Institute for Operations Research and the Management Sciences

Convexity Cuts and Cut Search
Author(s): Fred Glover
Source: *Operations Research*, Vol. 21, No. 1, Mathematical Programming and Its Applications (Jan. - Feb., 1973), pp. 123-134
Published by: INFORMS
Stable URL: http://www.jstor.org/stable/169093
Accessed: 20-02-2017 18:43 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://about.jstor.org/terms

INFORMS is collaborating with JSTOR to digitize, preserve and extend access to *Operations Research*

This content downloaded from 132.174.250.143 on Mon, 20 Feb 2017 18:43:52 UTC
All use subject to http://about.jstor.org/terms

Convexity Cuts and Cut Search

Fred Glover

University of Colorado, Boulder, Colorado

(Received July 21, 1970)

This note focuses on two new and related cut strategies for integer programming: the 'convexity-cut' and 'cut-search' strategies. The fundamental notions underlying the convexity-cut approach are due to RICHARD D. YOUNG and EGON BALAS, whose 'hypercylindrical' and 'intersection' cuts provide the conceptual starting points for the slightly more general framework developed here. We indicate the utility of our framework (and hence the importance of the original Young and Balas ideas) by specifying a variety of new cuts that can be obtained from it. The second new strategy, cut search, shares with the convexity-cut strategy the notion of generating a cut by passing a hyperplane through the terminal endpoints of edges extended from the vertex of a cone. However, whereas the convexity-cut approach determines the extensions of these edges by reference to a convex set that contains the vertex of the cone in its interior, the cut-search approach determines these extensions by reference to associations between certain 'proxy' sets of points (e.g., collections of hyperplanes) and 'candidate solutions' to the integer program. The cut-search approach typically involves more work than the convexity-cut approach, but offers the chance to identify feasible solutions in the process, and can sometimes also yield somewhat stronger cuts than the convexity cuts.

THIS PAPER generalizes the important new 'hypercylindrical' and 'intersection' cut ideas of YOUNG[15] and BALAS[1] to provide a broader class of cuts called *convexity cuts*, and describes several specific new subclasses of this general class. In addition, it describes a new approach to cut generation, called *cut search*, that appears to provide stronger cuts than the convexity cuts, although at greater computational effort. An auxiliary feature of the cut-search approach is the generation of a set of candidate solutions that can be tested for feasibility.

NOTATION

THE MIXED integer programming (MIP) problem will be written:

$$\text{maximize } x_0=a_{00}+\textstyle\sum_{j\epsilon N} a_{0j}(-t_j), \qquad \text{subject to:}$$

$$x_i=a_{i0}+\textstyle\sum_{j\epsilon N} a_{ij}(-t_j), \quad i\epsilon M; \quad x_i\geqq 0, \quad i\epsilon M=\{1,\cdots,m\};$$

$$t_j\geqq 0, \quad j\epsilon N=\{1,\cdots,n\}; \quad \text{and} \quad x_i \text{ integer}, \quad i\epsilon I=\{1,\cdots,n'\}\subset N.$$

Without the integer restriction, this is the ordinary linear programming (LP) problem. The t_j, $j\epsilon N$, represent the current nonbasic variables and are assumed to be a subset of the x_i, $i\epsilon M$. In particular, the t_j of the initial tableau representation of the LP problem are identified by the first n equations $x_i=-1(-t_i)$, $i\epsilon N$.

This content downloaded from 132.174.250.143 on Mon, 20 Feb 2017 18:43:52 UTC
All use subject to http://about.jstor.org/terms

We also write the MIP problem in matrix notation as

$$\text{maximize } x_0 = a_{00} + A^0(-t), \qquad \text{subject to } x = A_0 + A(-t),$$
$$x \geqq 0, \qquad t \geqq 0, \qquad \text{and} \quad x_i \text{ integer for } i \epsilon I,$$

where superscripted vectors refer to row vectors and subscripted vectors refer to column vectors.

CONVEXITY CUTS

THE CONVEXITY CUTS apply to any mathematical programming problem whose constraints include or imply

$$y = B_0 - Bt, \qquad t \geqq 0, \tag{1}$$

and

$$y \epsilon S, \tag{2}$$

where S may be any of a rather broad class of sets, but in the present context will generally consist of those y satisfying one or more of the three conditions $y \geqq 0$, y_i integer for $i \epsilon P$, or $y_i = 0$ or 1 for $i \epsilon P$ (where P is some subset of the index set Q for the components of y).

Thus, for example, (1) and (2) may simply be another way of formulating the constraints of the MIP problem (with $y = x$); or, alternatively, (1) may be obtained by identifying each y_i as an integer combination of the x_i for $i \epsilon I$, and (2) may stipulate that y_i is integer for each $i \epsilon Q$. The general characterization of convexity cuts is given by the following result. (*See* Note 1.)

CONVEXITY-CUT LEMMA. *Let R be any convex set whose interior contains B_0 but no points of S. Then for any constants $t_j^* > 0$ such that $B_0 - B_j t_j^* \epsilon R$ for all $j \epsilon N$, the cut*

$$\textstyle\sum_{j \epsilon N} (1/t_j^*) t_j \geqq 1 \tag{3}$$

is consistent with all y satisfying (1) *and* (2). (*By convention, $1/t_j^* = 0$ if $t_j^* = \infty$.*)

Proof. Consider any y that is a convex combination of the points B_0, $B_0 - B_j t_j^*$, $j \epsilon N$; i.e., $y = \lambda_0 B_0 + \sum_{j \epsilon N} \lambda_j (B_0 - B_j t_j^*)$, where $\lambda_0 + \sum_{j \epsilon N} \lambda_j = 1$ and $\lambda_j \geqq 0$ for $j = 0$ and $j \epsilon N$. From the assumptions, $\sum_{j \epsilon N} \lambda_j < 1$ implies y is in the interior of R. But the set of convex combinations satisfying this latter inequality is just the set of y satisfying (1) and $\sum_{j \epsilon N} (1/t_j^*) t_j < 1$ (equating t_j with $\lambda_j t_j^*$). Hence all y satisfying (1) that are also in S (and therefore not in the interior of R) must satisfy (3).

Note that, by the foregoing lemma, if (1) and (2) are implied by the constraints of the MIP problem, then (3) is a valid cut for the MIP problem, and, moreover, (3) excludes $x = A_0$ from the set of feasible continuous solutions. Also, the stipulation of the lemma that R contains B_0 in its interior is not strictly necessary; i.e., B_0 may be on the boundary of R as long as there is a neighborhood of B_0 such that all $y(\neq B_0)$ satisfying (1) in this neighborhood are in the interior of R.

The following easily proved results indicate the scope of the convexity cuts and also provide guidelines for creating acceptable sets R relative to which these cuts may be defined.

Remark 1. Assume (1) and (2) correspond to the constraints of the MIP prob-

This content downloaded from 132.174.250.143 on Mon, 20 Feb 2017 18:43:52 UTC
All use subject to http://about.jstor.org/terms

lem (with $y=x$) and moreover, (2) implies (1). (This of course can always be ensured.) Then the convexity cuts defined by (3) include all half spaces $H(x)\geqq 0$ such that $H(A_0)\,(<0)$ minimizes $H(x)$ over the cone defined by (1), and such that $H(x^*)\geqq 0$ for all x^* that are feasible for the MIP problem. [In particular, (3) includes all half spaces corresponding to the faces of the convex hull of feasible solutions to the MIP problem for which $H(x)$ achieves its minimum over the cone at A_0.]

Remark 2. Assume $y\epsilon S$ implies $y_i\geqq 1$ or $y_i\leqq 0$, $i\epsilon P$. Then $y_i^2\geqq y_i$, and hence $\sum_{i\epsilon P} h_i(y_i^2-y_i)\geqq 0$ for any nonnegative numbers h_i, $i\epsilon P$. Thus, the convex set of y satisfying

$$\sum_{i\epsilon P} h_i(y_i^2-y_i)\leqq 0 \tag{4}$$

contains no points of S in its interior. (The set constitutes a sphere when all h_i assume the same value, and more generally constitutes an ellipsoid when some of the h_i differ in value.) Moreover, this convex set will contain $y=B_0$ in its interior if $1\geqq b_{i0}\geqq 0$ (where b_{i0} is the ith component of B_0) for all $i\epsilon P$, and $1>b_{i0}>0$ for some $i\epsilon P$ such that $h_i>0$. Thus, under the stated assumptions, the convex set defined by (4) may serve as an acceptable R for determining a convexity cut.

Remark 3. It is easy to create y_i and b_{i0} satisfying the assumptions of Remark 2 for any problem containing the constraints $x=A_0-At$, $t\geqq 0$ and "$x_i\geqq f_i$ or $x_i\leqq g_i$" for $i\epsilon P$, provided $f_i\geqq a_{i0}\geqq g_i$ for all $i\epsilon P$ and $f_i>a_{i0}>g_i$ for some $i\epsilon P$. To do this one defines $y_i=(x_i-g_i)/(f_i-g_i)$, which, from the equation $x_i=a_{i0}-\sum_{j\epsilon N} a_{ij}t_j$, gives $b_{i0}=(a_{i0}-g_i)/(f_i-g_i)$ and $b_{ij}=a_{ij}/(f_i-g_i)$ for $j\epsilon N$.

The specific application of these comments to the MIP problem is elaborated in Remark 5.

Remark 4. While Remark 2 identifies a set of ellipsoids that can serve as R without taking into account any restrictions except those imposed by the discrete character of the y_i variables, it is also possible to identify intersections of ellipsoids that can serve as R by explicit reference to additional constraining relations (whose structure determines the structure of the convex sets whose intersection defines R). Specifically, assume $y\epsilon S$ implies that $y_i\leqq 0$ or $y_i\geqq 1$ for $i\epsilon P$ and that the inequality

$$\sum_{i\epsilon P} h_{ri}y_i\geqq h_{r0} \tag{5}$$

holds for at least one $r=1,\cdots,r'$. Furthermore, suppose that for each $r\leqq r'$, the solution $y=B_0$ either violates (5) or satisfies it with equality. Then if $1\geqq b_{i0}\geqq 0$ for all $i\epsilon P$, and if for each $r\leqq r'$ there is at least i such that $h_{ri}\neq 0$ and $1>b_{i0}>0$, then R can be given by the intersection of the convex sets

$$\sum_{i\epsilon P} |h_{ri}|z_{ri}^2\leqq h'_{r0}, \qquad (r=1,\cdots,r') \tag{6}$$

where h'_{r0} is h_{r0} minus the sum of the negative h_{ri}, $i\epsilon P$, and where the variables z_{ri} are given by $z_{ri}=y_i$ if $h_{ri}\geqq 0$, and $z_{ri}=1-y_i$ if $h_{ri}<0$. [The restrictions indicated for $y=B_0$ assure that B_0 is in the interior of r, but may be replaced by any other conditions that imply (6) is satisfied as a strict inequality for $y=B_0$ and all $r\leqq r'$.]

Remark 5. If the statement of an MIP problem does not otherwise include the type of constraining relations indicated in Remark 4, it is always possible to generate such relations using the ideas of Remark 3 if $x=A_0$ does not assign x_i an integer value for all $i\epsilon I$. Let P be any subset of I for which a_{i0} is not integer for at least one $i\epsilon P$, and define $y_i=x_i-[a_{i0}]+\delta_i$, where $\delta_i=0$ if a_{i0} is not an integer and other-

This content downloaded from 132.174.250.143 on Mon, 20 Feb 2017 18:43:52 UTC
All use subject to http://about.jstor.org/terms

wise $\delta_i=0$ or 1, as desired ($[a_{i0}]$ denotes the greatest integer $\leqq a_{i0}$). Then each y_i, $i\epsilon P$, satisfies the conditions of Remark 2, and furthermore, for $b_{i0}=a_{i0}-[a_{i0}]+\delta_i$ at least one of $\sum_{i\epsilon P} y_i \geqq [\sum_{i\epsilon P} b_{i0}]+1$ and $\sum_{i\epsilon P} - y_i \geqq -[\sum_{i\epsilon P} b_{i0}]$ must be violated for $y=B_0$, and neither can be satisfied except as an equality. Yet at least one of these inequalities is valid. Consequently, by Remark 4 it follows that R may be given by the inequalities

$$\sum_{i\epsilon P} y_i^2 \leqq \gamma_1 \quad \text{and} \quad \sum_{i\epsilon P} (1-y_i)^2 \leqq \gamma_2, \tag{7}$$

where $\gamma_1=[\sum_{i\epsilon P} b_{i0}]+1$ and $\gamma_2=|P|+1-\gamma_1$, and where $|P|$ is the order of P. The derivation of (7) is of course valid for any integer $\gamma_1 \leqq |P|$ and $\gamma_2=|P|+1-\gamma_1$, and usually there is more than one such γ_1 for which R is given by (7) contains $y=B_0$ in its interior. Also, (7) can be generalized in the obvious way by replacing y_i with $h_i y_i$, where h_i is an integer, and further replacing y_i by $y_i'=1-y_i$ if h_i is negative.

Remark 6. An alternative way to take advantage of Remark 4 is to note that every MIP problem either contains or can be made to contain constraints of the form

$$\sum_{i\epsilon P} h_i x_i = K, \tag{8}$$

where $i\epsilon P$ implies x_i is integer. [For example, if (8) is an inequality and the h_i are integers, then it can be made into an equality of the desired form by adding an integer slack variable and enlarging P.] By using the approach of Remark 5, (7) becomes $\sum_{i\epsilon P} h_i y_i = K'$, where each y_i is $\leqq 0$ or $\geqq 1$. Then, if b_{i0} is not 0 or 1 for some $i\epsilon P$ (which occurs whenever the corresponding a_{i0} is not integer), and if $h_i \neq 0$ for this i, it follows by Remark 4 that R can be given by

$$\sum_{i\epsilon P} |h_i| z_i^2 \leqq K'', \tag{9}$$

where K'' equals K' minus the sum of the negative h_i and where $z_i=y_i$ or $1-y_i$ according to the sign of h_i.

Two special cases of interest are: (i) equation (8) is $x_i+w_i=K$, where w_i is the slack variable for the upper-bound inequality $x_i \leqq K$; and (ii) $P=I$, $h_i=1$ for all i, and each x_i is a 0-1 variable. Case (i) [in which K'' of (9) is always equal to 1] gives the cuts due to RALPH GOMORY,[12] and case (ii) gives the cuts due to R. D. Young.[15]

Remark 7. The creation of ellipsoids to serve as R can be accomplished in a manner different from that in Remark 2 (disregarding constraining relations of the type accommodated in Remarks 4–6). In particular, assume $y\epsilon S$ implies $|y_i| \geqq K_i$ for all $i\epsilon P$. Then if $|b_{i0}| \leqq K_i$ for all $i\epsilon P$ and $|b_{i0}| < K_i$ for some $i\epsilon P$, then R can be given by

$$\sum_{i\epsilon P} h_i y_i^2 \leqq \sum_{i\epsilon P} h_i K_i^2 \tag{10}$$

for any positive numbers h_i, $i\epsilon P$. (The restrictions involving b_{i0} serve only to assure that $y=B_0$ will be in the interior of R, and are more stringent than necessary.)

Remark 8. Remark 7 can always be implemented for the MIP problem if $x=A_0$ does not yield x_i integer for all $i\epsilon I$. For example, suppose that x_i is not integer for at least one $i\epsilon P \subset I$, and define $y_i=x_i-[a_{i0}]+\delta_i-\frac{1}{2}$ (where δ_i is as in Remark 5), for $i\epsilon P$. Then x_i integer implies $|y_i| \geqq \frac{1}{2}$, but $|b_{i0}| \leqq \frac{1}{2}$ and $|b_{i0}| < \frac{1}{2}$ if a_{i0} is not an integer (where $b_{i0}=a_{i0}-[a_{i0}]+\delta_i-\frac{1}{2}$). Thus (10) becomes

$$\sum_{i\epsilon P} h_i y_i^2 \leqq \left(\sum_{i\epsilon P} h_i\right)/4. \tag{11}$$

This content downloaded from 132.174.250.143 on Mon, 20 Feb 2017 18:43:52 UTC
All use subject to http://about.jstor.org/terms

The Balas cuts[1] are given by (11) when $P=I$ and all $h_i=1$, and the Gomory cuts[12] are given by (11) when P has exactly one element.

It is interesting to note that Remark 7 is equivalent to Remark 2 upon scaling and translating the y_i variables (as in Remark 3) and hence (11) gives the same convexity cuts as (4). We have stated the latter remark in addition to the former to point up a connection between the cut derivations of Balas and Young, which develop these remarks (Balas Remark 7 and Young Remark 2) under the assumption $P=I$ and all $h_i=1$ (where Young limits consideration to $y_i=0$ or 1). (*See* Note 2.)

Remark 9. Implementation of Remarks 5, 6, and 8 to determine values t_j^* that give the convexity cut (3) may be handled as follows. In each instance, the y_i variables are given by $y_i-x_i-K_i$ for some constant K_i, so that the equations $y_i=b_{i0}-\sum_{j\epsilon N} b_{ij}t_i$, $i\epsilon P$, hold with $b_{i0}=a_{i0}-K_i$ and $b_{ij}=a_{ij}$ for all $j\epsilon N$.

The set R in Remarks 5, 6, and 8 is given by the intersection of one or more convex sets having the form

$$\sum_{i\epsilon P_1} h_iy_i^2+\sum_{i\epsilon P_2} h_i(1-y_i)^2\leqq h_0, \quad (12)$$

where all $h_i\geqq 0$ and P_1 and P_2 constitute a partition of P. The largest value of t_j for which $y=B_0-B_jt_j$ satisfies (12) is obtained by requiring (12) to hold as an equality with y_i replaced by $b_{i0}-b_{ij}t_j$. This gives the quadratic equation $\alpha t_j^2-\beta t_j+\gamma=0$, where $\alpha=\sum_{i\epsilon P} h_ib_{ij}^2$, $\beta=-2\sum_{i\epsilon P} h_ib_{i0}b_{ij}+2\sum_{i\epsilon P_2} h_ib_{ij}$ and

$$\gamma=\sum_{i\epsilon P} h_ib_{i0}^2+\sum_{i\epsilon P_2} h_i(1-2b_{i0})-h_0.$$

The sought-after value of t_j is the positive root of this quadratic, which exists and is unique under the stated assumptions unless $b_{ij}=0$ for all $i\epsilon P$, in which case $t_j=\infty$. [Recall that the roots of the quadratic are given by $t_j=(-\beta\pm\sqrt{\beta^2-4\alpha\gamma})/2\alpha$.] Thereupon the minimum over these positive values of t_j for each inequality (12) whose intersection determines R gives the appropriate value of t_j^*. (There are two such inequalities for Remark 5, and one for Remarks 6 and 8.)

Remark 10. The assumptions of Remark 7 can alternatively be exploited by allowing R to be given by $\sum_{i\epsilon P} h_i|y_i|\leqq\sum_{i\epsilon P} h_iK_i$. Moreover, the resulting cut must be uniformly stronger than the cut based on Remark 7. These observations were first made (independently) by V. Joseph Bowman and David Sommer. However, a substantially more complicated procedure is required to determine the t_j^* values for the indicated choice of R than for the R of Remark 7. (The specification of such a procedure is developed in Balas, Bowman, Glover, and Sommer.[3])

Remark 11. The strength of the convexity cut increases as the values of the t_j^* increase, and hence (roughly) as the absolute values of the b_{ij} become closer to 0. (The cut becomes uniformly stronger if a 'blocking subset' of the b_{ij} strictly decreases in absolute value without changing sign. It may or may not become stronger if such b_{ij} do not decrease in absolute value but change their sign.) Strengthening of the convexity cuts in this manner can be accomplished for the cuts of Remarks 4–8 and 10 by allowing the y_i to be defined from the x_i in two steps. The first step gives an 'intermediate' y_i equation as an integer combination of the x_i equations for $i\epsilon I$. The second step defines y_i and b_{i0} as before, but with the intermediate y_i equation providing the coefficients and variables in place of an original x_i equation. In particular, since each integer valued t_j variable is identified by an equation $x_i=-(-t_j)$ for some $i\epsilon I$, these equations can be used to make the absolute value of b_{ij} less than 1 for each such t_j and all $i\epsilon P$. This corresponds to

This content downloaded from 132.174.250.143 on Mon, 20 Feb 2017 18:43:52 UTC
All use subject to http://about.jstor.org/terms

the approach used by Gomory to strengthen the mixed integer cuts of reference 12. Since stronger cuts are thus obtained from integer nonbasic variables, it is reasonable to modify the convexity cut when possible so that its slack variable will be integer valued. For the pure integer problem, one way to do this is of course to use (3) as a source inequality for the Gomory all-integer cut, reference 13. Balas proposes such an 'integerizing' of his cuts[1] (although not as part of a strategy for obtaining stronger cuts in the manner indicated here). A reasonable alternative might be to express the convexity cut first in terms of the initial tableau, and then to take an all-integer cut from this expression, whereupon the all-integer cut could be reflected back to the current tableau. (The all-integer cut must sufficiently mirror the convexity cut to be violated by $t=0$ in the current tableau.)

Remark 12. Assume $y \epsilon S$ implies y_i is integer for $i \epsilon P$. Any closed convex set whose boundary contains all points of a unit hypercube (with integer vertices) cannot contain an integer point in its interior, and hence such a convex set is acceptable in the role of R if its interior contains B_0 (and the unit hypercube is defined relative to the y_i for $i \epsilon P$). Also, if $y \epsilon S$ further implies $1 \geqq y_i \geqq 0$ for all $i \epsilon P$, then any convex set is acceptable in the role of R whose interior contains B_0 but no points y such that $y_i=0$ or 1 for all $i \epsilon P$.

The foregoing remarks give rise to some speculations about possible equivalences and dominances among certain classes of convexity cuts:

1. The class of convexity cuts given by Remark 5 is equivalent to the class of convexity cuts given by Remark 6 when the h_i are integers and at least one $h_i=1$ or -1 in equation (8).

2. The class of cuts given by Remark 7 is equivalent to the class of cuts given by Remark 5, if the y_i of Remark 7 are permitted to be taken from integer combinations of the x_i, $i \epsilon I$, as indicated in Remark 11.

These speculations have not, to this time, been demonstrated to be either true or false. Two additional speculations, originally motivated by geometric example, have since been demonstrated to be false. However, the only counterexamples presently known (due to E. G. P. Harran) occur for certain restricted 'obtuse angle' situations. This gives rise to the interesting question of whether it may not be possible to assert that these speculations are 'almost always' true by showing that the conditions under which they are false can be narrowly confined, or circumvented. (For example, in Harran's examples, it is possible to transform the space to an 'acute angle' coordinate system that yields stronger cuts and in which the counterexample no longer holds.) These additional speculations are:

3. If $H_1 t \geqq 1$ is the strongest convexity cut obtained for $R=\{y\colon F_1(y) \leqq K_1\}$ and if $H_2 t \geqq 1$ is the strongest convexity cut obtained for $R=\{y\colon F_2(y) \leqq K_2\}$, where F_1 and F_2 are convex functions, then the cut $(\lambda_1 H_1+\lambda_2 H_2)t \geqq \lambda_1+\lambda_2$ with $\lambda_1, \lambda_2>0$, is at least as strong as the strongest convexity cut obtained for $R=\{y\colon \lambda_1 F_1(y)+\lambda_2 F_2(y) \leqq \lambda_1 K_1+\lambda_2 K_2\}$.

4. The cuts given by Remark 7 are dominated by positive linear combinations of the Gomory cuts.[12] Speculation 4 is a special case of the preceding one. (*See* Note 3.)

CUT SEARCH

THE IDEAS OF cut search are based on associating certain 'proxy' sets of points (usually hyperplanes) with points in the feasible solution space, and probing the

This content downloaded from 132.174.250.143 on Mon, 20 Feb 2017 18:43:52 UTC
All use subject to http://about.jstor.org/terms

space with the edges of the cone (1) for these proxy sets. (*See* Note 4.) Once an appropriate collection of these sets has thus been encountered, the associated 'solution points' are tested for feasibility. The probing is terminated when bounds indicate that one of the candidate solutions is optimal, or when a selected cut is reached off the limit. In the latter case a cut is adjoined, based upon the information generated by the probe. Thus, cut search differs from the convexity-cut approach in that no convex set that contains B_0 and excludes points of S is specified in advance, or even necessarily ever identified. (The cut may in fact exclude points of S contained in the cone, but only if these points have been located via the proxy sets and considered as candidates for optimality.)

A key result underlying our initial application of cut-search ideas is stated as follows.

FIRST CUT-SEARCH LEMMA. *Assume y' is contained in the truncated cone of points satisfying both* (1) *and*

$$\sum_{j\epsilon N} (1/t_j^*)t_j \leqq 1, \tag{13}$$

where $t_j^*>0$ for all $j\epsilon N$. Then every hyperplane $L(y-y')=0$ through y' (for L a nonzero row vector) intersects at least one of the edges of the truncated cone incident at B_0 (one of the line segments $y=B_0-B_jt_j$, $t_j^*>t_j>0$).

Proof. Let $t'\geqq 0$ be a vector satisfying (13) for which $y'=B_0-Bt'$. The representation of $L(y-y')=0$ in terms of the t variables (for y contained in the cone) is $H(t-t')=0$, where $H=LB$. If $Ht'=0$, it is clear that $y=B_0$ lies on the hyperplane $L(y-y')=0$, and the lemma is proved. Thus, suppose $Ht'\neq 0$, and, more specifically, $Ht'>0$ (replacing H by $-H$, if necessary). The assumptions imply that the set $J=\{j\epsilon N: n_jt_j'>0\}$ cannnot be empty (where h_j denotes the jth component of H). Furthermore, identify an index $k\epsilon J$ such that $h_kt_k^*=\max_{j\epsilon J}\{h_jt_j^*\}$, and define $\alpha=Ht'/h_k(\alpha>0)$. Note that the hyperplane $L(y-y')=0$ contains the point $y=B_0-B_k\alpha$, and consequently intersects the half line $y=B_0-B_kt_k$, $t_k\geqq 0$. Thus, to establish the lemma it must be shown that $\alpha\leqq t_k^*$. By assumption, $\sum t_j'/t_j^*\leqq 1$, and hence $\sum (t_k^*/t_j^*)t_j'\leqq t_k^*$. But $\alpha=\sum (h_j/h_k)t_j'$, and the definition of the index k implies $h_j/h_k\leqq t_k^*/t_j^*$ for all j. The conclusion $\alpha\leqq t_k^*$ thus follows from the nonnegativity of the t_j', completing the proof. (The proof also holds for $J=\{j\epsilon N: h_j>0\}$.)

Remark 13. One application of this cut-search lemma arises by letting the proxy set of points corresponding to a point $y\epsilon S$ be the collection of coordinate hyperplanes passing through y. Relative to this collection of hyperplanes, the statement of the lemma becomes: if $y=y'$ lies in the truncated cone defined by (1) and (13), with $t_j^*>0$ for all $_j\epsilon N$, then for each $i\epsilon Q$ (the index set of y) there is some $j\epsilon N$ and some number t_j', $t_j^*\geqq t_j'\geqq 0$, such that $y_i'=b_{i0}-b_{ij}t_j'$.

To apply this observation to the pure integer programming problem (where $I=N$), assume $B_0=A_0$ and $B=A$ (hence y corresponds to x). Further, suppose $y\epsilon S$ is equivalent to the statement that y is a feasible (integer) solution to the pure IP problem. Then all $y'\epsilon S$ can be discovered by parametrically increasing t_j (in discrete increments) on each of the half lines $B_0-B_jt_j$, $t_j\geqq 0$, $j\epsilon N$, where the successive values of t_j, denoted $t_j^1, t_j^2, \cdots$, are given by

$$t_j^1=\min\{t_j: t_j\geqq 0 \quad \text{and} \quad b_{i0}-b_{ij}t_j \text{ is a nonnegative integer for some } i\epsilon N\}$$

and

$$t_j^{r+1}=\min\{t_j: t_j>t_j^r \quad \text{and} \quad b_{i0}-b_{ij}t_j \text{ is a nonnegative integer for some } i\epsilon N\}. \qquad (r\geqq 1)$$

This content downloaded from 132.174.250.143 on Mon, 20 Feb 2017 18:43:52 UTC
All use subject to http://about.jstor.org/terms

If there is no value of t_j that satisfies the indicated restrictions for t_j^1 or t_j^{r+1}, then t_j^1 or t_j^{r+1} (as appropriate) is defined to equal ∞. Also, if y_i has an upper bound, then $b_{i0}-b_{ij}t_j$ may be required to satisfy this bound in the foregoing definitions.

A 'parametric search' procedure for identifying candidate solution points is then given by the following instructions:

1. Begin with $t_j=t_j^1$ for all $j\epsilon N$. Create two integer numbers U_i and L_i for each $i\epsilon N$, with the interpretation that all integer values of y_i currently found are precisely those in the interval $U_i \geqq y_i \geqq L_i$. Since $[b_{i0}]+1 \geqq b_{i0}-b_{ij}t_j^1 \geqq [b_{i0}]$, the initial U_i and L_i values are readily determined for all $i\epsilon N$ such that $b_{i0}-b_{ij}t_j^1$ is an integer for some $j\epsilon N$. By convention, if U_i and L_i are not thus determined for some $i\epsilon N$, then let $U_i=[b_{i0}]$ and $L_i=[b_{i0}]+1$ for this i.

2. If $U_i<L_i$ for some $i\epsilon N$, go to Step 3. Otherwise, identify all integer vectors $y \geqq 0$ (if any exist that were not examined on a previous execution of this step) such that $U_i \geqq y_i \geqq L_i$ for all $i\epsilon N$, and record these as feasible solutions to the IP problem. (The condition $y \geqq 0$ can be tested in the initial tableau from the indicated assignment of values to the y_i for $i\epsilon N$.)

3. Select any $j\epsilon N$ such that the current value t_j^r of t_j is finite (r depends on j). Increase t_j^r to its next value t_j^{r+1} (i.e., increment r by 1) where the definition of t_j^{r+1} given earlier is modified to further stipulate that $b_{i0}-b_{ij}t_j=U_i+1$ or L_i-1, according to whether b_{ij} is negative or positive. If now $t_j^r=\infty$ (for r incremented by 1), repeat Step 3, selecting another j. Otherwise, identify the new L_i or U_i value that results (for one or more $i\epsilon N$) as a consequence of increasing t_j. Then return to Step 2. (If all $t_j^r=\infty$, then the best feasible solution found is optimal and the procedure stops. Other stopping criteria can be specified by determining bounds on the objective function by reference to the cut of Remark 14 to follow.)

The number of y vectors to examine for nonnegativity in Step 2 will be small (perhaps = 0) for the first several executions of this step, but then will tend to grow rapidly. On the other hand, Step 3 permits t_j to bypass a number of irrelevant values by the modified definition of t_j^{r+1}. The significance of this in the cutting context is provided by the next result.

SECOND CUT-SEARCH LEMMA. *Assume $L \leqq B_0-B_jt_j \leqq U$ for all t_j satisfying $0 \leqq t_j \leqq t_j'$. Let $S^*=\{y: L \leqq y \leqq U\}$, and further assume that vectors L' and U' are known for which: $L' \leqq L$, $U' \geqq U$; $L_i'<L_i$ and $U_i'>U_i$ for all finite L_i and U_i; and $y\epsilon S-S^*$ implies $y_i \geqq U_i'>U_i$ or $y_i \leqq L_i'<L_i$ for at least one i. Finally, for each j, let t_j^* be the largest value of t_j such that $L' \leqq B_0-B_jt_j \leqq U'$. Then the cut (3) is satisfied by all y in the cone (1) for which $y\epsilon S-S^*$.*

Proof. First, note that $t_j^*>t_j'$ for all $t_j'<\infty$, and $L' \leqq B_0-B_jt_j \leqq U'$ for all t_j satisfying $0 \leqq t_j \leqq t_j^*$. Suppose the lemma is wrong and there is a $y^0\epsilon S-S^*$ in the cone (1) for which $\sum (1/t_j^*)t_j^0<1$ (where $y^0=B_0-Bt^0$, $t^0 \geqq 0$). We replace all infinite t_j^* by finite numbers that yield $\sum (1/t_j^*)t_j^0=\delta$ for some positive $\delta<1$. Defining $t_j''=\delta t_j^*$, we obtain $t_j''<t_j^*$ for all t_j^* such that $t_j^*>t_j'$. Thus, $b_{i0}-b_{ij}t_j<U_i'$ for all t_j such that $0 \leqq t_j \leqq t_j''$ and for all i such that $U_i<U_i'$. Similarly, $b_{i0}-b_{ij}t_j>L_i'$ for all t_j such that $0 \leqq t_j \leqq t_j''$ and for all i such that $L_i'<L_i$. Thus, y^0 lies in the truncated cone defined by (1) and $\sum (1/t_j'')t_j \leqq 1$, and furthermore, none of the edges of this truncated cone incident at B_0 are intersected by a hyperplane of the form $y_i=\theta$, where $\theta \leqq L_i'<L_i$ or $\theta \geqq U_i'>U_i$. But by assumption y^0 must be contained in at least one such hyperplane. The first cut-search lemma thus provides a contradiction, implying that y cannot be in the truncated cone. This completes the proof.

This content downloaded from 132.174.250.143 on Mon, 20 Feb 2017 18:43:52 UTC
All use subject to http://about.jstor.org/terms

It should be noted that this lemma also holds by replacing inequalities such as $L \leqq y \leqq U$ and $L \leqq B_0 - Bt \leqq U$ with $L \leqq Fy \leqq U$, $L \leqq F(B_0 - Bt) \leqq U$, etc., where F is a suitably dimensioned matrix. In this case the variable y_i is replaced by $F^i y_i$, where F^i is the ith row of R (except for the components of y in the statement $y \epsilon S - S^*$).

Remark 14. A useful consequence of the second cut-search lemma in application to the procedure of Remark 13 is the implication that, after completing the execution of Step 2, one can introduce the cut (3) with $t_j^* = t_j^{r+1}$ for all $j \epsilon N$ (using the modified definition of t_j^{r+1} indicated in Step 3). This is significant, since once $L_j \leqq U_j$ for all j, the parametric search procedure accumulates new y vectors to be

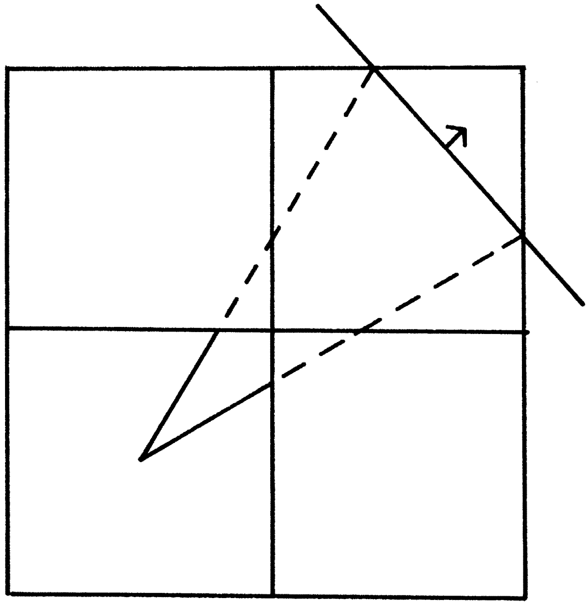

Figure 1

examined at Step 2 each time t_j^r is increased to t_j^{r+1} *for a single* j at Step 3. But the cut is valid by increasing t_j^r to t_j^{r+1} for *all* j, and does not require examination of new integer y vectors before it is enforced. The situation is illustrated in Fig. 1.

The cone (1) is shown with its vertex B_0 in the lower left square. The edges of the cone are extended to the first intersecting coordinate hyperplanes (corresponding to the values t_j^1), and the integer point determined by these hyperplanes is examined for feasibility. (It is clearly feasible relative to the cone, but other problem constraints that may be applicable have not been shown.) If the search procedure is stopped at this point, then the edges of the cone may be extended as indicated by the dotted lines to yield the cut depicted by the line that passes diagonally through the upper right square. (*See* Note 5.)

Remark 15. There are several ways that the procedures of Remarks 13 and 14 can be systematized for greater efficiency. For example, after adjoining the cut and reoptimizing to a new linear programming extreme point (as with a variety of

This content downloaded from 132.174.250.143 on Mon, 20 Feb 2017 18:43:52 UTC
All use subject to http://about.jstor.org/terms

standard cutting methods), the U_i and L_i values from previous extreme points can be used to help bypass the examination of some of the integer y vectors and t_j values applicable to the current extreme point. Also, as the t_j values are increased, the cut of Remark 14 can be kept 'updated' in terms of the y_i variables (by reference to the initial tableau) and the y vectors first checked for feasibility with respect to this cut. If the initial tableau representation of the cut is $\sum_{i\epsilon N} h_i y_i \leqq h_0$, then a simple 'implicit enumeration' procedure that begins with $y_i = U_i$ for $h_i < 0$ and $y_i = L_i$ for $h_i \geqq 0$, and that 'backtracks' whenever the inequality is violated, makes it possible to avoid examination of many of the y vectors altogether. (Note that to avoid reexamining y vectors already examined on previous steps, one automatically excludes an assignment of values to the y_i variables that satisfies $U_i \geqq y_i \geqq L_i$ for all $i\epsilon N$ relative to the U_i and L_i of the preceding execution of Step 2.) One can also pass the coordinate hyperplanes $y_k = b_{k0} - 1$ and $y_k = b_{k0} + 1$ through the cone to identify the edge intersections (if any) that minimize or maximize the variables y_i, $i \neq k$. The utility of such information is that it is applicable to all coordinate hyperplanes $y_k = y_k^*$ that intersect the cone. Finally, one can specify alternative coordinate systems for the integers and apply the procedure relative to one or more of these new systems. (*See* Note 6.)

Remark 16. A cut-search procedure patterned after the foregoing can be implemented for the general MIP problem by solving the linear program that results by assigning the integer values to the y_i, $i\epsilon I$, indicated at Step 2 of Remark 13. However, such an approach can be improved on by creating constraints involving only the integer variables. Such constraints automatically result from successive attempts to solve the linear programs, as noted by BENDERS.[4] However, rather than solve a pure integer program involving these constraints each time a new one is created (as in Benders' 'partitioning' procedure) one can instead use these constraints in the present context to supply a check for the feasibility of integer vectors examined at Step 2. It would seem reasonable to create a number of such constraints initially by generating feasible solutions to the inequalities: $\sum_{i\epsilon M'} a_{ij}\lambda_i \geqq 0$, $j \notin I$; $\lambda_i \geqq 0$, $i\epsilon M' = M - N$, where the a_{ij} are from the initial tableau for the MIP problem. In particular, one might generate solutions that are optimal or near optimal to the linear program whose objective is to minimize $\sum_{i\epsilon M'} a_{i0}\lambda_i$ subject to the foregoing inequalities and $\sum_{i\epsilon M'} \lambda_i = 1$. Each feasible solution $\lambda_i = \lambda_i^*$ then gives the constraint $\sum_{j\epsilon I} a_j^* t_j \leqq a_0^*$ for $a_j^* = \sum_{i\epsilon M'} a_{ij}\lambda_i^*$, $j = 0$ and $j\epsilon I$.

Remark 17. Another application of the cut-search ideas can be made in the context of the following observation. Assume $y\epsilon S$ implies $y\epsilon S_r$ for at least one $r = 1, \cdots, r'$, where S_r is the set of points in a half space or on a hyperplane. Then, if $B_0 \notin S_r$ for all $r \leqq r'$, an acceptable value of t_j^* for the cut (3) is given by the smallest positive value t_j such that $B_0 - B_j t_j \epsilon S_r$ and $S_r \cap S \neq \emptyset$ for some $r \leqq r'$ (where $\emptyset$ denotes the empty set). One example of a possible application of this remark arises for the pure 0-1 IP problem by noting the one-to-one correspondence between the vertices of the hypercube and the half spaces

$$\sum_{i\epsilon N} \delta_i h_i y_i \geqq \delta_0, \qquad (14)$$

where each δ_i, $k\epsilon N$ may be -1 or 1, the h_i are positive constants, and δ_0 is the sum of the positive $\delta_i h_i$. [The intersection of the 2^n half spaces that result by reversing the inequality sign of (14) corresponds to the region R of Remark 10.] The details of determining whether $B_0 - B_j t_j$ is contained in S_r [for S_r given by (14)] are beyond

This content downloaded from 132.174.250.143 on Mon, 20 Feb 2017 18:43:52 UTC
All use subject to http://about.jstor.org/terms

the scope of this paper, but such a determination can be made efficiently for the first few S_r encountered as t_j is increased from 0.

NOTES

1. The first use of the ideas underlying this lemma appears to occur in the procedure of Hoang Tui[14] for minimizing a concave function over a convex polytope, which has subsequently been adapted to mixed 0–1 programming by Ragavachari.[13] Connections between the work of Tui and Young are developed in Glover and Klingman,[10] where it is shown how to modify Tui's procedure to make it finite for Young's problem. (A finiteness proof has not been developed for Tui's procedure in the context of references 13 and 14.)

2. More recently, Balas and Young have allowed for the possibility of assigning the h_i coefficients values other than 1. For these and other generalizations, *see* references 1, 2, 16.

3. Interesting connections between the convexity cuts developed by Balas and convex combinations of Gomory cuts are established by Claude-Alain Burdet in reference 6.

4. Related 'enumerative inequality' approaches for integer programming have been developed by Burdet[7] that likewise succeed in obtaining strengthened cuts. For other approaches to obtaining strengthened cuts, *see also* V. J. Bowman and G. Nemhauser.[5]

5. The cut can be further strengthened in this example by the following observation. If there is an index i and an integer k such that $k \leqq b_{i0} - b_{ij}t_j^* \leqq k+1$ for all j, then each t_j^* can be increased to the largest value that permits the foregoing inequality to remain satisfied.

6. Some particularly interesting consequences of these remarks can be inferred for 'positive' integer corodinate systems, i.e., for systems defined relative to the integer coordinates of a vector $w = My$, where M is an all integer matrix with an all integer inverse, and $MB \leqq 0$. Details of these consequences are developed in reference 9.

REFERENCES

1. Egon Balas, "The Intersection Cut—A New Cutting Plane for Integer Programming," Management Science Research Report No. 187, Carnegie-Mellon University, October, 1969, published in *Opns. Res.* **19,** 19–39 (1971).
2. ———, "Intersection Cuts from Maximal Extensions of the Ball and Octahedron," Graduate School of Industrial Administration, Carnegie-Mellon University, Nov. 1970, presented at the 7th Mathematical Programming Symposium, The Hague, 1970.
3. ———, V. Joseph Bowman, Fred Glover, and David Sommer, "An Intersection Cut from the Dual of the Unit Hypercube," WP #90-69-7, Carnegie-Mellon University, April, 1970, published in *Opns. Res.* **19,** 40–44 (1971).
4. J. F. Benders, "Partitioning Procedures for Solving Mixed-Variables Programming Problems," *Numerische Mathematik* **4,** 238–252 (1962).
5. V. J. Bowman and G. Nemhauser, "Deep Cuts and Integer Programming," Oregon State University Technical Report #8, Feb. 1968.
6. Claude-A. Burdet, "A Class of Cuts and Related Algorithms in Integer Programming," Management Science Research Report #220, Carnegie-Mellon University, Sept. 1970.
7. ———, "Enumerative Inequalities in Integer Programming," Graduate School of Industrial Administration, Carnegie-Mellon University, presented at the 7th Mathematical Programming Symposium, The Hague, 1970.
8. Fred Glover, "Convexity Cuts," working paper, School of Business, University of Texas, Austin, December, 1969.

This content downloaded from 132.174.250.143 on Mon, 20 Feb 2017 18:43:52 UTC
All use subject to http://about.jstor.org/terms

9. ———, "Cut-Search Methods in Integer Programming," MSRS 70-3, School of Business, University of Colorado, September, 1970, presented at the 7th Mathematical Programming Symposium, The Hague, 1970.
10. ——— AND D. KLINGMAN, "Concave Programming Applied to a Special Class of 0-1 Integer Programs," Series in Applied Mathematics for Management, WP 69-30, AMM-11, The University of Texas at Austin, January, 1969, published in *Opns. Res.* **21,** 135–140 (1973).
11. RALPH GOMORY, "An All-Integer Integer Programming Algorithm," in MUTH AND THOMPSON (eds.), *Industrial Scheduling*, Prentice Hall, New York, 1963.
12. ———, "An Algorithm for the Mixed Integer Problem," RM-2597, The Rand Corporation, 1960.
13. M. RAGHAVACHARI, "On the Zero-One Integer Programming Problem," Technical Report No. 20, Department of Statistics, Carnegie-Mellon University, Pittsburgh, Pennsylvania, published in *Opns. Res.* **17,** 680–685 (1969).
14. HOANG, TUI, "Concave Programming Under Linear Constraints" (Russian), *Doklady Akademii Nauk SSSR*, 1964. English translation in *Soviet Mathematics*, 1437–1440 (1964).
15. RICHARD D. YOUNG, "New Cuts for a Special Class of 0 1 Integer Programs," Research Report in Applied Mathematics and Systems Theory, Rice University, Houston, Texas, October, 1968.
16. ———, "Hypercylindrically Deduced Cuts in Zero One Integer Programs," Rice University, Houston, Texas, June, 1970.

This content downloaded from 132.174.250.143 on Mon, 20 Feb 2017 18:43:52 UTC
All use subject to http://about.jstor.org/terms