RUTCOR RESEARCH REPORT

DEPTH-OPTIMIZED CONVEXITY CUTS

Jonathan Eckstein[a] Mikhail Nediak[b]

RRR 23-2003, MAY, 2003; REVISED NOVEMBER, 2003

RUTCOR
Rutgers Center for
Operations Research
Rutgers University
640 Bartholomew Road
Piscataway, New Jersey
08854-8003
Telephone: 732-445-3804
Telefax: 732-445-5472
Email: rrr@rutcor.rutgers.edu
http://rutcor.rutgers.edu/∼rrr

[a]Rutgers Business School and RUTCOR, Rutgers University, 640 Bartholomew Road, Piscataway NJ 08540 USA, jeckstei@rutcor.rutgers.edu

[b]Department of Mathematics and Statistics, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4K1 Canada, nediakm@math.mcmaster.ca

Rutcor Research Report
RRR 23-2003, May, 2003; Revised November, 2003

Depth-Optimized Convexity Cuts

Jonathan Eckstein   Mikhail Nediak

Abstract. This paper presents a general, self-contained treatment of *convexity* or *intersection* cuts. It describes two equivalent ways of generating a cut — via a convex set or a concave function — and a notion of strong cuts. We then characterize the structure of the sets and functions that generate strong cuts. We then specialize the framework to the case of mixed-integer linear programming (MIP). For this case, we formulate two kinds of the deepest cut generation problem, via sets or via functions. We then consider some special cases of the deepest cut generation problem which are amenable to efficient computation. We conclude with computational tests of one of these procedures on a large set of MIPLIB problems.

Acknowledgements: This research was supported in part by National Science Foundation grant CCR-9902092. The second author gratefully acknowledges partial support from DIMACS and SHARCNET. We would also like to thank John Forrest of IBM for his assistance in using the COIN software library and anonymous referees for helpful comments and suggestions.

1 Introduction

This paper studies a class of cutting planes known as *convexity* or *intersection* cuts. Such cuts were first formulated by Tuy [21] in 1964, as part of an algorithm for minimizing a concave function over a linear polyhedron. Raghavachari [18] then directly applied this method to 0-1 integer programming. Due to problems with convergence of the straightforward version of Tuy's method [22], this approach does not appear to have been studied further. Convexity cuts themselves were studied in the integer programming literature in the early seventies; see for example [1, 2, 3, 13]. An overview of this early work can be found in [14]. Convexity cuts are related to disjunctive cuts, which subsequently attracted attention in the integer programming literature [4, 5, 17, 20]. We became interested in convexity cuts because they arose naturally in the context of a general-purpose merit-function-based heuristic for mixed integer programming (MIP) [16].

Our main practical motivation in this paper is to efficiently generate cuts for the standard, general MIP problem

$$\begin{aligned} \min \quad & \mathbf{c}^\top \mathbf{x} && (1)\\ \text{s.t.} \quad & A\mathbf{x} = \mathbf{b} && (2)\\ & \mathbf{l} \le \mathbf{x} \le \mathbf{u} && (3)\\ & x_i \in \mathbb{Z},\ \forall i \in I, && (4) \end{aligned}$$

where A is an $m \times n$ matrix, $I \subseteq \{1, \ldots, n\}$ specifies the indices of the integer variables, and $l_i,\ u_i \in \mathbb{Z}$ for all $i \in I$. For future reference, let P denote the set of all $\mathbf{x} \in \mathbb{R}^n$ satisfying (2) and (3).

This paper revisits the theory of convexity/intersection cuts in a general theoretical setting, and also specializes this setting to the case of MIP. Section 2 gives a general theoretical framework for convexity cuts based on the notion of admissible sets, and briefly discusses its application to MIP. It gives a simple way of partial-ordering cuts by their strength, and characterizes the admissible sets that can generate cuts that are strongest in terms of this order. Section 3 relates this theory to the concept of admissible functions, the context in which convexity cuts originally arose. Section 4 takes the general framework laid out in Sections 2 and 3, and specializes it in more detail to MIP problems, giving several different deepest cut generation problems — auxiliary mathematical programs that select the best possible members from restricted collections of possible convexity cuts. Using the theory of the earlier sections, we construct these collections so they consist only of cuts that could be the deepest ones within some restricted class. From these collections, we attempt to maximize the distance of the cut from a given fractional starting point, measured using a variety of norms. Section 4.1 gives one possible approach within this general framework, using the notion of admissible functions, and 4.2 gives another, employing admissible sets.

Finally, Section 5 gives computational results for the procedure of Section 4.2, showing how much it can close the integrality gap of a large set of problems in the MIPLIB collection, with comparisons to mixed-integer Gomory cuts.

2 Convexity cuts and admissible sets

Consider a polyhedral set $Q \subset \mathbb{R}^n$, and let K be a finitely-generated cone with a vertex at $\bar{\mathbf{x}} \in \mathbb{R}^n$ such that $Q \subseteq K$. Suppose there are k edges of K emanating from $\bar{\mathbf{x}}$, and denote them by

$$\mathbf{x}^{(i)}(t) = \bar{\mathbf{x}} + t\boldsymbol{\xi}^{(i)},\ i = 1, \dots, k. \tag{5}$$

We do not assume a specific form of the polyhedron Q or cone K, except that $k \leq n$. By definition, a finitely generated cone K is a set of the following form:

$$\left\{ \bar{\mathbf{x}} + \sum_{i=1}^{k} \mu_i \boldsymbol{\xi}^{(i)} : \mu_i \geq 0, i = 1, \dots, k \right\}. \tag{6}$$

Note that while $\bar{\mathbf{x}}$ may be a vertex of Q, we do not necessarily assume that to be the case.

Example 2.1. Consider the situation where $Q = P$, the LP relaxation polytope of (1)-(4). Assuming that A is a full-rank matrix, let the point $\bar{\mathbf{x}}$ be a *basic* (but not necessarily feasible) solution of (2)-(3). The point $\bar{\mathbf{x}}$ then corresponds to a partition of the variable indices into three sets J_B, J_L and J_U, denoting basic, nonbasic at the lower bound, and nonbasic at the upper bound, respectively. Partitioning the columns of A into basic and nonbasic parts B and N, we obtain (after reordering indices) a representation for $\bar{\mathbf{x}}$ of the form

$$\bar{\mathbf{x}} = \begin{pmatrix} B^{-1}(\mathbf{b} - N\mathbf{g}) \\ \mathbf{g} \end{pmatrix}, \tag{7}$$

where

$$g_i = \begin{cases} l_i, & i \in J_L \\ u_i, & i \in J_U \end{cases}.$$

Let the entire nonbasic index set be denoted by $J_N = J_L \cup J_U$, and $\mathbf{e}^{(i)}$ denote the i^{th} cannonical unit vector in $\mathbb{R}^{n-m}$. The $n - m$ vectors giving the edge directions are, in this case, of the form

$$\boldsymbol{\xi}^{(i)} = s_i \begin{pmatrix} -B^{-1}N\mathbf{e}^{(i)} \\ \mathbf{e}^{(i)} \end{pmatrix}, \tag{8}$$

where

$$s_i = \begin{cases} +1, & i \in J_L \\ -1, & i \in J_U \end{cases},$$

and i is in the set of non-basic indices J_N.

Lemma 2.2. *Assume that A in (2) is a full-rank matrix, and consider the polyhedron P given by (2)-(3), along with the cone K with vertex $\bar{\mathbf{x}}$ given by (7) and edge directions given by (8) for all $i \in J_N$. Then $P \subseteq K$.*

Proof. Immediate, since $K = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b},\ \mathbf{x}_L \geq \mathbf{l}_L,\ \mathbf{x}_U \leq \mathbf{u}_U\}$. □

Consider now some set $Y \subseteq \mathbb{R}^n$ which contains the points that we "do not wish to cut"; that is, our goal is to cut away as much as possible of the cone K in the vicinity of $\bar{\mathbf{x}}$ without eliminating any members of Y. In the case of MIP, one may think of Y as some set of $\mathbf{x} \in \mathbb{R}^n$ satisfying the integrality conditions (4), as in Example 2.6 below. In our general theoretical framework, however, we let Y be an arbitrary set.

We are interested in valid inequalities for $K \cap Y$ which are not satisfied by $\bar{\mathbf{x}}$. Consider a closed convex set $X \in \mathbb{R}^n$ such that $\bar{\mathbf{x}} \in \operatorname{int} X$ and $(\operatorname{int} X) \cap Y = \emptyset$. We will call such a set *admissible* for Y and $\bar{\mathbf{x}}$. Let J denote the subset of indices of the edges of K which intersect the boundary of X, and $t_i > 0$ be such that $\mathbf{x}^{(i)}(t_i)$, $i \in J$ are the corresponding intersection points. We say that a pair $(\boldsymbol{\alpha}, \beta) \in \mathbb{R}^{n+1}$ specifies a *convexity cut* (an inequality $\boldsymbol{\alpha}^\top \mathbf{x} \geq \beta$) corresponding to $\bar{\mathbf{x}}$ and X if

$$\begin{aligned} \boldsymbol{\alpha}^\top \bar{\mathbf{x}} &< \beta && (9)\\ \boldsymbol{\alpha}^\top \mathbf{x}^{(i)}(t_i) &= \beta,\ i \in J && (10)\\ \boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} &= 0,\ i \notin J && (11) \end{aligned}$$

Remark 2.3. Note that by (9) we have $0 < t_i < \infty$ for all $i \in J$. One can introduce the variable $\tau_i = 1/t_i > 0$ for $i \in J$ and $\tau_i = 0$ for $i \notin J$. With this notation, conditions (10) and (11) can be rewritten as a single system of equations

$$\boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} = \tau_i(\beta - \boldsymbol{\alpha}^\top \bar{\mathbf{x}}),\ i = 1, \ldots, k. \tag{12}$$

Remark 2.4. Note also that, given J and t_i, $i \in J$, the conditions (10)-(11), or equivalently (12), specify a system of linear equations for the parameters $(\boldsymbol{\alpha}, \beta)$ defining a cut.

Proposition 2.5. *Convexity cuts are valid inequalities for $K \cap Y$; that is, if X is admissible for Y and $\bar{\mathbf{x}}$, and $(\boldsymbol{\alpha}, \beta)$ specifies a convexity cut for X and $\bar{\mathbf{x}}$, then $\boldsymbol{\alpha}^\top \mathbf{x} \geq \beta$ for all $\mathbf{x} \in K \cap Y$.*

Proof. See Appendix A. □

Example 2.6. In the case of the problem (1)-(4), we chose Y as follows: without loss of generality, assume that $I = \{1, \ldots, p\}$, $p \leq n$. Let $Y' \subseteq \mathbb{Z}^p$, and $Y = Y' \times \mathbb{R}^{n-p}$. The set Y' typically corresponds to all possible assignments of the integer variables which satisfy the simple bounds (3), *i.e.*,

$$Y' = \{\mathbf{y} \in \mathbb{Z}^p : l_i \leq y_i \leq u_i,\ i = 1, \ldots, p\}.$$

We can also choose Y' so that it excludes assignments of the integer variables that are known to be infeasible due to special problem structure, or assignments that have already been explicitly examined and eliminated from further consideration. Making the set Y smaller in this way can lead to deeper cuts.

For an arbitrary pair $(\boldsymbol{\alpha}, \beta) \in \mathbb{R}^{n+1}$, let $H(\boldsymbol{\alpha}, \beta) = \{\mathbf{x} \in \mathbb{R}^n : \boldsymbol{\alpha}^\top \mathbf{x} \geq \beta\}$. We say that a valid inequality $\bar{\boldsymbol{\alpha}}^\top \mathbf{x} \geq \bar{\beta}$ is *strong relative to* $\boldsymbol{\alpha}^\top \mathbf{x} \geq \beta$ if $H(\bar{\boldsymbol{\alpha}}, \bar{\beta}) \cap K \subseteq H(\boldsymbol{\alpha}, \beta) \cap K$. This relation induces a partial order on the $(\boldsymbol{\alpha}, \beta)$ which may be written $(\bar{\boldsymbol{\alpha}}, \bar{\beta}) \succeq_K (\boldsymbol{\alpha}, \beta)$. The inequality $\bar{\boldsymbol{\alpha}}^\top \mathbf{x} \geq \bar{\beta}$ is said to be *stronger* than $\boldsymbol{\alpha}^\top \mathbf{x} \geq \beta$ if the inclusion $H(\bar{\boldsymbol{\alpha}}, \bar{\beta}) \cap K \subset H(\boldsymbol{\alpha}, \beta) \cap K$ is strict, and *strongest* if there is no inequality that is stronger. Finally, we say that the two inequalities are *equally strong* if each is strong relative to the other, that is, $(\bar{\boldsymbol{\alpha}}, \bar{\beta}) \succeq_K (\boldsymbol{\alpha}, \beta)$ and $(\boldsymbol{\alpha}, \beta) \succeq_K (\bar{\boldsymbol{\alpha}}, \bar{\beta})$.

Proposition 2.7. *Let X and $\bar{X}$ be admissible sets. Let J be the set of indices of edges of K intersecting the boundary of X, and suppose t_i, $i \in J$ give the corresponding intersection points. Also, let $\bar{J}$ and $\bar{t}_i$, $i \in \bar{J}$ be similarly defined for $\bar{X}$. Then a convexity cut corresponding to $\bar{X}$ is strong relative to the one corresponding to X if and only if $\bar{J} \subseteq J$ and $\bar{t}_i \geq t_i$, $i \in \bar{J}$.*

Proof. See Appendix B. □

The following immediate corollary of the Proposition 2.7 allows us to identify equivalence classes of equally strong convexity cuts:

Corollary 2.8. *Using the above notation, two convexity cuts given by $(\bar{\boldsymbol{\alpha}}, \bar{\beta})$ and $(\boldsymbol{\alpha}, \beta)$ are equally strong if and only if $\bar{J} = J$ and $\bar{t}_i = t_i$, $i \in J$.*

This result is natural in light of Remark 2.4.

The following proposition shows that the strength of convexity cuts is monotonic with respect to the inclusion relation of the corresponding admissible sets:

Proposition 2.9. *Let X and $\bar{X}$ be admissible sets. If $X \subseteq \bar{X}$, then a convexity cut corresponding to $\bar{X}$ is strong relative to the one corresponding to X.*

Proof. Immediate from the Proposition 2.7, as $X \subseteq \bar{X}$ implies that $\bar{J} \subseteq J$ and $\bar{t}_i \geq t_i$, $i \in \bar{J}$. □

Note that we are primarily interested in generating strongest cuts. The following proposition describes a subclass of admissible sets which is sufficient to generate all strongest convexity cuts:

Proposition 2.10. *For any admissible set X there exists a mapping $\mathbf{d} : Y \to \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that the set $\bar{X}(\mathbf{d})$ of the form*

$$\bar{X}(\mathbf{d}) = \{\mathbf{x} : \mathbf{d}^\top(\mathbf{y})\mathbf{x} \geq \mathbf{d}^\top(\mathbf{y})\mathbf{y}, \ \forall \mathbf{y} \in Y\} \tag{13}$$

is admissible, and $X \subseteq \bar{X}(\mathbf{d})$.

Proof. For any $\mathbf{y} \in Y$, since $\mathbf{y} \notin \operatorname{int} X$ and $\operatorname{int} X$ is convex, the separation theorem (see Theorem 11.2 of [19]) asserts that there exist a hyperplane passing through $\mathbf{y}$ such that $\operatorname{int} X$ is contained in one of the associated open halfspaces. In other words, there exists $\mathbf{d}(\mathbf{y}) \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that

$$\operatorname{int} X \subseteq \{\mathbf{x} : \mathbf{d}^\top(\mathbf{y})\mathbf{x} > \mathbf{d}^\top(\mathbf{y})\mathbf{y}\}.$$

Choosing one such $\mathbf{d}(\mathbf{y})$ for each $\mathbf{y} \in Y$, one may construct $\bar{X}(\mathbf{d})$ as in (13). Since X is closed and convex, we have for all $\mathbf{y} \in Y$ that

$$X \subseteq \{\mathbf{x} : \mathbf{d}^\top(\mathbf{y})\mathbf{x} \geq \mathbf{d}^\top(\mathbf{y})\mathbf{y}\}$$

by Theorem 6.3 of [19]. Thus, it follows that $X \subseteq \bar{X}(\mathbf{d})$.

Suppose that there exists some $\mathbf{x} \in \operatorname{int} \bar{X}(\mathbf{d})$ and $\mathbf{y} \in Y$ such that $\mathbf{d}^\top(\mathbf{y})\mathbf{x} = \mathbf{d}^\top(\mathbf{y})\mathbf{y}$. Then, $\mathbf{d}(\mathbf{y}) \neq \mathbf{0}$ and the continuity of the linear form $\mathbf{d}^\top(\mathbf{y})\mathbf{x}$ together imply that there is a point $\mathbf{x}'$ with $\mathbf{d}^\top(\mathbf{y})\mathbf{x}' < \mathbf{d}^\top(\mathbf{y})\mathbf{y}$ arbitrarily close to $\mathbf{x}$. Since $\mathbf{x}'$ is in the complement of $\bar{X}(\mathbf{d})$, it follows that $\mathbf{x} \in \operatorname{bd} \bar{X}(\mathbf{d})$, a contradiction. Therefore,

$$\operatorname{int} \bar{X}(\mathbf{d}) \subseteq \{\mathbf{x} : \mathbf{d}^\top(\mathbf{y})\mathbf{x} > \mathbf{d}^\top(\mathbf{y})\mathbf{y},\ \forall \mathbf{y} \in Y\},$$

and Y does not intersect $\operatorname{int} \bar{X}(\mathbf{d})$. Furthermore, the set $\bar{X}(\mathbf{d})$ is closed and convex, since it is an intersection of closed halfspaces. Finally, note that $\bar{\mathbf{x}} \in \operatorname{int} X \subseteq \operatorname{int} \bar{X}(\mathbf{d})$. Thus, $\bar{X}(\mathbf{d})$ is an admissible set. □

Propositions 2.9 and 2.10 imply that, since $X \subseteq \bar{X}(\mathbf{d})$, the cut generated from $\bar{X}(\mathbf{d})$ is strong relative to the one from X. In fact, we have the following corollary:

Corollary 2.11. *For any convexity cut that is strongest, there exists a set of the form (13) that generates it.*

Proof. Given Y and K, consider any strongest convexity cut and a corresponding admissible set X. By Proposition 2.10, there will be a cut from a set $\bar{X} \supseteq X$ of the form (13), which is strong relative to the one from X. Since the given cut is strongest, both have to be equally strong, which, in turn, implies that $\bar{X}$ and X result in the same set J and parameters t_i, $i \in J$ by Corollary 2.8. But then the given cut can also be generated from $\bar{X}$. □

In many applications — in particular, MIP where some of the variables are continuous — the set Y may have special structure. The proposition below specializes Proposition 2.10 to this case:

Proposition 2.12. *If $Y = Y' \times \mathbb{R}^{n-p}$ for some $p < n$ and $Y' \subseteq \mathbb{R}^p$, then for any admissible set X, there exists a mapping $\mathbf{d}' : Y' \to \mathbb{R}^p \setminus \{\mathbf{0}\}$ such that the set*

$$\bar{X}(\mathbf{d}') = \{\mathbf{x} : \sum_{i=1}^{p} d_i'(\mathbf{y})(x_i - y_i) \geq 0,\ \forall \mathbf{y} \in Y'\} \tag{14}$$

is admissible, and $X \subseteq \bar{X}(\mathbf{d}')$.

Proof. Consider the closure of the projection of X onto $\mathbb{R}^p$,

$$X' = \operatorname{cl} \{\mathbf{z} \in \mathbb{R}^p : \exists \mathbf{w} \in \mathbb{R}^{n-p} : (\mathbf{z}, \mathbf{w}) \in X\}$$

and the projection $\bar{\mathbf{x}}'$ of $\bar{\mathbf{x}}$ onto $\mathbb{R}^p$. We now claim that X' is admissible for Y' and $\bar{\mathbf{x}}'$. To establish this claim, note first that X' is closed and convex. The interior of X' contains the projection of $\operatorname{int} X$ (a rectangular neighborhood of any point in $\mathbb{R}^n$ is mapped to a rectangular neighborhood of its projection onto $\mathbb{R}^p$), so $\bar{\mathbf{x}}' \in \operatorname{int} X'$. Thus, to prove admissibility of X', we only need to prove that $Y' \cap \operatorname{int} X' = \emptyset$. Suppose that there exists a $\mathbf{y}' \in Y' \cap \operatorname{int} X'$; then there exist a neighborhood $N_{\mathbf{y}'}$ of $\mathbf{y}'$ contained in X'. Pass a line through $\bar{\mathbf{x}}'$ and $\mathbf{y}'$ and consider a point $\tilde{\mathbf{x}}' \in N_{\mathbf{y}'} \setminus \{\mathbf{y}'\}$ on it, such that, for some small enough $\lambda \in (0, 1)$, we have $\mathbf{y}' = \lambda\bar{\mathbf{x}}' + (1 - \lambda)\tilde{\mathbf{x}}'$:

$$\tilde{\mathbf{x}}' = \mathbf{y}' + \frac{\lambda}{1-\lambda}(\mathbf{y}' - \bar{\mathbf{x}}').$$

Consider any preimage $\tilde{\mathbf{x}} \in X$ of $\tilde{\mathbf{x}}'$ under projection, and an open neighborhood $N_{\bar{\mathbf{x}}}$ of $\bar{\mathbf{x}}$, entirely contained in X (such a neighborhood exists since $\bar{\mathbf{x}}$ is interior point of X). Note that the Minkowski sum $\lambda N_{\bar{\mathbf{x}}} + (1 - \lambda)\tilde{\mathbf{x}}$ of $N_{\bar{\mathbf{x}}}$ and $\{\tilde{\mathbf{x}}\}$ is an open set, since $N_{\bar{\mathbf{x}}}$ is open, and is contained in X, since X is convex. But then the point $\mathbf{y} = \lambda\bar{\mathbf{x}} + (1 - \lambda)\tilde{\mathbf{x}} \in Y$ is an interior point of X, contradicting X's admissibility. Thus, X' is admissible.

Construct an admissible set $\bar{X}'(\mathbf{d}') \subseteq \mathbb{R}^p$ for Y' and $\bar{\mathbf{x}}'$ of the form (13) containing X' ($\bar{X}'(\mathbf{d}')$ exists by Proposition 2.10). So, $\bar{X}(\mathbf{d}') = \bar{X}'(\mathbf{d}') \times \mathbb{R}^{n-p}$ is admissible, contains X, and is of the form (14). □

We immediately get the equivalent of Corollary 2.11:

Corollary 2.13. *If $Y = Y' \times \mathbb{R}^{n-p}$ for some $p < n$ and $Y' \subseteq \mathbb{R}^p$, then for any convexity cut that is strongest, there exists a set of the form (14) that generates it.*

3 An alternative description: admissible functions

In this section, we will show how convexity cuts can be equivalently obtained from concave functions rather than convex sets. Convexity cuts were originally formulated in this context [21], and it may be a more natural approach in some applications [16].

We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is *admissible* for $Y \subset \mathbb{R}^n$ and $\bar{\mathbf{x}} \in \mathbb{R}^n$ if it is continuous, concave, $f(\bar{\mathbf{x}}) > 0$, and $f(\mathbf{y}) \leq 0$ for all $\mathbf{y} \in Y$.

Proposition 3.1. *If $f(\cdot)$ is admissible for $\bar{\mathbf{x}}$ and Y, then $X = \{\mathbf{x} : f(\mathbf{x}) \geq 0\}$ is an admissible set for $\bar{\mathbf{x}}$ and Y, as defined in Section 2.*

Proof. First, note that X is convex and closed by the concavity and continuity of $f(\cdot)$. Second, continuity implies $\bar{\mathbf{x}} \in \operatorname{int} X$, since there exist a neighborhood $N_{\bar{\mathbf{x}}}$ of $\bar{\mathbf{x}}$ such that $f(\mathbf{x}) > 0$ for all $\mathbf{x} \in N_{\bar{\mathbf{x}}}$. Third, we claim that $f(\mathbf{x}) > 0$ for all $\mathbf{x} \in \operatorname{int} X$. Suppose this is not the case, and there exists $\mathbf{x} \in \operatorname{int} X$ such that $f(\mathbf{x}) = 0$. For $0 < \lambda < 1$, consider the point

$$\mathbf{x}_\lambda = \mathbf{x} + \frac{\lambda}{1-\lambda}(\mathbf{x} - \bar{\mathbf{x}}),$$

and note that $\mathbf{x} = \lambda\bar{\mathbf{x}} + (1 - \lambda)\mathbf{x}_\lambda$. Then, by concavity,

$$0 = f(\mathbf{x}) \geq \lambda f(\bar{\mathbf{x}}) + (1 - \lambda)f(\mathbf{x}_\lambda)$$

and, for all $0 < \lambda < 1$,

$$f(\mathbf{x}_\lambda) \leq \frac{-\lambda f(\bar{\mathbf{x}})}{1-\lambda} < 0$$

which implies that $\mathbf{x}_\lambda \notin X$. Note also that

$$\|\mathbf{x}_\lambda - \mathbf{x}\| = \frac{\lambda}{1-\lambda}\|\mathbf{x} - \bar{\mathbf{x}}\|,$$

so by taking small enough $\lambda > 0$, we get a point $\mathbf{x}_\lambda \notin X$ arbitrarily close to $\mathbf{x}$. This implies that $\mathbf{x}$ is a boundary point of X, a contradiction. Thus, $\operatorname{int} X \cap Y = \emptyset$. □

The Proposition 3.1 allows us to generate a convexity cut from an admissible function $f(\cdot)$ by simply using its corresponding admissible set. Proposition 2.9 can be reformulated in terms of the admissible functions as follows:

Proposition 3.2. *Let $f(\cdot)$ and $\bar{f}(\cdot)$ be admissible functions such that $f(\mathbf{x}) \leq \bar{f}(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$. Then a convexity cut generated from $\bar{f}(\cdot)$ is strong relative to the one from $f(\cdot)$.*

Proof. Immediate. □

We can also describe in terms of admissible functions the class (13) of admissible sets sufficient for generating all strongest cuts. We say that a set X and a function $f(\cdot)$ *correspond* if $X = \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \geq 0\}$:

Proposition 3.3. *For any admissible set $\bar{X}(\mathbf{d})$ of the form (13) there exists a mapping $\bar{\mathbf{d}} : Y \to \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that the function*

$$f(\mathbf{x}|\bar{\mathbf{d}}) = \inf_{\mathbf{y}\in Y}\{\bar{\mathbf{d}}^\top(\mathbf{y})(\mathbf{x}-\mathbf{y})\} \tag{15}$$

is admissible, and $\bar{X}(\mathbf{d})$ and $f(\mathbf{x}|\bar{\mathbf{d}})$ correspond. Moreover, for any admissible function of the form (15), there is a corresponding admissible set of the form (13).

Proof. See Appendix C. □

When $Y = Y' \times \mathbb{R}^{n-p}$ for some subset Y' of $\mathbb{R}^p$, similarly to Proposition 2.12, we can obtain a smaller class of admissible functions sufficient for generation of all strongest cuts:

Corollary 3.4. *If $Y = Y' \times \mathbb{R}^{n-p}$ for some $p < n$ and $Y' \subseteq \mathbb{R}^p$, then for any admissible set $\bar{X}(\mathbf{d}')$ of the form (14) there exists a mapping $\bar{\mathbf{d}}' : Y \to \mathbb{R}^p \setminus \{\mathbf{0}\}$, such that the function*

$$f(\mathbf{x}|\bar{\mathbf{d}}') = \inf_{\mathbf{y}\in Y'}\left\{\sum_{i=1}^{p} \bar{d}'_i(\mathbf{y})(x_i - y_i)\right\} \tag{16}$$

is admissible, and $\bar{X}(\mathbf{d}')$ and $f(\mathbf{x}|\bar{\mathbf{d}}')$ correspond. Moreover, for any admissible function of the form (16), there is a corresponding admissible set of the form (14).

Proof. Also see Appendix C. □

4 Specialization of convexity cuts to the case of MIP

This section specializes our general convexity cut framework to the case of MIP, as already suggested in Examples 2.1 and 2.6. In particular, it gives a few notable subclasses of strongest convexity cuts and formulates ways of selecting cuts from these subclasses through a "deepest cut selection problem". The following two subsections present the deepest cut generation problem, formulated in terms of admissible functions and admissible sets. Although the two approaches are equivalent, it is more convenient to generate certain classes of cuts through the admissible functions rather than admissible sets, and *vice versa*.

4.1 Deepest cut generation via admissible functions

Corollary 3.4 describes the class (16) of admissible functions that allow one to generate all strongest convexity cuts. In the following discussion, we will assume the case of 0-1 MIP, *i.e.* $Y' = \{0,1\}^p$, and present a way to treat general MIP at the end of this subsection. Even with this restriction to the binary case, the number of parameters describing a function of the form (16) is $p2^p$. Thus, to obtain a more tractable class of admissible functions, we will impose the additional restriction that they are separable, *i.e.*, they are of the form

$$f(\mathbf{x}) = \sum_{i=0}^{p} f_i(x_i). \tag{17}$$

Proposition 4.1. *Let $Y' = \{0,1\}^p$. Then for any separable admissible function (17) there exists an admissible function of the form*

$$\bar{f}(\mathbf{x}) = \sum_{j=1}^{p} \gamma_j \min\left\{\frac{x_j}{\tilde{x}_j}, \frac{1-x_j}{1-\tilde{x}_j}\right\} \tag{18}$$

with some $\boldsymbol{\gamma} \in \mathbb{R}^p_+$, $\tilde{\mathbf{x}} \in [0,1]^p$, such that $\bar{f}(\mathbf{x}) \geq f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$.

Proof. Consider a particular component $f_i(x_i)$ of $f(\mathbf{x})$ and let $\delta_i(0)$ and $\delta_i(1)$ be subgradients of $f_i(x_i)$ at 0 and 1 respectively. Then, by definition,

$$\begin{aligned} f_i(0) + x\delta_i(0) &\geq f_i(x) \\ f_i(1) + (x-1)\delta_i(1) &\geq f_i(x) \end{aligned}$$

for all $x \in \mathbb{R}$. These inequalities imply

$$\begin{aligned} x\delta_i(0) &\geq f_i(x) - \max\{f_i(0), f_i(1)\} \\ (x-1)\delta_i(1) &\geq f_i(x) - \max\{f_i(0), f_i(1)\} \end{aligned}$$

for all $x \in \mathbb{R}$. Let

$$\bar{f}_i(x) = \min\{x\delta_i(0), (x-1)\delta_i(1)\},$$

and note that we then have

$$\bar{f}_i(x) \geq f_i(x) - \max\{f_i(0), f_i(1)\}$$

for all $x \in \mathbb{R}$. It follows that

$$\bar{f}(\mathbf{x}) = \sum_{i=1}^{p} \bar{f}_i(x_i) \geq f(\mathbf{x}) - \sum_{i=1}^{p} \max\{f_i(0), f_i(1)\}.$$

Due to a separability of $f(\cdot)$, we have

$$\max\{f(\mathbf{y}) : \mathbf{y} \in Y\} = \sum_{i=1}^{p} \max\{f_i(0), f_i(1)\}.$$

Due to f's admissibility, $f(\mathbf{y}) \leq 0$ for all $\mathbf{y} \in Y$, and consequently

$$\sum_{i=1}^{p} \max\{f_i(0), f_i(1)\} \leq 0.$$

Thus, $\bar{f}(\mathbf{x}) \geq f(\mathbf{x})$.

Admissibility of $\bar{f}(\mathbf{x})$ is immediate. To see that it can be transformed to the form (18) take

$$\gamma_i = \frac{\delta_i(0)\delta_i(1)}{\delta_i(0) + \delta_i(1)}$$

and

$$\tilde{\mathbf{x}}_i = \begin{cases} \frac{1}{2}, & \gamma_i = 0 \\ \frac{\delta_i(1)}{\delta_i(0)+\delta_i(1)}, & \gamma_i \neq 0. \end{cases}$$

for $i = 1, \ldots, p$. □

The above proposition, together with Proposition 3.2, implies that functions of the the form (18) are sufficient to generate all strongest convexity cuts that can be obtained from separable admissible functions. Furthermore, we also have the following proposition:

Proposition 4.2. *Any function $f(\cdot)$ of the form (18), where $\boldsymbol{\gamma} \in \mathbb{R}^p \setminus \{\mathbf{0}\}$ and $\tilde{\mathbf{x}} \in (0,1)^p$, is admissible for $Y = \{0,1\}^p \times \mathbb{R}^{n-p}$ and any $\bar{\mathbf{x}} \in [0,1]^p \times \mathbb{R}^{n-p}$ with at least one fractional component $\bar{x}_i \in (0,1)$ where i is such that $\gamma_i > 0$. It also belongs to the class (16), where the mapping $\bar{\mathbf{d}} : \{0,1\}^p \to \mathbb{R}^p \setminus \{\mathbf{0}\}$ is given by*

$$\bar{d}_i(\mathbf{y}) = \frac{\gamma_i}{\tilde{x}_i - y_i}, \ \forall \mathbf{y} \in \{0,1\}^p, \ i = 1, \ldots, p. \tag{19}$$

Proof. First, observe that $\bar{f}(\mathbf{y}) = 0$ for all $\mathbf{y} \in Y = \{0,1\}^p \times \mathbb{R}^{n-p}$ and any function $\bar{f}(\cdot)$ of the form (18). Second, since there exists an index $j^* \in I$ such that $\bar{x}_{j^*}$ is fractional with $\gamma_{j^*} > 0$, we have

$$\bar{f}(\bar{\mathbf{x}}) \geq \gamma_{j^*} \min \left\{ \frac{\bar{x}_{j^*}}{\tilde{x}_{j^*}}, \frac{1-\bar{x}_{j^*}}{1-\tilde{x}_{j^*}} \right\} > 0.$$

Admissibility follows.

Consider now a function $f(\cdot|\bar{\mathbf{d}})$ of the form (16), where $\bar{\mathbf{d}}$ is given by (19):

$$f(\mathbf{x}|\bar{\mathbf{d}}) = \min_{\mathbf{y}\in\{0,1\}^p} \left\{ \sum_{i=1}^{p} \gamma_i \frac{x_i - y_i}{\tilde{x}_i - y_i} \right\},$$

where we have replaced "inf" with "min" by the finiteness of $Y' = \{0,1\}^p$. The min operation can be taken before the summation, since each summand depends only on the i^{th} components of $\mathbf{x}$, $\tilde{\mathbf{x}}$ and $\mathbf{y}$:

$$f(\mathbf{x}|\mathbf{d}) = \sum_{i=1}^{p} \gamma_i \min_{y_i\in\{0,1\}} \left\{ \frac{x_i - y_i}{\tilde{x}_i - y_i} \right\} = \bar{f}(\mathbf{x}).$$

□

Note that the class of functions (18) is quite general, as the corresponding admissible sets will include all the "band" sets of the form

$$X_i = \{\mathbf{x} : 0 \leq x_i \leq 1\}, \quad i = 1, \ldots, p. \tag{20}$$

To obtain such an admissible set, we would select $\gamma_i = 1$ and $\gamma_j = 0$ for all $j \neq i$. As stated in [4], under certain normalization conditions, disjunctive cuts for 0-1 MIP coincide with convexity cuts obtained from the X_i, and thus the class of convexity cuts generated by functions of the form (18) should include standard disjunctive cuts. A detailed discussion of the relation between disjunctive cuts obtained from single-variable disjunctions of the form $(x_i = 0) \vee (x_i = 1)$, intersection cuts from the band sets X_i, and mixed-integer Gomory cuts can be found in [8, 11]. It is not clear at this stage, however, if there exists a similar relationship between more general admissible sets generating strongest cuts and disjunctive cuts from multivariate disjunctions, and we see this as an interesting opportunity for future research. Another benefit of using admissible functions from the separable class is that the problem of selecting an admissible function reduces to the selection of at most $2p$ parameters γ_i, $\tilde{x}_i$, $i = 1, \ldots, p$. Moreover, one could even keep the $\tilde{x}_i$'s fixed, and still get a collection of admissible sets that includes all the band sets X_i.

We now construct a procedure for selecting an admissible function generating the "best" convexity cuts. One natural criterion is the distance from the vertex $\bar{\mathbf{x}}$ of the cone to the hyperplane $L(\boldsymbol{\alpha}, \beta) = \{\mathbf{x} : \boldsymbol{\alpha}^\top \mathbf{x} = \beta\}$ defining the cut. We measure this distance with an arbitrary norm $\nu(\cdot)$ on $\mathbb{R}^n$:

$$d_\nu(\bar{\mathbf{x}}, L(\boldsymbol{\alpha}, \beta)) = \inf\{\nu(\mathbf{x} - \bar{\mathbf{x}}) : \boldsymbol{\alpha}^\top \mathbf{x} = \beta\}. \tag{21}$$

Examples of possible norms include the s-norm, for $1 \le s < \infty$

$$\nu(\mathbf{x}) = \left(\sum_{i=1}^{n} |x_i|^s \right)^{1/s}$$

and the ∞-norm

$$\nu(\mathbf{x}) = \max_{i=1,\ldots,n} |x_i|.$$

The following lemma and its corollary present a closed-form expression for $d_\nu(\bar{\mathbf{x}}, L(\boldsymbol{\alpha}, \beta))$ in terms of the data and the polar norm

$$\nu^\circ(\boldsymbol{\alpha}) = \inf \left\{ \frac{\nu(\mathbf{z})}{\boldsymbol{\alpha}^\top \mathbf{z}} : \mathbf{z} \neq \mathbf{0} \right\}.$$

Lemma 4.3. $d_\nu(\mathbf{0}, L(\boldsymbol{\alpha}, 1)) = (\nu^\circ(\boldsymbol{\alpha}))^{-1}$.

Proof. In this special case, the distance d_ν^* from $\mathbf{0}$ to $L(\boldsymbol{\alpha}, 1)$ is

$$d_\nu^* = \inf\{\nu(\mathbf{y}) : \boldsymbol{\alpha}^\top \mathbf{y} = 1\}.$$

Performing the change of variables $\mathbf{y} = \lambda \mathbf{z}$, $\nu(\mathbf{z}) = 1$, $\lambda \ge 0$,

$$d_\nu^* = \lambda^* = \inf \left\{ \lambda \ge 0 : \exists \mathbf{z} : \boldsymbol{\alpha}^\top \mathbf{z} = \frac{1}{\lambda},\ \nu(\mathbf{z}) = 1 \right\}.$$

Note that $\lambda^* = 1/\mu^*$, where

$$\mu^* = \sup \left\{ \mu \ge 0 : \exists \mathbf{z} : \boldsymbol{\alpha}^\top \mathbf{z} = \mu,\ \nu(\mathbf{z}) = 1 \right\} = \sup \left\{ \frac{\boldsymbol{\alpha}^\top \mathbf{z}}{\nu(\mathbf{z})} : \mathbf{z} \neq 0 \right\} = \nu^\circ(\boldsymbol{\alpha}).$$

Thus $d_\nu^* = (\nu^\circ(\boldsymbol{\alpha}))^{-1}$. $\square$

Corollary 4.4.

$$d_\nu(\bar{\mathbf{x}}, L(\boldsymbol{\alpha}, \beta)) = \frac{|\beta - \boldsymbol{\alpha}^\top \bar{\mathbf{x}}|}{\nu^\circ(\boldsymbol{\alpha})} \tag{22}$$

Proof. Note that the formula holds when $\beta = \boldsymbol{\alpha}^\top \bar{\mathbf{x}}$. Transform the problem in the definition (21) of $d_\nu(\bar{\mathbf{x}}, L(\boldsymbol{\alpha}, \beta))$ by the change of variables $\mathbf{y} = \mathbf{x} - \bar{\mathbf{x}}$ to

$$d_\nu(\bar{\mathbf{x}}, L(\boldsymbol{\alpha}, \beta)) = \inf\{\nu(\mathbf{y}) : \boldsymbol{\alpha}^\top \mathbf{y} = \beta - \boldsymbol{\alpha}^\top \bar{\mathbf{x}}\}.$$

Assume without loss of generality that $\beta - \boldsymbol{\alpha}^\top \bar{\mathbf{x}} > 0$ and divide both sides of the equality by this value. Then

$$d_\nu(\bar{\mathbf{x}}, L(\boldsymbol{\alpha}, \beta)) = \inf \left\{ \nu(\mathbf{y}) : \left(\frac{\boldsymbol{\alpha}}{\beta - \boldsymbol{\alpha}^\top \bar{\mathbf{x}}} \right)^\top \mathbf{y} = 1 \right\} = \left(\nu^\circ \left(\frac{\boldsymbol{\alpha}}{\beta - \boldsymbol{\alpha}^\top \bar{\mathbf{x}}} \right) \right)^{-1}.$$

The statement follows by the homogeneity property of the norm. $\square$

Remark 4.5. It is well known (see, for example, Section 15 of [19]) that the polar of a s-norm is a t-norm and vice versa, where $1/s + 1/t = 1$ for $s, t > 1$, and $s = \infty$ for $t = 1$. We can also consider more general class of norms by scaling individual coordinates. For example, for an arbitrary strictly positive vector $\mathbf{w}$, consider a generalized 1-norm

$$\nu(\mathbf{x}) = \sum_{i=1}^{n} w_i |x_i|. \tag{23}$$

Its polar is the following generalized ∞-norm

$$\nu^\circ(\mathbf{x}) = \max_{i=1,\dots,n} \left| \frac{x_i}{w_i} \right|. \tag{24}$$

We can now formulate the *deepest convexity cut selection problem* as that of finding a deepest cut in the set of all convexity cuts generated by some given class of admissible functions $\mathcal{F}$. One possible formulation is

$$\begin{aligned}
\sup\nolimits_{\boldsymbol{\alpha},\beta,\mathbf{t},f(\cdot)} \quad & \frac{\beta - \boldsymbol{\alpha}^\top \bar{\mathbf{x}}}{\nu^\circ(\boldsymbol{\alpha})} && (25)\\
\text{s.t.} \quad & f(\bar{\mathbf{x}} + t_i \boldsymbol{\xi}^{(i)}) \geq 0,\ i = 1, \dots, k && (26)\\
& \boldsymbol{\alpha}^\top (\bar{\mathbf{x}} + t_i \boldsymbol{\xi}^{(i)}) = \beta,\ i = 1, \dots, k && (27)\\
& f(\cdot) \in \mathcal{F} && (28)\\
& \mathbf{t} > 0 && (29)
\end{aligned}$$

Remark 4.6. Note that, when $k < n$, the system of equations (27), viewed as a system on $\boldsymbol{\alpha}$ for fixed $\mathbf{t}$, is underdetermined. Thus, selecting $\mathbf{t}$ does not determine the cut uniquely. All the cuts corresponding to the same $\mathbf{t}$ will be equally strong (Corollary 2.8), although the objective function value (25) may be different. Since the usefulness of all such cuts is likely to be similar, we may impose $n - k$ additional equations on $\boldsymbol{\alpha}$ so that there is a one-to-one correspondence between $\mathbf{t}$ and $\boldsymbol{\alpha}$. This modification will effectively change our problem into the one of selecting a convexity cut equivalence class that delivers the deepest representative satisfying the additional $n - k$ constraints.

The formulation (25)-(29) still has the difficulty, however, that we only use condition (10) in the convexity cut definition. Thus, we implicitly assume that all rays of the cone K intersect with admissible set X corresponding to $f(\cdot)$. This assumption may result in the problem only having solution in the limit when some of the t_i's tend to ∞. One may be able to fix the formulation by switching to a parameterization of convexity cuts as in Remark 2.3 ($\tau_i = 1/t_i$ for all i such that the edge $\mathbf{x}^{(i)}(t)$ intersects with a boundary of the admissible set and $\tau_i = 0$, otherwise). In general, this change of parameterization can be done whenever $\mathcal{F}$ is a set of piecewise linear functions of the form (16). Note, however, that in this case the pair of constraints (26),(28) can be equivalently restated in terms of admissible sets (14). We postpone this general treatment until the next subsection and, instead, focus on the case when $f(\cdot)$ is separable piecewise linear of the form (18) with some additional restrictions.

To simplify the formulation, we impose a restriction that

$$\tilde{x}_j = \begin{cases} \bar{x}_j, & \bar{x}_j \in (0,1) \\ \frac{1}{2}, & \bar{x}_j \in \{0,1\} \end{cases} \tag{30}$$

and $\gamma_j = 0$ for all j such that $\bar{x}_j \in \{0,1\}$. The restricted function $\bar{f}(\mathbf{x})$ is still admissible as long as $\bar{\mathbf{x}}$ has at least one fractional component among the the first p. In fact, we can assume, without loss of generality, that all $\bar{x}_j$, $j = 1, \ldots, p$ are fractional (otherwise, relabel components of $\bar{\mathbf{x}}$ so that the first p' are fractional, and use p' instead of p in the problem formulation). Finally, we assume that $\sum_{j=0}^{p} \gamma_j = 1$ to bound the maximum value of the function and also ensure that it is not constant zero. Now, consider the value of $f(\cdot)$ along a particular edge $\mathbf{x}^{(i)}(t)$:

$$\begin{aligned} f(\bar{\mathbf{x}} + t\boldsymbol{\xi}^{(i)}) &= \sum_{j=0}^{p} f_j(\bar{x}_j + t\xi_j^{(i)}) \\ &= \sum_{j:\xi_j^{(i)}>0} \gamma_j \frac{1-(\bar{x}_j + t\xi_j^{(i)})}{1-\bar{x}_j} + \sum_{j:\xi_j^{(i)}<0} \gamma_j \frac{\bar{x}_j + t\xi_j^{(i)}}{\bar{x}_j} + \sum_{j:\xi_j^{(i)}=0} \gamma_j \\ &= \sum_{j:\xi_j^{(i)}>0} \gamma_j \left(1 - t\frac{\xi_j^{(i)}}{1-\bar{x}_j}\right) + \sum_{j:\xi_j^{(i)}<0} \gamma_j \left(1 + t\frac{\xi_j^{(i)}}{\bar{x}_j}\right) + \sum_{j:\xi_j^{(i)}=0} \gamma_j \\ &= -t\left(\sum_{j:\xi_j^{(i)}>0} \gamma_j \frac{\xi_j^{(i)}}{1-\bar{x}_j} - \sum_{j:\xi_j^{(i)}<0} \gamma_j \frac{\xi_j^{(i)}}{\bar{x}_j}\right) + \sum_{j=0}^{p} \gamma_j. \end{aligned}$$

We introduce vectors $\boldsymbol{\eta}^{(i)}$ defined via

$$\eta_j^{(i)} = \begin{cases} \frac{\xi_j^{(i)}}{1-\bar{x}_j}, & \xi_j^{(i)} > 0 \\ -\frac{\xi_j^{(i)}}{\bar{x}_j}, & \xi_j^{(i)} < 0 \\ 0, & \xi_j^{(i)} = 0 \end{cases} \tag{31}$$

Recalling that $\sum_{j=0}^{p} \gamma_j = 1$, we get

$$f(\bar{\mathbf{x}} + t\boldsymbol{\xi}^{(i)}) = 1 - t\boldsymbol{\gamma}^{\top}\boldsymbol{\eta}^{(i)}.$$

After introduction of $\tau_i = 1/t_i$, the constraints (26) can be rewritten as

$$\boldsymbol{\gamma}^{\top}\boldsymbol{\eta}^{(i)} \le \tau_i,\ i = 1, \ldots, k,$$

and the whole problem (25)-(29) as

$$\begin{array}{ll} \max_{\boldsymbol{\alpha},\beta,\boldsymbol{\tau},\boldsymbol{\gamma}} & \dfrac{\beta - \boldsymbol{\alpha}^\top \bar{\mathbf{x}}}{\nu^\circ(\boldsymbol{\alpha})} \\ \text{s.t.} & \boldsymbol{\gamma}^\top \boldsymbol{\eta}^{(i)} \le \tau_i,\ i = 1, \ldots, k \\ & \boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} = \tau_i(\beta - \boldsymbol{\alpha}\bar{\mathbf{x}}),\ i = 1, \ldots, k \\ & \sum_{j=0}^{p} \gamma_j = 1,\ \boldsymbol{\gamma} \ge 0 \\ & \boldsymbol{\tau} \ge 0 \end{array}$$

Note that the value of the objective function is invariant under the scale transformations of the pair $(\boldsymbol{\alpha}, \beta)$. Thus, we can impose an arbitrary normalization condition such as, for example, $\beta - \boldsymbol{\alpha}\bar{\mathbf{x}} = 1$. Then, we will be able to eliminate $\boldsymbol{\tau}$ from the problem, since

$$\boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} = \tau_i(\beta - \boldsymbol{\alpha}\bar{\mathbf{x}}) = \tau_i.$$

After taking a reciprocal of the objective and switching to minimization, we obtain:

$$\begin{array}{ll} \min_{\boldsymbol{\alpha},\boldsymbol{\gamma}} & \nu^\circ(\boldsymbol{\alpha}) \qquad (32) \\ \text{s.t.} & \boldsymbol{\gamma}^\top \boldsymbol{\eta}^{(i)} \le \boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)},\ i = 1, \ldots, k \qquad (33) \\ & \boldsymbol{\gamma}^\top \mathbf{1} = 1,\ \boldsymbol{\gamma} \ge 0 \qquad (34) \end{array}$$

Note that constraints of this problem are linear. Now we need to make a specific choice of $\nu(\cdot)$ or, equivalently, $\nu^\circ(\cdot)$. It is well-known that if $\nu^\circ(\cdot)$ is the 1- or ∞-norm, then the problem of minimizing the norm can be replaced by a minimization of a linear objective subject to additional linear constraints, and thus the overall problem will be linear. Another potentially interesting choice is to use 2-norm, which will result in a quadratic programming problem.

We conclude this section by a detailed analysis of the 1- and ∞-norm cases when K arises from a feasible solution to the LP relaxation (Example 2.1). The number of rays k forming a cone K in this case is $n - m$. We will label these rays by the indices of the corresponding non-basic variables $i \in J_N$. Note that with this cone structure, for a given $\boldsymbol{\tau}$, the number of variables in

$$\boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} = \tau_i, i \in J_N$$

is n, while the number of constraints is $n - m$. Following the motivation of Remark 4.6, we impose m additional constraints on $\boldsymbol{\alpha}$, so that only one cut is considered for each equivalence class. In particular, it is convenient to assume that $\alpha_j = 0,\ j \in J_B$, because in this case

$$\boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} = \boldsymbol{\alpha}_B^\top \boldsymbol{\xi}_B^{(i)} + \alpha_i s_i = \alpha_i s_i.$$

Strictly speaking, a problem obtained in this way is not equivalent to the original one. However, this assumption makes a problem simpler, which is important in practice.

First, let $\nu(\cdot)$ be the ∞-norm, in which case $\nu^\circ(\cdot)$ is the 1-norm. This choice results in (32)-(34) reducing to

$$\begin{array}{ll} \min_{\boldsymbol{\alpha},\boldsymbol{\gamma}} & \sum_{i\in J_N} |\alpha_i| \\ \text{s.t.} & \boldsymbol{\gamma}^\top \boldsymbol{\eta}^{(i)} \le \alpha_i s_i,\ i \in J_N \\ & \boldsymbol{\gamma}^\top \mathbf{1} = 1,\ \boldsymbol{\gamma} \ge 0. \end{array}$$

Introducing $z_i = \alpha_i s_i$, and noting that $\boldsymbol{\gamma} \ge 0$ and $\boldsymbol{\eta}^{(i)} \ge 0$ together imply $\boldsymbol{\gamma}^\top \boldsymbol{\eta}^{(i)} \ge 0$, $i \in J_N$, we get

$$\begin{array}{ll} \min_{\boldsymbol{\gamma},\mathbf{z}} & \sum_{i\in J_N} z_i \\ \text{s.t.} & \boldsymbol{\gamma}^\top \boldsymbol{\eta}^{(i)} \le z_i,\ i \in J_N \\ & \boldsymbol{\gamma}^\top \mathbf{1} = 1,\ \boldsymbol{\gamma} \ge 0, \end{array}$$

which can be further reduced to

$$\begin{array}{ll} \min_{\boldsymbol{\gamma}} & \gamma^\top \left(\sum_{i\in J_N} \boldsymbol{\eta}^{(i)} \right) \\ \text{s.t.} & \boldsymbol{\gamma}^\top \mathbf{1} = 1,\ \boldsymbol{\gamma} \ge 0. \end{array}$$

The solution of this problem will be given by $\gamma_j = 1$ for j such that

$$\sum_{i\in J_N} \boldsymbol{\eta}_j^{(i)} \tag{35}$$

is minimized. The resulting cut will coincide with a convexity cut from a "band" set $X_j = \{\mathbf{x} : 0 \le x_j \le 1\}$. Once the index j is known, the cut coefficients $\boldsymbol{\alpha}$ are obtained by using the relations

$$\alpha_i s_i = z_i = \boldsymbol{\gamma}^\top \boldsymbol{\eta}^{(i)} = \eta_j^{(i)}.$$

Thus, $\alpha_i = s_i \eta_j^{(i)}$. This result also suggests that a natural criterion for comparing convexity cuts from the band sets X_j is their depth with respect to the ∞-norm, rather than with respect to the typically used 2-norm.

In the case when $\nu(\cdot)$ is the 1-norm (thus, $\nu^\circ(\cdot)$ is the ∞-norm), the problem reduces to

$$\min_{\boldsymbol{\gamma},z} \quad z \tag{36}$$
$$\text{s.t.} \quad z - \boldsymbol{\gamma}^\top \boldsymbol{\eta}^{(i)} \ge 0,\ i \in J_N \tag{37}$$
$$\boldsymbol{\gamma}^\top \mathbf{1} = 1,\ \boldsymbol{\gamma} \ge 0 \tag{38}$$

While it is impossible to write an analytic solution, we can easily construct a feasible basis. The number of constraints in the problem is $n-m+1$, and the number of structural variables is $p+1$ with $p \leq m$ (recall that we assumed that $\gamma_j = 0$ for all variables currently on their bounds). We will make basic one of the γ_j's (to be selected arbitrarily), z, and the slacks of all the constraints in (37), except the one corresponding to the smallest $\eta_j^{(i)}$. Thus, we have selected $n-m+1$ variables, all constraints are satisfied, and we have a feasible basis.

Notice that the problem has smaller dimension than the original LP relaxation, and will have a sparsity structure corresponding to the simplex tableau for the current vertex $\bar{\mathbf{x}}$. By comparison, the LP typically used to generate disjunctive inequalities has roughly twice the number of constraints and variables of the original problem.

To conclude this subsection, we note the following:

Remark 4.7. All cut generation procedures developed in this section also apply to the general MIP case. To see that, it is sufficient to use $x_i - \lfloor \bar{x}_i \rfloor$ instead of x_i in the function (18) and to select $\tilde{x}_i = \bar{x}_i - \lfloor \bar{x}_i \rfloor$ for all fractional $\bar{x}_i$'s.

4.2 Deepest cut generation via admissible sets

The deepest convexity cut generation problem (25)-(29) can also be formulated in terms of admissible sets of the form (14) by replacing constraints (26) and (28) with inequalities of the form

$$\mathbf{d}^\top(\mathbf{y})(\bar{\mathbf{x}} - \mathbf{y} + t_i \boldsymbol{\xi}^{(i)}) \geq 0, \tag{39}$$

where $i = 1, \ldots, k$, $\mathbf{y} \in Y' \times \{0\}^{n-p}$ and $\mathbf{d}(\mathbf{y}) \in (\mathbb{R}^p \times \{0\}^{n-p}) \setminus \{\mathbf{0}\}$ (see Corollary 2.13). In order to use inner product notation in (39), note that we have embedded the elements of the set Y' and the mapping $\mathbf{d}(\cdot)$ into $\mathbb{R}^n$ by extending them with $n-p$ components having value zero. To make the set of constraints (39) manageable, we must assume that Y' is finite (for example, l_i and u_i are finite for $i = 1, \ldots, p$). We must also ensure that $\bar{\mathbf{x}}$ is in the interior of the set corresponding to the mapping $\mathbf{d}(\cdot)$. By the finiteness of Y', this restriction is equivalent to requiring that

$$\mathbf{d}^\top(\mathbf{y})(\bar{\mathbf{x}} - \mathbf{y}) > 0, \ \mathbf{y} \in Y' \times \{0\}^{n-p}.$$

Notice that $\mathbf{d}(\mathbf{y})$ can be arbitrarily normalized, so we can impose a constraint

$$\mathbf{d}^\top(\mathbf{y})(\bar{\mathbf{x}} - \mathbf{y}) = 1, \ \mathbf{y} \in Y' \times \{0\}^{n-p}.$$

This constraint will also ensure that $\mathbf{d}(\mathbf{y}) \neq \mathbf{0}$.

We also switch to the τ_i-parameterization for the same reasons as in (25)-(29). For $\tau_i > 0$, the inequality (39) can be rewritten as

$$\mathbf{d}^\top(\mathbf{y})(\bar{\mathbf{x}} - \mathbf{y}) + \frac{\mathbf{d}^\top(\mathbf{y})\boldsymbol{\xi}^{(i)}}{\tau_i} \geq 0,$$

which, upon noticing that $\mathbf{d}^\top(\mathbf{y})(\bar{\mathbf{x}} - \mathbf{y}) = 1$, can be written as

$$\tau_i + \mathbf{d}^\top(\mathbf{y})\boldsymbol{\xi}^{(i)} \geq 0.$$

The case of $\tau_i = 0$ corresponds to a situation when the ith edge does not intersect the boundary of admissible set, and is still represented by the same inequality above. The parameters of the cut $(\boldsymbol{\alpha}, \beta)$ are related to the vector $\boldsymbol{\tau}$ through (12):

$$\boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} = \tau_i(\beta - \boldsymbol{\alpha}^\top \bar{\mathbf{x}}),\ i = 1, \ldots, k.$$

Together with a normalization assumption $\beta - \boldsymbol{\alpha}^\top \bar{\mathbf{x}} = 1$, this leads to a condition $\boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} = \tau_i$ for $i = 1, \ldots, k$. Through this sequence of transformations, we (similarly to the previous section) arrive at the formulation

$$\begin{aligned}
\min_{\boldsymbol{\alpha}, \mathbf{d}(\cdot)} \quad & \nu^\circ(\boldsymbol{\alpha}) && && (40)\\
\text{s.t.} \quad & \boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} + \mathbf{d}^\top(\mathbf{y})\boldsymbol{\xi}^{(i)} \ge 0, && \mathbf{y} \in Y' \times \{0\}^{n-p},\ i = 1, \ldots, k && (41)\\
& \mathbf{d}^\top(\mathbf{y})(\bar{\mathbf{x}} - \mathbf{y}) = 1, && \mathbf{y} \in Y' \times \{0\}^{n-p}. && (42)
\end{aligned}$$

When (40)-(42) is applied to the cone K arising from the basic feasible solution $\bar{\mathbf{x}}$ of LP relaxation, we can take advantage of the simplifications of the previous sections, such as $\boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} = s_i \alpha_i,\ i \in J_N$. If we also take $\nu(\cdot)$ to be the 1-norm (and therefore $\nu^\circ(\cdot)$ to be the ∞-norm) and $Y' = \{0, 1\}^p$, (40)-(42) becomes

$$\begin{aligned}
\min_{z, \boldsymbol{\alpha}, \mathbf{d}(\cdot)} \quad & z && \\
\text{s.t.} \quad & z \ge |\alpha_i s_i|, && i \in J_N\\
& \alpha_i s_i + \mathbf{d}^\top(\mathbf{y})\boldsymbol{\xi}^{(i)} \ge 0, && \mathbf{y} \in Y' \times \{0\}^{n-p},\ i \in J_N\\
& \mathbf{d}^\top(\mathbf{y})(\bar{\mathbf{x}} - \mathbf{y}) = 1, && \mathbf{y} \in Y' \times \{0\}^{n-p}.
\end{aligned}$$

Note that when $\mathbf{d}^\top(\mathbf{y})\boldsymbol{\xi}^{(i)} \ge 0$, we can set $\alpha_i = 0$, and the corresponding constraint on z will not be binding. Thus, we may replace the formulation above with

$$\begin{aligned}
\min_{z, \mathbf{d}(\cdot)} \quad & z && && (43)\\
\text{s.t.} \quad & z + \mathbf{d}^\top(\mathbf{y})\boldsymbol{\xi}^{(i)} \ge 0, && \mathbf{y} \in Y' \times \{0\}^{n-p},\ i \in J_N && (44)\\
& \mathbf{d}^\top(\mathbf{y})(\bar{\mathbf{x}} - \mathbf{y}) = 1, && \mathbf{y} \in Y' \times \{0\}^{n-p}, && (45)
\end{aligned}$$

where we subsequently let

$$\begin{aligned}
\alpha_i &= s_i \max\left\{0, \max_{\mathbf{y} \in Y' \times \{0\}^{n-p}} \left\{-\mathbf{d}^\top(\mathbf{y})\boldsymbol{\xi}^{(i)}\right\}\right\} && (46)\\
\beta &= \boldsymbol{\alpha}^\top \bar{\mathbf{x}} + 1, && (47)
\end{aligned}$$

using the normalization $\beta - \boldsymbol{\alpha}^\top \bar{\mathbf{x}} = 1$ to recover β.

However, the difficulty remains that $p2^p$ parameters are needed to describe the mapping $\mathbf{d}(\cdot)$ (and even more parameters when general integer variables are present). Thus, we have to impose some additional restrictions on $\mathbf{d}(\cdot)$ to reduce this enormous number of parameters and make the above formulation practical. One possibility is to fix vectors $\mathbf{d}(\mathbf{y})$ for most

$\mathbf{y} \in Y' \times \{0\}^{n-p}$ at some simple values that would satisfy (42) and vary only some small subcollection. Another, simpler approach is to restrict $d_i = 0$ for most of dimensions i. For example, we may let vary any *pair* of dimensions of $\mathbf{d}(\mathbf{y})$ that correspond to fractional components of $\bar{\mathbf{x}}$ (by restricting all but the jth component to zero, we get an admissible set of the form X_j from the previous section).

We now illustrate this latter possibility by picking indices $q, r \in \{1, \ldots, p\}$ so that $\bar{x}_q$ and $\bar{x}_r$ are fractional, and requiring that

$$d_j(\mathbf{y}) = 0,\ j \neq q, r,\ \mathbf{y} \in Y' \times \{0\}^{n-p}. \tag{48}$$

Lemma 4.8. *Suppose that the problem (43)-(45) is feasible. Then the problem (43)-(45) with the additional constraint (48) has an optimal solution $(z, \mathbf{d}(\cdot))$ such that*

$$\begin{array}{rcl} d_q(\mathbf{y}) & = & d_q(\mathbf{y}'), \\ d_r(\mathbf{y}) & = & d_r(\mathbf{y}'), \end{array} \tag{49}$$

whenever $y_q = y'_q$ and $y_r = y'_r$.

Proof. Consider an arbitrary optimal solution $(z, \mathbf{d}(\cdot))$ to (43)-(45) with the additional constraint (48). For each pair of values $v, w \in \{0, 1\}$, choose some

$$\mathbf{y}'(v, w) \in \operatorname{Argmin}\{\min\{\mathbf{d}^\top(\mathbf{y})\boldsymbol{\xi}^{(i)} : i \in J_N\} : \mathbf{y} \in Y' \times \{0\}^{n-p},\ y_q = v,\ y_r = w\}.$$

Consider a mapping $\mathbf{d}'(\cdot)$ defined by $\mathbf{d}'(\mathbf{y}) = \mathbf{d}(\mathbf{y}'(v, w))$ for all $\mathbf{y} \in Y' \times \{0\}^{n-p}$ such that $y_q = v$ and $y_r = w$. Note that $(z, \mathbf{d}'(\cdot))$ is a feasible solution to (43)-(45) also satisfying (48) and (49). Thus, the objective (43) can not be improved by considering $\mathbf{d}(\cdot)$ that does not satisfy (49). $\square$

Condition (49) implies that we can consider mappings $\mathbf{d}(\mathbf{y})$ which depend only on $y_q = v$ and $y_r = w$. We denote such mappings as $\mathbf{d}(v, w)$. The formulation becomes

$$\begin{array}{rll} \min_{z,\mathbf{d}(\cdot)} & z & \quad (50) \\ \text{s.t.} & z + d_q(v, w)\xi_q^{(i)} + d_r(v, w)\xi_r^{(i)} \geq 0, & v, w \in \{0, 1\},\ i \in J_N \quad (51) \\ & d_q(v, w)(\bar{x}_q - v) + d_r(v, w)(\bar{x}_r - w) = 1, & v, w \in \{0, 1\}. \quad (52) \end{array}$$

This formulation only has 9 variables, but can be simplified further if we exclude $d_r(v, w)$ by using the normalization constraint (52) on $\mathbf{d}(v, w)$, as follows:

$$d_r(v, w) = \frac{1 - d_q(v, w)(\bar{x}_q - v)}{\bar{x}_r - w}.$$

Thus, we get

$$\min_{z, d_q(\cdot)} z \tag{53}$$

$$\text{s.t. } z + d_q(v, w)\frac{(\bar{x}_r - w)\xi_q^{(i)} - (\bar{x}_q - v)\xi_r^{(i)}}{\bar{x}_r - w} \geq -\frac{\xi_r^{(i)}}{\bar{x}_r - w}, \quad v, w \in \{0, 1\},\ i \in J_N. \tag{54}$$

This final formulation (53)-(54) has 5 variables and $4(n-m)$ constraints with 2 nonzeros each. This simple structure suggests that we will be able to solve the problem quite efficiently. In fact, we can decompose (53)-(54) into 4 independent subproblems, one for each possible value of (v, w), and then let z be the largest objective value of these four problems. Notice that, for a fixed pair, the problem is that one of finding a minimum of a piecewise-linear function, for which there exists a specialized algorithm running in randomized expected linear time (see, for example, Section 4.4 of [12]). We test the quality of the cuts generated in this manner in Section 5.

Remark 4.9. The formulation (50)-(52) was developed for the $\{0, 1\}$ case. To apply it in the general-integer case, we must allow v and w take arbitrary integer values allowed by the bounds

$$\begin{aligned} v &\in \mathbb{Z},\ l_q \leq v \leq u_q, \\ w &\in \mathbb{Z},\ l_r \leq w \leq u_r. \end{aligned}$$

These additional values of v and w may greatly increase the number of variables and constraints in the cut generation problem: for every possible value of v and w, we need two variables to specify $\mathbf{d}(v, w)$, a normalization constraint for $\mathbf{d}(v, w)$ as in (52), and $|J_N|$ constraints as in (51) to ensure the validity of the cut. However, we can keep the formulation simple by slightly reducing the generality of the admissible sets considered. Specifically, we add constraints

$$d_q(\lfloor \bar{x}_q \rfloor, w) \geq 0, \quad w \in \{\lfloor \bar{x}_r \rfloor, \lceil \bar{x}_r \rceil\} \tag{55}$$

$$d_q(\lceil \bar{x}_q \rceil, w) \leq 0, \quad w \in \{\lfloor \bar{x}_r \rfloor, \lceil \bar{x}_r \rceil\} \tag{56}$$

whenever x_q is a general integer variable, and similar constraints

$$d_r(v, \lfloor \bar{x}_r \rfloor) \geq 0, \quad v \in \{\lfloor \bar{x}_q \rfloor, \lceil \bar{x}_q \rceil\} \tag{57}$$

$$d_r(v, \lceil \bar{x}_r \rceil) \leq 0, \quad v \in \{\lfloor \bar{x}_q \rfloor, \lceil \bar{x}_q \rceil\} \tag{58}$$

on $d_r(v, w)$ if x_r is general integer. One can then guarantee that constraints of the type (51) will never have to be active for values of $v \notin \{\lfloor \bar{x}_q \rfloor, \lceil \bar{x}_q \rceil\}$ and $w \notin \{\lfloor \bar{x}_r \rfloor, \lceil \bar{x}_r \rceil\}$, by choosing $\mathbf{d}(v, w)$ to point in the same direction as $\mathbf{d}(\lfloor \bar{x}_q \rfloor, \lfloor \bar{x}_r \rfloor)$ for $v \leq \lfloor \bar{x}_q \rfloor$, $w \leq \lfloor \bar{x}_r \rfloor$, to point in the same direction as $\mathbf{d}(\lfloor \bar{x}_q \rfloor, \lceil \bar{x}_r \rceil)$ for $v \leq \lfloor \bar{x}_q \rfloor$, $w \geq \lceil \bar{x}_r \rceil$, and so forth. These modifications are minor, and keep the computational complexity of cut generation similar to (50)-(52).

Remark 4.10. To conclude this section, we briefly describe how one can strengthen cut coefficients by using integrality information for the non-basic variables. This technique is based on the Theorem 1 of [7], which, when used in conjunction with convexity cuts from the band sets X_i, $i \in I$, defined by (20), leads to Gomory mixed-integer cuts.

Suppose that $d_q(v, w), d_r(v, w)$ satisfy the restrictions (55)-(58), regardless of whether x_q, x_r are binary or general integer. In this case, one can see that there must exist a $\mathbf{y} \in Y$

such that $\mathbf{d}^\top(\mathbf{y})\boldsymbol{\xi}^{(i)} \leq 0$, and, consequently that the inner maximum in (46) is nonnegative. Thus, after solving (50)-(52), the formula (46) for the cut coefficients simplifies to

$$\alpha_i = s_i \max_{\substack{v=\lfloor \bar{x}_q \rfloor, \lceil \bar{x}_q \rceil \\ w=\lfloor \bar{x}_r \rfloor, \lceil \bar{x}_r \rceil}} \left\{ -d_q(v,w)\xi_q^{(i)} - d_r(v,w)\xi_r^{(i)} \right\}, \quad i \in J_N. \tag{59}$$

We now apply Theorem 1 of [7], first fixing $d_q(\cdot,\cdot)$ and $d_q(\cdot,\cdot)$, and then defining the theorem's mapping π via (59). The theorem then allows us to reduce the cut coefficient α_i via

$$\alpha_i = s_i \inf_{m_q, m_r \in \mathbb{Z}} \left\{ \max_{\substack{v=\lfloor \bar{x}_q \rfloor, \lceil \bar{x}_q \rceil \\ w=\lfloor \bar{x}_r \rfloor, \lceil \bar{x}_r \rceil}} \left\{ -d_q(v,w)(\xi_q^{(i)} - m_q) - d_r(v,w)(\xi_r^{(i)} - m_r) \right\} \right\},$$

for every $i \in J_N \cap I$. Since computing true infimum in the above expression may be difficult, we try to strengthen coefficients by testing only the possibilities $m_q \in \{\lfloor \xi_q^{(i)} \rfloor, \lceil \xi_q^{(i)} \rceil\}$ and $m_r \in \{\lfloor \xi_r^{(i)} \rfloor, \lceil \xi_r^{(i)} \rceil\}$. Note that when both $\xi_q^{(i)}$ and $\xi_r^{(i)}$ are integer, this rule is guaranteed to reduce α_i zero. Optimal coefficient strengthening, possibly incorporated into depth optimization, is an interesting problem reserved for possible future research.

5 Numerical experiments

In practice, one typically adds cuts to a MIP problem not one-by-one, but in groups or "rounds"; see for example [5, 6]. The following practical issues have to be addressed in this situation:

- The cuts have to be sufficiently deep on average.
- The procedure has to be computationally more efficient than generating optimized cuts one-by-one, with intervening reoptimization of the strengthened LP relaxation.
- The cuts must have normal vectors that are not too close to one another.

The first issue may be addressed by using depth-optimized cuts generated by one of the approaches of Section 4. As for the second issue, generation of a single cut should also be quite efficient if we use either

1. ∞-norm deepest cuts from separable admissible functions (see Section 4.1), or
2. 1-norm deepest cuts from admissible sets corresponding to pairs of integer variables which are fractional at the LP relaxation solution (see Section 4.2).

One problem with the first approach is that, given the current simplex tableau, we can only generate a single cut. To get more than one cut at a time, for example, one could apply the cut generation procedure at the current vertex and all neighboring vertices. In this case, LP degeneracy would actually be beneficial, since some of the neighboring vertices would in fact be identical to the current one, but with a different simplex tableau. These tableaux would generate different cuts that would be sure to eliminate the current vertex. Workable implementation of such a procedure would require an efficient pivot-level interface to an "industrial-strength" simplex library, and we do not attempt it here.

On the other hand, the variable-pair approach of Section 4.2 may readily lead to large sets of cuts by simply selecting many different pairs of variables. To make sure the cuts have sufficiently distinct normal vectors, we employ a heuristic rule that only accepts cuts whose normals form a sufficiently large angle with normals of those already generated; see for example [4].

We now describe a simple computational experiment to test the effectiveness of variable-pair convexity cuts on the integrality gaps of realistic mixed-integer problems. For each problem, we allowed a maximum of 5 rounds of cuts. After the cuts are applied, we reoptimize the LP and discard all new cuts with slacks greater than 10^{-5}. In general, one might also be interested in saving deleted cuts and giving them a "a second chance" if they subsequently become violated. We plan to examine this possibility in future experiments. Presently, each cut generation iteration proceeds as follows:

1. We identify all variables that violate the integrality constraints by more than 0.05.

2. We form all pairs of these variables, and put the pairs on a queue in decreasing order of combined fractionality measure:

$$\min\left\{(\bar{x}_q - v)^2 + (\bar{x}_r - w)^2 : v \in \{\lfloor \bar{x}_r \rfloor, \lceil \bar{x}_r \rceil\}, w \in \{\lfloor \bar{x}_q \rfloor, \lceil \bar{x}_q \rceil\}\right\}.$$

   Thus, pairs like $(0.5, 0.5)$ would appear at the head of the queue, and pairs like $(0.05, 0.05)$ or $(0.95, 0.95)$ would come last.

3. We extract pairs from the head of the queue, and generate corresponding cuts by solving (50)-(52), modified according to Remark 4.9 when there are general integer variables. Note that when the number of fractional variables is large, some of them may be disproportionately represented in the pairs at the head of the queue. Thus, we skip a pair when one of its variables has already been used more than $2.5M/n_f$ times, where M is the maximum number of cuts per round and n_f denotes the number of "sufficiently" fractional integer variables identified in step 1. Note that if we attempted to generate cuts from *all* sufficiently fractional variable pairs, we would have $M = n_f(n_f-1)/2$, and the number of pairs involving a particular variable x_j would be $n_f - 1 = 2M/n_f$. Thus, the rule that no variable can appear more than $2.5M/n_f$ times keeps the "diversity" of the chosen pairs reasonably close to that of all possible pairs.

4. Coefficients of the cut are strengthened according to the heuristic rule of Remark 4.10 whenever possible.

5. We place the resulting cuts in a cut queue ordered by decreasing 1-norm depth (the inverse of the optimal value of (50)-(52)). Ties are broken using ∞-norm depth. Generation proceeds until the variable-pair queue is empty or 2000 cuts have been generated.

6. Finally, we select at most $M = 500$ cuts from the head of the cut queue. However, if the sine of the angle between normal vectors of the new cut and any cut which is already selected is less than 0.01 in absolute value, we discard the cut instead.

7. We adjoin the resulting cuts to the LP, reoptimize it, and discard all inactive cuts.

Below, we will refer to the cuts obtained by this procedure as *strengthened variable-pair deepest convexity* (SVPDC) cuts.

We implemented this procedure in C++, using the COIN open-source software library [10, 15]. The cut generator's interface is consistent with COIN/CGL's `CglCutGenerator`. For optimizing and reoptimizing the linear programs, we used the COIN/CLP simplex solver originated by John Forrest (see [10]). The program was compiled with `g++ 3.2` and executed on a 2.4 GHz Pentium 4 system with 1 Gb of RAM, running Mandrake Linux 9.1.

To test the procedure, we used the MIPLIB 3.0 suite of test problems [9], with the exception of three problems we ruled out as too large: `dano3mip`, `nw04`, and `fast0507`; we also excluded `mitre` due to numerical instability. On most of the remaining problems, the procedure executed for the full 5 iterations. On four problems, the procedure terminated earlier because there were fewer than two variables deviating from integrality by at least 0.05. These problems were `air03`, `bell3a`, `mod010` and `rentacar`.

Figure 1 gives a graphical summary of the improvement in the integrality gap. In the figure, 100% corresponds to the entire integrality gap being eliminated, while 0% corresponds to no improvement in the gap. The black bars correspond to the total improvement at termination, and the white bars to the improvement after the first round of cuts. Problems `dsbmip`, `enigma` and `noswot` do not appear, since they have zero integrality gap. Complete gap improvement information appears in Table 1. This table also shows running time in seconds, excluding solution of initial LP relaxation, for all 61 problems.

As one can see from the figure and the table, variable-pair cuts significantly reduced the gap on a number of 0-1 as well as general-integer problems (Table 2 summarizes the MIPLIB problems' characteristics; for more detailed information, see [9]). The best case is `air03`, for which the gap is completely eliminated after one iteration. On two other problems, `10teams` and `mod010`, the gap was eliminated after 5 rounds. On the other hand, six 0-1 problems appeared completely unaffected by the cuts.

To comparatively evaluate the performance of SVPDC cuts, we performed an additional experiment. We generated a Gomory mixed-integer cut from every variable deviating from integrality by at least 0.05 (using COIN's `CglGomory` cut generator) and compared the improvements in the integrality gap after one round. Our choice of Gomory cuts was motivated by two factors: first, they are one of the best-known and most effective classes of cuts; and second, if one generates convexity cuts from the simple band sets X_i defined by (20), and then applies the strengthening technique of [7] alluded to in Remark 4.10, one obtains mixed-integer Gomory cuts.

We performed two comparisons: SVPDC cuts versus Gomory cuts, and a combination of SVPDC cuts and Gomory cuts versus Gomory cuts alone. Figure 2 shows both comparisons, with a "positive" bar indicating an advantage to SVPDC cuts. Clearly, there are a large number of problems for which SVPDC cuts alone underperform Gomory mixed-integer cuts. One possible explanation is that our coefficient-strengthening method for SVPDC cuts is not optimal, whereas it is provably optimal for Gomory cuts. However, there are six problems for which both a combination of cut generators and the SVPDC cuts alone produce significantly better results than Gomory cuts. On a number of additional problems, there is incremental benefit to using both generators.

These computational tests involve only the "root" LPs of the problems. Possible future research could study how the cuts perform when embedded in a general branch-and-cut framework.

References

[1] Egon Balas. Intersection cuts — a new type of cutting planes for integer programming. *Operations Research*, 19:19–39, 1971.

[2] Egon Balas. Integer programming and convex analysis: Intersection cuts from outer polars. *Mathematical Programming*, 2:330–382, 1972.

[3] Egon Balas, V. Joseph Bowman, Fred Glover, and David Sommer. An intersection cut from the dual of the unit hypercube. *Operations Research*, 19:40–44, 1971.

[4] Egon Balas, Sebastián Ceria, and Gérard Cornuéjols. A lift-and-project cutting plane algorithm for mixed 0-1 programs. *Mathematical Programming*, 58(3):295–324, 1993.

[5] Egon Balas, Sebastián Ceria, and Gérard Cornuéjols. Mixed 0-1 programming by lift-and-project in a branch-and-cut framework. *Management Science*, 42(9):1229–1246, 1996.

[6] Egon Balas, Sebastián Ceria, Gérard Cornuéjols, and N. Natraj. Gomory cuts revisited. *Operations Research Letters*, 19(1):1–9, 1996.

[7] Egon Balas and Robert G. Jeroslow. Strengthening cuts for mixed integer programs. *European Journal of Operational Research*, 4(4):224–234, 1980.

[8] Egon Balas and Michael Perregaard. A precise correspondence between lift-and-project cuts, simple disjunctive cuts, and mixed integer Gomory cuts for 0-1 programming. *Mathematical Programming*, 94(2-3, Ser. B):221–245, 2003. The Aussois 2000 Workshop in Combinatorial Optimization.

[9] Robert E. Bixby, Sebastián Ceria, Cassandra M. McZeal, and Martin W. P. Savelsberg. An updated mixed integer programming library: MIPLIB 3.0. Technical Report 98-3, Department of Computational and Applied Mathematics, Rice University, 1998.

[10] COmputational INfrastructure for Operations Research. http://www-124.ibm.com/developerworks/opensource/coin/.

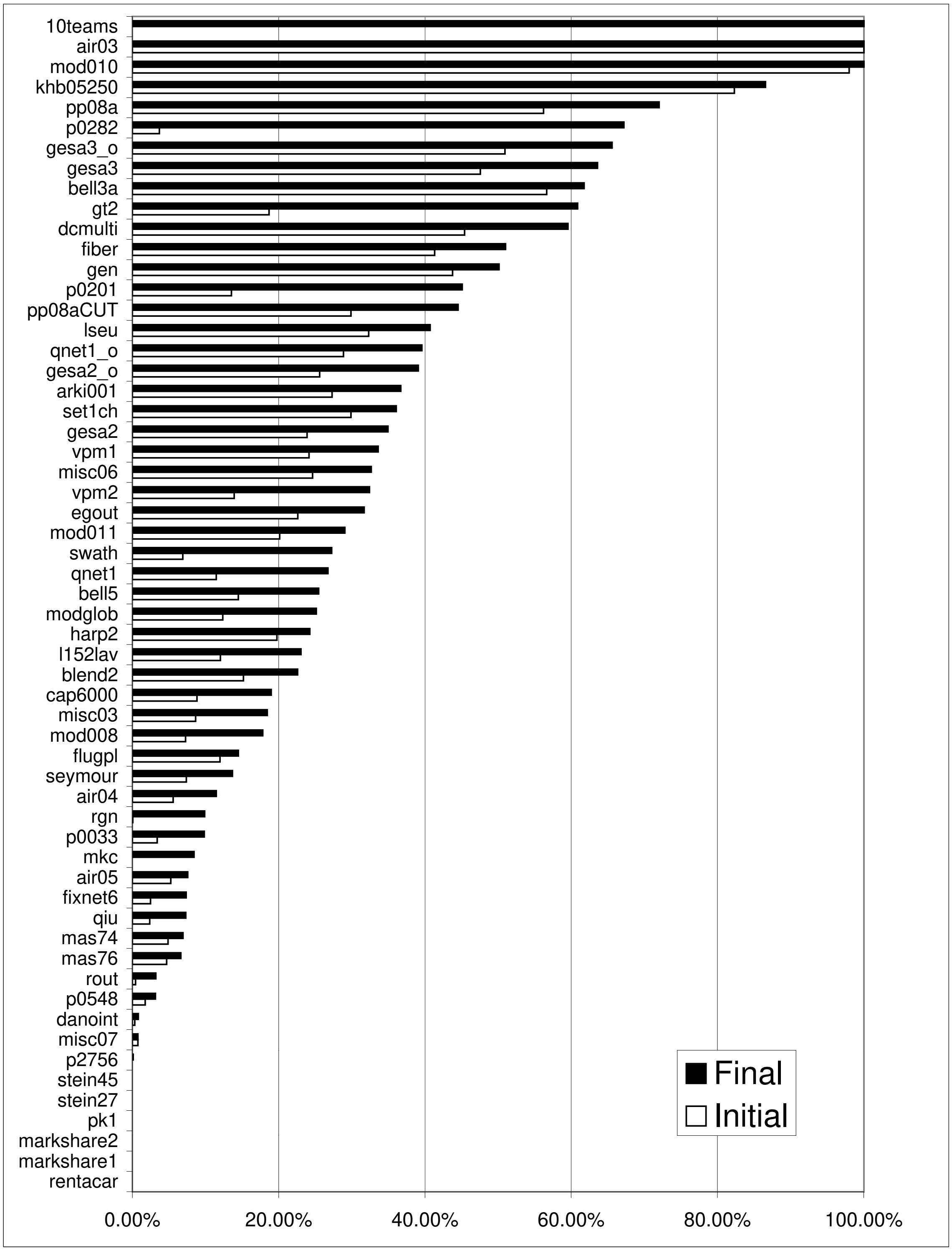

Figure 1: Percentage of the gap eliminated by SVPDC cuts for 58 MIPLIB problems

Name	Initial	Final	Time(s)	Name	Initial	Final	Time(s)
10teams	0.00%	100.00%	58.38	misc03	8.62%	18.40%	0.71
air03	100.00%	100.00%	25.37	misc06	24.60%	32.64%	0.52
air04	5.54%	11.44%	236.03	misc07	0.72%	0.72%	4.20
air05	5.22%	7.53%	186.15	mkc	0.00%	8.39%	59.08
arki001	27.28%	36.68%	21.81	mod008	7.23%	17.80%	0.07
bell3a	56.63%	61.73%	0.02	mod010	97.99%	100.00%	20.67
bell5	14.45%	25.44%	0.20	mod011	20.12%	29.02%	152.76
blend2	15.16%	22.56%	0.14	modglob	12.34%	25.13%	1.48
cap6000	8.81%	18.95%	0.94	noswot	N/A	N/A	0.38
danoint	0.28%	0.79%	9.35	p0033	3.36%	9.80%	0.02
dcmulti	45.40%	59.53%	10.02	p0201	13.51%	45.08%	3.71
dsbmip	N/A	N/A	4.39	p0282	3.67%	67.17%	0.73
egout	22.60%	31.66%	0.26	p0548	1.72%	3.13%	1.43
enigma	N/A	N/A	0.07	p2756	0.01%	0.08%	0.28
fiber	41.30%	51.00%	10.10	pk1	0.00%	0.00%	0.14
fixnet6	2.44%	7.33%	0.31	pp08aCUTS	29.85%	44.51%	6.24
flugpl	11.96%	14.45%	0.02	pp08a	56.20%	72.01%	5.66
gen	43.75%	50.10%	3.41	qiu	2.34%	7.26%	37.13
gesa2	23.86%	34.92%	8.74	qnet1	11.44%	26.71%	21.58
gesa2_o	25.57%	39.06%	12.98	qnet1_o	28.85%	39.55%	18.10
gesa3	47.56%	63.56%	9.50	rentacar	0.00%	0.00%	0.26
gesa3_o	50.91%	65.57%	13.25	rgn	0.02%	9.85%	0.28
gt2	18.67%	60.85%	0.07	rout	0.41%	3.17%	11.68
harp2	19.73%	24.24%	39.12	set1ch	29.85%	36.06%	16.41
khb05250	82.28%	86.53%	3.39	seymour	7.35%	13.65%	43.26
l152lav	12.00%	23.02%	55.29	stein27	0.00%	0.00%	0.19
lseu	32.27%	40.69%	0.11	stein45	0.00%	0.00%	1.04
markshare1	0.00%	0.00%	0.03	swath	6.87%	27.21%	101.26
markshare2	0.00%	0.00%	0.03	vpm1	24.12%	33.58%	1.61
mas74	4.86%	6.89%	0.13	vpm2	13.88%	32.38%	2.57
mas76	4.66%	6.59%	0.11				

Table 1: Gap improvements and run times for 62 MIPLIB problems

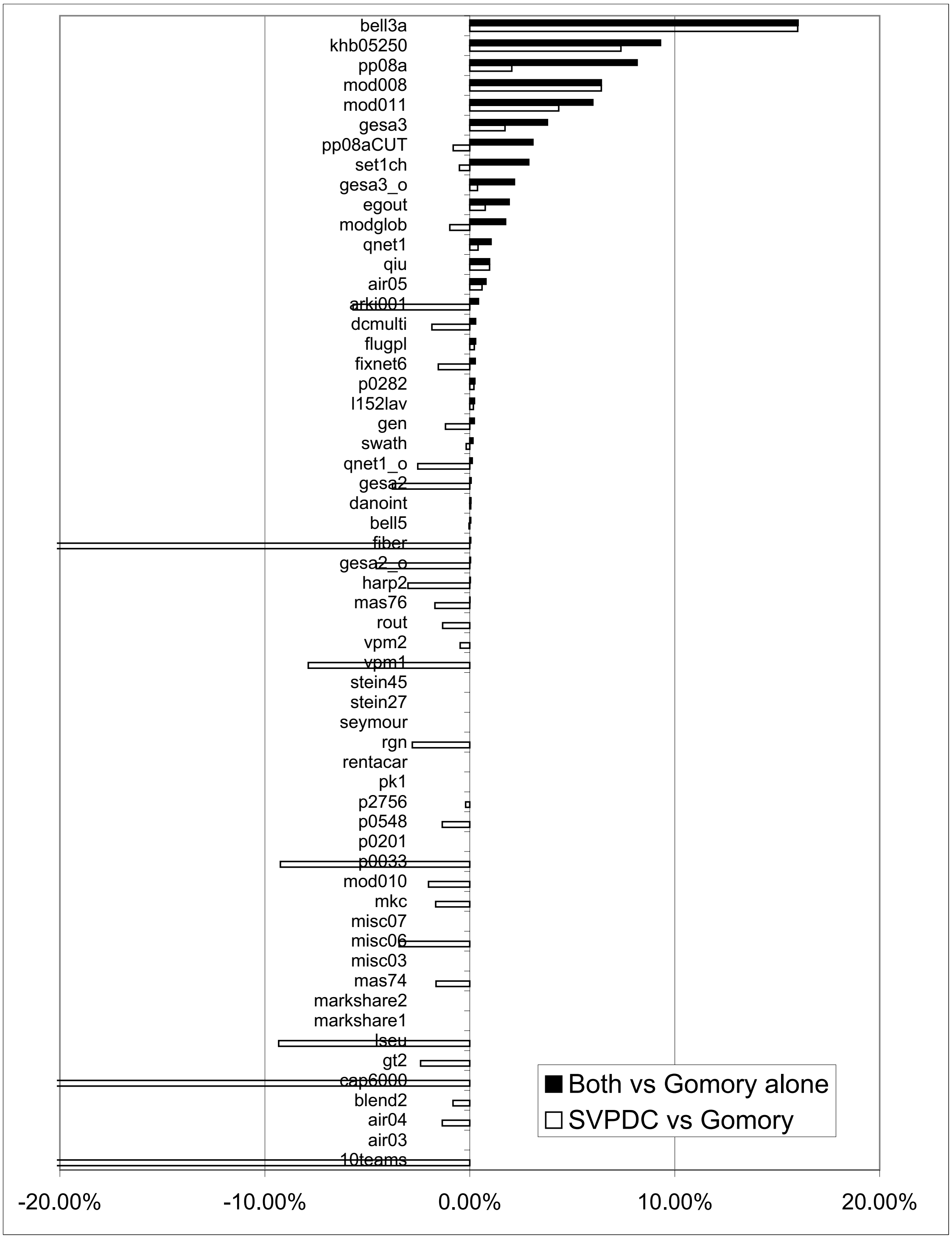

Figure 2: Comparison of the gap eliminated by a single round of various cut generator combinations for 58 MIPLIB problems

Name	Rows	Columns	Integer	0-1	Continuous
10teams	230	2025	1800	ALL	225
air03	124	10757	10757	ALL	0
air04	823	8904	8904	ALL	0
air05	426	7195	7195	ALL	0
arki001	1048	1388	538	415	850
bell3a	123	133	71	39	62
bell5	91	104	58	30	46
blend2	274	353	264	231	89
cap6000	2176	6000	6000	ALL	0
dano3mip	3202	13873	552	ALL	13321
danoint	664	521	56	ALL	465
dcmulti	290	548	75	ALL	473
dsbmip	1182	1886	192	160	1694
egout	98	141	55	ALL	86
enigma	21	100	100	ALL	0
fast0507	507	63009	63009	ALL	0
fiber	363	1298	1254	ALL	44
fixnet6	478	878	378	ALL	500
flugpl	18	18	11	0	7
gen	780	870	150	144	720
gesa2	1392	1224	408	240	816
gesa2_o	1248	1224	720	384	504
gesa3	1368	1152	384	216	768
gesa3_o	1224	1152	672	336	480
gt2	29	188	188	24	0
harp2	112	2993	2993	ALL	0
khb05250	101	1350	24	ALL	1326
l152lav	97	1989	1989	ALL	0
lseu	28	89	89	ALL	0
markshare1	6	62	50	ALL	12
markshare2	7	74	60	ALL	14
mas74	13	151	150	ALL	1
mas76	12	151	150	ALL	1
misc03	96	160	159	ALL	1
misc06	820	1808	112	ALL	1696
misc07	212	260	259	ALL	1
mitre	2054	10724	10724	ALL	0
mkc	3411	5325	5323	ALL	2
mod008	6	319	319	ALL	0
mod010	146	2655	2655	ALL	0
mod011	4480	10958	96	ALL	10862
modglob	291	422	98	ALL	324
noswot	182	128	100	75	28
nw04	36	87482	87482	ALL	0
p0033	16	33	33	ALL	0
p0201	133	201	201	ALL	0
p0282	241	282	282	ALL	0
p0548	176	548	548	ALL	0
p2756	755	2756	2756	ALL	0
pk1	45	86	55	ALL	31
pp08a	136	240	64	ALL	176
pp08aCUTS	246	240	64	ALL	176
qiu	1192	840	48	ALL	792
qnet1	503	1541	1417	1288	124
qnet1_o	456	1541	1417	1288	124
rentacar	6803	9557	55	ALL	9502
rgn	24	180	100	ALL	80
rout	291	556	315	300	241
set1ch	492	712	240	ALL	472
seymour	4944	1372	1372	ALL	0
stein27	118	27	27	ALL	0
stein45	331	45	45	ALL	0
swath	884	6805	6724	ALL	81
vpm1	234	378	168	ALL	210
vpm2	234	378	168	ALL	210

Table 2: Characteristics of the MIPLIB problems

[11] Gérard Cornuéjols and Yanjun Li. Elementary closures for integer programs. *Operations Research Letters*, 28(1):1–8, 2001.

[12] Mark de Berg, Marc van Kreveld, Mark Overmars, and Otfried Schwarzkopf. *Computational geometry: algorithms and applications*. Springer-Verlag, Berlin, second, revised edition, 2000.

[13] Fred Glover. Convexity cuts and cut search. *Operations Research*, 21:123–134, 1973.

[14] Robert G. Jeroslow. Cutting-plane theory: disjunctive methods. In *Studies in integer programming (Proc. Workshop, Bonn, 1975)*, pages 293–330. Annals of Discrete Mathematics, Vol. 1. North-Holland, Amsterdam, 1977.

[15] Robin Lougee-Heimer. The common optimization interface for operations research: Promoting open-source software in the operations research community. *IBM Journal of Research and Development*, 47(1):57–66, 2003. http://www.research.ibm.com/journal/rd47-1.html.

[16] Mikhail Nediak and Jonathan Eckstein. Pivot, Cut, and Dive: A heuristic for 0-1 mixed integer programming. RUTCOR Research Report 53-2001. Rutgers University, Piscataway, NJ.

[17] Jonathan H. Owen and Sanjay Mehrotra. A disjunctive cutting plane procedure for general mixed-integer linear programs. *Mathematical Programming*, 89(3):437–448, 2001.

[18] M. Raghavachari. On connections between zero-one integer programming and concave programming under linear constraints. *Operations Research*, 17:680–684, 1969.

[19] R. Tyrrell Rockafellar. *Convex analysis*. Princeton University Press, Princeton, N.J., 1970.

[20] Hanif D. Sherali and Warren P. Adams. A hierarchy of relaxations and convex hull characterizations for mixed-integer zero-one programming problems. *Discrete Applied Mathematics*, 52(1):83–106, 1994.

[21] Hoang Tuy. Concave programming under linear constraints. *Soviet Mathematics*, pages 1437–1440, 1964.

[22] Philip B. Zwart. Nonlinear programming: counterexamples to two global optimization algorithms. *Operations Research*, 21(6):1260–1266, 1973.

A Proof of Proposition 2.5

To prove the statement, we will first need the following lemma:

Lemma A.1. *Any vector $\boldsymbol{\alpha}$ satisfying (9)-(11) specifies a supporting hyperplane for K at $\bar{\mathbf{x}}$.*

Proof. We need to prove that $\boldsymbol{\alpha}^\top(\mathbf{x}-\bar{\mathbf{x}}) \geq 0$ for all $\mathbf{x} \in K$. Note that combining (9) and (10) with the definition of the edges (5) we get:

$$\boldsymbol{\alpha}^\top \bar{\mathbf{x}} < \beta = \boldsymbol{\alpha}^\top \mathbf{x}^{(i)}(t_i) = \boldsymbol{\alpha}^\top \bar{\mathbf{x}} + t_i \boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)}$$

for all $i \in J$. As $t_i > 0$ for $i \in J$, we have

$$\boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} > 0,\ i \in J.$$

Together with (11), this implies

$$\boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} \geq 0,\ i = 1, \ldots, k.$$

Since any point $\mathbf{x}$ of K is a nonnegative combination of $\boldsymbol{\xi}^{(i)}$'s plus $\bar{\mathbf{x}}$ we have $\boldsymbol{\alpha}^\top \mathbf{x} \geq \boldsymbol{\alpha}^\top \bar{\mathbf{x}}$. □

The proof of the proposition may now proceed:

Proof. Consider any $\mathbf{x}' \in K$ such that $\boldsymbol{\alpha}^\top \mathbf{x}' = \beta' < \beta$ (*i.e.*, $\mathbf{x}'$ is cut off). Since Lemma A.1 implies $\{\mathbf{x} : \boldsymbol{\alpha}^\top \mathbf{x} = \boldsymbol{\alpha}^\top \bar{\mathbf{x}}\}$ is a supporting hyperplane for K at $\bar{\mathbf{x}}$, we have $\boldsymbol{\alpha}^\top(\mathbf{x}' - \bar{\mathbf{x}}) \geq 0$, which implies that $\boldsymbol{\alpha}^\top \bar{\mathbf{x}} \leq \beta'$. By the continuity of the functions $\boldsymbol{\alpha}^\top \mathbf{x}^{(i)}(t)$, $i \in J$, they take the intermediate value $\beta' \in [\boldsymbol{\alpha}^\top \bar{\mathbf{x}}, \beta)$ at some points $t_i' \in [0, t_i)$. For each $i \in J$, either $t_i' = 0$, in which case $\mathbf{x}^{(i)}(t_i') = \bar{\mathbf{x}} \in \operatorname{int} X$, or $\mathbf{x}^{(i)}(t_i') \in \operatorname{int} X$, since it is a strict convex combination of $\bar{\mathbf{x}} \in \operatorname{int} X$ and $\mathbf{x}^{(i)}(t_i) \in X$, $i \in J$ (refer to Theorem 6.1 of [19]). It is straightforward to see that all $\mathbf{x} \in \{\mathbf{x} : \boldsymbol{\alpha}^\top \mathbf{x} = \beta'\} \cap K$ are of the form

$$\mathbf{x} = \sum_{i \in J} \lambda_i \mathbf{x}^{(i)}(t_i') + \sum_{i \in J^c} \mu_i \boldsymbol{\xi}^{(i)},\ \sum_{i \in J} \lambda_i = 1,\ \lambda_i \geq 0,\ \mu_i \geq 0.$$

Next, notice that

$$\sum_{i \in J} \lambda_i \mathbf{x}^{(i)}(t_i') \in \operatorname{int} X,$$

as it is a convex combination of points in $\operatorname{int} X$. Since the $\boldsymbol{\xi}^{(i)}$, $i \in J^c$, are directions of recession of $\operatorname{int} X$, any such $\mathbf{x} \in \operatorname{int} X$. But $\operatorname{int} X \cap Y = \emptyset$ so $\mathbf{x}' \notin Y$. Thus, the inequality $\boldsymbol{\alpha}^\top \mathbf{x} \geq \beta$ is valid for $K \cap Y$. □

B Proof of Proposition 2.7

To establish the proposition, we need the following lemma:

Lemma B.1. *Let $(\boldsymbol{\alpha}, \beta)$ specify a convexity cut corresponding to X, then $H(\boldsymbol{\alpha}, \beta) \cap K$ has the following form, where J^c denotes $\{1, \ldots, k\} \backslash J$:*

$$\left\{ \bar{\mathbf{x}} + \sum_{i \in J} t_i \lambda_i \boldsymbol{\xi}^{(i)} + \sum_{i \in J^c} \mu_i \boldsymbol{\xi}^{(i)} : \sum_{i \in J} \lambda_i \geq 1,\ \lambda_i \geq 0,\ i \in J,\ \mu_j \geq 0,\ j \in J^c \right\}. \tag{60}$$

Proof. First, we verify the $\supseteq$ inclusion. Consider a point $\mathbf{x}$ in the set (60). It can be seen that $\mathbf{x}$ is in $H(\boldsymbol{\alpha}, \beta)$ by multiplying the expression in the definition of a set (60) by $\boldsymbol{\alpha}^\top$ and using that $t_i \boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} = \beta - \boldsymbol{\alpha}^\top \bar{\mathbf{x}} > 0$, $i \in J$ and $\boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} = 0$, $i \in J^c$:

$$\begin{aligned}\boldsymbol{\alpha}^\top \mathbf{x} &= \boldsymbol{\alpha}^\top \bar{\mathbf{x}} + \sum_{i \in J} t_i \lambda_i \boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} + \sum_{i \in J^c} \mu_i \boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} \\ &= \boldsymbol{\alpha}^\top \bar{\mathbf{x}} + \sum_{i \in J} \lambda_i (\beta - \boldsymbol{\alpha}^\top \bar{\mathbf{x}}) \geq \boldsymbol{\alpha}^\top \bar{\mathbf{x}} + 1 \cdot (\beta - \boldsymbol{\alpha}^\top \bar{\mathbf{x}}) = \beta.\end{aligned}$$

On the other hand, $\mathbf{x} \in K$ by the definition of K.

The $\subseteq$ inclusion is verified as follows: note that any $\mathbf{x} \in K$ has to be of the form (60), but not necessarily restricted by $\sum_{i \in J} \lambda_i \geq 1$. We have to show that any $\mathbf{x}$ of this form that also has $\boldsymbol{\alpha}^\top \mathbf{x} \geq \beta$ must also satisfy this additional restriction. Indeed, multiplying the expression for any $\mathbf{x}$ in (60) by $\boldsymbol{\alpha}^\top$, we get

$$\boldsymbol{\alpha}^\top \mathbf{x} = \boldsymbol{\alpha}^\top \bar{\mathbf{x}} + \sum_{i \in J} \lambda_i (\beta - \boldsymbol{\alpha}^\top \bar{\mathbf{x}}) = \left(1 - \sum_{i \in J} \lambda_i\right) \boldsymbol{\alpha}^\top \bar{\mathbf{x}} + \sum_{i \in J} \lambda_i \beta \geq \beta.$$

The latter inequality is equivalent to

$$\left(1 - \sum_{i \in J} \lambda_i\right) \boldsymbol{\alpha}^\top \bar{\mathbf{x}} \geq \left(1 - \sum_{i \in J} \lambda_i\right) \beta.$$

If $1 > \sum_{i \in J} \lambda_i$, this last relation implies $\boldsymbol{\alpha}^\top \bar{\mathbf{x}} \geq \beta$, a contradiction. $\square$

The proof of the proposition is then as follows:

Proof. Let $\bar{\boldsymbol{\alpha}}^\top \mathbf{x} \geq \bar{\beta}$ define the cut corresponding to $\bar{X}$. To prove the "if" part, suppose that $\bar{J} \subseteq J$ and $\bar{t}_i \geq t_i$, $i \in \bar{J}$. Consider any $\mathbf{x} \in H(\bar{\boldsymbol{\alpha}}, \bar{\beta}) \cap K$. By Lemma B.1, $\mathbf{x}$ is of the form

$$\begin{aligned}\mathbf{x} &= \bar{\mathbf{x}} + \sum_{i \in \bar{J}} \bar{t}_i \bar{\lambda}_i \boldsymbol{\xi}^{(i)} + \sum_{i \in \bar{J}^c} \bar{\mu}_i \boldsymbol{\xi}^{(i)} \\ &= \bar{\mathbf{x}} + \sum_{i \in \bar{J}} \frac{\bar{t}_i}{t_i} t_i \bar{\lambda}_i \boldsymbol{\xi}^{(i)} + \sum_{i \in J \setminus \bar{J}} \frac{\bar{\mu}_i}{t_i} t_i \boldsymbol{\xi}^{(i)} + \sum_{i \in J^c} \bar{\mu}_i \boldsymbol{\xi}^{(i)},\end{aligned}$$

with $\sum_{i \in \bar{J}} \bar{\lambda}_i \geq 1$, $\bar{\lambda}_i \geq 0$, $i \in \bar{J}$ and $\bar{\mu}_j \geq 0$, $j \in \bar{J}^c$. Let $\lambda_i = (\bar{t}_i / t_i) \bar{\lambda}_i$, $i \in \bar{J}$, $\lambda_i = \bar{\mu}_i / t_i$, $i \in J \setminus \bar{J}$, and $\mu_i = \bar{\mu}_i$, $i \in J^c$. Then, we have

$$\mathbf{x} = \bar{\mathbf{x}} + \sum_{i \in J} t_i \lambda_i \boldsymbol{\xi}^{(i)} + \sum_{i \in J^c} \mu_i \boldsymbol{\xi}^{(i)}$$

and

$$\sum_{i \in J} \lambda_i = \sum_{i \in \bar{J}} \frac{\bar{t}_i}{t_i} \bar{\lambda}_i + \sum_{i \in J \setminus \bar{J}} \frac{\bar{\mu}_i}{t_i} \geq \sum_{i \in \bar{J}} \bar{\lambda}_i \geq 1,$$

since $\bar{t}_i/t_i \geq 1$ and $\bar{\mu}_i/t_i \geq 0$. Thus, $\mathbf{x} \in H(\boldsymbol{\alpha}, \beta) \cap K$, by Lemma B.1, and $\bar{\boldsymbol{\alpha}}^\top \mathbf{x} \geq \bar{\beta}$ is indeed strong relative to $\boldsymbol{\alpha}^\top \mathbf{x} \geq \beta$.

To prove the "only if" part, suppose that the inequality $\bar{\boldsymbol{\alpha}}^\top \mathbf{x} \geq \bar{\beta}$ is strong relative to $\boldsymbol{\alpha}^\top \mathbf{x} \geq \beta$. If $\bar{J} \not\subseteq J$, then there exists an index $i \in \bar{J} \cap J^c$. The corresponding intersection parameter $\bar{t}_i \geq 0$ is defined by $\bar{\boldsymbol{\alpha}}^\top(\bar{\mathbf{x}} + \bar{t}_i \boldsymbol{\xi}^{(i)}) = \bar{\beta}$, which means that the point $\bar{\mathbf{x}} + \bar{t}_i \boldsymbol{\xi}^{(i)}$ is not cut off by the cut corresponding to $\bar{X}$. On the other hand, a condition $i \in J^c$ implies that $\boldsymbol{\alpha}^\top \boldsymbol{\xi}^{(i)} = 0$ and

$$\boldsymbol{\alpha}^\top(\bar{\mathbf{x}} + \bar{t}_i \boldsymbol{\xi}^{(i)}) = \boldsymbol{\alpha}^\top \bar{\mathbf{x}} < \beta,$$

i.e., $\bar{\mathbf{x}} + \bar{t}_i \boldsymbol{\xi}^{(i)}$ is cut off by the cut corresponding to X, a contradiction of the hypothesis. Thus, $\bar{J} \subseteq J$. Similarly, if there is $i \in \bar{J}$ such that $\bar{t}_i < t_i$, we get a contradiction since the point $\bar{\mathbf{x}} + \bar{t}_i \boldsymbol{\xi}^{(i)}$ will violate $\boldsymbol{\alpha}^\top \mathbf{x} \geq \beta$ but not $\bar{\boldsymbol{\alpha}}^\top \mathbf{x} \geq \bar{\beta}$. □

C Proofs of Proposition 3.3 and Corollary 3.4

Consider an admissible set X of the form (13) and the corresponding mapping $\mathbf{d}(\cdot)$. We construct $\bar{\mathbf{d}}(\cdot)$ by a simple normalization of $\mathbf{d}$, namely

$$\bar{\mathbf{d}}(\mathbf{y}) = (1/\mathbf{d}^\top(\mathbf{y})(\bar{\mathbf{x}} - \mathbf{y}))\mathbf{d}(\mathbf{y})$$

for all $\mathbf{y} \in Y$. It follows that

$$\bar{\mathbf{d}}^\top(\mathbf{y})(\bar{\mathbf{x}} - \mathbf{y}) = 1 \quad \forall \mathbf{y} \in Y. \tag{61}$$

With this choice of $\bar{\mathbf{d}}(\cdot)$, the function $f(\mathbf{x}|\bar{\mathbf{d}})$ is concave, since it is the pointwise infimum of a collection of concave functions. Also, it takes a positive value at $\bar{\mathbf{x}}$ and is non-positive at any point of Y. To prove admissibility of $f(\mathbf{x}|\bar{\mathbf{d}})$, we only need to establish its continuity.

Note that by admissibility of X, we have $\bar{\mathbf{x}} \in \operatorname{int} X$. Then there exists $\delta > 0$ such that all $\mathbf{x} \in \mathbb{R}^n$ with $\|\mathbf{x} - \bar{\mathbf{x}}\| \leq \delta$ are contained in X, that is,

$$\begin{aligned} \|\mathbf{x} - \bar{\mathbf{x}}\| \leq \delta,\ \mathbf{y} \in Y \quad &\Rightarrow \quad \bar{\mathbf{d}}^\top(\mathbf{y})(\mathbf{x} - \mathbf{y}) \geq 0 \\ &\Rightarrow \quad \mathbf{d}^\top(\mathbf{y})(\mathbf{x} - \mathbf{y}) \geq 0. \end{aligned}$$

Together with (61), it follows that

$$\begin{aligned} \bar{\mathbf{d}}^\top(\mathbf{y})(\mathbf{x} - \mathbf{y}) &= \bar{\mathbf{d}}^\top(\mathbf{y})(\mathbf{x} - \bar{\mathbf{x}}) + \bar{\mathbf{d}}^\top(\mathbf{y})(\bar{\mathbf{x}} - \mathbf{y}) \\ &= \bar{\mathbf{d}}^\top(\mathbf{y})(\mathbf{x} - \bar{\mathbf{x}}) + 1 \geq 0. \end{aligned}$$

Consider now

$$\mathbf{x} = \bar{\mathbf{x}} - \delta \frac{\bar{\mathbf{d}}(\mathbf{y})}{\|\bar{\mathbf{d}}(\mathbf{y})\|}.$$

Then

$$\bar{\mathbf{d}}^\top(\mathbf{y})(\mathbf{x} - \bar{\mathbf{x}}) + 1 = 1 - \delta\|\bar{\mathbf{d}}(\mathbf{y})\| \geq 0,$$

and it follows that $\|\bar{\mathbf{d}}(\mathbf{y})\| \leq 1/\delta$ for all $\mathbf{y} \in Y$.

Consider an arbitrary point $\mathbf{x}$ and any $\mathbf{y} \in Y$. The Schwartz inequality implies that

$$|\bar{\mathbf{d}}^\top(\mathbf{y})(\mathbf{x}-\bar{\mathbf{x}})| \leq \|\bar{\mathbf{d}}(\mathbf{y})\| \cdot \|\mathbf{x}-\bar{\mathbf{x}}\| \leq \frac{\|\mathbf{x}-\bar{\mathbf{x}}\|}{\delta}.$$

Then one obtains

$$\begin{aligned} |\bar{\mathbf{d}}^\top(\mathbf{y})(\mathbf{x}-\mathbf{y})| &= |\bar{\mathbf{d}}^\top(\mathbf{y})(\mathbf{x}-\bar{\mathbf{x}}) + \bar{\mathbf{d}}^\top(\mathbf{y})(\bar{\mathbf{x}}-\mathbf{y})| \\ &\leq |\bar{\mathbf{d}}^\top(\mathbf{y})(\mathbf{x}-\bar{\mathbf{x}})| + 1 \\ &\leq \frac{\|\mathbf{x}-\bar{\mathbf{x}}\|}{\delta} + 1. \end{aligned}$$

It follows that $f(\mathbf{x}|\bar{\mathbf{d}})$ is finite on all of $\mathbb{R}^n$. Concavity and finiteness imply continuity (see Corollary 10.1.1 of [19]).

Consider now an arbitrary function of the form (15). The corresponding level set is defined by

$$\inf_{\mathbf{y}\in Y}\{\bar{\mathbf{d}}^\top(\mathbf{y})(\mathbf{x}-\mathbf{y})\} \geq 0$$

or, equivalently, by

$$\bar{\mathbf{d}}^\top(\mathbf{y})(\mathbf{x}-\mathbf{y}) \geq 0,\ \forall \mathbf{y} \in Y,$$

which, in turn, defines a set of the form (13) by simply taking $\mathbf{d} = \bar{\mathbf{d}}$. The proof of Proposition 3.3 is now complete.

As for Corollary 3.4, the same proof can be used without significant modification if we start from a set of the form (14) and a mapping $\bar{\mathbf{d}}' : Y \to \mathbb{R}^p \setminus \{\mathbf{0}\}$. In that case, we take the infimum over Y' instead of Y, since the value of linear forms will be the same, given $\mathbf{y}' \in Y'$, for all $\mathbf{y} \in \{\mathbf{y}'\} \times \mathbb{R}^{n-p}$. Thus, we arrive at a function of the form (16).