July 1979.

[4] E. Balas and M. W. Padberg, "Set Partitioning: A Survey." SIAM Review, 18, 1976, p. 710-760.

[5] M. Bellmore and H. D. Ratliff, "Set Covering and Involutory Bases." Management Science, 18, 1971, p. 194-206.

[6] V. J. Bowman and J. Starr, "Set Covering by Ordinal Cuts. I: Linear Objective Functions." MSRR # 321, Carnegie-Mellon University, June 1973.

[7] N. Christofides and S. Korman, "A Computational Survey of Methods for the Set Covering Problem." Management Science, 21, 1975, p. 591-599.

[8] R. J. Duffin, "Infinite Programs," in Linear Inequalities and Related Systems. H. W. Kuhn and A. W. Tucker (editors), Princeton University Press, 1956, p. 157-170.

[9] R. Garfinkel and G. L. Nemhauser, "Optimal Set Covering: A Survey." A. M. Geoffrion (editor), Perspectives on Optimization. Addison-Wesley, 1972, p. 164-183.

[10] L. B. Kovács and I. Dienes, "Maximal Direction - Complete Paths and Their Application to a Geological Problem: Setting Up Stratigraphic Units." Paper presented at the 9th International Symposium in Mathematical Programming, Budapest, August 1976.