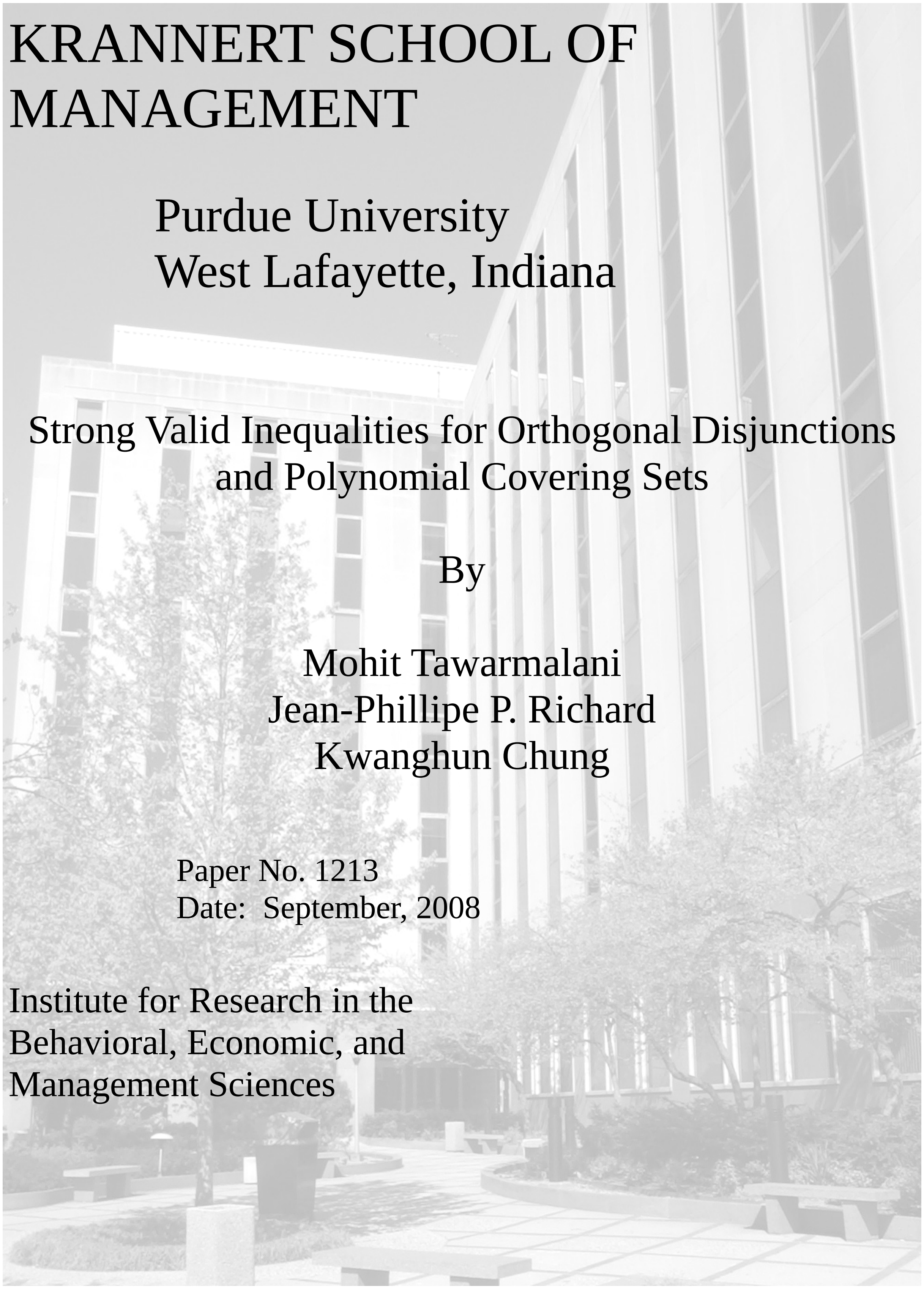

# KRANNERT SCHOOL OF MANAGEMENT

Purdue University
West Lafayette, Indiana

## Strong Valid Inequalities for Orthogonal Disjunctions and Polynomial Covering Sets

By

Mohit Tawarmalani
Jean-Phillipe P. Richard
Kwanghun Chung

Paper No. 1213
Date: September, 2008

Institute for Research in the Behavioral, Economic, and Management Sciences

# Strong Valid Inequalities for Orthogonal Disjunctions and Polynomial Covering Sets*

Mohit Tawarmalani [1], Jean-Philippe P. Richard [2], Kwanghun Chung [2]

September 3, 2008

**Abstract**

In this paper, we develop a convexification tool that enables construction of convex hulls for orthogonal disjunctive sets using convex extensions and disjunctive programming techniques. A distinguishing feature of our technique is that, unlike most applications of disjunctive programming, it does not require the introduction of new variables in the relaxation. We develop and apply a toolbox of results that help in checking the technical assumptions under which the convexification tool can be employed. We demonstrate its applicability in integer programming by deriving the intersection cut for mixed-integer polyhedral sets and the convex hull of certain mixed/pure-integer bilinear sets. We then develop a key result that extends the applicability of the convexification tool to relaxing nonconvex inequalities, which are not naturally disjunctive, by providing sufficient conditions for establishing the convex extension property over the non-negative orthant. Then, we illustrate the convexification tool by developing convex hulls for certain polynomial covering sets with non-negative variables. We specialize the results to bilinear covering sets and use them to derive a tight relaxation of the bilinear covering sets over a hypercube. We use the orthogonally disjunctive characterization to show that the derived relaxation is at least as tight as the standard factorable relaxation for the same inequality, and derive necessary and sufficient conditions under which it is strictly tighter. Finally, we present a preliminary computational study on a set of randomly generated bilinear covering sets that indicates that the derived relaxation is substantially tighter than the factorable relaxation.

## 1 Introduction and Motivation

Finding globally optimal solutions to nonconvex problems is a challenging problem that has received much attention in the last few decades; see Neumaier [19] for a survey of the existing solution methods. Nonlinear branch-and-bound is one such method that has been implemented successfully in various global optimization software; see Adjiman et al. [1], Sahinidis and Tawarmalani [24], LINDO Systems Inc. [17], and Belotti et al. [8]. The branch-and-bound method typically bounds the nonconvex optimization problem by solving its convex relaxations over successively refined partitions (see Falk and Soland [13] and Horst and Tuy [15]). For factorable problems–problems involving functions that can be written as recursive sums and products of univariate functions–McCormick [18] proposed a composition theorem that allows automatic construction of convex relaxations provided that tight concave and/or convex envelopes are known for the intrinsic nonlinear terms. McCormick's relaxation is an instance of a commonly used technique for deriving convex relaxations for nonconvex problems that relaxes inequalities of the form $f(x) \geq r$ by $\bar{f}(x) \geq r$, where $\bar{f}(x)$ is a concave overestimator of the function $f(x)$. There is a significant amount of literature that develops techniques for deriving tight overestimators for various classes of functions; see Tawarmalani and Sahinidis [28] and Bliek et al. [9] for a more detailed treatment. However, the current literature rarely considers the right-hand-side of the inequality. More precisely, the above technique relaxes the hypograph of

*Krannert Working Paper

[1] Krannert School of Management, Purdue University. Email: mtawarma@purdue.edu. Corresponding Author.

[2] Department of Industrial and Systems Engineering, University of Florida.

$f(x)$ instead of relaxing the appropriate upper-level set. As a result, the derived relaxations can be weak. For an illustration of the difference, consider the set $S$ defined as:

$$S = \left\{(x, y, z) \in \mathbb{R}^3_+ \;\middle|\; xy + z \geq r\right\},$$

where $r > 0$. It can be easily seen that $S$ is not convex since both $\left(\sqrt{r}, \sqrt{r}, 0\right)$ and $(0, 0, r)$ belong to $S$ while their convex combination with a weight of $\frac{1}{2}$ on each point does not. Therefore, if the constraint defining $S$ was to appear as one of the constraints in a problem, local optimization techniques would not be guaranteed to find a globally optimal solution for the problem. However, because this set belongs to the general family of factorable programs, it can be relaxed using McCormick's scheme. More generally, if traditional techniques were used to derive a convex relaxation of $S$, a concave overestimator $\bar{f}$ of the function $f(x, y, z) = xy+z$ would first be obtained. Observe that the concave envelope of this function over the non-negative orthant is infinite as long as $x$ and $y$ are both positive. The resulting convex relaxation of $S$ is $\left\{(x, y, z) \in \mathbb{R}^3_+ \;\middle|\; x, y > 0\right\} \cup \left\{(x, y, z) \in \mathbb{R}^3_+ \mid z \geq r,\, xy = 0\right\}$. If in addition, the concave overestimator is required to be upper-semicontinuous, as is typically the case, or even if the relaxation is required to be a closed set, then the relaxation would be $\mathbb{R}^3_+$. In other words, standard relaxation schemes will essentially drop the defining constraint.

In this paper, we propose a scheme that produces tighter convex approximations by considering the right-hand-side of the constraint. In particular, for the set $S$ presented above, our scheme produces the following convex relaxation

$$\mathrm{RS} = \left\{(x, y, z) \in \mathbb{R}^3_+ \;\middle|\; \sqrt{\frac{xy}{r}} + \frac{z}{r} \geq 1\right\},$$

which is a much tighter approximation than $\mathbb{R}^3_+$. Considering this simple example, we can make three interesting observations. First, the relaxation, RS, is nonlinear. This is in contrast to current implementations of nonlinear branch-and-bound that typically construct linear relaxations for multivariate terms (see Tawarmalani and Sahinidis [31]). Second, the form of the nonlinear cut is surprising as it applies different functions to the different terms of the initial inequality. For $S$, the first term is modified using a square-root after being divided by $r$, while the second is simply divided by $r$. Third, RS is not only a convex relaxation of $S$, but it is in fact (as will be shown later) the convex hull of $S$. These observations generalize to many polynomial covering sets. Surprisingly, the convex hull for these sets can be expressed in a simple form without introducing new variables while developing the concave envelope of the corresponding polynomial can be much harder.

The convex hull representation for polynomial covering sets, even though interesting in itself, arises from a much more general theory of orthogonal disjunctions that we develop in this paper. To provide an example, consider the set $S$ again. We will show that the convex hull of $S$ is determined by the points of $S$ that either belong to the half-plane $(x, y, 0)$, where $(x, y) \in \mathbb{R}^2_+$ or to the half-line $(0, 0, z)$, where $z \in \mathbb{R}_+$. In other words, the set $S$ satisfies the convex extension property (see Tawarmalani and Sahinidis [29]) in which the important subsets belong to orthogonal subspaces. Because such a convex extension property holds, it is natural to expect that one could build a higher dimensional description of the convex hull of $S$ using disjunctive programming arguments; see Rockafellar [22] and Balas [5]. Disjunctive programming has been used to develop tight relaxations and cutting planes in integer, nonlinear, and robust optimization; see [4, 26, 27, 10, 30, 7, 3, 25]. Unlike this paper, the literature on disjunctive programming formulations, however, is mostly focused on naturally disjunctive sets. Cutting planes based on disjunctive formulations, are typically linear and derived by solving separation problems over extended formulations; see Cornuéjols and Lemaréchal [11]. One interesting observation in this paper is that, as long as the disjunctive terms are orthogonal and a few technical conditions are satisfied, there is no need to introduce additional variables. Furthermore, the convex hull of $S$ can be easily expressed in closed-form using the representations of the convex hull of $S$ in each of the two orthogonal subspaces, namely $\sqrt{\frac{xy}{r}} \geq 1$ and $\frac{z}{r} \geq 1$. We establish a much more general set of conditions under which the argument evoked above is correct, allowing the use of both right-hand-side and left-hand-side information in the derivation of convex relaxations for nonlinear programming. Our results rely on the ability to prove that a convex extension property holds over orthogonal disjunctions and the ability to derive closed form expressions of convex hulls (possibly in a higher dimensional space) over each of the subspaces. We also describe large families of problems for which our construction applies and yields stronger convex approximations than those currently used in the nonlinear branch-and-bound solvers.

Branch-and-bound codes typically require bounds on variables for constructing relaxations. Even though this requirement cannot be relaxed in general, the set $S$ is an instance where tight relaxations can be constructed in the absence of upper bounds on the variables. As branch-and-bound proceeds, the partitions are typically refined, and the algorithm imposes upper and lower bounds on the variables consistent with the relevant partition. Therefore, subsequent to branching, one may wish to relax a subset of $S$ such as follows:

$$S' = \left\{(x, y, z) \in [l_x, u_x] \times [l_y, u_y] \times [l_z, u_z] \;\middle|\; xy + z \geq r\right\}.$$

We show that by convexifying the orthogonal disjunctions we can construct relaxations for bilinear covering sets that are at least as tight as McCormick's relaxation. The relaxations are strictly tighter when the bounds on the variables are loose or the variable bounds form a hypercube whose lower corner is almost feasible. For relaxations used in the branch-and-bound algorithm, the second feature is very desirable since the algorithm often constructs small boxes around near-optimal solutions before deriving a bound that is sufficiently strong to fathom the search.

In Section 2, we describe a tool to obtain the convex hull of orthogonal disjunctive sets. The result can be evoked under certain technical conditions. We provide tools to verify these assumptions. We also provide counterexamples to show the need for some of the assumptions. The intersection/split cut for mixed integer linear sets is shown to be a special case of our general convexification tool. We illustrate the application of the tool in nonlinear integer programming by convexifying a bilinear pure/mixed-integer set. Nonconvex inequalities in continuous variables are not naturally disjunctive. For such inequalities, we establish sufficient conditions under which the convex extension property holds over the non-negative orthant. For problems with bounded variables, the convex extension property might not hold even when it holds with unbounded variables. However, our procedure can typically be adapted to yield strong cuts even in the presence of bounds. The idea is to tighten the nonlinear inequality outside of the bounds in such a way that the convex extension property continues to hold and then to apply our main convexification result. In Section 3, we obtain convex hull formulations of certain polynomial covering sets. These convex hull representations have many applications in factorable nonlinear programming. In addition, they serve as an example of the general theoretical framework of Section 2. In particular, we prove concavity of the ratio of certain monomials over the non-negative orthant, verify that the convex extension property holds for the polynomial covering sets, and apply our convexification tool over orthogonal disjunctions. We then refine our results in the context of bilinear covering sets to build new relaxations. In this context, we incorporate bound information, using the ideas of Section 2, by considering a linear extension of the defining bilinear function outside the hypercube formed by the variable bounds. We show that our relaxation is at least as tight as the factorable relaxation and provide precise conditions under which it is strictly tighter. In Section 4, we present the results of a preliminary computational study on randomly generated instances of the bilinear covering set. Our computations demonstrate that the quality of the relaxation derived in Section 3 for the bilinear covering set is significantly better than the standard factorable relaxation. We summarize the contributions of this work in Section 5 and conclude with remarks and directions for future research.

## 2 Convexification of Orthogonal Disjunctive Sets

In this section, we first introduce and prove a general result that exposes the closed-form convex hull inequality description of the disjunctive union of a finite number of sets defined over subspaces that are orthogonal to each other. This result also applies to non-disjunctive sets provided that their convex hulls are entirely defined by their restrictions over a finite number of orthogonal subspaces. Next, we illustrate the utility of this result in finding convex hull descriptions. Simultaneously, we discuss the need for certain seemingly technical assumptions in the statement of the result. In particular, we discuss each one of the four assumptions of the theorem and describe, with examples, situations where they are satisfied. For some of the assumptions, we establish sufficient conditions that are simple to verify. We also show later that the cuts that yield the convex hull, under the specified technical conditions, continue to produce valid inequalities even when some of the conditions are not satisfied. Throughout, we demonstrate the generality and applicability of our convexification result by deriving new convex hull descriptions of various continuous, mixed, and pure integer bilinear covering sets, and providing an alternate derivation of the classic intersection cut derived in the integer programming literature. Applications of this result to polynomial covering sets will be discussed in Section 3.

In the following, given a set $S$, we represent its convex hull by $\operatorname{conv}(S)$, its closure by $\operatorname{cl}(S)$, and its projection on the space of $z$ variables by $\operatorname{proj}_z S$. For a closed convex set, $S$, $0^+(S)$ denotes the set of its recession directions. When we display equations, we sometimes write $\min \begin{Bmatrix} f(z) \\ g(z) \end{Bmatrix}$ to denote $\min\{f(z), g(z)\}$.

**Theorem 2.1.** *Let $S \subseteq \mathbb{R}^{\sum_i d_i}$ and for all $i \in N = \{1, \ldots, n\}$, let $S_i \subseteq S$. Let the points $z$ of $S$ be written as $z = (z_1, \ldots, z_i, \ldots, z_n) \in S$, where $z_i \in \mathbb{R}^{d_i}$. Assume that:*
*(A1) if $(z_1, \ldots, z_i, \ldots, z_n) \in S_i$, then $z_j = 0$ for $\forall j \neq i$,*
*(A2) for any $z \in S$, there exists $\chi_i \in \operatorname{conv}(S_i)$, $i \in I \subseteq N$, such that $z \in \operatorname{conv}\big(\bigcup_{i\in I} \chi_i\big)$,*
*(A3) $\operatorname{conv}(S_i) \subseteq \operatorname{proj}_z A_i \subseteq \operatorname{cl}\big(\operatorname{conv}(S_i)\big)$, where, for each $i \in \{1, \ldots, n\}$,*

$$A_i = \left\{ \big(0, z_i, u_i, 0\big) \;\middle|\; \begin{array}{lll} t_i^{j_i}(z_i, u_i) \geq 1, & \forall j_i \in J_i, \\ v_i^{k_i}(z_i, u_i) \geq -1, & \forall k_i \in K_i, \\ w_i^{l_i}(z_i, u_i) \geq 0, & \forall l_i \in L_i \end{array} \right\}. \tag{1}$$

*Assume that $t_i^{j_i}$, $v_i^{k_i}$, and $w_i^{l_i}$ are positively-homogenous functions, i.e., for $\lambda > 0$,*

$$\lambda t_i^{j_i}\left(\frac{(z_i, u_i)}{\lambda}\right) = t_i^{j_i}(z_i, u_i),\ \lambda v_i^{k_i}\left(\frac{(z_i, u_i)}{\lambda}\right) = v_i^{k_i}(z_i, u_i),\ \lambda w_i^{l_i}\left(\frac{(z_i, u_i)}{\lambda}\right) = w_i^{l_i}(z_i, u_i).$$

*(A4) $\operatorname{proj}_z C_i$ is a subset of the recession cone of $\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right)$, i.e., for all $i$,*

$$\operatorname*{proj}_z C_i \subseteq 0^+\left(\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right)\right)$$

*where*

$$C_i = \left\{ \big(0, z_i, u_i, 0\big) \;\middle|\; \begin{array}{ll} t_i^{j_i}(z_i, u_i) \geq 0, & \forall j_i \in J_i, \\ v_i^{k_i}(z_i, u_i) \geq 0, & \forall k_i \in K_i, \\ w_i^{l_i}(z_i, u_i) \geq 0, & \forall l_i \in L_i \end{array} \right\}.$$

*Let*

$$X = \left\{ (z, u) \;\middle|\; \begin{array}{ll} \sum_{i\in N} t_i^{j_i}(z_i, u_i) \geq 1, & \forall j_i \in J_i, \\ \sum_{i\in I} v_i^{k_i}(z_i, u_i) \geq -1, & \forall I \subseteq N, \forall k_i \in K_i, \\ t_i^{j_i}(z_i, u_i) + v_i^{k_i}(z_i, u_i) \geq 0, & \forall i, \forall j_i \in J_i, \forall k_i \in K_i, \\ t_i^{j_i}(z_i, u_i) \geq 0, & \forall i, \forall j_i \in J_i, \\ w_i^{l_i}(z_i, u_i) \geq 0, & \forall i, \forall l_i \in L_i \end{array} \right\}. \tag{2}$$

*Then, $\operatorname{conv}(S) \subseteq \operatorname{proj}_z X \subseteq \operatorname{cl}\operatorname{conv}(S)$. If in addition, $\forall i \in N$, $\operatorname{proj}_z A_i$ is closed and $\operatorname{proj}_z C_i = 0^+\big(\operatorname{cl}\operatorname{conv}(S_i)\big)$, then $\operatorname{proj}_z X = \operatorname{cl}\operatorname{conv}(S)$.*

Before proving Theorem 2.1, we briefly comment on its assumptions, its practical importance, and its applicability. In Assumption (A2), we impose that any point in $S$ can be expressed as a convex combination of points in some of the $S_i$s. This implies that only the subsets $S_i$s are needed when computing the convex hull of $S$. In Assumption (A1), we require that these subsets are orthogonal to each other and aligned along the principal axes. In Assumption (A3), we require that an inequality description of the convex hull of each one of the sets $S_i$ be known. Note that this inequality description might make use of an extended formulation (using the additional variables $u_i$). The assumption that the right-hand-sides of all the inequalities are either 1, 0, or $-1$ is without loss of generality as inequalities with nonzero right-hand-sides can be rescaled to satisfy this assumption. Note also that Theorem 2.1 requires that all inequalities be defined using positively-homogeneous functions. We will comment further on this assumption after the proof of the theorem and describe a technique for satisfying this assumption in Section 3. We will also show later that this assumption is typically not needed to prove the validity of the cuts derived in Theorem 2.1. In Assumption (A4), we impose, in essence, that the recession directions of each one of the sets $A_i$ are also the recession directions for the closure convex hull of the union of the $S_i$s. Under these four assumptions, we show that an inequality description of the convex hull of $S$ can be obtained by combining in a systematic way the inequalities arising in the convex hull descriptions of the $S_i$s. Note however that, for reasons that will be described later, this inequality description might describe a superset of the desired convex hull. However, the superset will never be larger than the closure convex hull of $S$, which is sufficient for all practical purposes.

*Proof.* Claim 1: We claim that $\mathrm{conv}(S) = \mathrm{conv}\left(\bigcup_{i=1}^n S_i\right)$. We first show that $\mathrm{conv}(S)$ contains $\mathrm{conv}\left(\bigcup_{i=1}^n S_i\right)$. Clearly, for all $i$, $S_i \subseteq S$. Therefore, $S \supseteq \bigcup_{i=1}^n S_i$ and, so, $\mathrm{conv}(S) \supseteq \mathrm{conv}\left(\bigcup_{i=1}^n S_i\right)$. Now, we show that (A2) implies that $\mathrm{conv}(S) \subseteq \mathrm{conv}\left(\bigcup_{i=1}^n S_i\right)$. Let $z \in S$. There exists $I \subseteq N$ and $\chi_i \in \mathrm{conv}(S_i)$ such that $z \in \mathrm{conv}\big(\bigcup_{i\in I} \chi_i\big) \subseteq \mathrm{conv}\left(\bigcup_{i=1}^n S_i\right)$. Claim 1 is thus proved and, therefore, we can use disjunctive programming techniques to compute the convex hull of $S$. Using these techniques, we now show that it is possible to construct, in a closed-form, a set $X$ that contains $\mathrm{conv}\left(\bigcup_{i=1}^n S_i\right)$ and is itself contained in $\mathrm{cl}\big(\mathrm{conv}\left(\bigcup_{i=1}^n S_i\right)\big)$.

For $T \subseteq N$, we define

$$R_T(\lambda_T) = \left\{ (z_T, u_T) \;\middle|\; \begin{array}{ll} \sum_{i\in T} t_i^{j_i}(z_i,u_i) \geq \lambda_T & \forall j_i \in J_i \\ \sum_{i\in I} v_i^{k_i}(z_i,u_i) \geq -\lambda_T & \forall I \subseteq T, \forall k_i \in K_i \\ t_i^{j_i}(z_i,u_i) + v_i^{k_i}(z_i,u_i) \geq 0 & \forall i, \forall j_i \in J_i, \forall k_i \in K_i \\ t_i^{j_i}(z_i,u_i) \geq 0 & \forall i, \forall j_i \in J_i \\ w_i^{l_i}(z_i,u_i) \geq 0 & \forall i, \forall l_i \in L_i \end{array} \right\}.$$

In the remainder of this proof, whenever $T$ is a singleton, say $\{i\}$, we will denote it as $i$ itself. Also, we define

$$Q = \left\{ (\lambda, z, u) \;\middle|\; \begin{array}{ll} \lambda_i \geq 0 & \forall i \in N \\ (z_i,u_i) \in R_i(\lambda_i) & \forall i \in N \\ \sum_{i=1}^n \lambda_i = \lambda_{1,\ldots,n} = 1 & \end{array} \right\}.$$

We next prove that $X = \mathrm{proj}_{z,u}\, Q$ and $\mathrm{conv}(S) \subseteq \mathrm{proj}_z\, Q \subseteq \mathrm{cl}\,\mathrm{conv}(S)$. Clearly, together these results imply that $\mathrm{conv}(S) \subseteq \mathrm{proj}_z\, X \subseteq \mathrm{cl}\,\mathrm{conv}(S)$. First, we prove that $X = \mathrm{proj}_{z,u}\, Q$. Given two disjoint subsets $A$ and $B$ of $N$, we consider

$$W = \left\{ (\lambda_A, \lambda_B, \lambda_{A\cup B}, z_A, u_A, z_B, u_B) \;\middle|\; \begin{array}{l} \lambda_A \geq 0 \\ (z_A, u_A) \in R_A(\lambda_A) \\ \lambda_B \geq 0 \\ (z_B, u_B) \in R_B(\lambda_B) \\ \lambda_A + \lambda_B = \lambda_{A\cup B} \end{array} \right\},$$

and

$$P = \left\{ (\lambda_{A\cup B}, z_{A\cup B}, u_{A\cup B}) \;\middle|\; \begin{array}{l} \lambda_{A\cup B} \geq 0 \\ (z_{A\cup B}, u_{A\cup B}) \in R_{A\cup B}(\lambda_{A\cup B}) \end{array} \right\}.$$

A straightforward sequential application of the following claim shows that when $\lambda_1, \ldots, \lambda_n$ are projected out from $Q$ we obtain $R_N(1) = X$.

Claim 2: If $z_{A\cup B} = (z_A, z_B)$ and $u_{A\cup B} = (u_A, u_B)$, then $P$ is the set obtained when $\lambda_A$ and $\lambda_B$ are projected out from $W$. Note that since $A$ and $B$ are disjoint and $z_{A\cup B} \in \mathbb{R}^{|\sum_{i\in A} d_i + \sum_{i\in B} d_i|} = \mathbb{R}^{|\sum_{i\in A} d_i|} \times \mathbb{R}^{|\sum_{i\in B} d_i|}$, the definitions of $z_{A\cup B}$ and, similarly, $u_{A\cup B}$ are dimensionally consistent. Claim 2 is verified by first substituting $\lambda_B = \lambda_{A\cup B} - \lambda_A$ and then projecting $\lambda_A$ out using Fourier-Motzkin elimination; see Theorem 1.4 in [33]. We substitute $\lambda_B = \lambda_{A\cup B} - \lambda_A$ in $W$ to obtain:

$$\begin{array}{l} \lambda_A \geq 0 \\ (z_A, u_A) \in R_A(\lambda_A) \\ \lambda_{A\cup B} - \lambda_A \geq 0 \\ (z_B, u_B) \in R_B(\lambda_{A\cup B} - \lambda_A). \end{array}$$

On the one hand, note that the inequalities

$$t_i^{j_i}(z_i, u_i) + v_i^{k_i}(z_i, u_i) \geq 0, \tag{3}$$

$$t_i^{j_i}(z_i, u_i) \geq 0, \tag{4}$$

$$w_i^{l_i}(z_i, u_i) \geq 0 \tag{5}$$

for all $i \in A \cup B$, $j_i \in J_i$, $k_i \in K_i$, and $l_i \in L_i$ remain untouched during projection since they are independent of $\lambda_A$. On the other hand, the inequalities containing $\lambda_A$ can be rewritten as:

$$\min \left\{ \begin{array}{c} \sum_{i\in A} t_i^{j_i}(z_i, u_i) \\ \lambda_{A\cup B} + \min_{B'\subseteq B} \sum_{i\in B'} v_i^{k_i}(z_i, u_i) \end{array} \right\} \geq \lambda_A \geq \max \left\{ \begin{array}{l} \lambda_{A\cup B} - \sum_{i\in B} t_i^{j_i}(z_i, u_i) \\ -\min_{A'\subseteq A} \sum_{i\in A'} v_i^{k_i}(z_i, u_i) \end{array} \right\}$$

so that Fourier-Motzkin elimination is simple to perform. Observe that the constraints $\lambda_{A\cup B} - \lambda_A \geq 0$ and $\lambda_A \geq 0$ are represented in the above system respectively when $A' = \emptyset$ and $B' = \emptyset$. Projecting $\lambda_A$ out of the system, we obtain:

$$\sum_{i\in A\cup B} t_i^{j_i}(z_i, u_i) \geq \lambda_{A\cup B} \tag{6}$$

$$\sum_{i\in A} t_i^{j_i}(z_i, u_i) + \sum_{i\in A'} v_i^{k_i}(z_i, u_i) \geq 0 \quad \forall A' \subseteq A, j_i \in J_i, k_i \in K_i \quad \text{(redundant)} \tag{7}$$

$$\sum_{i\in B} t_i^{j_i}(z_i, u_i) + \sum_{i\in B'} v_i^{k_i}(z_i, u_i) \geq 0 \quad \forall B' \subseteq B, j_i \in J_i, k_i \in K_i \quad \text{(redundant)} \tag{8}$$

$$\sum_{i\in A'\cup B'} v_i^{k_i}(z_i, u_i) \geq -\lambda_{A\cup B} \qquad \forall B' \subseteq B, A' \subseteq A. \tag{9}$$

Inequalities (3) for $i \in A'$ and (4) for $i \in A \backslash A'$ imply (7), showing that (7) is redundant. Similarly, Inequality (8) can be shown to be redundant. Observe that $\lambda_{A \cup B} \geq 0$ can be shown to be represented in (9) by selecting $A' = B' = \emptyset$. Therefore, the set obtained by projecting $\lambda_A$ and $\lambda_B$ out of $W$ is given by (3), (4), (5), (6), and (9), which is exactly the definition of $P$. We have thus proved Claim 2. By applying this result sequentially, we obtain that $X = \text{proj}_{z,u} Q$.

We now prove that $\text{conv}(S) \subseteq \text{proj}_z Q \subseteq \text{cl}\big(\text{conv}(S)\big)$. We first show that if $z \in \text{conv}\left(\bigcup_{i=1}^n S_i\right)$, it can be extended to a point that belongs to $Q$ by suitably defining $(\lambda, u)$. If $z \in \text{conv}\left(\bigcup_{i=1}^n S_i\right)$, then, by (A1), there exist $\lambda_i$ and $z_i'$ such that

$$z = (z_1, \ldots, z_i, \ldots, z_n) = \sum_{i=1}^{n} \lambda_i (0, z_i', 0),$$

where, for each $i$, $\lambda_i \geq 0$, $(0, z_i', 0) \in \text{conv}(S_i)$, and the multipliers sum up to one, *i.e.*, $\sum_{i=1}^n \lambda_i = 1$. We reindex $S_i$ so that the sets containing the points associated with non-zero multipliers are indexed from 1 to $t$. Then, $(z, u) = \sum_{i=1}^t \lambda_i (0, z_i', u_i', 0)$, where $(0, z_i', u_i', 0) \in A_i$, $\lambda_i > 0$ for $i = 1, \ldots, t$, and $\sum_{i=1}^t \lambda_i = 1$. Such a representation exists since $z$ is expressible as a convex combination of points in $\text{conv}(S_i)$ which can be extended to belong to $A_i$, the representation of a superset of $\text{conv}(S_i)$, possibly in a higher dimensional space. Observe that $\lambda_i z_i' = z_i$ and $\lambda_i u_i' = u_i$. Observe further that $R_i(1)$ is the same as $A_i$, except that it is defined in a lower-dimensional space. Since $(z_i', u_i') \in R_i(1)$ for each $i \in \{1, \ldots, t\}$, it is clear that

$$\begin{array}{ll}
t_i^{j_i}(z_i', u_i') \geq 1 & \forall j_i \in J_i \\
v_i^{k_i}(z_i', u_i') \geq -1 & \forall k_i \in K_i \\
t_i^{j_i}(z_i', u_i') + v_i^{k_i}(z_i', u_i') \geq 0 & \forall j_i \in J_i, \forall k_i \in K_i \\
t_i^{j_i}(z_i', u_i') \geq 0 & \forall j_i \in J_i \\
w_i^{l_i}(z_i', u_i') \geq 0 & \forall l_i \in L_i.
\end{array}$$

After substituting $(z_i', u_i') = \left(\frac{z_i}{\lambda_i}, \frac{u_i}{\lambda_i}\right)$ for each $i \in \{1, \ldots, t\}$ and multiplying both sides of the inequalities by the positive value $\lambda_i$, we obtain:

$$\begin{array}{ll}
\lambda_i t_i^{j_i}\left(\dfrac{z_i}{\lambda_i}, \dfrac{u_i}{\lambda_i}\right) \geq \lambda_i & \forall j_i \in J_i \\
\lambda_i v_i^{k_i}\left(\dfrac{z_i}{\lambda_i}, \dfrac{u_i}{\lambda_i}\right) \geq -\lambda_i & \forall k_i \in K_i \\
\lambda_i t_i^{j_i}\left(\dfrac{z_i}{\lambda_i}, \dfrac{u_i}{\lambda_i}\right) + \lambda_i v_i^{k_i}\left(\dfrac{z_i}{\lambda_i}, \dfrac{u_i}{\lambda_i}\right) \geq 0 & \forall j_i \in J_i, \forall k_i \in K_i \\
\lambda_i t_i^{j_i}\left(\dfrac{z_i}{\lambda_i}, \dfrac{u_i}{\lambda_i}\right) \geq 0 & \forall j_i \in J_i \\
\lambda_i w_i^{l_i}\left(\dfrac{z_i}{\lambda_i}, \dfrac{u_i}{\lambda_i}\right) \geq 0 & \forall l_i \in L_i.
\end{array}$$

Since $t_i^{j_i}$, $v_i^{k_i}$ and $w_i^{l_i}$ are positively-homogenous by (A3), and $\lambda_i > 0$, the above system of inequalities can be rewritten as:

$$\begin{array}{ll}
t_i^{j_i}(z_i, u_i) \geq \lambda_i & \forall j_i \in J_i \\
v_i^{k_i}(z_i, u_i) \geq -\lambda_i & \forall k_i \in K_i \\
t_i^{j_i}(z_i, u_i) + v_i^{k_i}(z_i, u_i) \geq 0 & \forall j_i \in J_i, \forall k_i \in K_i \\
t_i^{j_i}(z_i, u_i) \geq 0 & \forall j_i \in J_i \\
w_i^{l_i}(z_i, u_i) \geq 0 & \forall l_i \in L_i,
\end{array}$$

which implies that $(z_i, u_i) \in R_i(\lambda_i)$. Therefore, it follows that, for each $i \in \{1, \ldots, t\}$, $(\lambda_i, z_i, u_i)$ is such that $\lambda_i > 0$ and $(z_i, u_i) \in R_i(\lambda_i)$. Additionally, we set $(z_i, u_i) = 0$ for $t < i \leq n$. Since $t_i^{j_i}(0, 0) = \lambda t_i^{j_i}\left(\frac{0}{\lambda}, \frac{0}{\lambda}\right)$ for $\lambda > 0$, it follows that $t_i^{j_i}(0, 0) = 0$. Similarly, for all $i$, $j_i \in J_i$, $k_i \in K_i$, and $l_i \in L_i$, $t_i^{j_i}(0,0) = w_i^{l_i}(0,0) = v_i^{k_i}(0,0) = 0$. It follows that $(0,0) \in R_i(0)$. In other words, for each $i \in N$, $(\lambda_i, z_i, u_i)$ is such that $\lambda_i \geq 0$ and $(z_i, u_i) \in R_i(\lambda_i)$. Therefore, $(\lambda, z, u) \in Q$. Now, we show that if $(\lambda, z, u) \in Q$ then $z \in \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right)$. Clearly, if $(\lambda, z, u) \in Q$ and $\lambda_i > 0$, then by positive homogeneity of $t_i^{j_i}$, $v_i^{k_i}$, and $w_i^{l_i}$, it follows that $\frac{(z_i,u_i)}{\lambda_i} \in R_i(1)$. As before, then $\left(0, \frac{z_i}{\lambda_i}, \frac{u_i}{\lambda_i}, 0\right) \in A_i$. Assume without loss of generality, by reindexing $S_i$ if necessary, that $\lambda_i > 0$ for $i = 1, \ldots, t$ and $\lambda_i = 0$ for $i = t+1, \ldots, n$. Then, it follows easily that $(z_1, u_1, \ldots, z_t, u_t, 0, 0) \in \operatorname{conv}\left(\bigcup_{i=1}^n A_i\right)$ since it can be expressed as a convex combination of points in $\bigcup_{i=1}^t A_i$. Since $\operatorname{proj}_z \operatorname{conv}\left(\bigcup_{i=1}^n A_i\right) \subseteq \operatorname{conv}\left(\bigcup_{i=1}^n \operatorname{proj}_z A_i\right)$ and, by (A3), $\operatorname{proj}_z A_i \subseteq \operatorname{cl}\operatorname{conv}(S_i)$, it follows that $(z_1, \ldots, z_t, 0) \in \operatorname{conv}\left(\bigcup_{i=1}^n \operatorname{cl}(\operatorname{conv}(S_i)\right) \subseteq \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right)$. Now, since $\lambda_{t+1} = 0$, then by (A4), $(0, z_{t+1}, 0) \in 0^+\left(\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right)\right)$. Therefore, $(z_1, \ldots, z_t, z_{t+1}, 0) \in \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right)$. By induction, $z \in \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right)$.

We now prove the last part of the theorem. For this, we assume that, for every $i$, $\operatorname{proj}_z A_i$ is closed and $\operatorname{proj}_z C_i = 0^+\left(\operatorname{cl}\operatorname{conv}(S_i)\right)$. Since the sets $S_i$ are orthogonal, there do not exist vectors $\psi_i = (0, z_i, 0) \in \operatorname{proj}_z C_i$, not all zero, such that $\sum_{i=1}^n \psi_i = 0$. Define $T_i(\lambda_i) = \lambda_i \operatorname{cl}\operatorname{conv}(S_i)$ for $\lambda_i > 0$ and $T_i(0) = 0^+(\operatorname{cl}\operatorname{conv}(S_i))$. Then, by Theorem 9.8 in [22], it follows that $\bigcup_{i=1}^n \{z \mid \sum_{i=1}^n \lambda_i = 1, z_i \in T_i(\lambda_i)\}$, denoted hereafter as $T$, equals $\operatorname{cl}\operatorname{conv}(S)$. If $\bar{z} \in T$, then there exists a $\lambda$ such that $\bar{z}_i \in T_i(\lambda_i)$. If $\lambda_i > 0$, then $\frac{\bar{z}_i}{\lambda_i} \in \operatorname{cl}\operatorname{conv}(S_i)$, and therefore, there exists $u_i$ such that $\frac{(\bar{z}_i,u_i)}{\lambda_i} \in A_i$. On the other hand, if $\lambda_i = 0$, there exists $u_i$ such that $(\bar{z}_i, u_i) \in C_i$. Since $A_i$ and $C_i$ (restricted to the space of $z_i$ and $u_i$ variables) are $R_i(1)$ and $R_i(0)$ respectively, it follows that $(\lambda, \bar{z}, u) \in Q$ and so $\bar{z} \in \operatorname{proj}_z X$ and $\operatorname{cl}\operatorname{conv}(S) \subseteq \operatorname{proj}_z X$. However, we already showed that $\operatorname{proj}_z X \subseteq \operatorname{cl}\operatorname{conv}(S)$ and, therefore, $\operatorname{proj}_z X$ is equal to $\operatorname{cl}\operatorname{conv}(S)$. □

We now discuss the result and the assumptions of Theorem 2.1 in more detail. Considering first the result of this theorem, one might initially think that the stronger result that $\operatorname{proj}_z X = \operatorname{conv}(S)$ holds. We show with examples that $\operatorname{proj}_z X$ can be different from $\operatorname{conv}(S)$ and from $\operatorname{cl}\operatorname{conv}(S)$. In that sense, the result of Theorem 2.1 is as tight as possible. We consider first an example where $\operatorname{conv}(S) \subsetneq \operatorname{proj}_z X$.

**Example 2.2.** *Consider the set* $S \subseteq \mathbb{R}^2_+$*, defined as* $S = S_1 \cup S_2$*, where* $S_1 = \left\{(z_1, 0) \;\middle|\; 1 \leq z_1 \leq 2\right\}$ *and* $S_2 = \left\{(0, z_2) \;\middle|\; z_2 \geq 1\right\}$*. It can be easily verified that* $\operatorname{conv}(S) = \left\{(z_1, z_2) \;\middle|\; z_1 + z_2 \geq 1, z_1 \geq 0, z_1 < 2, z_2 \geq 0\right\} \cup \left\{(2,0)\right\}$ *as is shown in Figure 1. Observe that* $\operatorname{conv}(S)$ *is not closed. We now*

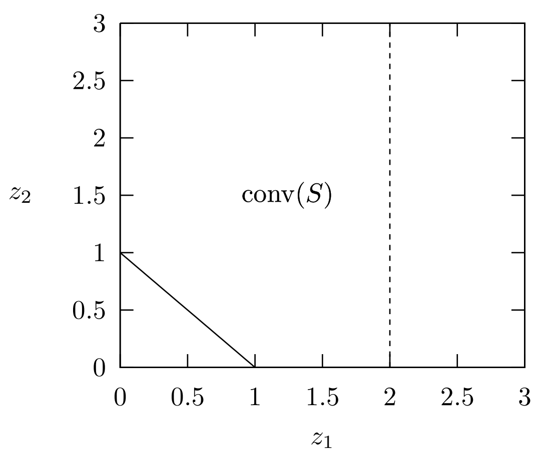


Figure 1: Illustration of Theorem 2.1 and that $\operatorname{conv}(S) \subsetneq \operatorname{proj}_z X$

*apply the convexification tool of Theorem 2.1 to* $S$ *and derive a set* $X$ *that contains* $\operatorname{conv}(S)$ *but is no larger than* $\operatorname{cl}\operatorname{conv}(S)$*. First, we verify that the set* $S$ *satisfies the assumptions of Theorem 2.1. Clearly, (A1) and (A2) hold by the definition of* $S$*. Next, it is easy to verify that* $\operatorname{conv}(S_1) = \left\{\left(z_1, 0\right) \;\middle|\; z_1 \geq 1, -\frac{1}{2}z_1 \geq -1\right\}$ *and* $\operatorname{conv}(S_2) = \left\{\left(0, z_2\right) \;\middle|\; z_2 \geq 1\right\}$*. Since* $z_1$*,* $-\frac{1}{2}z_1$*, and* $z_2$ *are linear, and, therefore, positively-homogeneous, (A3) clearly holds. Finally, since* $C_1 = \left\{(0,0)\right\} \subseteq$

*$0^+\big(\mathrm{cl\,conv}(S)\big)$ and $C_2 = \big\{(0, z_2) \bigm| z_2 \geq 0\big\} \subseteq 0^+\big(\mathrm{cl\,conv}(S_2)\big) \subseteq 0^+\big(\mathrm{cl\,conv}(S)\big)$, then (A4) also holds. Applying Theorem 2.1, we obtain that $X = \big\{\big(z_1, z_2\big) \bigm| z_1 + z_2 \geq 1, z_1 \leq 2, z_1 \geq 0, z_2 \geq 0\big\}$. In fact, since, for each $i$, $C_i = 0^+\left(\mathrm{cl\,conv}(S_i)\right)$ and $\mathrm{conv}(S_i)$ is closed, it follows from Theorem 2.1 and is apparent for this example that $X = \mathrm{cl\,conv}(S)$. This example illustrates that $X$ may contain $\mathrm{conv}(S)$ as a strict subset.* □

We now consider an example where $\mathrm{proj}_z X \subsetneq \mathrm{cl\,conv}(S)$.

**Example 2.3.** *Consider the set $S = \bigcup_{i=1}^n S_i$, where $S_i = \mathrm{proj}_z \big\{(0, z_i, u_i, 0) \in \mathbb{R}^{2n}_+ \bigm| \sqrt{z_i u_i} \geq 1\big\} = \big\{(0, z_i, 0) \bigm| z_i > 0\big\}$. Clearly, (A1) and (A2) hold by the definition of $S$. Since $\sqrt{z_i u_i}$ is positively-homogeneous, (A3) is also satisfied. Observe that $\mathrm{proj}_z C_i = \mathrm{proj}_z \big\{(0, z_i, u_i, 0) \in \mathbb{R}^{2n}_+ \bigm| \sqrt{z_i u_i} \geq 0\big\} = \big\{(0, z_i, 0) \bigm| z_i \geq 0\big\} \subseteq 0^+\big(\mathrm{cl\,conv}(S)\big)$. Therefore, (A4) holds. Applying Theorem 2.1, we obtain that $X = \big\{(z, u) \in \mathbb{R}^{2n}_+ \bigm| \sum_{i=1}^n \sqrt{z_i u_i} \geq 1\big\}$. If, for any $i$, $z_i > 0$ then there exists $u$ such that $(z, u) \in X$. Further, for all $u$, it is easy to see that $(0, u) \notin X$. Therefore, $\mathrm{proj}_z X = \big\{z \in \mathbb{R}^n_+ \bigm| \sum_{i=1}^n z_i > 0\big\}$. This example illustrates that if $\mathrm{proj}_z A_i$ is not closed then $\mathrm{proj}_z X$ may not be closed either and that, in some cases, $\mathrm{proj}_z X \subsetneq \mathrm{cl\,conv}(S)$.* □

In the above example, we exploit the fact that $\mathrm{proj}_z A_i$s are not closed to show that $\mathrm{proj}_z X$ may not be closed either. Instead, if $\mathrm{proj}_z A_i$s were closed for all $i$ then, as shown in Theorem 2.1, $\mathrm{proj}_z X$ would typically be closed as well.

We now turn our attention to Assumption (A1) in Theorem 2.1. Assumption (A1) requires that the sets $S_i$ be oriented along orthogonal principal subspaces. A weaker assumption however suffices to prove the theorem. Consider $L_i$, for $i \in \{1, \dots, n\}$, to be linear subspaces of $\mathbb{R}^{\sum_{i=1}^n d_i}$, where $L_i$ has dimension $d_i$. Further, assume that a vector $z_i \in L_i$ cannot be expressed as a linear combination of vectors in $\{L_1, \dots, L_{i-1}, L_{i+1}, \dots, L_n\}$. In this case, it is possible to construct a matrix $B$ whose columns form a basis for $\mathbb{R}^{\sum_{i=1}^n d_i}$ where the columns, that are indexed from $1 + \sum_{i=1}^{j-1} d_i$ to $\sum_{i=1}^{j} d_i$, form a basis for $L_j$. Then, define new variables $s$ such that $s = B^{-1}z$. If $z \in S_j \subseteq L_j$, it follows that $s_k \neq 0$ only if $1 + \sum_{i=1}^{j-1} d_i \leq k \leq \sum_{i=1}^{j} d_i$. Therefore, the theorem now applies to the transformed space of $s$ variables. This observation leads to the following simple derivation of the intersection cut in integer programming.

**Example 2.4.** *Consider a polyhedral cone $P = \{x \mid Ax \leq b\}$, where $A \in \mathbb{R}^{n\times n}$ is an invertible matrix. Let $X$ be the set of points that satisfy the disjunction $\pi^T x \leq \pi_0^1 \vee \pi^T x \geq \pi_0^2$, where $\pi_0^1 < \pi_0^2$. We are interested in deriving the convex hull of $P \cap X$. Observe that this setting can be used to derive all intersection/split cuts (see Balas [6]). Introducing the slack variables $\mu$ and defining $\gamma = \pi^T A^{-1}$, $\gamma_0^1 = \gamma b - \pi_0^2$, and $\gamma_0^2 = \gamma b - \pi_0^1$, we reduce the above problem into one involving convexification of $\mathcal{M} = \{\mu \mid \mu \geq 0, \gamma\mu \leq \gamma_0^1 \vee \gamma\mu \geq \gamma_0^2\}$. We assume without loss of generality that, for each $i$, $\gamma_i \neq 0$. The reformulation of the problem in the space of the slack variables, after suitable translation, is an example of the orthogonalization discussed above. Here, $\mu$ corresponds to $-s$ and $x$ corresponds to $z$. The matrix $B$ equals $A^{-1}$ and its columns are the extreme rays of $P$. Since $\mu \geq 0$ is the recession cone for $\mathcal{M}$, whenever it contains a feasible point, if $\mu = 0$ is feasible to $\mathcal{M}$, then $\mathrm{conv}(\mathcal{M}) = \{\mu \mid \mu \geq 0\}$. Define $p_i = \frac{\gamma_0^1}{\gamma_i}$ and $q_i = \frac{\gamma_0^2}{\gamma_i}$. If $\mu = 0$ is not feasible to $\mathcal{M}$, then $\gamma_0^1 < 0$ and $\gamma_0^2 > 0$. It follows that, for each $i$, exactly one of $p_i$ or $q_i$ is greater than $0$. Since $\mu_i \geq 0$ is a recession direction for $\mathrm{conv}(\mathcal{M})$ and the extreme points of $\mathcal{M}$ have at most one non-zero, it follows that:*

$$\mathrm{conv}(\mathcal{M}) = \bigcup_{i=1}^n \big\{(0, \dots, 0, \mu_i, 0, \dots, 0) \bigm| \mu_i \geq \max\{p_i, q_i\}\big\}.$$

*Now, applying Theorem 2.1, it follows that:*

$$\mathrm{conv}(\mathcal{M}) = \left\{\mu \;\middle|\; \sum_{i=1}^n \frac{\mu_i}{\max\{p_i, q_i\}} \geq 1, \mu \geq 0\right\}.$$

*Substituting back $\mu$, $p_i$, and $q_i$ in the above, we obtain:*

$$\mathrm{conv}(\mathcal{M}) = \left\{x \;\middle|\; \sum_{i=1}^n \frac{(b - Ax)_i}{\max\left\{\frac{\pi^T A_{\cdot i}^{-1} b - \pi_0^2}{\pi^T A_{\cdot i}^{-1}}, \frac{\pi^T A_{\cdot i}^{-1} b - \pi_0^1}{\pi^T A_{\cdot i}^{-1}}\right\}} \geq 1, Ax \leq b\right\}.$$ □

We next discuss Assumption (A3). This assumption requires that the convex hulls of the sets $S_i$ be known, possibly in a higher dimensional space, and that the functions $t_i^{j_i}$, for all $j_i \in J_i$, $v_i^{k_i}$, for all $k_i \in K_i$, and $w_i^{l_i}$, for all $l_i \in L_i$, used in the description of the convex hulls be positively-homogenous. In the ensuing example, we show that a simple transformation might suffice to transform the natural inequality description of $\mathrm{conv}(S_i)$ into one that uses positively-homogenous functions. We also illustrate that it is necessary to make the assumption that the functions are positively-homogenous.

**Example 2.5.** *Let $S = \bigcup_{i=1}^n S_i$, where $S_i = \left\{(0, x_i, y_i, 0) \in \mathbb{R}^{2n}_+ \;\middle|\; x_i y_i \geq r\right\}$ and $r > 0$. Clearly, (A1) and (A2) hold by the definition of $S$. Since $S_i$ is already closed and convex, $\mathrm{cl\,conv}(S_i) = S_i$, i.e., $\mathrm{cl\,conv}(S_i) = \left\{(0, x_i, y_i, 0) \in \mathbb{R}^{2n}_+ \;\middle|\; \frac{1}{r} x_i y_i \geq 1\right\}$. The above representation of $\mathrm{cl\,conv}(S_i)$ does not directly satisfy (A3) since $\frac{1}{r} x_i y_i$ is not a positively-homogenous function of $(x_i, y_i)$. However, $\mathrm{cl\,conv}(S_i)$ may be rewritten as $\mathrm{cl\,conv}(S_i) = \left\{(0, x_i, y_i, 0) \in \mathbb{R}^{2n}_+ \;\middle|\; \sqrt{\frac{1}{r} x_i y_i} \geq 1\right\}$, an expression that uses the function, $\sqrt{\frac{1}{r} x_i y_i}$, which is positively-homogenous in $(x_i, y_i)$. With this representation, (A3) is satisfied. Since $C_i = \left\{(0, x_i, y_i, 0) \in \mathbb{R}^{2n}_+ \;\middle|\; \sqrt{x_i y_i} \geq 0\right\} = 0^+(\mathrm{cl\,conv}(S_i))$, (A4) is satisfied. Therefore, Theorem 2.1 implies that $X = \mathrm{cl\,conv}(S) = \left\{(x, y) \in \mathbb{R}^{2n}_+ \;\middle|\; \sum_{i=1}^n \sqrt{x_i y_i} \geq \sqrt{r}\right\}$. Observe finally that the transformation to positively-homogenous functions is necessary and not an artifact of the proof technique. In fact, if we use the original definition of $\mathrm{cl\,conv}(S_i)$, when applying Theorem 2.1, and disregard the lack of positive-homogeneity, the resulting set would be $X' = \left\{(x, y) \in \mathbb{R}^{2n}_+ \;\middle|\; \sum_{i=1}^n x_i y_i \geq r\right\}$. The set $X'$ is nonconvex and does not even contain $\mathrm{conv}(S)$. To see this, let $r = 1$ and $n = 2$. Note that $(x_1, y_1, x_2, y_2) = (0.5, 0.5, 0.5, 0.5)$ is expressible as a convex combination of the two points in $S$, namely, $(1, 1, 0, 0) \in S_1$ and $(0, 0, 1, 1) \in S_2$. Therefore $(0.5, 0.5, 0.5, 0.5)$ belongs to $\mathrm{conv}(S)$. However, it does not satisfy the defining inequality of $X'$ whereas it does satisfy the defining inequality of $X$.* □

If $\lambda_i t_i^{j_i}\left(\frac{z_i}{\lambda_i}, \frac{u_i}{\lambda_i}\right) \leq t_i^{j_i}(z_i, u_i)$ for all $\lambda \in (0, 1]$, then $X$ still outer-approximates $\mathrm{cl\,conv}(S)$. Intuitively, while performing Fourier-Motzkin elimination, $\lambda_i t_i^{j_i}\left(\frac{z_i}{\lambda_i}, \frac{u_i}{\lambda_i}\right) \leq t_i^{j_i}(z_i, u_i)$ ensures that $X$ is contained in the closure convex hull of the disjunctive union of $S_i$, whereas $\lambda_i t_i^{j_i}\left(\frac{z_i}{\lambda_i}, \frac{u_i}{\lambda_i}\right) \geq t_i^{j_i}(z_i, u_i)$ ensures that $X$ is contained in $\mathrm{cl\,conv}\left(\bigcup_{i=1}^n S_i\right)$. Similar statements can also be made about $v_i^{k_i}(z_i, u_i)$ and $w_i^{l_i}(z_i, u_i)$. The latter of these conditions will be explored further in Proposition 2.15 to derive sufficient conditions that help verify a slightly relaxed version of (A2).

We now turn our attention to Assumption (A4). This assumption might appear quite technical and might also seem difficult to verify in practice. However, this is not the case. We show next that by simply requiring that the functions $t_i^{j_i}$, $v_i^{k_i}$, and $w_i^{l_i}$ are concave, in addition to being positively-homogenous, Assumption (A4) is automatically satisfied. Concavity of $t_i^{j_i}$, $v_i^{k_i}$, and $w_i^{l_i}$ is not an important restriction since the convexity of a positively-homogenous function's upper-level set implies concavity over the region of interest.

**Proposition 2.6.** *If, for all $i$, $j_i \in J_i$, $k_i \in K_i$, and $l_i \in L_i$, the functions $t_i^{j_i}$, $v_i^{k_i}$, and $w_i^{l_i}$, as defined in Theorem 2.1, are concave in addition to being positively-homogeneous, and the sets $S_i$ are not empty, then $\mathrm{proj}_z\, C_i \subseteq 0^+\left(\mathrm{cl\,conv}\left(\bigcup_{i=1}^n S_i\right)\right)$, i.e., Assumption (A4) is satisfied. Moreover, if the upper-level set of a positively-homogenous function is convex, then the function is concave, wherever it is positive. More precisely, if $W = \left\{(z, u) \mid t(z, u) \geq 1\right\}$ is convex and $t(z, u)$ is positively-homogenous, then $D = \{(z, u) \mid t(z, u) > 0\}$ is convex and $t(z, u)$ is concave over $D$. If, in addition, $\mathrm{cl}(D)$ is locally simplicial or more specially, polyhedral, and $t(z, u)$ is continuous then $t(z, u)$ is concave over $\mathrm{cl}(D)$.*

*Proof.* Let $(0, z_i, 0) \in S_i$. By Assumption (A3), there exists $u_i$ such that $(0, z_i, u_i, 0) \in A_i$. Consider $(0, z_i', u_i', 0) \in C_i$ and $\alpha > 0$. Then, by positive homogeneity and concavity of $t_i^{j_i}$, it follows that

$$t_i^{j_i}(z_i + \alpha z_i', u_i + \alpha u_i') \geq t_i^{j_i}(z_i, u_i) + t_i^{j_i}(\alpha z_i', \alpha u_i') = t_i^{j_i}(z_i, u_i) + \alpha t_i^{j_i}(z_i', u_i') \geq t_i^{j_i}(z_i, u_i) \geq 1.$$

The first inequality holds because of Theorem 4.7 in [22], the first equality because $t_i^{j_i}$s are positively-homogenous, the second inequality because $(0, z_i', u_i', 0) \in C_i$ and $\alpha > 0$, and the last inequality because $(0, z_i, u_i, 0) \in A_i$. Similarly, $v_i^{k_i}(z_i + \alpha z_i', u_i + \alpha u_i') \geq -1$ and $w_i^{l_i}(z_i + \alpha z_i', u_i + \alpha u_i') \geq 0$. Therefore, $(z_i + \alpha z_i', u_i + \alpha u_i') \in A_i$ and so, for all $\alpha > 0$, $(0, z_i + \alpha z_i', 0) \in \operatorname{cl}\operatorname{conv}(S_i) \subseteq \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right)$. Since $(0, z_i, 0) \in \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right)$, it follows by Theorem 8.3 in [22] that $(0, z_i', 0) \in 0^+\big(\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right)\big)$.

If $W$ is convex, then $W_K = \big\{(\lambda, x) \mid \lambda > 0, x = \lambda(z, u),\ t(z, u) \geq 1\big\}$ is the smallest convex cone containing $\big\{(1, x) \mid x \in W\big\}$. Exploiting the positive homogeneity of $t$, we may rewrite $W_K$ as:

$$W_K = \{(\lambda, x) \mid \lambda > 0, t(x) \geq \lambda\}.$$

Now, $D$ is the projection of $W_K$ in the space of $x$ and is therefore convex. Further, the hypograph of $t(z, u)$ over $D$ is $\{(r, x) \mid r \leq t(x), x \in D\} = \{(r, x) \mid r \leq \lambda \leq t(x), \lambda > 0\}$, which is convex if $W_K$ is convex. The last statement of the proposition follows from Theorems 10.3 and 20.5 in [22]. □

Even when some of the technical assumptions of Theorem 2.1 are not satisfied, it is typically the case that $X$ yields an outer-approximation of $\operatorname{conv}(S)$. To see this, observe that Proposition 2.6 shows that the functions $t_i^{j_i}$, $v_i^{k_i}$, and $w_i^{l_i}$ are concave, if they are positively-homogenous, as is assumed in Theorem 2.1, and their upper-level sets are convex. However, if concavity of these functions is known, then the outer-approximation of $\operatorname{conv}(S)$ by $\operatorname{proj}_z X$ can be shown under relatively mild assumptions.

**Proposition 2.7.** *Let $S \subseteq \mathbb{R}^{\sum_i d_i}$ and, for all $i \in N = \{1, \ldots, n\}$, let $S_i \subseteq S$. Let the points $z$ of $S$ be written as $z = (z_1, \ldots, z_i, \ldots, z_n) \in S$, where $z_i \in \mathbb{R}^{d_i}$. Assume that Assumption (A1) of Theorem 2.1 holds. Further, assume that $\operatorname{proj}_z A_i$, where $A_i$ is as defined in (1), yields an outer-approximation of $\operatorname{conv}(S_i)$ and that, for all $i \in N$, $j_i \in \{1, \ldots, J_i\}$, $k_i \in \{1, \ldots, K_i\}$, and $l_i \in \{1, \ldots, L_i\}$, $t_i^{j_i}(0, 0)$, $v_i^{k_i}(0, 0)$, and $w_i^{l_i}(0, 0)$ are non-negative. Then, $\operatorname{proj}_z(X)$, where $X$ is as defined in (2), outer-approximates $\bigcup_{i=1}^n S_i$. If, in addition, Assumption (A2) of Theorem 2.1 holds and $X$ is convex (for example, if the functions $t_i^{j_i}$, $v_i^{k_i}$, and $w_i^{l_i}$ are concave), then $\operatorname{proj}_z X \supseteq \operatorname{conv}(S)$.*

*Proof.* If Assumption (A1) is satisfied, then the sets $S_i$, for $i \in N$, are orthogonal. It can be easily verified that, if $t_i^{j_i}(0, 0)$, $v_i^{k_i}(0, 0)$, and $w_i^{l_i}(0, 0)$ are non-negative, then every constraint defining $X$ is valid for all $S_i$, where $i \in N$. Therefore, $\operatorname{proj}_z X \supseteq \bigcup_{i=1}^n S_i$. If Assumption (A2) is satisfied as well, then Claim 1 in the proof of Theorem 2.1 holds. Therefore, $\operatorname{conv}(S) = \operatorname{conv}\left(\bigcup_{i=1}^n S_i\right)$. Further, if $X$ is convex, so is $\operatorname{proj}_z X$. Since $\operatorname{proj}_z X \supseteq \bigcup_{i=1}^n S_i$, it follows that $\operatorname{proj}_z X \supseteq \operatorname{conv}\left(\bigcup_{i=1}^n S_i\right) = \operatorname{conv}(S)$. □

When the constituent functions $t_i^{j_i}$, $v_i^{k_i}$, and $w_i^{l_i}$ are concave, the result of Proposition 2.7 could also be derived using disjunctive programming. We verify Proposition 2.7 using this approach, since it more clearly reveals the source of the difference between the outer-approximation of Proposition 2.7 and the convex hull identified in Theorem 2.1. For example, one can assert that $\sum_{i \in N} t_i^{j_i}(z_i, u_i) \geq 1$, by simply noticing that if $\lambda_i > 0$ for $i \in \{1, \ldots, t\}$ then:

$$\begin{aligned}
1 &= \sum_{i=1}^t \lambda_i \\
&\leq \sum_{i=1}^t \lambda_i t_i^{j_i}\left(\frac{z_i, u_i}{\lambda_i}\right) + \sum_{i=t+1}^n t_i^{j_i}(z_i, u_i) \\
&\leq \sum_{i=1}^t \lambda_i \left(t_i^{j_i}\left(\frac{z_i, u_i}{\lambda_i}\right) + \sum_{i' \in N, i' \neq i} t_{i'}^{j_{i'}}\left(\frac{0, 0}{\lambda_i}\right)\right) + \sum_{i=t+1}^n t_i^{j_i}(z_i, u_i) \leq \sum_{i=1}^n t_i^{j_i}(z_i, u_i),
\end{aligned} \tag{10}$$

where the first inequality follows by summing the inequalities $\lambda_i \leq \lambda_i t_i^{j_i}\left(\frac{z_i, u_i}{\lambda_i}\right)$ for $i \in \{1, \ldots, t\}$ and $t_i^{j_i}(z_i, u_i) \geq 0$ for $i \in \{t+1, \ldots, n\}$, the second inequality follows since $t_{i'}^{j_{i'}}(0, 0) \geq 0$, and the third inequality from the concavity of $\sum_{i=1}^t t_i^{j_i}(z_i, u_i)$. Similarly, $\sum_{i \in T} v_i^{j_i}(z_i, u_i) \geq -1$, by realizing, in addition, that $-\sum_{i \in T} \lambda_i \geq -1$.

Proposition 2.7 provides a simple proof of the validity of the constraints defining $X$ for $\operatorname{conv}(S)$. In fact, if the primary purpose of deriving $X$ is to develop a convex outer-approximation, then

Proposition 2.7 can often replace Theorem 2.1. Therefore, many of the results derived in Section 3, can be proven, often with a weaker assertion that the derived set provides a convex outer-approximation rather than the convex hull, by using Proposition 2.7 in place of Theorem 2.1. However, there are cases, see Theorem 3.3 for example, when Proposition 2.7 provides a valid convex outer-approximation, whereas it is not apparent how to apply Theorem 2.1 to derive the convex-hull representation in the space of the original problem variables. Nevertheless, the insights gained from Theorem 2.1 are very useful. For example, we illustrate next that the search for a representation of $\operatorname{conv}(S_i)$ using positively-homogenous functions can substantially improve the relaxation. This insight will play an important role for the relaxations we derive in Section 3.

**Example 2.8.** *Consider $S = \bigcup_{i=1}^n S_i$, where, for each $i \in \{1, \ldots, n\}$, let*

$$S_i = \left\{(0, \ldots, 0, z_i, 0, \ldots, 0) \in \mathbb{R}^n_+ \;\middle|\; \sqrt{z_i} \geq 1\right\}.$$

*Proposition 2.7 shows that*

$$X' = \left\{(z_1, \ldots, z_n) \in \mathbb{R}^n_+ \;\middle|\; \sum_{i=1}^n \sqrt{z_i} \geq 1\right\}$$

*is a convex outer-approximation of $\operatorname{conv}(S)$. Note that the square-root function used in expressing $S_i$ is concave, but not positively-homogenous. Instead, if $S_i$s are represented equivalently as*

$$S_i = \left\{(0, \ldots, 0, z_i, 0, \ldots, 0) \in \mathbb{R}^n_+ \;\middle|\; z_i \geq 1\right\},$$

*then Theorem 2.1 yields the convex hull of $S$, which is*

$$X = \left\{(z_1, \ldots, z_n) \in \mathbb{R}^n_+ \;\middle|\; \sum_{i=1}^n z_i \geq 1\right\}.$$

*Clearly, by construction, $X = \operatorname{conv}(S) \subseteq X'$. In this particular example, the inclusion of $X$ in $X'$ can also be verified using the subadditivity of the square-root function for non-negative variables. This example illustrates that it often helps to find representations of convex hulls of $S_i$ using positively-homogenous functions, even when equivalent representations exist using concave functions.*

Finally, we focus on the convex extension property which is the basis of Assumption (A2). We first formally define the notion of a convex extension for orthogonal disjunctive sets. This definition is adapted from Tawarmalani and Sahinidis [29].

**Definition 2.9.** *Let $S_i \subseteq S$ for $i \in N = \{1, \ldots, n\}$. We say that $S$ has the convex extension property for orthogonal disjunctive sets $S_i$ if (A1) and a slightly relaxed form of (A2) hold. More specifically, $S$ has the convex extension property if every point $z$ in $S$ can be expressed as a convex combination of points $\chi_i$ in $\operatorname{cl}\operatorname{conv}(S_i)$ and/or a conic combination of rays $\psi_i$ in $0^+\big(\operatorname{cl}\operatorname{conv}(S_i)\big)$, i.e., for $i \in I \subseteq N$, there exist $\lambda_i \geq 0$ and $\mu_i \geq 0$, that satisfy $\sum_{i \in I} \lambda_i = 1$, such that*

$$z = \sum_{i \in I} \lambda_i \chi_i + \sum_{i \in I} \mu_i \psi_i. \tag{11}$$

The convex extension property in Definition 2.9 is more general than Assumption (A2) in Theorem 2.1, in that it allows the use of non-negative multiples of recession directions in the expression of $z$. Since $\chi_i + \frac{\mu_i}{\lambda_i}\psi_i \in \operatorname{cl}\operatorname{conv}(S_i)$, it may seem that the recession directions in (11) are not necessary. However, this is not true since $\lambda_i$ may be zero even when $\mu_i$ is not. This technicality is often important in practical applications. Nevertheless, it can be observed that even if (A2) is replaced with (11), Theorem 2.1 holds with slight modifications, as discussed below. Instead of $\operatorname{conv}(S) = \operatorname{conv}\left(\bigcup_{i=1}^n S_i\right)$, as was proved in Claim 1, we can only establish that (11) implies

$$\operatorname{cl}\operatorname{conv}(S) = \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right). \tag{12}$$

In fact, (12) is equivalent to (11). On the one hand, since, for each $i \in \{1, \ldots, n\}$, $S_i \subseteq S$ it follows that $\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right) \subseteq \operatorname{cl}\operatorname{conv}(S)$. On the other hand, since $S_i$s are orthogonal, by Theorem 9.8 in [22],

$$\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right) = \bigcup\left\{\lambda_1 \operatorname{cl}\operatorname{conv}(S_1) + \cdots \lambda_n \operatorname{cl}\operatorname{conv}(S_n) \;\middle|\; \lambda_i \geq 0^+,\ \sum_{i=1}^n \lambda_i = 1\right\}, \tag{13}$$

where the notation $\lambda_i \geq 0^+$ means that $\lambda_i \operatorname{cl}\operatorname{conv}(S_i)$ is taken to be $0^+(\operatorname{cl}\operatorname{conv}(S_i))$ rather than $\{0\}$ when $\lambda_i = 0$. Observe that (11) is another way to represent the set on the right-hand-side of (13) since if $\lambda_i > 0$ then $\chi_i + \frac{\mu_i}{\lambda_i}\psi_i \in \operatorname{cl}\operatorname{conv}(S_i)$. Otherwise, $\psi_i \in 0^+(\operatorname{cl}\operatorname{conv}(S_i))$. Now, if we assume (11), or equivalently, (12), the proof of Theorem 2.1 shows that $\operatorname{cl}\operatorname{proj}_z X = \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right)$, and, therefore, by (12), $\operatorname{cl}\operatorname{proj}_z X = \operatorname{cl}\operatorname{conv}(S)$. In this case, the last statement of Theorem 2.1 can often be used to establish closedness of $\operatorname{proj}_z X$. Note that $\operatorname{proj}_z A_i$ is closed whenever $\operatorname{conv}(S_i)$ is closed. Therefore, in most practical situations, it suffices to establish (11) instead of Assumption (A2) in Theorem 2.1. Similarly, if Assumption (A2) is replaced with (11) in Proposition 2.7, it can be easily established that $\operatorname{cl}\operatorname{conv}(S) \subseteq \operatorname{cl}\operatorname{proj}_z X$. This is because $\operatorname{cl}\operatorname{conv}(S) = \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right) \subseteq \operatorname{cl}\operatorname{conv}(\operatorname{proj}_z X) = \operatorname{cl}\operatorname{proj}_z X$, where the first equality follows from the equivalence of (11) and (12), the first containment since $\bigcup_{i=1}^n S_i \subseteq \operatorname{proj}_z X$, and the last equality since $\operatorname{proj}_z X$ is convex.

We next present a nontrivial set for which it can be proved from first principles that the convex extension property holds for orthogonal disjunctive sets. This set appears in a nonconvex formulation of the trim-loss problem proposed by Harjunkoski et al. [14]. The model is designed to determine the best way to cut a finite number of large rolls of a raw-material into smaller products using a certain number of cutting patterns. Let $I$ be the index set of products and the $J$ be the index set of the cutting patterns that are to be chosen. The demand for a product $i$ is known *a priori* and is denoted by $n_{i,\text{order}}$. For each $(i,j) \in I \times J$, let $n_{ij} \in \mathbb{Z}_+$ be the decision variable that specifies the number of products to type $i$ produced in the cutting pattern $j$ and, for each $j \in J$, let $m_j \in \mathbb{Z}_+$ be the number of times the cutting pattern $j$ is used. The following bilinear constraints model that the demand for each product is met:

$$\sum_{j=1}^{J} m_j n_{ij} \geq n_{i,\text{order}}, \text{ for } i = 1, \ldots, I, \tag{14}$$

In Proposition 2.10, we show that the bilinear integer sets defined by the constraint (14) satisfy the convex extension property for disjunctive orthogonal sets. We use this result along with Theorem 2.1 to obtain the convex hull of integer bilinear covering sets in Proposition 2.11.

**Proposition 2.10.** *Consider a bilinear integer knapsack set*

$$B^I = \left\{(x_1, y_1, x_2, y_2) \in \mathbb{Z}_+^2 \times \mathbb{Z}_+^2 \;\middle|\; x_1 y_1 + x_2 y_2 \geq r\right\}.$$

*where $r > 0$. Then, $B^I$ has the convex extension property (11) with respect to the orthogonal disjunctive sets*

$$B_1^I = \left\{(x_1, y_1, 0, 0) \in \mathbb{Z}_+^2 \times \mathbb{Z}_+^2 \;\middle|\; x_1 y_1 \geq r\right\},$$
$$B_2^I = \left\{(0, 0, x_2, y_2) \in \mathbb{Z}_+^2 \times \mathbb{Z}_+^2 \;\middle|\; x_2 y_2 \geq r\right\}.$$

*Proof.* Let $(x_1, y_1, x_2, y_2) \in B^I$. We show that there exist (i) certain subsets $I$ and $I'$ of $\{1, 2\}$, (ii) for each $i \in I$, a finite $j_i$, (iii) for each $i \in I'$, a finite $j'_i$, (iv) for each $i \in I$ and $j \in \{1, \ldots, j_i\}$, a point $\chi_{i,j} \in B_i^I$, and (v) for each $i \in I'$ and $j \in \{1, \ldots, j'_i\}$, a ray $\psi_{i,j}$ of $B_i^I$, such that

$$(x_1, y_1, x_2, y_2) = \sum_{i \in I} \sum_{j=1}^{j_i} \lambda_{i,j} \chi_{i,j} + \sum_{i \in I'} \sum_{j=1}^{j'_i} \mu_{i,j} \psi_{i,j}, \tag{15}$$

where the multipliers are such that (a) $\sum_{i \in I} \sum_{j=1}^{j_i} \lambda_{i,j} = 1$, (b) for each $i \in I$ and $j \in \{1, \ldots, j_i\}$, $\lambda_{i,j} \geq 0$, and (c) for each $i \in I'$ and $j \in \{1, \ldots, j'_i\}$, $\mu_{i,j} \geq 0$.

We assume without loss of generality that $x_1 \leq y_1 \leq y_2$ and $x_2 \leq y_2$ since the variables $x_1$, $y_1$, $x_2$, and $y_2$ can be renamed such that the largest variable is called $y_2$ and the largest variable in the other pair is called $y_1$. Note first that if $x_1 = 0$, it suffices to choose $I = \{2\}$, $I' = \{1\}$, $j_2 = 1$, $j_1' = 1$ with $\chi_{2,1} = (0, 0, x_2, y_2)$ and $\psi_{1,1} = (0, 1, 0, 0)$ to show that (11) holds. Therefore, we assume in the remainder of this proof that $x_1 \geq 1$ and, consequently, $x_1 y_1 \geq 1$. We consider two cases.

**Case 1:** $x_2 \geq x_1 y_1$. In this case, we choose $I = \{1, 2\}$, $I' = \{2\}$, and $j_1 = j_2 = j_2' = 1$. Consider the points $\chi_{1,1} = ((y_2 + 1)\, x_1, (y_2 + 1)\, y_1, 0, 0)$ and $\chi_{2,1} = (0, 0, x_2, y_2 + 1)$, and the ray $\psi_{2,1} = (0, 0, 1, 0)$. Clearly, $\chi_{1,1} \in B_1^I$, since $(y_2+1)^2 x_1 y_1 \geq x_1 y_1 + y_2^2 x_1 y_1 \geq x_1 y_1 + y_2^2 \geq x_1 y_1 + x_2 y_2 \geq r$. Similarly, $\chi_{2,1} \in B_2^I$, since $x_2\,(y_2 + 1) \geq x_2 y_2 + x_2 \geq x_2 y_2 + x_1 y_1 \geq r$. It is easily verified that

$$(x_1, y_1, x_2, y_2) = \frac{1}{y_2 + 1}\chi_{1,1} + \frac{y_2}{y_2 + 1}\chi_{2,1} + \frac{x_2}{y_2 + 1}\psi_{2,1}$$

which shows that (15) is feasible.

**Case 2:** $x_2 \leq x_1 y_1 - 1$. In this case, we choose $I = \{1, 2\}$, $I' = \{1, 2\}$, $j_2 = 1$, and $j_1 = j_1' = j_2' = 2$ with $\chi_{1,1} = (x_1 + \alpha, y_1, 0, 0)$, $\chi_{1,2} = (x_1, y_1 + \beta, 0, 0)$, $\chi_{2,1} = (0, 0, x_2, y_2 + \delta)$, $\psi_{1,1} = (1, 0, 0, 0)$, $\psi_{1,2} = (0, 1, 0, 0)$, $\psi_{2,1} = (0, 0, 1, 0)$, and $\psi_{2,2} = (0, 0, 0, 1)$, where $\alpha = \left\lceil \frac{x_2 y_2}{y_1} \right\rceil$, $\beta = \left\lceil \frac{x_2 y_2}{x_1} \right\rceil$, and $\delta = \left\lceil \frac{x_1 y_1}{x_2} \right\rceil$. It follows from the way $\alpha$, $\beta$, and $\delta$ are defined that $\chi_{1,1}$ and $\chi_{1,2}$ belong to $B_I^1$ whereas $\chi_{2,1}$ belongs to $B_I^2$. We need to prove that (15) has a feasible solution. Eliminating $\mu_{i,j}$ and using $\lambda_{2,1} = 1 - \lambda_{1,1} - \lambda_{1,2}$ to eliminate $\lambda_{2,1}$, (15) reduces to the following system:

$$\begin{aligned}
\lambda_{1,1}(x_1 + \alpha) + \lambda_{1,2}(x_1) &\leq x_1 \\
\lambda_{1,1}(y_1) + \lambda_{1,2}(y_1 + \beta) &\leq y_1 \\
(1 - \lambda_{1,1} - \lambda_{1,2})\, x_2 &\leq x_2 \quad \text{(redundant)} \\
(1 - \lambda_{1,1} - \lambda_{1,2})\,(y_2 + \delta) &\leq y_2 \\
\lambda_{1,1} + \lambda_{1,2} &\leq 1 \\
\lambda_{1,1} &\geq 0 \\
\lambda_{1,2} &\geq 0.
\end{aligned} \tag{16}$$

Projecting out $\lambda_{1,1}$ using Fourier-Motzkin elimination, we obtain

$$\max\left\{0, \frac{\alpha\delta - x_1 y_2}{\alpha\,(y_2 + \delta)}\right\} \leq \lambda_{1,2} \leq \min\left\{1, \frac{y_1}{y_1 + \beta}, \frac{y_1 y_2}{\beta\,(y_2 + \delta)}\right\}.$$

Since $\beta\delta = \left\lceil \frac{x_2 y_2}{x_1} \right\rceil \left\lceil \frac{x_1 y_1}{x_2} \right\rceil \geq \frac{x_2 y_2}{x_1} \frac{x_1 y_1}{x_2} = y_1 y_2$, it follows that:

$$\frac{y_1 y_2}{\beta(y_2 + \delta)} = \frac{1}{\frac{\beta}{y_1}\left(1 + \frac{\delta}{y_2}\right)} \leq \frac{1}{\frac{\beta}{y_1} + 1} = \frac{y_1}{y_1 + \beta} = \min\left\{1, \frac{y_1}{y_1 + \beta}\right\}.$$

Moreover, since $\alpha\delta = \left\lceil \frac{x_2 y_2}{y_1} \right\rceil \left\lceil \frac{x_1 y_1}{x_2} \right\rceil \geq y_2 x_1$, it follows that $0 \leq \frac{\alpha\delta - x_1 y_2}{\alpha(y_2 + \delta)}$ and (16) is feasible if $\alpha\beta\delta \leq \alpha y_1 y_2 + \beta x_1 y_2$. We consider two cases:

**Case 2.1:** $x_2 = 1$. In this case, $\alpha = \left\lceil \frac{y_2}{y_1} \right\rceil$, $\beta = \left\lceil \frac{y_2}{x_1} \right\rceil$, and $\delta = x_1 y_1$. There exist $f_\alpha, f_\beta \in [0, 1)$ such that $\alpha = \frac{y_2}{y_1} + f_\alpha$ and $\beta = \frac{y_2}{x_1} + f_\beta$. We observe that

$$\begin{aligned}
\alpha\beta\delta &= \left(\frac{y_2}{y_1} + f_\alpha\right)\left(\frac{y_2}{x_1} + f_\beta\right) x_1 y_1 \\
&= y_1 y_2 \left(\frac{y_2}{y_1} + f_\alpha\right) + x_1 y_2 \left(\frac{y_1}{y_2} f_\alpha f_\beta + f_\beta\right) \\
&\leq y_1 y_2 \left(\frac{y_2}{y_1} + f_\alpha\right) + x_1 y_2 \left(\frac{y_2}{x_1} + f_\beta\right) \\
&= \alpha y_1 y_2 + \beta x_1 y_2
\end{aligned}$$

where the inequality holds because $x_1 \leq y_1 \leq y_2$ implies that $x_1 y_1 f_\alpha f_\beta \leq x_1 y_1 \leq y_2^2$.

**Case 2.2:** $x_2 \geq 2$. For $(u,v) \in \mathbb{Z}_+^2$, we define $\bar{l}(u,v) = u - l$ where $l$ is the only integer in the interval $\{0,\ldots,v-1\}$ that is such that $u = qv + l$ for some $q \in \mathbb{Z}_+$, *i.e.*, $l$ is the remainder when $u$ is divided by $v$. Using this notation, it is easy to verify that $\alpha = \frac{x_2y_2+\bar{l}(x_2y_2,y_1)}{y_1}$, $\beta = \frac{x_2y_2+\bar{l}(x_2y_2,x_1)}{x_1}$, and $\delta = \frac{x_1y_1+\bar{l}(x_1y_1,x_2)}{x_2}$. Now observe that:

$$\begin{aligned}
\frac{\delta}{y_2} = \frac{x_1y_1 + \bar{l}\,(x_1y_1,x_2)}{x_2y_2} \quad &\leq \quad \frac{x_1y_1 + x_2 - 1}{x_2y_2} \\
&= \quad \frac{x_1y_1}{x_2y_2}\left(1 + \frac{x_2-1}{x_1y_1}\right) \\
&\leq \quad \frac{x_1y_1}{x_2y_2}\left(1 + \frac{x_2-1}{x_2+1}\right) \\
&= \quad \frac{1}{x_2y_2}\left(\frac{x_1y_1}{1+\frac{1}{x_2}} + \frac{x_1y_1}{1+\frac{1}{x_2}}\right) \\
&\leq \quad \frac{1}{x_2y_2}\left(\frac{x_1y_1}{1+\frac{y_1-1}{x_2y_2}} + \frac{x_1y_1}{1+\frac{x_1-1}{x_2y_2}}\right) \\
&\leq \quad \frac{x_1y_1}{x_2y_2 + \bar{l}\,(x_2y_2,y_1)} + \frac{x_1y_1}{x_2y_2 + \bar{l}\,(x_2y_2,x_1)} = \frac{x_1}{\alpha} + \frac{y_1}{\beta},
\end{aligned}$$

where the first inequality holds because $\bar{l}\,(x_1y_1,x_2) \leq x_2 - 1$, the second inequality because $x_2 \leq x_1y_1 - 1$, the third inequality holds since $y_1 \leq y_2$ implies $\frac{y_1-1}{y_2} \leq 1$ and $x_1 \leq y_2$ implies that $\frac{x_1-1}{y_2} \leq 1$, and the fourth inequality holds since $y_1 - 1 \geq \bar{l}\,(x_2y_2,y_1)$ and $x_1 - 1 \geq \bar{l}\,(x_2y_2,x_1)$. Therefore, $\alpha\beta\delta \leq \alpha y_1 y_2 + \beta x_1 y_2$.

In summary, for $(x_1,y_1,x_2,y_2) \in B^I$, (15) is feasible, and, therefore, (11) holds for $B^I$. $\square$

We now apply the result of Proposition 2.10 in conjunction with Theorem 2.1 to obtain the following result that describes the convex hull of (14).

**Proposition 2.11.** *Let*

$$B^I = \left\{(x,y) \in \mathbb{Z}_+^n \times \mathbb{Z}_+^n \;\middle|\; \sum_{i=1}^n x_i y_i \geq r\right\}, \tag{17}$$

*where $r > 0$ and, for each $i \in \{1,\ldots,n\}$, define:*

$$B_i^I = \left\{(x,y) \in B^I \;\middle|\; (x_j,y_j) = (0,0), \forall j \neq i\right\}.$$

*Let the convex hull of $B_i^I$ be represented by:*

$$\operatorname{conv}(B_i^I) = \left\{(0,0,x_i,y_i,0,0) \in \mathbb{R}_+^n \times \mathbb{R}_+^n \;\middle|\; l^j(x_i,y_i) \geq 1, \forall j \in J\right\},$$

*where $l^j(x_i,y_i) = \alpha_j x_i + \beta_j y_i$. Then,*

$$\operatorname{conv}(B^I) = \left\{(x,y) \in \mathbb{R}_+^n \times \mathbb{R}_+^n \;\middle|\; \sum_{i=1}^n l^{j_i}(x_i,y_i) \geq 1, \forall j_i \in J\right\}. \tag{18}$$

*Proof.* We prove this result by applying Theorem 2.1. Let $z_i = (x_i,y_i)$. Assumption (A1) holds by the definition of $B_i^I$. The convex extension property, (11), follows from a sequential application of Proposition 2.10. Assumption (A3) is satisfied since $\operatorname{conv}\left(B_i^I\right)$ is closed and the functions $l^j(x_i,y_i)$ are positively-homogeneous. Further, since $0^+\left(\operatorname{cl}\operatorname{conv}\left(B_i^I\right)\right) = \mathbb{R}_+^n \times \mathbb{R}_+^n$, it follows that

$$C_i = \left\{(0,0,x_i,y_i,0,0) \in \mathbb{R}_+^n \times \mathbb{R}_+^n \;\middle|\; l^j(x_i,y_i) \geq 0, \forall j \in J\right\} \subseteq 0^+\left(\operatorname{cl}\operatorname{conv}\left(B_i^I\right)\right).$$

Therefore, (A4) holds. Now, by Theorem 2.1 and the discussion following Definition 2.9, it follows that

$$\operatorname{cl}\operatorname{conv}(B^I) = X = \left\{ (x,y) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ \;\middle|\; \sum_{i=1}^n l^{j_i}(x_i, y_i) \geq 1, \forall j_i \in J \right\},$$

where the closure operation is not needed on $X$ since it is a closed set, being an intersection of closed half-spaces. In fact, $X$ is polyhedral, since there are only finitely many half-spaces in its expression. Now, consider the closed sets $B_i^{I'} = \left\{ (x,y) \in \mathbb{Z}^{2n}_+ \;\middle|\; x_i y_i \geq r \right\}$. Observe that $B_i^I \subseteq B_i^{I'} \subseteq B^I$. Now, by Corollary 9.8.1 in [22], $\operatorname{conv}\left(\bigcup_{i=1}^n B^{I'}\right)$ is closed. Since

$$\operatorname{conv}\left(B^I\right) \subseteq \operatorname{cl}\operatorname{conv}(B^I) \subseteq \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n B_i^{I'}\right) = \operatorname{conv}\left(\bigcup_{i=1}^n B_i^{I'}\right) \subseteq \operatorname{conv}\left(B^I\right),$$

where the second containment holds since $B_i^I \subseteq B_i^{I'}$ and because the discussion following Definition 2.9 argues that $\operatorname{cl}\operatorname{conv}(B^I) = \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n B_i^{I'}\right)$, the first equality since $\operatorname{conv}\left(\bigcup_{i=1}^n B_i^{I'}\right)$ is closed, and the third containment since $B_i^{I'} \subseteq B^I$. Therefore, the equality holds throughout, and the result follows. ☐

Observe that, even though $\operatorname{conv}(B^I)$ is closed, $\operatorname{conv}\left(\bigcup_{i=1}^n B_i^I\right)$ is not closed. Proposition 2.11 shows that $\operatorname{conv}(B^I)$ has exponentially many facets. In particular, if $B_i^I$ has $|J|$ facets, there are $|J|^n$ inequalities in the description of $\operatorname{conv}(B^I)$. We note, however, that separation is not difficult to perform as the coefficients of each pair of variables can be determined independently. Since there is an obvious pseudo-polynomial algorithm to compute the facets of $\operatorname{conv}(B_i^I)$, it is clearly possible to separate the facets of $\operatorname{conv}(B^I)$ in pseudo-polynomial time.

**Example 2.12.** *Consider the set*

$$B^I = \left\{ (x,y) \in \mathbb{Z}^2_+ \times \mathbb{Z}^2_+ \;\middle|\; x_1 y_1 + x_2 y_2 \geq 10 \right\}. \tag{19}$$

*It is easily verified that for both $i \in \{1,2\}$*

$$\operatorname{conv}\left(B_i^I\right) = \left\{ (0, x_i, y_i, 0) \in \mathbb{R}^4_+ \;\middle|\; y_i \geq 1, 10x_i + 2y_i \geq 30, x_i + y_i \geq 7, 2x_i + 10y_i \geq 30, x_i \geq 1 \right\}.$$

*It follows from Proposition 2.11 that the convex hull of $B^I$ has 25 inequalities and is represented by*

$$\operatorname{conv}(B^I) = \left\{ (x,y) \in \mathbb{R}^2_+ \times \mathbb{R}^2_+ \;\middle|\; \left\{ \begin{array}{rr} & y_1 \\ \frac{5}{15}x_1 & +\frac{1}{15}y_1 \\ \frac{1}{7}x_1 & +\frac{1}{7}y_1 \\ \frac{1}{15}x_1 & +\frac{5}{15}y_1 \\ x_1 & \end{array} \right\} + \left\{ \begin{array}{rr} & y_2 \\ \frac{5}{15}x_2 & +\frac{1}{15}y_2 \\ \frac{1}{7}x_2 & +\frac{1}{7}y_2 \\ \frac{1}{15}x_2 & +\frac{5}{15}y_2 \\ x_2 & \end{array} \right\} \geq 1 \right\}, \tag{20}$$

*where each pair of coefficients for $(x_1, y_1)$ can be matched with each pair of coefficients for $(x_2, y_2)$.*

Similarly, the convex hull characterization for a variety of bilinear sets can be obtained using the result of Theorem 2.1. In particular, we study now the mixed integer variant while we study the continuous variant in more detail in Section 3.

**Proposition 2.13.** *Let*

$$B^M = \left\{ (x,y) \in \mathbb{Z}^n_+ \times \mathbb{R}^n_+ \;\middle|\; \sum_{i=1}^n a_i x_i y_i \geq r \right\}, \tag{21}$$

*where $r > 0$, and, for each $i \in \{1, \ldots, n\}$, $a_i > 0$. Define, for each $i \in \{1, \ldots, n\}$,*

$$B_i^M = \left\{ (x,y) \in B^M \;\middle|\; (x_j, y_j) = (0,0), \forall j \neq i \right\}.$$

*Let the convex hull of $B_i^M$ be represented by:*

$$\operatorname{conv}\left(B_i^M\right) = \left\{(0,0,x_i,y_i,0,0) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ \;\middle|\; l^j(x_i,y_i) \geq 1, \forall j \in J_i\right\},$$

*where $l^j(x_i,y_i) = \alpha_j x_i + \beta_j y_i$. Then,*

$$\operatorname{conv}\left(B^M\right) = \left\{(x,y) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ \;\middle|\; \sum_{i=1}^n l^{j_i}(x_i,y_i) \geq 1, \forall j_i \in J_i\right\}. \tag{22}$$

*Proof.* Because the verification of the convex extension property is the only technical part of the proof that is significantly different from that of $B^I$, we only discuss the proof of this property next. Because induction can be used, it suffices to prove the result when $n=2$. Let $(x_1,y_1,x_2,y_2) \in B^M$. We show that there exist (i) subsets $I$ and $I'$ of $\{1,2\}$, (ii) for each $i \in I$, a point $\chi_i \in B_i^M$, and (iii) for each $i \in I'$, a ray $\psi_i$ of $B_i^M$, such that

$$(x_1,y_1,x_2,y_2) = \sum_{i\in I} \lambda_i \chi_i + \sum_{i \in I'} \mu_i \psi_i, \tag{23}$$

where the multipliers satisfy the following conditions: (a) $\sum_{i\in I} \lambda_i = 1$, (b) for all $i \in I$, $\lambda_i \geq 0$, and (c) for all $i \in I'$, $\mu_i \geq 0$.

Note first that, if $x_2 = 0$, it suffices to choose $I = \{1\}$, $I' = \{2\}$, $\chi_1 = (x_1,y_1,0,0)$, and $\psi_2 = (0,0,0,1)$ to show that (11) holds. Similarly, if $y_2 = 0$, it suffices to choose $I = \{1\}$, $I' = \{2\}$, $\chi_1 = (x_1,y_1,0,0)$, and $\psi_2 = (0,0,1,0)$ to show that (11) holds. We assume without loss of generality that $x_1y_1 \geq x_2y_2$ since the pair of variables $(x_1,y_1)$ and $(x_2,y_2)$ can be interchanged along with their respective coefficients $a_1$ and $a_2$. Therefore, in addition to the positivity of $x_2$ and $y_2$, we may also assume in the remainder of this proof that $x_1 \geq 1$ and $y_1 > 0$. Define $\chi_1 = \left(x_1, y_1 + \frac{a_2x_2y_2}{a_1x_1}, 0, 0\right)$, $\chi_2 = \left(0,0,x_2,y_2+\frac{a_1x_1y_1}{a_2x_2}\right)$, $\psi_1 = (x_1,0,0,0)$, and $\psi_2 = (0,0,x_2,0)$. It can be easily verified that

$$(x_1,y_1,x_2,y_2) = \frac{a_1x_1y_1}{a_1x_1y_1 + a_2x_2y_2}(\chi_1 + \psi_2) + \frac{a_2x_2y_2}{a_1x_1y_1+a_2x_2y_2}(\chi_2 + \psi_1)$$

which shows that the convex extension property (11) holds. $\square$

Propositions 2.11 and 2.13 illustrate both the fact that the convex extension property used in Theorem 2.1 holds in surprising settings and that this property might not always be trivial to verify. We next present in Theorem 2.14 and Proposition 2.15 conditions under which the convex extension property over orthogonal disjunctive sets can be shown to hold. These conditions are satisfied by many polynomial covering inequalities as we will discuss later in Section 3.

**Theorem 2.14.** *Consider a function $g(z_1,\ldots,z_n) : \mathbb{R}_+^{\sum_i d_i} \mapsto \mathbb{R}$, where $z_i \in \mathbb{R}^{d_i}_+$, and the set $G = \left\{z \in \mathbb{R}_+^{\sum_i d_i} \;\middle|\; g(z_1,,\ldots,z_n) \geq r\right\}$, where $r > 0$. Let $G_i = G \cap \left\{(0,\ldots,0,z_i,0,\ldots,0) \;\middle|\; z_i \in \mathbb{R}^{d_i}_+\right\}$ and $g_i(z_i) = g(0,\ldots,0,z_i,0,\ldots,0)$. If there exist functions $h_i : \mathbb{R}^{d_i}_+ \mapsto \mathbb{R}^{k_i}$ and $f : \mathbb{R}^{\sum_i k_i} \mapsto \mathbb{R}$ such that:*
*(S1) $g(z) \leq f\big(h_1(z_1),\ldots,h_n(z_n)\big)$, where $f$ is a convex function,*
*(S2) $f(y_1) > f(y_2)$ whenever $y_1 \geq y_2$ and at least one component of $y_1$ is larger than the corresponding component of $y_2$,*
*(S3) $g_i(z_i) = f(0,\ldots,0,h_i(z_i),0,\ldots,0)$,*
*(S4) For all $i$, $h_i(0) = 0$ and, for $\lambda \in (0,1]$, $\lambda h_i\left(\frac{z_i}{\lambda}\right) \geq h_i(z_i)$, and*
*(S5) For all $i$, $h_i(z_i) \leq 0$ implies that $(0,z_i,0) \in 0^+\big(\operatorname{cl}\operatorname{conv} G_i\big)$,*
*are satisfied over $\mathbb{R}_+^{\sum_{i=1}^n d_i}$ then the convex extension property, (11), holds for the set $G$. Assume that, for each $i \in \{1,\ldots,n\}$, $\operatorname{conv}(G_i)$ is closed. Define $G'_i = \operatorname{conv}(G_i) + \sum_{i'\neq i} 0^+(\operatorname{conv} G_{i'})$. If, for all $i$, $G'_i \subseteq \operatorname{conv}(G)$, then $\operatorname{conv}(G)$ is closed.*

*Proof.* Let $z \in G$ and $y(z) = \big(h_1(z_1), \ldots, h_n(z_n)\big)$. In the following, we sometimes denote $h_i(z_i)$ as $y_i(z)$ to emphasize that it is the $i^{\text{th}}$ component of $y(z)$. Let $T = \{i \mid h_i(z_i) \le 0\}$. Then, by (S5), for each $i \in T$, $(0, z_i, 0) \in 0^+\big(\operatorname{cl}\operatorname{conv} G_i\big)$. If $z - \sum_{i\in T}(0, z_i, 0) \in \operatorname{cl}\operatorname{conv}\big(\bigcup_{i=1}^n G_i\big)$, then so does $z$. We now show that $z' = z - \sum_{i\in T}(0, z_i, 0) \in \operatorname{cl}\operatorname{conv}\big(\bigcup_{i=1}^n G_i\big)$. Let $\delta$ be a subgradient of $f$ at $y(z')$. Then, (S2) implies that $\delta > 0$. Otherwise, suppose that $\delta_i \le 0$. Let $e_i$ denote the $i^{\text{th}}$ unit vector and choose $\epsilon > 0$. Observe that $f(y(z') - \epsilon e_i) \ge f(y(z')) - \epsilon\langle \delta, e_i\rangle = f\big(y(z')\big) - \epsilon\delta_i \ge f\big(y(z')\big)$, a contradiction to (S2). Clearly, for each $i \notin T$, $h_i(z'_i) = h_i(z_i)$. By construction, for each $i \in T$, $z'_i = 0$ and, therefore, $h_i(z'_i) = 0 \ge h_i(z_i)$. In other words, $y(z') = \max\{y(z), 0\}$. Observe that (S1) and (S2) together imply that $f(y(z')) \ge f(y(z)) \ge g(z) \ge r$.

First, consider the case where $\langle \delta, y(z')\rangle = 0$. Then, $y(z') = 0$. Consider any $i$ and a $\bar{z}_i \in \mathbb{R}^{d_i}_+$. On the one hand, assume that $h_i(\bar{z}_i) > 0$. Then, $g(0, \bar{z}_i, 0) = g_i(\bar{z}_i) = f(0, h_i(\bar{z}_i), 0) > f(0) = f(y(z')) \ge r$, where the first equality follows from the definition of $g_i$, the second from (S3), and the first inequality from (S2) and $h_i(\bar{z}_i) > 0$. Therefore, $(0, \bar{z}_i, 0) \in G \subseteq \operatorname{cl}\operatorname{conv}(G)$. On the other hand, assume that $h_i(\bar{z}_i) \le 0$. Then, by (S5), we know that $(0, \bar{z}_i, 0) \in 0^+\big(\operatorname{cl}\operatorname{conv} G_i\big)$. Since $g_i(0) = f(0) = f(y(z')) \ge r$, it follows that $0 \in G_i$. Combining $0 \in G_i$ and $(0, \bar{z}_i, 0) \in 0^+\big(\operatorname{cl}\operatorname{conv} G_i\big)$, we can conclude that $(0, \bar{z}_i, 0) \in \operatorname{cl}\operatorname{conv}(G_i) \subseteq \operatorname{cl}\operatorname{conv}(G)$. In other words, regardless of the sign of $h_i(z_i)$, it follows that $(0, \bar{z}_i, 0) \in \operatorname{cl}\operatorname{conv}(G)$. Since $\bar{z}_i$ was arbitrarily chosen in $\mathbb{R}^{d_i}_+$, it follows that $\mathbb{R}^{\sum_i d_i}_+ \subseteq \operatorname{cl}\operatorname{conv}\big(\bigcup_{i=1}^n G_i\big) \subseteq \operatorname{cl}\operatorname{conv}(G) \subseteq \mathbb{R}^{\sum_i d_i}_+$. Since equality holds throughout, (12), or equivalently (11) holds for $G$.

Now, consider the case when $\langle \delta, y(z')\rangle > 0$. Define $\lambda_i = \frac{\delta_i y_i(z')}{\langle \delta, y(z')\rangle}$. Since $\delta_i$ and $y_i(z')$ are non-negative, it follows that $\lambda_i \ge 0$ and $\sum_{i=1}^n \lambda_i = 1$. Define $I = \big\{i \mid \lambda_i > 0\big\}$ and observe that $|I| \ge 1$. The following chain of implication holds

$$i \notin I \Rightarrow y_i(z') = 0 \Rightarrow i \in T \Rightarrow z'_i = 0,$$

where the first implication follows since $\delta_i > 0$; the second because, for each $i \notin T$, $y_i(z') > 0$; and the third by the construction of $z'$. Therefore, $z' = \sum_{i\in I} z''_i$, where $z''_i = \big(0, \ldots, 0, z'_i, 0, \ldots, 0\big)$. For each $i \in I$, let $\chi_i = \frac{z''_i}{\lambda_i}$. Observe that $z' = \sum_{i\in I}\lambda_i\chi_i$, *i.e.*, $z'$ can be expressed as a convex combination of $\chi_i$ for $i \in I$. The following shows that, for all $i \in I$, $\chi_i \in G_i$:

$$\begin{aligned}
g(\chi_i) = g_i\left(\frac{z'_i}{\lambda_i}\right) &= f\big(y(\chi_i)\big) \ge f\left(\frac{1}{\lambda_i}y(z''_i)\right) \ge f\big(y(z')\big) + \delta_i\frac{\langle \delta, y(z')\rangle}{\delta_i y_i(z')}y_i(z'') - \sum_{j=1}^n \delta_j y_j(z') \\
&= f\big(y(z')\big) + \delta_i\frac{\langle \delta, y(z')\rangle}{\delta_i y_i(z')}y_i(z') - \sum_{j=1}^n \delta_j y_j(z') = f\big(y(z')\big) \ge r.
\end{aligned}$$

The first equality follows from the definition of $g_i$, the second equality from (S3), the first inequality follows since $f$ is non-decreasing by (S2) and $h_i\big(\frac{z'_i}{\lambda_i}\big) \ge \frac{1}{\lambda_i}h_i(z'_i)$, the second inequality because $\delta$ is a subgradient of $f$ at $y(z')$, and the third equality because $y_i(z'') = h_i(z'_i) = y_i(z')$. Since $z = z' + \sum_{i\in T}(0, z_i, 0)$, where, for each $i \in T$, $(0, z_i, 0) \in 0^+\left(\operatorname{cl}\operatorname{conv}(G_i)\right)$ it follows that (11) holds for $G$.

We now prove the last statement of the theorem. Consider an arbitrary $i \in N$. Clearly, $G'_i$, as defined in the statement of the theorem, is convex. We argue that it is closed as well. By Corollary 9.1.1 in [22], $G'_i$ is closed if there do not exist $(0, z_i, 0) \in \operatorname{conv}(G_i)$ and, for $i' \in N\backslash\{i\}$, $(0, z_{i'}, 0) \in 0^+\left(\operatorname{conv} G_{i'}\right)$, not all zero, such that $\sum_{i=1}^n (0, z_i, 0) = 0$. But, the orthogonal vectors $(0, z_i, 0)$ sum to zero if and only if each of the vectors is zero. Therefore, $G'_i$ is closed. Again by Corollary 9.1.1 in [22], $0^+(G'_i) = \sum_{i=1}^n 0^+\left(\operatorname{conv} G_i\right)$. Since the recession directions of $G'_i$ are independent of $i$, it follows by Corollary 9.8.1 in [22] that $\operatorname{conv}\left(\bigcup_{i=1}^n G_i\right)$ is closed. Now,

$$\operatorname{conv}(G) \subseteq \operatorname{cl}\operatorname{conv}(G) = \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n G_i\right) \subseteq \operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n G'_i\right) = \operatorname{conv}\left(\bigcup_{i=1}^n G'_i\right) \subseteq \operatorname{conv}(G),$$

where the first equality follows from the equivalence of (11) and (12), the second containment follows since $G_i \subseteq G'_i$, the second equality follows since $\operatorname{conv}\left(\bigcup_{i=1}^n G'_i\right)$ is closed and the third containment follows since $G'_i \subseteq \operatorname{conv}(G)$. $\square$

A special case of Theorem 2.14 deserves special attention since it finds many applications. In this case, $f(\cdot)$ is the summation operator and $h_i(z_i) = g_i(z_i)$, $g(z)$ is subadditive, $g_i(z_i)$ satisfies (S4), and $g_i(z_i)$ eventually increases to infinity along every direction in the non-negative orthant. Observe that Assumption (S2) is satisfied trivially in this case. Further, Assumption (S1) reduces to $g(z) \leq \sum_{i=1}^{n} g_i(z_i)$, which is satisfied by all subadditive functions since $z = \sum_{i=1}^{n}(0,\ldots,0,z_i,0,\ldots,0)$ and $g_i(z_i) = g(0,\ldots,0,z_i,0,\ldots,0)$. Many of the applications we discuss in Section 3 will invoke Theorem 2.14 in such a setup.

The main challenge in applying Theorem 2.14 in practical situations is verifying Assumption (S4). However, when $h_i(z_i)$ is derived from other functions using operations such as summations, minimizations, or maximizations, then (S4) can often be established easily by studying the same properties for the functions used in the derivation of $h_i(z_i)$. To see this, first note that the assumption is satisfied trivially by any linear function. If $h(z) = w\left(p_1(z),\ldots,p_K(z)\right)$, for all $k \in \{1,\ldots,K\}$, $p_k(z)$ satisfies (S4), $w$ satisfies (S4), $w$ is isotonic, *i.e.*, $w(y_1) \geq w(y_2)$ if $y_1 \geq y_2$, and $w(0,\ldots,0) = 0$, then $h(z)$ satisfies (S4) as well. Clearly, $h(0) = w(p_1(0),\ldots,p_k(0)) = w(0,\ldots,0) = 0$ and:

$$\lambda h\left(\frac{z}{\lambda}\right) = \lambda w\left(p_1\left(\frac{z}{\lambda}\right),\ldots,p_k\left(\frac{z}{\lambda}\right)\right) \geq \lambda w\left(\frac{1}{\lambda}p_1(z),\ldots,\frac{1}{\lambda}p_k(z)\right) \geq w\left(p_1(z),\ldots,p_k(z)\right) = h(z),$$

where the first inequality follows since $w$ is isotonic and $p_k(z)$ obeys (S4); and the second inequality because $w$ obeys (S4). If $w$ satisfies (S4) only over the non-negative orthant, then $p_k(z)$ must be non-negative as well. In particular, $\sum_{k=1}^{K} p_k(z)$ satisfies the assumption as long as, for all $k$, $p_k(z)$ satisfies the assumption. Now, consider $h(z) = \mathrm{op}_y\, p(y,z)$, where op is an operator such as min or max that satisfies $\mathrm{op}_y\, f_1(y) \geq \mathrm{op}_y\, f_2(y)$ if, for all $y$, $f_1(y) \geq f_2(y)$ and $\lambda\, \mathrm{op}_y\, f(y) \geq \mathrm{op}_y\, \lambda f(y)$ for $\lambda \in (0,1]$. In addition, assume that $\lambda p\left(y, \frac{z}{\lambda}\right) \geq p(y,z)$ for $\lambda \in (0,1]$. Then,

$$\lambda h\left(\frac{z}{\lambda}\right) = \lambda \operatorname*{op}_y p\left(y,\frac{z}{\lambda}\right) \geq \lambda \operatorname*{op}_y \frac{1}{\lambda}p(y,z) \geq \operatorname*{op}_y p(y,z) = h(z),$$

for $\lambda \in (0,1]$. In particular, if $h(z) = \min\bigl(p_1(z),\ldots,p_K(z)\bigr)$ and, for all $\lambda \in (0,1]$, $p_k(z) \leq \lambda p_k\left(\frac{z}{\lambda}\right)$ then $h(z) \leq \lambda h\left(\frac{z}{\lambda}\right)$.

Theorem 2.14 establishes the convex extension property over the non-negative orthant. As will be discussed later, we are often interested in the covering set $G$ discussed in Theorem 2.14, but with the added restriction that the variables belong to a compact convex set. In that case, Theorem 2.14 can clearly be used to construct a relaxation. However, it also points to a construction that can make this relaxation stronger as the compact convex set becomes smaller. Intuitively, it suffices to underestimate the original function outside of the region of interest by extending it linearly in such a way that the modified function still satisfies (S4). We next describe this procedure more precisely by showing that if $z \in C$, where $C$ is a nonempty closed convex subset of $\mathbb{R}^d_+$, and $p(z)$ is a function that satisfies Assumption (S4), then it is possible to underestimate $p(z)$ and still satisfy (S4). For this, we partition $\mathbb{R}^d_+$ as follows. Given $z \in \mathbb{R}^d_+$, we define $\Lambda(z) = \{\lambda \mid \lambda z \in C\}$ and define $\lambda_{\inf}(z)$ and $\lambda_{\sup}(z)$ as the infimum and supremum over $\lambda$ in $\Lambda(z)$ respectively. Define $\mathcal{L} = \{z \in \mathbb{R}^d_+ \backslash C \mid \lambda_{\inf}(z) < 1\}$, $\mathcal{S} = \{z \in \mathbb{R}^d_+ \backslash C \mid \lambda_{\sup}(z) > 1\}$, and $\mathcal{N} = \{z_i \in \mathbb{R}^d_+ \backslash 0 \mid \forall\lambda \geq 0, \lambda z \notin C\}$. Note that $\mathbb{R}^d_+ = \mathcal{L} \cup \mathcal{S} \cup \mathcal{N} \cup C\backslash\{0\} \cup \{0\}$. If $z \in \mathcal{N}$, it does not belong to $C$, $\mathcal{L}$, or $\mathcal{S}$. Otherwise, there exists $\lambda$ such that $\lambda z \in C$. Therefore, $\Lambda(z) \neq \emptyset$. If $z \in C$, it does not belong to $\mathcal{L}$ or $\mathcal{S}$. Otherwise, if $z \in \mathcal{L}$, *i.e.*, $\lambda_{\inf}(z) < 1$, then $\lambda_{\sup}(z) < 1$ (since the set $C$ is convex and $z \notin C$) and, therefore, $z \notin \mathcal{S}$. Mutual exclusivity of $\mathcal{L}$ and $\mathcal{S}$ can be easily verified. Now define:

$$h(z) = \begin{cases} 0 & \text{if } z = 0, \\ p(z) & \text{if } z \in C\backslash\{0\}, \\ \dfrac{1}{\lambda_{\sup}(z)}p\left(\lambda_{\sup}(z)z\right) & \text{if } z \in \mathcal{L}, \\ -\infty & \text{otherwise.} \end{cases} \tag{24}$$

We show that $h(z)$ satisfies Assumption (S4). By definition, $h(0) = 0$. We need to verify that $\lambda h(\frac{z}{\lambda}) \geq h(z)$ for all $\lambda \in (0,1]$. If $z \in \mathcal{N} \cup \mathcal{S}$, the verification is straightforward. Consider $z \in \mathcal{L}$.

Then, for an arbitrary $\lambda \in (0, 1]$,

$$h\left(\frac{z}{\lambda}\right) = \frac{1}{\lambda_{\sup}\left(\frac{z}{\lambda}\right)} p\left(\lambda_{\sup}\left(\frac{z}{\lambda}\right)\frac{z}{\lambda}\right) = \frac{1}{\lambda\lambda_{\sup}(z)} p(\lambda_{\sup}(z)z) = \frac{1}{\lambda} h(z) \tag{25}$$

since it follows from the definition of $\lambda_{\sup}(\cdot)$ that $\lambda_{\sup}(\frac{z}{\lambda}) = \lambda\lambda_{\sup}(z)$. If $z \in C$, then for an arbitrary $\lambda \in \left[\frac{1}{\lambda_{\sup}(z)}, 1\right]$

$$h\left(\frac{z}{\lambda}\right) = p\left(\frac{z}{\lambda}\right) \geq \frac{1}{\lambda} p(z) = \frac{1}{\lambda} h(z) \tag{26}$$

since in this case $\frac{z}{\lambda} \in C$. If $\lambda \in \left[0, \frac{1}{\lambda_{\sup}(z)}\right)$ then $\frac{z}{\lambda} \in \mathcal{L}$. The argument proceeds in two steps by showing that

$$h\left(\frac{z}{\lambda}\right) = \frac{1}{\lambda\lambda_{\sup}(z)} h(\lambda_{\sup}(z)z) \geq \lambda h(z),$$

where the first equality follows from (25) since $\lambda\lambda_{\sup}(z) \leq 1$ and the first inequality follows from (26).

Clearly, one could define $h(z)$ as any function that satisfies Assumption (S4) over $\mathcal{N}$. That is however not the case over $\mathcal{S}$. If $p(z)$ is continuous or $C$ is closed, then $h(\lambda_{\inf}(z)z) = p(\lambda_{\inf}(z)z)$, which must be at least as large as $\lambda_{\inf}(z)h(z)$. Therefore,

$$h(z) \leq \frac{1}{\lambda_{\inf}(z)} p(\lambda_{\inf}(z)z). \tag{27}$$

It turns out that we could define $h(z)$ to be any function that satisfies Assumption (S4) over $\mathcal{S}$ as long as it satisfies (27). This can be verified as follows. If $z \in \left(\frac{1}{\lambda_{\inf}(z)}, 1\right]$ then the condition is satisfied by the definition of the function. At $\lambda = \frac{1}{\lambda_{\inf}(z)}$, the assumption is satisfied because of (27). If $\lambda \in \left(\frac{1}{\lambda_{\sup}(z)}, \frac{1}{\lambda_{\inf}(z)}\right)$ then the result follows from $\lambda = \lambda\lambda_{\inf}(z)\frac{1}{\lambda_{\inf}(z)}$, using (26) for $\lambda\lambda_{\inf}(z)$ and using (27) for $\frac{1}{\lambda_{\inf}(z)}$. Finally, if $\lambda \in \left(0, \frac{1}{\lambda_{\inf}(z)}\right]$, the result follows by expressing $\lambda = \lambda\lambda_{\sup}(z)\frac{\lambda_{\inf}(z)}{\lambda_{\sup}(z)}\frac{1}{\lambda_{\inf}(z)}$ and using (25) for $\lambda\lambda_{\sup}(z)$, (26) for $\frac{\lambda_{\inf}(z)}{\lambda_{\sup}(z)}$ and (27) for $\frac{1}{\lambda_{\inf}(z)}$. Similarly, it is reasonable, to define $h(z)$ over $\mathcal{L}$ as any function that satisfies the assumption and is at least as large as $\frac{1}{\lambda_{\sup}(z)} p\left(\lambda_{\sup}(z)z\right)$.

Theorem 2.1 also points to an interesting set of sufficient conditions that can be used to verify the convex extension property. The primary difference from the conditions in Theorem 2.14 is that Proposition 2.15 does not impose a structure on the original set, $S$. Instead, it constructs a set $X$ whose projection in the $z$-space is contained within $\operatorname{cl}\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right)$, using a construction similar to Theorem 2.1, and then leaves it to the user to verify that $X$ outerapproximates $S$. This technique may be useful when $S$ is defined by more than one inequality. Also, note that the special case of Theorem 2.14, where $f(\cdot)$ is a summation operator, $h_i(z_i) = g_i(z_i)$, $g(z)$ is subadditive, and $g_i(z_i)$ tends to infinity along every direction in the non-negative orthant, can also be seen to have the convex extension property using Proposition 2.15.

**Proposition 2.15.** *For a set $S$ and its subsets $S_i \subseteq S$ for $i \in N = \{1, \ldots, n\}$, let $z_i \in \mathbb{R}^{d_i}$ and $z = (z_1, \ldots, z_i, \ldots, z_n) \in S \subseteq \mathbb{R}^{\sum_i d_i}$. Assume that (A1) and (A4) are satisfied as in Theorem 2.1 and the sets $A_i$ and $X$ are as defined in (1) and (2) respectively. If, in addition, the following assumptions are satisfied:*
*(N1) $S_i \subseteq \operatorname{proj}_z A_i \subseteq \operatorname{cl}\big(\operatorname{conv}(S_i)\big)$,*
*(N2) $t_i^{j_i}$, $v_i^{k_i}$, and $w_i^{l_i}$ are such that for all $0 < \lambda \leq 1$,*

$$\lambda t_i^{j_i}\left(\frac{(z_i, u_i)}{\lambda}\right) \geq t_i^{j_i}(z_i, u_i),\ \lambda v_i^{k_i}\left(\frac{(z_i, u_i)}{\lambda}\right) \geq v_i^{k_i}(z_i, u_i),\ \lambda w_i^{l_i}\left(\frac{(z_i, u_i)}{\lambda}\right) \geq w_i^{l_i}(z_i, u_i),$$

*(N3) $S \subseteq \operatorname{cl}\operatorname{proj}_z X$.*
*Then, (11) holds for $S$.*

*Proof.* Here, Fourier-Motzkin elimination shows, as it did in the proof of Theorem 2.1, that $X = \mathrm{proj}_{z,u} Q$. We will now show that $\mathrm{proj}_z X = \mathrm{proj}_z Q \subseteq \mathrm{cl\,conv}\left(\bigcup_{i=1}^n S_i\right)$. The proof is again similar to that for Theorem 2.1 except that the positive homogeneity is replaced by the weaker inequalities assumed in (N2). Even then, if $(\lambda, z, u) \in Q$ and $0 < \lambda_i \leq 1$, it follows that $\frac{z_i, u_i}{\lambda_i} \in R_i(1)$ since the inequalities are satisfied in the same manner as:

$$t_i^{j_i}(z_i, u_i) \geq \lambda_i \text{ and } \lambda_i t_i^{j_i}\left(\frac{z_i, u_i}{\lambda_i}\right) \geq t_i^{j_i}(z_i, u_i) \Rightarrow \lambda_i t_i^{j_i}\left(\frac{z_i, u_i}{\lambda_i}\right) \geq t_i^{j_i}(z_i, u_i) \geq \lambda_i \Rightarrow t_i^{j_i}\left(\frac{z_i, u_i}{\lambda_i}\right) \geq 1.$$

Clearly, $\mathrm{cl\,conv}\left(\bigcup_{i=1}^n S_i\right) \subseteq \mathrm{cl\,conv}(S)$ and we have assumed that $S \subseteq \mathrm{proj}_z X$. Observe that $\mathrm{cl\,conv}(S) \subseteq \mathrm{cl\,conv}(\mathrm{proj}_z X) \subseteq \mathrm{cl\,conv}\left(\bigcup_{i=1}^n S_i\right) \subseteq \mathrm{cl\,conv}(S)$ and, therefore, equality holds throughout. $\square$

Observe that Assumptions (N1) and (N2) are less restrictive than (A3) in Theorem 2.1 since $\mathrm{proj}_z A_i$ may be a nonconvex subset of $\mathrm{conv}(S_i)$ and the positive homogeneity is relaxed. Here, it is not necessary to use $t_i^{j_i}(z_i, u_i)$, $v_i^{k_i}(z_i, u_i)$ and $w_i^{l_i}(z_i, u_i)$ as the underestimators in Assumption (N2). Rather, any function of $(z_i, u_i)$ that underestimates $\lambda_i t_i^{j_i}\left(\frac{z_i, u_i}{\lambda_i}\right)$, $\lambda_i v_i^{k_i}\left(\frac{z_i, u_i}{\lambda_i}\right)$, and $\lambda_i w_i^{l_i}\left(\frac{z_i, u_i}{\lambda_i}\right)$ for all $\lambda_i \in (0, 1]$ suffices. As long as the set $C_i$ defined using these functions inner-approximates the recession cone of $\mathrm{cl\,conv}(S)$, a suitable set $X$ can be derived by projecting out the $\lambda$ variables and Assumption (N3) can be posed in terms of this set. Instead of exploring this extension further, we will retain in the remainder of this paper that $t_i^{j_i}(z_i, u_i)$, $v_i^{k_i}(z_i, u_i)$, and $w_i^{l_i}(z_i, u_i)$ are themselves the underestimating functions since it keeps the notation simpler while still conveys the main ideas.

# 3 Nonlinear Valid Inequalities for Polynomial Covering Sets

In this section, we apply the results of Section 2 to develop valid inequalities and convex hull representations for a variety of polynomial covering sets. Then, we consider the bilinear covering sets in greater detail. In particular, we find a convex relaxation of the bilinear covering set when the variables are restricted to be in a hypercube that is at least as tight as the standard factorable relaxation, and is, in fact, strictly tighter in many instances. The results in this section give applications of the theoretical framework built in Section 2 and provide many techniques to build better relaxations for nonlinear programs.

We will show that the conditions of Theorem 2.14 are satisfied for large classes of polynomial covers. In fact, we will allow the powers of the variables to be any real numbers larger than or equal to one. Then, we will apply Theorem 2.1 to construct the convex hull representation for the polynomial covering set over the non-negative orthant. To construct the inequality description of the convex hull, we will need convex hull descriptions of the orthogonally disjunctive sets, which will be defined by a single inequality. This inequality is an upper-level set of a function that generalizes the geometric mean. We show that the function in consideration is concave and, therefore, its upper-level set is convex. The concavity of this function does not seem to have been studied in the literature. We first provide a proof of this result, for it may have other applications.

**Theorem 3.1.** *Consider $f(x) = \frac{\prod_{i=1}^n x_i^{a_i}}{(r+\sum_{i=1}^n b_i x_i)^a}$, where $a \geq 0$, $r > 0$, $\sum_{i=1}^n a_i \leq 1$, and for each $i \in \{1, \ldots, n\}$, $a_i$ and $b_i$ are non-negative. Let $I = N \cap \{i \mid b_i > 0\}$ and assume that $a \leq \min\{a_i \mid i \in I\}$. Then, $f(x)$ is concave over $\mathbb{R}_+^n$.*

*Proof.* We may assume that $|I| > 0$, otherwise the result follows directly from Cauchy-Schwarz inequality (see discussion following Theorem 4.5 in [22]). We may further restrict ourselves to the case where, for all $i \in I$, $a_i$ equals $a$. Otherwise, consider

$$g(x, y) = \prod_{i \in N \setminus I} x_i^{a_i} \prod_{i \in I}^{n} y_i^{a_i - a} \frac{\prod_{i \in I} x_i^a}{(r + \sum_{i \in I} b_i x_i)^a}.$$

Observe that $g(x, y)$ satisfies our assumption, and if $g(x, y)$ is concave, then so is $f(x)$ since $f(x) = g(x, x)$. Therefore, we may assume that $f(x) = \frac{\prod_{i=1}^n x_i^a}{(r+\sum_{i \in I} b_i x_i)^a} \prod_{i \in N \setminus I} x_i^{a_i}$. We can also assume that $I = N$ and $a = \frac{1}{|I|} = \frac{1}{n}$. Otherwise, partition the variables into $x_I$ and $x_{N\setminus I}$. Then, consider the function $g(x_I) = \frac{\prod_{i\in I} x_i^{\frac{1}{|I|}}}{(r+\sum_{i\in I} b_i x_i)^{\frac{1}{|I|}}}$. Let $h(y, x_{N\setminus I}) = y^{|I|a} \prod_{i\in N\setminus I} x_i^{a_i}$. Then, $f(x)$ equals $h(g(x_I), x_{N\setminus I})$. If $g(x_I)$, a function that satisfies our assumptions, is concave, then $f(x)$ is concave since $h$ is jointly concave and isotonic (does not decrease when the argument increases) in $(y, x_{N\setminus I})$. The verification, although standard, is included here for completeness (see for example Theorem 5.1 in [22] for a similar argument in one dimension). Given $x^a$ and $x^b$ belonging to $\mathbb{R}^n_+$ and an arbitrary $\lambda \in [0, 1]$:

$$
\begin{aligned}
&f(\lambda x^a + (1-\lambda)x^b) = h\big(g(\lambda x_I^a + (1-\lambda)x_I^b), \lambda x^a_{N\setminus I} + (1-\lambda)x^b_{N\setminus I}\big)\\
&\geq h\big(\lambda g(x_I^a) + (1-\lambda)g(x_I^b), \lambda x^a_{N\setminus I} + (1-\lambda)x^b_{N\setminus I}\big) = h\big(\lambda(g(x_I^a), x^a_{N\setminus I}) + (1-\lambda)(g(x_I^b), x^b_{N\setminus I})\big)\\
&\geq \lambda h\big(g(x_I^a), x^a_{N\setminus I}\big) + (1-\lambda)h\big(g(x_I^b), x^b_{N\setminus I}\big) = \lambda f(x^a) + (1-\lambda)f(x^b),
\end{aligned}
$$

where the first inequality follows since $h$ is non-decreasing in $y$ and $g(x_I)$ is concave; and the second inequality since $h(y, x_{N\setminus I})$ is concave. Therefore, we only need to prove that, for all $n \in N$, $g(n, x) = \left(\frac{\prod_{i=1}^n x_i}{r+\sum_{i=1}^n b_i x_i}\right)^{\frac{1}{n}}$ is concave in $x$ over $\mathbb{R}^n_+$, assuming that $r > 0$ and $b_i \geq 0$. We do this by induction on $n$. For the basis of induction, we observe that $g(1, x) = \frac{x}{r+b_1 x}$ is concave since $\frac{\partial^2 g(1,x)}{\partial x^2} = \frac{-2b_1 r}{(r+b_1 x)^3} \leq 0$. Also, note that the right scalar multiplication $yg\left(1, \frac{x}{y}\right)$ is concave in $(x, y)$ as long as $y > 0$ (see discussion on page 35 of Rockafellar [22]). In particular, it follows by setting $r = 1$ that $u(b, x, y) := \frac{xy}{bx+y}$ is concave in $(x, y)$ as long as $b \geq 0$, $x \geq 0$, and $y > 0$. This function will be used in the inductive step. For the inductive step, we assume that $g(n, x)$ is concave over $\mathbb{R}^n_+$ when $r > 0$ and $b_i \geq 0$, and prove that $g(n+1, x)$ is concave over $\mathbb{R}^{n+1}_+$ under the same conditions. Observe that:

$$
\begin{aligned}
g(n+1, x) &= \left(\frac{\prod_{i=1}^{n+1} x_i}{r + \sum_{i=1}^{n+1} b_i x_i}\right)^{\frac{1}{n+1}} = \left(\frac{\prod_{i=1}^{n} x_i^{\frac{1}{n}}}{\left(r + \sum_{i=1}^{n} b_i x_i\right)^{\frac{1}{n}}}\right)^{\frac{n}{n+1}} \left(\frac{(r + \sum_{i=1}^{n} b_i x_i) x_{n+1}}{r + \sum_{i=1}^{n+1} b_i x_i}\right)^{\frac{1}{n+1}}\\
&= g(n, x)^{\frac{n}{n+1}} u\left(b_{n+1}, x_{n+1}, r + \sum_{i=1}^{n} b_i x_i\right)^{\frac{1}{n+1}}.
\end{aligned}
$$

Now, $u\big(b_{n+1}, x_{n+1}, r + \sum_{i=1}^n b_i x_i\big)$ is concave in $\big(x_{n+1}, r + \sum_{i=1}^n b_i x_i\big)$ since $r > 0$, $b_{n+1} \geq 0$, and $x_{n+1} \geq 0$. Further, $g(n, x)$ is concave by our induction hypothesis. Since $r + \sum_{i=1}^n b_i x_i > 0$ and $x_{n+1} \geq 0$, $g(n+1, x)$ can be expressed as a composition of a concave function that is isotone, *i.e.*, $x^{\frac{n}{n+1}} y^{\frac{1}{n}}$, with concave functions, *i.e.*, $u(\cdot)$ and $g(n, x)$. In other words, $g(n+1, x)$ is concave over $\mathbb{R}^n_+$ and the result is proven. □

The concavity of the geometric mean is typically proven in one of two ways. The first proof technique exploits the convexity of $x^2$ or, equivalently, Cauchy-Schwarz inequality to show that the Hessian of the geometric mean is negative semidefinite. The second proof technique uses the AM-GM inequality and the concavity of the $\log(\cdot)$ function to establish Jensen's inequality. However, the above proofs do not generalize to show the concavity of the function in Theorem 3.1 because the denominator destroys the multiplicative nature of the geometric mean. We mentioned in the introduction that the traditional approach of relaxing the left-hand-side of the polynomial covering set by a concave overestimator is often difficult to carry out. The reason is that it is not clear how to express the hypograph of a polynomial using a convex inequality. Consider for example the inequality $p(x) = \prod_{i=1}^n x_i^{\frac{a_i}{a}} - \sum_{i=1}^n b_i x_i \geq r$. Observe that, using Theorem 3.1, there is a simple way to write this inequality as a convex inequality. However, convexity is not preserved if $r$ is allowed to vary in this inequality. In fact, the function in Theorem 3.1 is strictly convex as a function of $r$ for a given $x$. Therefore, relaxing the hypograph of $p(x)$ to a convex set is not straightforward, whereas the upper-level set of $p(x)$ is easily expressed as a single convex inequality.

Incidentally, the convexity of the function in Theorem 3.1 as a function of $r$ also shows that if one replaces the linear term in the denominator with another variable, the function is rendered nonconvex. Standard factorable relaxation schemes carry out such a transformation whenever they encounter a function of this type and, as a result, may yield weak relaxations. Therefore, the function in Theorem 3.1 should be added to list of primitive concave functions if nonlinear relaxations are being built using factorable programming techniques. It was shown in [31] that if the concavity of a function follows from the composition rules then, to achieve the same relaxation quality, overestimating the concave function directly may require exponentially more linear overestimators compared to overestimating the functions involved in the composition. However, it is clear that the concavity proved in Theorem 3.1 does not directly follow from composition rules. Rather, the inductive proof of Theorem 3.1 provides a recipe for introducing intermediate variables in a manner that the concavity of the function will follow using composition rules. Such a recipe, if followed during the construction of polyhedral outer-approximation, will yield a tighter overestimator with fewer inequalities than overestimating the function directly.

The function $\frac{\prod_{i=1}^n x_i^{a_i}}{(r+\sum_{i=1}^n b_i x_i)^a}$, however, unlike the geometric mean, is not positively-homogenous. In Theorem 2.1, positive homogeneity of the function plays a critical role in finding the convex hull of the disjunctive set, as was discussed in Example 2.5. Unfortunately, this means that it will not in general be possible to write the convex hull inequality without introducing new variables. However, when the $b_i$s are zero, we can rewrite the upper-level sets of the above function using a positively-homogenous function, and, therefore, express the convex hull inequality in the space of the original variables. Otherwise, we can still use the right-scalar multiplication to recover positive homogeneity by introducing an additional variable for each of the disjunctive sets. We demonstrate this technique in the next corollary. Later, we discuss techniques that allow to recover positive-homogeneity without introducing additional variables.

**Corollary 3.2.** *Consider*

$$f(x) = \begin{cases} \dfrac{\prod_{i=0}^n x_i^{a_i}}{\left(\sum_{i=0}^n b_i x_i\right)^a} & \exists j, \text{ such that } b_j x_j \neq 0, \\ 0 & \text{otherwise;} \end{cases}$$

*where $a \geq 0$, $\sum_{i=0}^n a_i = 1 + a$, and, for each $i \in \{0, \ldots, n\}$, $a_i$ and $b_i$ are non-negative. Further, assume that there exists $j \in \{0, \ldots, n\}$ such that $b_j > 0$. Define $I = N \cap \{i \mid b_i > 0\}$ and assume that $a \leq \min\{a_i \mid i \in I\}$. Then, $f(x)$ is concave and positively-homogenous over $\mathbb{R}_+^{n+1}$.*

*Proof.* Assume without loss of generality that $b_0 > 0$, *i.e.*, $0 \in I$. By substituting $a_0 = 1+a-\sum_{i=1}^n a_i$, the conditions in the statement of the Corollary on $a_0, \ldots, a_n$ can be restated in terms of $a_1, \ldots, a_n$ as: $a \geq 0$, $\sum_{i=1}^n a_i \leq 1$, and $a_i \geq a$ for all $i \in I$, while $a_i \geq 0$ for $i \notin I$. Then, it follows by considering the right-scalar multiplication of $\frac{\prod_{i=1}^n x_i^{a_i}}{\left(b_0+\sum_{i=1}^n b_i x_i\right)^a}$ (see discussion before Theorem 5.5 in [22]) that $f(x)$ is concave and positively-homogenous over $\mathbb{R}_+^{n+1}$. $\square$

Theorem 3.1 and its Corollary 3.2 allow us to build valid inequalities for polynomial covering sets. As mentioned before, the main idea is that as long as the function grows superlinearly, the convex extension property can be inferred from Theorem 2.14 and then Theorem 2.1 provides a recipe for constructing the convex hull from its restriction to the orthogonal subspaces spanned by certain subsets of the original variables. A representation of the convex hull in the orthogonal subspaces is easily obtained using Corollary 3.2. The following theorem formalizes this construction yielding valid inequalities for the polynomial covering sets.

**Theorem 3.3.** *Consider the set $S = \left\{ z \in \mathbb{R}_+^{\sum_{i=1}^n d_i} \;\middle|\; s(z) \geq r + \langle b, z \rangle \right\}$, where $b \geq 0$ and $r > 0$. Let $z_i = (z_{i1}, \ldots, z_{id_i})$ and $s(z) = \sum_{i=1}^n s_i(z_i)$, where $s_i(z_i) = \sum_{k=1}^{K_i} c_{ik} \prod_{j=1}^{d_i} z_{ij}^{a_{ijk}}$. Assume that the coefficients are non-negative, i.e., $c_{ik} > 0$. Further, assume that for all $(i,k)$, $\sum_{j=1}^{d_i} a_{ijk} > 1$. Let $m_i = \max_k \left\{ \sum_{j=1}^{d_i} a_{ijk} \;\middle|\; k = 1, \ldots, K_i \right\}$. Then, the convex extension property (11) holds for $S$. If,* *in addition, $b_{ij} > 0$ implies that, for all $k$, $a_{ijk} > 1$, then the following convex inequality is valid for* $\operatorname{conv}(S)$*:*

$$\sum_{i=1}^{n}\sum_{k=1}^{K_i} \frac{c_{ik}^{\frac{1}{m_i}} \prod_{j=1}^{d_i} z_{ij}^{\frac{a_{ijk}}{m_i}}}{(r+\sum_{j=1}^{d_i} b_{ij} z_{ij})^{\frac{1}{m_i}}} \geq 1. \tag{28}$$

*Define, by introducing variables, $z_{i0}$ for $i \in \{1,\dots,n\}$:*

$$\vartheta_k(z_{0i}, z_i) = \begin{cases} z_{i0}^{1+\frac{1}{m_i}-\sum_{j=1}^{d_i}\frac{a_{ijk}}{m_i}} \dfrac{c_{ik}^{\frac{1}{m_i}} \prod_{j=1}^{d_i} z_{ij}^{\frac{a_{ijk}}{m_i}}}{(r z_{i0}+\sum_{j=1}^{d_i} b_{ij} z_{ij})^{\frac{1}{m_i}}} & \text{if } z_{i0} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

*Then, the following provides a convex outer-approximation of the polynomial covering set that is at least as tight as (28):*

$$\operatorname{conv}(S) \subseteq \left\{ z \in \mathbb{R}_+^{\sum_{i=1}^n d_i} \;\middle|\; \sum_{i=1}^{n}\sum_{k=1}^{K_i} \vartheta_k\big(z_{0i}, z_i\big) \geq 1, \sum_{i\in I} z_{i0} = 1, \forall i\; z_{i0} \geq 0 \right\}. \tag{29}$$

*In addition, if $b_i = 0$, then it is not necessary to introduce $z_{i0}$. In particular, if $b$ itself is zero, i.e., for all $i$, $b_i$ equals zero, then the (closure) convex hull of $S$ is given as follows:*

$$\operatorname{conv}(S) = \operatorname{cl}\operatorname{conv}(S) = \left\{ z \in \mathbb{R}_+^{\sum_{i=1}^n d_i} \;\middle|\; \sum_{i=1}^{n} \left(\frac{c_{i1}\prod_{j=1}^{d_i} z_{ij}^{a_{ij1}}}{r}\right)^{\frac{1}{\sum_{j=1}^{d_i} a_{ij1}}} \geq 1 \right\}. \tag{30}$$

*Proof.* Let $g(z) = s(z) - \langle b, z\rangle$ and $g_i(z_i) = s_i(z_i) - \langle b_i, z_i\rangle$. We first use Theorem 2.14 to verify that the convex extension property holds. Let $h_i(z_i) = g_i(z_i)$ and let $f$ be the summation operator. Then, by construction, Assumption (S1), (S2), and (S3) are satisfied. Further, by definition, $h_i(0) = 0$ and for $\lambda \in (0,1]$,

$$\lambda h_i\left(\frac{z_i}{\lambda}\right) = \frac{1}{\lambda^{\sum_{k=1}^{K_i} a_{ijk}-1}} \sum_{k=1}^{K_i} c_{ik} \prod_{j=1}^{d_i} z_{ij}^{a_{ijk}} - \lambda\left\langle b_i, \frac{z_i}{\lambda}\right\rangle \geq \sum_{k=1}^{K_i} c_{ik} \prod_{j=1}^{d_i} z_{ij}^{a_{ijk}} - \langle b_i, z_i\rangle = h_i(z_i).$$

The first inequality follows since $\lambda \in (0,1]$, $\sum_{k=1}^{K_i} a_{ijk} > 1$, $z_{ij} \geq 0$, and $c_{ik} \geq 0$. Therefore, Assumption (S4) is satisfied. Since $m_i > 1$, it follows that, for any $z_i' \in \mathbb{R}_+^{d_i}$, there exists a sufficiently large $\lambda'$ such that, whenever $\lambda > \lambda'$, $s_i(\lambda z_i') - \langle b_i, \lambda z_i'\rangle \geq r$. Therefore, $z_i'$ is a recession direction of $0^+(\operatorname{cl}\operatorname{conv} S_i)$ and it follows by Theorem 2.14 that the convex extension property holds for $S$. Since $m_i > 1$, it follows that $(\cdot)^{\frac{1}{m_i}}$ is subadditive, and, therefore,

$$\operatorname{conv}(S_i) \subseteq O_i' := \left\{ (0, z_i, 0) \;\middle|\; \sum_{k=1}^{K_i} \frac{c_{ik}^{\frac{1}{m_i}} \prod_{j=1}^{d_i} z_{ij}^{\frac{a_{ijk}}{m_i}}}{(r+\sum_{j=1}^{d_i} b_{ij} z_{ij})^{\frac{1}{m_i}}} \geq 1 \right\}. \tag{31}$$

If $b_{ij} > 0$ implies that, for all $(j,k)$, $a_{ijk}$ is larger than one then, by Theorem 3.1, it follows that $O_i'$ is a convex set. This outer-approximation of $\operatorname{conv}(S_i)$ is not expressed in terms of positively-homogenous functions. However, it follows from Proposition 2.7 that (28) is still a valid inequality for the polynomial covering set. On the other hand, the inequality can be homogenized as follows:

$$\operatorname{conv}(S_i) \subseteq O_i = \left\{ (0, z_i, 0) \;\middle|\; \sum_{k=1}^{K_i} \vartheta_k(z_{i0}, z_i) \geq 1, z_{i0} = 1 \right\}.$$

Now, instead of computing $\operatorname{conv}\left(\bigcup_{i=1}^n S_i\right)$, we compute its outer-approximation, $\operatorname{conv}\left(\bigcup_{i=1}^n O_i\right)$. By Corollary 3.2, the left-hand-side in the defining inequality of $O_i$ is concave. Further, for $\lambda$

sufficiently large, $z_i = (\lambda, \dots, \lambda) \in O_i$. Therefore, by Proposition 2.6, it follows that Assumption (A4) is satisfied. Then, it follows that $\operatorname{conv}(S_i)$ is outer-approximated by the set on the right-hand-side of (29). It follows by Theorem 3.1 and an argument similar to (10) that (29) is at least as tight as (28). The inclusion is tight in (29) when, for $k \neq 1$, $c_{ik} = 0$ because the subadditivity of $(\cdot)^{\frac{1}{m_i}}$ is not exploited and, therefore, $\operatorname{conv}(S_i) = O_i$. If, for $k \neq 1$, $c_{ik} = 0$ and, in addition, $b_i = 0$, then the defining inequality of (31) is itself positively-homogenous and, consequently, $z_{i0}$ does not need to be introduced. If $b = 0$, the set $X$ in Theorem 2.1 corresponds to the right-hand-side of (30). Since the left-hand-side of the defining inequality of the set (30) is continuous, the set is closed. The closedness of the set also follows from Theorem 2.1, $C_i = \mathbb{R}^{d_i}_+ = 0^+(\operatorname{cl}\operatorname{conv} S_i)$ and that $A_i$s are closed. Let $S'_i = \left\{ z \in \mathbb{R}_+^{\sum_{i=1}^n d_i} \;\middle|\; s_i(z_i) \geq r \right\}$. Note that if $b = 0$ then $S'_i$ is closed convex subset of $S$. Then, by the last statement of Theorem 2.14, $\operatorname{conv}(S)$ is closed. □

The valid inequalities for the polynomial covering sets that are derived in Theorem 3.3 are surprisingly simple to construct. They are also widely applicable in factorable programming. In particular, consider the process of relaxing an inequality using recursive decomposition; see [18] and [23]. Assume that in an intermediate step, we encounter a polynomial inequality where the dominant terms have positive coefficients. Then, Theorem 3.3 can typically be used to relax the inequality by underestimating the remaining terms using linear underestimators following the current practice. Another, not so apparent, use of these inequalities is in objective function cuts. If the optimization involves maximizing a polynomial function and an upper-bound is found during the course of the algorithm then one may wish to restrict the search to better solutions. Theorem 3.3 provides a convex outer-approximation for such a restriction. The following example highlights the important procedural steps in constructing relaxations using Theorem 3.3.

**Example 3.4.** *Consider the following set: $S = \left\{(x, y, u, v, z) \in \mathbb{R}^5_+ \mid xy + u^2v + z \geq r\right\}$ for a given $r > 0$. The defining inequality satisfies the most stringent assumptions of Theorem 3.3. In particular, note that the left-hand-side is a sum of monomials, each with positive coefficients, i.e., for all $k \neq 1$, $c_{ik} = 0$. Further, there are no linear terms on the right-hand-side, i.e., for all $i$, $b_i = 0$. Therefore,* $\operatorname{conv}(S)$ *is given by (30). In particular, one starts with the convex hull descriptions of $xy \geq r$, $u^2v \geq r$, and $z \geq r$ over the non-negative orthant, which are defined by the inequalities $\left(\frac{xy}{r}\right)^{\frac{1}{2}} \geq 1$, $\left(\frac{u^2v}{r}\right)^{\frac{1}{3}} \geq 1$, and $\frac{z}{r} \geq 1$ respectively. Observe that the functions used in the convex-hull representations are scalar multiples of the geometric mean, and are, therefore, positively-homogenous and concave. The closure convex hull of $S$ is then defined by the upper-level set of the summation of these functions as follows:*

$$\operatorname{conv}(S) = \left\{ (x, y, u, v, z) \in \mathbb{R}^5_+ \;\middle|\; \left(\frac{xy}{r}\right)^{\frac{1}{2}} + \left(\frac{u^2v}{r}\right)^{\frac{1}{3}} + \frac{z}{r} \geq 1 \right\}. \qquad \square$$

It was discussed in Proposition 2.7 that positive-homogeneity of the constituent functions is not required to prove the validity of $X$. However, as in Example 2.8, if one can find a description of $\operatorname{conv}(S_i)$ that uses positively-homogenous functions then one can apply Theorem 2.1 to identify the convex hull of the orthogonal disjunctions, thus deriving a superior relaxation. The proof of Corollary 3.2 demonstrates a general technique for constructing such a homogenization using the right-scalar multiplication. However, this process suffers from the drawback that it introduces new variables in the relaxation. Instead, it may be possible to find a separating hyperplane without increasing the problem dimension and, thereby, circumvent the need to introduce new variables. Consider, for simplicity, the case of Theorem 2.1, where $A_i$ is not an extended formulation, *i.e.*, it does not need the additional $u_i$ variables. The case with the $u_i$ variables can be handled similarly. Now, consider a point $z'$ that does not belong to $\operatorname{cl}\operatorname{conv}(S)$. If it is possible to find, for all $i$, a $j'_i \in \operatorname{argmin}_j \left\{ t_i^j(z'_i) \;\middle|\; j = 1, \dots, J_i \right\}$, a $k'_i \in \operatorname{argmin}_k \left\{ v_i^k(z'_i) \mid k = 1, \dots, K_i \right\}$ and an $l'_i \in \operatorname{argmin}_l \left\{ w_i^l(z'_i) \mid l = 1, \dots, L_i \right\}$ then using the closed-form expression of X in (2), one can identify an inequality that separates $z'$ from $X$. For example, if an inequality of the form

$\sum_{i\in N} t_i^{j_i}(z_i) \geq 1$ violates $z_i'$, *i.e.*, $\sum_{i\in N} t_i^{j_i}(z_i') < 1$, then $\sum_{i\in N} t_i^{j_i'}(z_i') < 1$ as well, since, by the definition of $j_i'$, $t_i^{j_i'}(z_i') \leq t_i^{j_i}(z_i')$ for all $i$.

We now discuss another technique that can be used to find representations of the convex hull of each $S_i$ that uses positively-homogenous functions but does not require additional variables. The main idea is that one can homogenize the inequality using an extra variable and then maximize the resulting function over the introduced variable to derive a positively-homogenous function describing the set. We illustrate this idea by deriving a positively-homogenous function that describes the following bilinear covering set:

$$Q = \left\{(x, y) \in \mathbb{R}^2_+ \mid axy + bx + cy \geq r\right\},$$

where $a$, $b$, and $c$ are assumed to be non-negative. We assume without loss of generality that $r > 0$. Otherwise, $Q = \mathbb{R}^2_+$. We may also assume without loss of generality that $c \geq b$ and, consequently, assume that at least one of $a$ and $c$ is strictly positive. Then, for any feasible $(x, y)$, it follows that $ax + c > 0$. Therefore, $Q = \left\{(x, y) \in \mathbb{R}^2_+ \mid y \geq \frac{r-bx}{ax+c}\right\}$. First, we verify that the inequality is convex. Let $f(x) = \frac{r-bx}{ax+c}$. Since

$$\frac{\partial^2 f}{\partial x^2} = \frac{2a(bc + ar)}{(ax + c)^3}$$

is nonnegative if $x \geq 0$, $Q$ is expressed as the intersection of the epigraph of a convex function with the non-negative orthant. Therefore, $Q$ is convex. Also, note that the defining inequality of $Q$ is not positively-homogenous. We show how the above inequality can be homogenized without introducing any new variables in the formulation. To carry out this transformation, first homogenize the defining inequality, $axy + bx + cy \geq r$, using an additional variable $h$, that is restricted to be positive. This is accomplished by rewriting the defining inequality of $Q$ as $\frac{axy}{h} + bx + cy \geq rh$. Since $h$ is positive, we can multiply throughout by $h$, and express the above inequality as: $axy + bxh + cyh \geq rh^2$. This is a positively-homogenous inequality which defines $Q$ as long as $h$ is positive. Therefore, $Q$ can now be described by the inequalities:

$$axy + bxh + cyh \geq rh^2 \text{ and } h \geq 1.$$

In order for $(x, y, h)$ to satisfy the first inequality above, $h$ must be such that:

$$\frac{bx + cy - \sqrt{(bx + cy)^2 + 4arxy}}{2r} \leq h \leq \frac{bx + cy + \sqrt{(bx + cy)^2 + 4arxy}}{2r}.$$

It can be easily verified that the functions bounding $h$ are positively-homogenous. In fact, when the bounding functions on $h$ are derived from a positively-homogenous constraint, they must, in general, be positively-homogenous. This can be inferred because for each $(x, y, h)$ that satisfies a positively-homogenous constraint and an arbitrary $\lambda > 0$, it must be that $(\lambda x, \lambda y, \lambda h)$ satisfies the constraint as well. The lower bounding function is nonpositive. Therefore, the set $Q$ can be rewritten as:

$$\eta(x, y) = \frac{1}{2}\left(bx + cy + \sqrt{(bx + cy)^2 + 4arxy}\right) \geq r. \tag{32}$$

We have thus expressed $Q$ as the upper-level set of a positively-homogenous function without introducing any new variables. In fact, since Proposition 2.6 asserts that a positively-homogenous function whose upper-level set is convex, is concave, it follows from the convexity of $Q$ that $\eta(x, y)$ must be concave over the non-negative quadrant. In other words, we have established the following result.

**Proposition 3.5.** *Let $Q = \left\{(x, y) \in \mathbb{R}^2_+ \mid axy + bx + cy \geq r\right\}$, where $a$, $b$, $c$ are non-negative, and $r$ is strictly positive. Then, $Q$ has a convex description (upper level set of a concave function) that uses positively-homogenous functions. In particular, $Q = \left\{(x, y) \in \mathbb{R}^2_+ \mid \eta(x, y) \geq r\right\}$, where $\eta(x, y)$ is as defined in (32).* □

We now consider a more general bilinear covering set that reduces to $Q$ when restricted to any one of $n$ orthogonal subspaces. As long as the convex extension property holds, since Proposition 3.5 provides the inequality description for the convex hull in each of the orthogonal subspaces, we can use Theorem 2.1 to find the convex hull description of the bilinear covering set over the non-negative orthant. We formalize this argument in the following proposition.

**Proposition 3.6.** *Consider a bilinear covering set:*

$$B^R = \left\{ (x,y) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ \;\middle|\; \sum_{i=1}^n (a_i x_i y_i + b_i x_i + c_i y_i) \geq r \right\}.$$

*where, for each $i \in \{1,\ldots,n\}$, $a_i$, $b_i$ and $c_i$ are non-negative and $r$ is strictly positive. Let*

$$\eta_i(x_i,y_i) = \frac{1}{2}\left( b_i x_i + c_i y_i + \sqrt{(b_i x_i + c_i y_i)^2 + 4 a_i r x_i y_i} \right).$$

*Then,*

$$\operatorname{conv}\bigl(B^R\bigr) = X = \left\{ (x,y) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ \;\middle|\; \sum_{i=1}^n \eta_i(x_i,y_i) \geq r \right\}. \tag{33}$$

*Proof.* We may assume without loss of generality that, for each $i$, at least one of $a_i$, $b_i$, or $c_i$ is positive. First, we use Theorem 2.14 to show that the convex extension property (11) holds for $B^R$. Let $z_i = (x_i, y_i)$, $g_i(x_i,y_i) = h_i(x_i,y_i) = a_i x_i y_i + b_i x_i + c_i y_i$, $g(z) = \sum_{i=1}^n g_i(z_i)$, and define $f$ to be the summation operator. Then, by definition, (S1), (S2), and (S3) are satisfied. Clearly, $h_i(0) = 0$ and for $0 < \lambda \leq 1$,

$$\lambda h_i\left(\frac{z_i}{\lambda}\right) = \frac{a_i x_i y_i}{\lambda} + b_i x_i + c_i y_i \geq h_i(z_i).$$

Therefore, Assumption (S4) is satisfied as well. Observe that, if $(x_i', y_i') \geq 0$, then $h_i(x_i + x_i', y_i + y_i') \geq h_i(x_i, y_i)$. Therefore, if $z_i' = (x_i', y_i') \geq 0$ then $(0, z_i', 0) \in 0^+\bigl(\operatorname{cl}\operatorname{conv} G_i\bigr)$ and, consequently, Assumption (S5) is satisfied. Therefore, the convex extension property holds for $S$. By Proposition 3.5, it follows that the convex hull of $B_i^R = \{(0, x_i, y_i, 0) \mid a_i x_i y_i + b_i x_i + c_i y_i \geq r\}$ is defined by $\eta_i(x_i, y_i) \geq r$. Observe that $\eta_i(x_i,y_i)$ is a positively-homogenous function. Therefore, Assumption (A3) is satisfied. Finally, $\eta_i(x_i,y_i)$ is concave by Proposition 2.6 and since for sufficiently large $z_i$, $h_i(x_i,y_i) \geq r$, it follows by Proposition 2.6 that Assumption (A4) is satisfied as well. Then, by Theorem 2.1 and the discussion following Definition 2.9, the set $X$ in (33) is $\operatorname{cl}\operatorname{conv}(B^R)$. Let

$$B_i^{R'} = \left\{ (x,y) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ \;\middle|\; a_i x_i y_i + b_i x_i + c_i y_i \geq r \right\}.$$

Then, $B_i^{R'} \subseteq B^R$, and, therefore, by Theorem 2.14, $\operatorname{conv}\left(B^R\right)$ is closed. ∎

Consider the special case of Proposition 3.6 where $b_i = c_i = 0$. In this case, the convex hull inequality takes the following simple form: $\sum_{i=1}^n \sqrt{a_i x_i y_i} \geq \sqrt{r}$. This representation can also be derived using Theorem 3.3. We include here a direct short proof of this result. First, the validity of the inequality can be verified using the following argument:

$$\sum_{i=1}^n \sqrt{a_i x_i y_i} \geq \sqrt{\sum_{i=1}^n a_i x_i y_i} \geq \sqrt{r},$$

where the first inequality follows by the subadditivity of square-root over non-negative real numbers. Second, by Example 2.5, the above inequality defines the closure convex hull of the disjunctive union of $\{(x_i, y_i) \mid a_i x_i y_i \geq r\}$ over the non-negative orthant and, therefore, it must also be the closure convex hull of $\sum_{i=1}^n a_i x_i y_i \geq r$ over the same set. Note that in the argument, we did not employ Theorem 2.14. Instead, we replaced it with a proof that the convex hull of the disjunctive union of orthogonal restrictions of the set includes the original set. This illustrates a different technique, similar to the proof technique of Proposition 2.15, that may sometimes be useful in establishing the convex extension property.

However, the above technique for establishing validity fails for another special case of Proposition 3.6, where the defining inequality is $ax_1y_1 + bx_2 \geq r$ with $a > 0$, $b > 0$ and $r > 0$. A simpler variant of this set was mentioned in the introduction of the paper. Its convex hull over the non-negative orthant is defined by

$$\sqrt{\frac{ax_1y_1}{r}} + \frac{bx_2}{r} \geq 1. \tag{34}$$

Once again, this convex hull representation could also be derived as a consequence of Theorem 3.3. Note that the right-hand-side $r$ participates differently with different subsets of variables in this convex hull inequality. One could use subadditivity of the square-root function to instead derive the following valid inequality

$$\sqrt{\frac{ax_1y_1}{r}} + \sqrt{\frac{bx_2}{r}} \geq 1. \tag{35}$$

However, as expected, (35) is not as tight as (34). This can be seen by considering a point $(x_1, y_1, x_2)$ that is feasible to (34). If $\frac{bx_2}{r} \geq 1$, it follows that $\sqrt{\frac{bx_2}{r}} \geq 1$. Otherwise $\frac{bx_2}{r} < 1$, in which case

$$\sqrt{\frac{ax_1y_1}{r}} + \sqrt{\frac{bx_2}{r}} > \sqrt{\frac{ax_1y_1}{r}} + \frac{bx_2}{r} \geq 1.$$

Therefore, $(x_1, y_1, x_2)$ is feasible to (35) as well. In this case, the subadditivity of the square-root function is not sufficient to prove the convex extension property, and, thus, cannot replace Theorem 2.14. Without realizing the convex extension property *a priori*, even the form of the inequality (34) is not obvious. The key to deriving this convex hull is thus to realize that the convex hull is formed by restricting attention to orthogonal subspaces. The first subspace spans the $(x_1, y_1)$ variables and the second subspace spans $x_2$. Then, Theorem 2.1 quickly reveals the structure of the convex hull. Note that for this example, $\sqrt{\frac{bx_2}{r}} \geq 1$ as well as $\frac{bx_2}{r} \geq 1$ define the convex hull of the set restricted to $(0, 0, x_2)$. However, as the insight from Theorem 2.1 suggests, it is preferable to choose the latter representation since it uses a positively-homogenous function.

Most of our discussion to this point has focused on developing convex hulls of inequalities over the non-negative orthant. The main purpose of developing convex relaxations is that linear functions can typically be optimized over them quickly. As a result, relaxations are used to derive quickly computable bounds on the optimal objective value of a global optimization problem. However, when using such relaxations in a branch-and-bound algorithm, the relaxations are iteratively constructed over smaller partitions of the search space, which are obtained by branching on the variables. And so, the nodes of the branch-and-bound tree are typically associated with variable bounds which can be exploited to derive tighter relaxations. In our discussion that follows Theorem 2.14, we alluded to the fact that if the variables are restricted to belong to a compact convex set, the relaxations derived using orthogonal disjunctions may be tightened by altering the defining inequality outside the region of interest. We apply this technique in the context of the bilinear covering set to derive its relaxation over a hypercube.

**Theorem 3.7.** *Let $\mathcal{B} = \prod_{i=1}^n [l_i, u_i] \times \prod_{i=1}^n [L_i, U_i]$, where, for each $i$, $l_i$ and $L_i$ are non-negative. Consider the following bilinear covering set:*

$$B^R_{\mathcal{B}} = \left\{ (x, y) \in \mathcal{B} \;\middle|\; \sum_{i=1}^n (a_i x_i y_i + b_i x_i + c_i y_i) \geq r \right\},$$

*where $r > 0$ and, for each $i$, $a_i$, $b_i$, and $c_i$ are non-negative. Let $b'_i = b_i + a_i L_i$, $c'_i = a_i l_i + c_i$, and $r' = r - \sum_{i=1}^n (a_i l_i L_i + b_i l_i + c_i L_i)$. If $r' \leq 0$ then $\operatorname{conv}(B^R) = \mathcal{B}$. Otherwise, define*

$$\tau_i(x_i, y_i) = \frac{1}{2}\left( b'_i(x_i - l_i) + c'_i(y_i - L_i) + \sqrt{(b'_i(x_i - l_i) + c'_i(y_i - L_i))^2 + 4a_i r'(x_i - l_i)(y_i - L_i)} \right).$$

*Then,*

$$\operatorname{conv}\bigl(B^R_{\mathcal{B}}\bigr) \subseteq \left\{ (x,y) \in \mathcal{B} \;\middle|\; \sum_{i=1}^{n} \min \left\{ \begin{array}{c} \tau_i(x_i,y_i) \\ \\ (a_i(U_i - L_i) + b'_i)(x_i - l_i) + c'_i(y_i - L_i) \\ \\ b'_i(x_i - l_i) + (a_i(u_i - l_i) + c'_i)(y_i - L_i) \end{array} \right\} \geq r' \right\}. \tag{36}$$

*Proof.* First note that $(x,y) = (l_1,\ldots,l_n,L_1,\ldots,L_n)$ is feasible if and only if $r' \leq 0$. If it is feasible then, clearly, $\mathcal{B} = B^R_{\mathcal{B}} = \operatorname{conv}(B^R_{\mathcal{B}})$. Therefore, we may now assume that setting $x$ and $y$ to their lower bounds does not yield a feasible solution. We assume without loss of generality that, for all $i$, $l_i$ and $L_i$ are zero. Otherwise, we rewrite the defining inequality of $B^R_{\mathcal{B}}$ using the transformed variables $x'_i = x_i - l_i$ and $y'_i = x_i - L_i$. This inequality then becomes:

$$\sum_{i=1}^{n} \bigl(a_i x'_i y'_i + (b_i + a_i L_i)x'_i + (a_i l_i + c_i)y'_i\bigr) \geq r - \sum_{i=1}^{n}(a_i l_i L_i + b_i l_i + c_i L_i).$$

Since, setting the variables to the lower bound does not yield a feasible solution, it follows that:

$$r' = r - \sum_{i=1}^{n}(a_i l_i L_i + b_i l_i + c_i L_i) > 0.$$

Similarly, one can define $b'_i = b_i + a_i L_i$, $c'_i = a_i l_i + c_i$, and transform the above inequality to:

$$\sum_{i=1}^{n} \bigl(a_i x'_i y'_i + b'_i x'_i + c'_i y'_i\bigr) \geq r',$$

which is in the same form as the defining inequality of $B^R_{\mathcal{B}}$. Defining $u'_i = u_i - l_i$ and $U'_i = U_i - L_i$, we note that $(x'_i, y'_i) \in \prod_{i=1}^{n}[0,u'_i] \times \prod_{i=1}^{n}[0,U'_i]$, which is a hypercube of the assumed type.

Now, we only need to verify the result for $\mathcal{B} = \prod_{i=1}^{n}[0,u_i] \times \prod_{i=1}^{n}[0,U_i]$. The general result follows by appropriately substituting $x'_i$, $y'_i$, $b'_i$, $c'_i$, $u'_i$, $U'_i$ and $r'$ for $x_i$, $y_i$, $b_i$, $c_i$, $u_i$, $U_i$, and $r$ respectively. Even though the natural way to apply Theorem 2.14 is to define $h_i(z_i) = a_i x_i y_i + b_i x_i + c_i y_i$, it may be observed that such an application does not yield the desired result. However, the discussion following Theorem 2.14 provides an improvement that we will exploit. If the variables are bounded, we may modify the inequality outside the bounds without changing the set. The discussion following Theorem 2.14 shows a way that this change can be made without compromising Assumption (S4), which is required for the application of Theorem 2.14. Following this discussion, we underestimate $p_i(x_i,y_i) = a_i x_i y_i + b_i x_i + c_i y_i$ as in (24). Then, let $g_i(x_i,y_i) = h_i(x_i,y_i) = \min(a_i x_i y_i + b_i x_i + c_i y_i, (a_i U_i + b_i)x_i + c_i y_i, b_i x_i (a_i u_i + c_i) y_i)$. We rewrite $B^R_{\mathcal{B}}$ as:

$$B^R_{\mathcal{B}} = \left\{ (x,y) \in \mathcal{B} \;\middle|\; \sum_{i=1}^{n} g_i\left(x_i, y_i\right) \geq r \right\}. \tag{37}$$

Now, we ignore the bounds and construct the closure convex hull of $\sum_{i=1}^{n} g_i\left(x_i,y_i\right) \geq r$ over $\mathbb{R}^{2n}_{+}$. Assumptions (S1), (S2), and (S3) follow by assuming $f$ to be the summation operator. Since $g_i(x_i,y_i) \geq 0$, it follows that $\mathbb{R}^2_+$ is the recession cone of $\{(x_i,y_i) \mid g_i(x_i,y_i) \geq r\}$. Therefore, Assumption (S5) holds. By construction, $h_i(x_i,y_i)$ satisfies Assumption (S4). However, in this context, it is easier to verify that $h_i(x_i,y_i)$ satisfies Assumption (S4) by observing that $h_i(x_i,y_i)$ is expressed as a minimum of three functions, each of which satisfies the assumption. Therefore, the convex extension property, (11), holds for

$$\mathcal{B}^R_{\mathcal{B}'} = \left\{ (x,y) \in \mathbb{R}^{2n}_{+} \;\middle|\; \sum_{i=1}^{n} g_i\left(x_i, y_i\right) \geq r \right\}. \tag{38}$$

Assumption (A1) is clearly satisfied by $\mathcal{B}^R_{\mathcal{B}'}$. Since, by Proposition 3.5, $\tau_i(x_i, y_i) \geq r$ if and only if $a_i x_i x_i + b_i x_i + c_i y_i \geq r$, it is easy to find a representation of $\mathcal{B}^R_{\mathcal{B}'}$ restricted to the space of $(x_i, y_i)$ variables as an intersection of upper-level sets of positively-homogenous functions. It follows that the inequality in the proposition is the closure convex hull of (38). Since $B^R_{\mathcal{B}} \subseteq B^R_{\mathcal{B}'}$, the resulting convex hull is a convex outer-approximation of $B^R_{\mathcal{B}}$. $\square$

The relaxation of $\operatorname{conv}\left(B^R_{\mathcal{B}}\right)$ presented in (36) has $3^n$ constraints. However, it is not necessary to explicitly list the exponentially many constraints to exploit this relaxation. Instead, one can introduce variables $w_i$, $i = 1, \ldots, n$, and rewrite (36) as:

$$\left\{ (x,y) \in \mathcal{B} \;\middle|\; \min \left\{ \begin{matrix} \tau_i(x_i, y_i) \\ \\ (a_i(U_i - L_i) + b'_i)(x_i - l_i) + c'_i(y_i - L_i) \\ \\ b'_i(x_i - l_i) + (a_i(u_i - l_i) + c'_i)(y_i - L_i) \end{matrix} \right\} \geq w_i, \; \sum_{i=1}^{n} w_i \geq r' \right\}.$$

Given a point $(\bar{x}, \bar{y})$, that is not feasible to the above relaxation, it is also easy to find a separating inequality that violates $(\bar{x}, \bar{y})$. In particular, the separating inequality is $\sum_{i=1}^{n} \zeta_i(x_i, y_i) \geq r'$ where $\zeta_i$ is chosen among $\tau_i(x_i, y_i)$, $(a_i(U_i - L_i) + b'_i)(x_i - l_i) + c'_i(y_i - l_i)$, and $b'_i(x_i - l_i) + (a_i(u_i - l_i) + c'_i)(y_i - L_i)$ as the function that has the minimum value at $(\bar{x}_i, \bar{y}_i)$. The separating inequality is nonlinear. Instead, a separating hyperplane can be found by replacing $\zeta_i(x_i, y_i)$ by its linear overestimator, $\zeta_i(\bar{x}_i, \bar{y}_i) + \langle s, (x_i, y_i) - (\bar{x}_i, \bar{y}_i) \rangle$, where $s$ is the subgradient of $\zeta_i$ at $(\bar{x}_i, \bar{y}_i)$. The subgradient can be chosen to be the gradient of any of the above three functions that equals $\zeta_i$ at $(\bar{x}_i, \bar{y}_i)$.

The main trick in the proof of Theorem 3.7, other than the use of Theorems 2.1 and 2.14, is that the defining bilinear inequality was tightened outside the feasible region. This technique for handling bounds can also be used for polynomial covering inequalities and other inequalities that satisfy the assumptions of Theorem 2.14. The only obstacle to its application is the ability to construct tight convex relaxations in the orthogonal subspaces. However, the orthogonal subspaces typically contain a few variables and, therefore, convexifying the restricted inequality is a much simpler problem than constructing tight relaxations for the original inequality.

Even though the proposed technique for handling bounds is general in its applicability, the primary motivation we have offered for it is that this construction allows us to sidestep a theoretical roadblock, *i.e.*, it may not be possible to construct the relaxation by convexifying orthogonal disjunctions, if Assumption (S4) of Theorem 2.14 is not satisfied, because the convex extension property either does not hold, or, at least, is not easy to argue. Therefore, it is important to determine whether this technique for convexifying orthogonal disjunctions, even though provably effective with unbounded variables (as it often constructs convex hulls), will yield interesting relaxations when the variables are restricted to belong to a compact convex set, such as a box. Since the ability to construct tight relaxations over hypercubes is the key to the success of branch-and-bound algorithms, it is crucial in the context of algorithm development that such a property holds as otherwise only the initial few branches of the tree would see an improvement in relaxation quality after the derived cuts have been incorporated. In fact, for establishing convergence of branch-and-bound type algorithms, not only does one need tight relaxations, but the relaxations must converge to the original problem as the hypercubes are contracted down to a point.

In the seminal work of McCormick [18], the author proposed a factorable relaxation technique that produces relaxations which exhibit the above convergence property. This relaxation technique has, therefore, occupied a prominent place in the toolbox of algorithm designers. The key ingredient of this relaxation scheme is that bilinear envelopes (see [2]) are used to relax products of variables. Global optimization algorithms and software typically exploit this factorable programming technique to build relaxations for bilinear inequalities such as the one that defines $B^R_{\mathcal{B}}$. Using McCormick's factorable programming technique, one would obtain the following relaxation for $B^R_{\mathcal{B}}$:

$$\operatorname{conv}\left(B^R_{\mathcal{B}}\right) \subseteq \left\{ (x,y) \in \mathcal{B} \;\middle|\; \sum_{i=1}^{n} \min \left\{ \begin{array}{c} (a_i U_i + b_i)x_i + (a_i l_i + c_i)y_i - a_i l_i U_i \\ \\ (a_i L_i + b_i)x_i + (a_i u_i + c_i)y_i - a_i u_i L_i \end{array} \right\} \geq r \right\}. \tag{39}$$

Undoubtedly, McCormick's relaxation is not always the best possible relaxation and tighter relaxations have been developed for a variety of problems. The technique that is most often used to derive such tighter relaxations is one that relaxes a given inequality, $f(x) \geq r$, to $\bar{f}(x) \geq r$, where $\bar{f}(x)$ is a concave overestimator of $f(x)$ over the relevant domain (typically, a hypercube) of $x$ that is tighter than the polyhedral concave overestimator implicit in McCormick's relaxation; see Tawarmalani and Sahinidis [28] for examples of this technique. Unfortunately, this technique for improving McCormick's relaxation is bound to fail in the context of $B^R_{\mathcal{B}}$. This is because it can be easily argued, using the additivity of $f(x,y) = \sum_{i=1}^n \left(a_i x_i y_i + b_i x_i + c_i y_i\right)$ over $i$ and the concave envelope of the bilinear term, that (39) is the relaxation of $f(x,y) \geq r$, where $f(x,y)$ has been replaced with its concave envelope over $\mathcal{B}$; see Rikun [21] for a more general result of this type. It is also known (see Proposition 35 in Richard and Tawarmalani [20]) that for the set $B^K$, defined as $\left\{(x,y) \in [0,1]^{2n} \;\middle|\; \sum_{i=1}^n a_i x_i y_i \leq r\right\}$, McCormick's relaxation produces the convex hull of $B^K$ assuming that, for all $i$, $a_i \geq 0$. Observe that $B^K$ is more specialized than $B^R_{\mathcal{B}}$ in the bilinear function and variable bounds used, but differs from $B^R_{\mathcal{B}}$ in that the direction of the inequality is reversed. These results support the intuition that McCormick's relaxations are strong for bilinear inequalities and seem to indicate that it is improbable that a relaxation tighter than McCormick's relaxation can be easily found for $B^R_{\mathcal{B}}$.

However, a careful consideration of (36) reveals that this relaxation is in fact tighter than McCormick's relaxation. The geometric insight that drives this result is that the original inequality was tightened outside the region of interest using a linear extension of the bilinear function that matches the bilinear envelope. It turns out that McCormick's relaxation can also be visualized as arising from orthogonal disjunctions. Since Theorem 3.7 constructs the relaxation by convexifying disjunctions that are, along each subspace, at least as tight as the one implicit in (39), the resulting relaxation, (36), is at least as tight as (39). We first provide a direct algebraic verification of this fact and then explore the geometry of the relaxations further.

**Corollary 3.8.** *The relaxation (36) is tighter than McCormick's relaxation (39).*

*Proof.* First relax the right-hand-side of (36) as follows:

$$\left\{ (x,y) \in \mathcal{B} \;\middle|\; \sum_{i=1}^{n} \min \left\{ \begin{array}{c} (a_i(U_i - L_i) + b'_i)(x_i - l_i) + c'_i(y_i - L_i) \\ \\ b'_i(x_i - l_i) + (a_i(u_i - l_i) + c_i)(y_i - L_i) \end{array} \right\} \geq r' \right\}, \tag{40}$$

and rewrite the defining inequality of (40) by substituting $b'_i = b_i + a_i L_i$ and $c'_i = a_i l_i + c_i$ as follows:

$$\begin{aligned} &\sum_{i=1}^{n} \Big( -a_i l_i L_i - cL_i - bl_i + \\ &\qquad \min\left\{(a_i U_i + b_i)x_i + (a_i l_i + c_i)y_i - a_i l_i U_i, (a_i L_i + b_i)x_i + (a_i u_i + c_i)y_i - a_i u_i L_i\right\}\Big) \\ &\qquad \geq r - \sum_{i=1}^{n} \left(a_i l_i L_i + cL_i + bl_i\right) \end{aligned}$$

to see the equivalence with (39). □

To explore the geometry of (36), we will now consider an example, which illustrates the source of its improvement over (39). In fact, it can be argued that the relaxation of Proposition 3.6 does not necessarily improve over (39) and that, in order to observe dominance, the inequality must be tightened outside the region of interest. The next example will also reveal the reason that the linear extension of the inequality beyond the hypercube has the effect of resurrecting the power of the bilinear envelopes. In addition, the example serves as a guide for how similar arguments can be applied to other general contexts that deploy Theorems 2.1 and 2.14 for constructing relaxations.

**Example 3.9.** *Consider the set:*

$$B^S = \left\{ (x,y) \in \prod_{i=1}^{n} [0,u] \times \prod_{i=1}^{n} [0,u] \;\middle|\; \sum_{i=1}^{n} x_i y_i \geq r \right\}.$$

*Assume that $r < u^2$. Refer to Figure 2 for an illustration of the scenario presented. The relaxations of Proposition 3.6 and Theorem 3.7 are orthogonally disjunctive and since the set is symmetric for all $i$, the relaxations can be visualized in the space of $(x_i, y_i)$ variables, while setting the remaining variables to zero. In fact, McCormick's relaxation can also be visualized on this plot since it can also be viewed as orthogonally disjunctive. Let $H = \prod_{i=1}^{n}[0,u] \times \prod_{i=1}^{n}[0,u]$ and define*

$$S_i^p(x_i, y_i) = \left\{ (x,y) \in H \;\middle|\; x_i \geq \frac{r}{u}, x_i y_i \geq r, y_i \geq \frac{r}{u}, \text{ and, } \forall j \neq i,\ x_j = y_j = 0 \right\}$$
$$S_i^m(x_i, y_i) = \left\{ (x,y) \in H \;\middle|\; x_i \geq \frac{r}{u}, y_i \geq \frac{r}{u}, \text{ and, } \forall j \neq i,\ x_j = y_j = 0 \right\}$$
$$S_i^o(x_i, y_i) = \left\{ (x,y) \in H \mid x_i y_i \geq r, \text{ and, } \forall j \neq i,\ x_j = y_j = 0 \right\}.$$

*The relaxation of Theorem 3.7 is* $\operatorname{conv}\left(\bigcup_{i=1}^{n} S_i^p(x_i, y_i)\right) \cap H$*. For the example depicted in Figure 2, $S_i^p(x_i, y_i)$ corresponds to Region 1. McCormick's relaxation is* $\operatorname{conv}\left(\bigcup_{i=1}^{n} S_i^m(x_i, y_i)\right) \cap H$*. The improvement to this relaxation can be visualized by the difference between $S_i^p(x_i, y_i)$ and $S_i^m(x_i, y_i)$ which is depicted as Region 2 in Figure 2. Finally, the relaxation obtained in Proposition 3.6, where the bounds on $x_i$ and $y_i$ are ignored in the construction, is* $\operatorname{conv}\left(\bigcup_{i=1}^{n} S_i^o(x_i, y_i)\right) \cap H$*. The improvement that results from the tightening of the defining inequality outside the bounds, following the discussion after Theorem 2.14, has the effect of removing Region 3 in Figure 2. It should be noted that even though Region 3 is outside $H$, it impacts the relaxation within $H$ when the disjunctive convex hull is constructed.*

*To visualize the impact of relaxation tightening, consider minimizing $\sum_{i=1}^{n}(x_i + y_i)$ over $B^S$. Then, the optimal value for McCormick's relaxation (a lower bound on the optimal value over $B^S$) is $\frac{2r}{u}$, whereas the bound with the relaxations of Proposition 3.6 and Theorem 3.7 equals the optimal value over $B^S$, i.e., $2\sqrt{r}$, which by our assumption, (i.e., $r < u^2$) is strictly less than $\frac{2r}{u}$. For the example illustrated in Figure 2, the bound from McCormick's relaxation is* 5 *whereas the optimal objective value as well as the bound from the relaxations of Proposition 3.6 and Theorem 3.7 is* 10. ☐

Theorems 2.1 and 2.14 not only allow us to construct tight relaxations but also help in analyzing their tightness. Our next result analyzes the relative tightness of (36) and (39) using the orthogonally disjunctive characterization. In particular, it identifies the precise condition under which (36) is tighter than McCormick's relaxation (39). This condition generalizes the assumed condition relating $r$ and $u$, namely $r < u^2$, in the above example. In fact, Example 3.9 also contains the key geometric intuition for the proof of the next result.

**Proposition 3.10.** *Assume that $r$ and for each $i$, $a_i$, $b_i$, and $c_i$ are positive. Then, relaxation (36) is tighter than McCormick's relaxation (39) if and only if there exists an $i \in \{1, \ldots, n\}$ such that $a_i(u_i - l_i)(U_i - L_i) + b_i'(u_i - l_i) + c_i'(U_i - L_i) > r'$. In other words, (36) is tighter than McCormick's relaxation (39) when there is a feasible point that sets all pairs of variables, except for one at their lower bound, and the remaining pair is not set at its upper bound.*

*Proof.* As in the proof of Theorem 3.7, it suffices to verify the claim assuming that $b_i' = b_i$, $c_i' = c_i$, $r' = r$, $l_i = 0$, and $L_i = 0$. Consider the point $(\bar{x}, \bar{y})$, where $(\bar{x}_j, \bar{y}_j) = 0$ for $j \neq i$, $\bar{x}_i = \frac{u_i r}{a_i u_i U_i + b_i u_i + c_i U_i}$ and $\bar{y}_i = \frac{U_i r}{a_i u_i U_i + b_i u_i + c_i U_i}$. If $r < a_i u_i U_i + b_i u_i + c_i$, then $(\bar{x}, \bar{y})$ is feasible to (39).

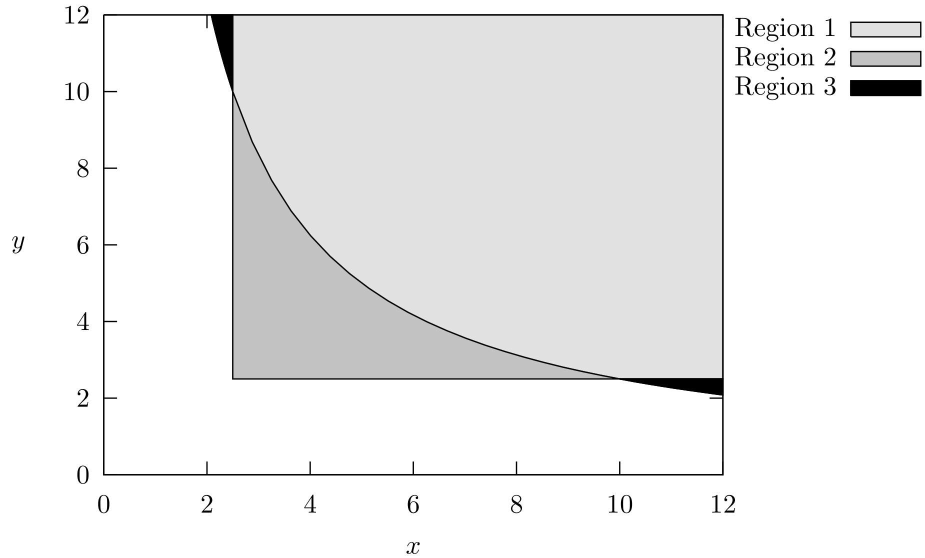


Figure 2: Various relaxations of $xy \geq 25$ over $0 \leq x, y \leq 10$

Note that $\tau_i(x_i, y_i) < r \Leftrightarrow a_i x_i y_i + b_i x_i + c_i y_i < r$. Since, $a_i\bar{x}_i\bar{y}_i + b_i\bar{x}_i + c_i\bar{y}_i < (a_iU_i + b_i)\bar{x}_i + c_i\bar{y}_i = r$, it follows that $\tau_i(\bar{x}_i, \bar{y}_i) < r$ and, therefore, $(\bar{x}_i, \bar{y}_i)$ is infeasible to (36).

If we show that $\tau_i(x_i, y_i) \geq r$ is redundant for each of the disjunctions, then it follows from the proof of Corollary 3.8 and Theorem 2.1 that (36) and (39) are identical. To prove this by contradiction, assume that $a_iu_iU_i + b_iu_i + c_iU_i < r$, *i.e.*, $\tau_i(u_i, U_i) < r$, but that there exists an $(\bar{x}_i, \bar{y}_i)$ such that $a_i\bar{x}_i\bar{y}_i + b_i\bar{x}_i + c_i\bar{y}_i < r$ and $\min\left((a_iU_i + b_i)\bar{x}_i + c_i\bar{y}_i, b_i\bar{x}_i + (a_iu_i + c_i)\bar{y}_i\right) \geq r$. Clearly, if $\bar{x}_i > u_i$ or $\bar{y}_i > U_i$, we obtain a contradiction since either $a_i\bar{x}_i\bar{y}_i + b_i\bar{x}_i + c_i\bar{y}_i > (a_iU_i + b_i)\bar{x}_i + c_i\bar{y}_i \geq r$ or $a_i\bar{x}_i\bar{y}_i + b_i\bar{x}_i + c_i\bar{y}_i > b_i\bar{x}_i + (a_iu_i + c_i)\bar{y}_i \geq r$. On the other hand, if $\bar{x}_i \leq u_i$ and $\bar{y}_i \leq U_i$, then $r \leq (a_iU_i + b_i)\bar{x}_i + c_i\bar{y}_i \leq a_iU_iu_i + b_iu_i + c_iU_i < r$, which is also a contradiction. $\square$

Proposition 3.10 formalizes the condition under which our relaxation improves McCormick's relaxation for $B^R_{\mathcal{B}}$. We next develop some insight into this condition. To place the relevance of this condition in the right perspective, it helps to realize that the infeasibility of $B^R_{\mathcal{B}}$ is simple to detect. More precisely, $B^R_{\mathcal{B}}$ is infeasible if and only if the solution obtained by setting the variables at their upper bounds is infeasible to the defining constraint. If this is indeed the case, then we verify that (39) will detect infeasibility. If for all $i$, $l_i \leq x_i \leq u_i$ and $L_i \leq y_i \leq U_i$, then

$$
\begin{aligned}
\sum_{i=1}^{n} (a_iU_i + b_i)\, x_i + (a_il_i + c_i)\, y_i - a_il_iU_i &= \sum_{i=1}^{n} a_iU_ix_i + b_ix_i + c_iy_i + a_il_i(y_i - U_i) \\
&\leq \sum_{i=1}^{n} a_iu_iU_i + b_iu_i + c_iU_i < r,
\end{aligned}
\tag{41}
$$

where the last inequality follows by the infeasibility of $B^R_{\mathcal{B}}$. Since the first expression in (41) is at least as large as the left-hand-side of the defining inequality of (39), it follows that (39) is infeasible whenever $B^R_{\mathcal{B}}$ is infeasible. Corollary 3.8 then proves that (36) will detect infeasibility as well.

Proposition 3.10 states that the new relaxation is tighter when: $a(u_i - l_i)(U_i - L_i) + b'_i(u_i - l_i) + c'_i(U_i - L_i) > r'$. Intuitively, this condition is true if the bounds on the variables are loose, *i.e.*, $u_i \gg l_i$ and $U_i \gg L_i$, or if $r'$ is small, *i.e.*, the solution obtained by setting variables at their lower bounds is almost feasible. Both of these cases are interesting in the context of branch-and-bound algorithms. Typically, branch-and-bound algorithms require bounds on variables in order to build relaxations. On the contrary, Theorem 3.7 provides a convex hull relaxation in the absence of upper bound (see Proposition 3.6). Further, the convergence of branch-and-bound algorithms in global optimization is often mired by the fact that they are $\epsilon$-convergent. In essence, branch-and-bound tends to produce many small boxes, close to feasible solutions, in an effort to prove tight bounds over these regions. The new relaxation helps address these deficiencies in current relaxations by offering the potential of improving relaxation quality close to feasible solutions.

# 4 Computational Behavior of Relaxations

In Section 2, we developed a theory for constructing convex hulls over orthogonal disjunctions and established conditions under which a relaxation for a nonconvex inequality constraint can be constructed using this approach. This theory was used to develop relaxations for polynomial, and in particular, bilinear cover inequalities in Section 3. Corollary 3.8 and Proposition 3.10 showed that the resulting relaxations are at least as tight as the McCormick relaxations for $B^R_{\mathcal{B}}$, and may be strictly tighter if the lower corner of $\mathcal{B}$ is close to a feasible solution or if the bounds are loose. Example 3.9 provided a concrete example where the difference between the two relaxations was found to be substantial. The purpose of this section is to carry out a preliminary computational study to provide numerical evidence regarding the strength of the resulting relaxations. The results here complement the theoretical results of the previous section in that they give an empirical measure of the improvement in relaxation quality on a set of randomly generated problem instances, in addition to the theoretical assertion that there is no deterioration.

We generate random instances of $B^R_{\mathcal{B}}$ as follows. The number of pairs of variables, $n$, is varied from 1 to 10. A uniform distribution from $a$ to $b$ will be denoted as $U[a, b]$. The coefficients for the bilinear cover are generated as follows $a_i \in U[1, 10]$, $b_i \in U[0, 10]$, $c_i \in U[0, 10]$, $r' \in U[1, 200n]$, $l_i \in U[0, 10]$, $u_i \in l_i + U[1, 10]$, $L_i \in U[0, 10]$, and $U_i \in L_i + U[1, 10]$. For each setting of $n$, 400 problems are generated. For each problem, the objective is to minimize $\sum_{i=1}^n (x_i + y_i)$. This objective does not restrict the generality of the study since there is sufficient flexibility in the problem generator to scale the problem variables rendering it unnecessary to choose objective function coefficients for the problem variables. The problems are generated and solved using GAMS 22.8 [12]. The objective is minimized over the bilinear cover using BARON 8.1 [24]. The nonlinear relaxation (36) is solved using KNITRO 5.1 [32] and (39) is solved using CPLEX 11.1 [16]. The absolute optimality tolerance is set as 0.0 and the relative tolerance is set to $10^{-5}$. The problems are solved on a Linux (Fedora Werewolf) PC with an INTEL 2.13 GHz dual-core processor and 3GB RAM.

Table 1 reports our computational results with the bilinear covering set. Here, $n$ is the number of pairs of variables; $n_{\text{inf}}$, $n_{\text{inforth}}$, and $n_{\text{infmc}}$ are the number of problems found to be infeasible using the bilinear inequality, (36), and (39) respectively; $n_{\text{orth}}$ and $n_{\text{mc}}$ are the number of problems that exhibit a non-zero relaxation gap with (36) and (39) respectively; and $p_{\text{gap}}$ is the average percentage of gap closed for the problems for which (39) exhibits a relaxation gap. The percentage of gap closed is defined as $1 - \frac{z_n - z_o}{z_n - z_m}$, where $z_n$, $z_o$, and $z_m$ are the optimal values over $B^R_{\mathcal{B}}$, (36), and (39) respectively. As was proved in the discussion following Proposition 3.10, $n_{\text{inf}} = n_{\text{inforth}} = n_{\text{infmc}}$. Further, since (36) defines the convex hull of $B^R_{\mathcal{B}}$ when restricted to one variable pair, it is expected that, for $n = 1$, $p_{\text{gap}} = 100\%$ and $n_{\text{orth}} = 0$. However, even when $n > 1$, it is interesting to note that the average gap closed by the new relaxation is substantial. Also, the number of problems with a non-zero gap is substantially smaller for (36) as compared to that for (39), *i.e.*, $n_{\text{orth}} \ll n_{\text{mc}}$.

# 5 Conclusions

In this paper, we developed a convexification tool for orthogonal disjunctions that does not introduce new variables. As an application, we provided a simple derivation of intersection cuts for mixed-integer polyhedral sets. The convexification tool was also shown to be useful in deriving cuts for a variety of nonconvex constraints; those that satisfy a key convex extension property. Verifying the convex extension property can be an arduous task. To address this difficulty, we provided a general set of conditions that are sufficient to establish the convex extension property. These conditions were then used to verify the convex extension property and find convex hull representations and convex relaxations for a variety of polynomial covering sets. The results were specialized and refined for the bilinear covering set. The relaxations of the bilinear covering set were shown to be at least

| $n$ | $n_{\text{inf}}$ | $n_{\text{inforth}}$ | $n_{\text{infmc}}$ | $n_{\text{orth}}$ | $n_{\text{mc}}$ | $p_{\text{gap}}$ (%) |
|---|---|---|---|---|---|---|
| 1 | 21 | 21 | 21 | 0 | 272 | 100.0 |
| 2 | 3 | 3 | 3 | 35 | 301 | 91.5 |
| 3 | 1 | 1 | 1 | 43 | 311 | 91.0 |
| 4 | 0 | 0 | 0 | 77 | 334 | 83.5 |
| 5 | 1 | 1 | 1 | 90 | 317 | 78.0 |
| 6 | 0 | 0 | 0 | 106 | 331 | 73.0 |
| 7 | 0 | 0 | 0 | 127 | 345 | 70.9 |
| 8 | 0 | 0 | 0 | 148 | 343 | 63.6 |
| 9 | 0 | 0 | 0 | 151 | 336 | 58.0 |
| 10 | 0 | 0 | 0 | 158 | 344 | 60.6 |

Table 1: Performance of relaxations on the bilinear covering set

as tight as the standard factorable relaxation of McCormick. The precise condition under which the relaxation developed is strictly tighter was identified by exploiting the geometry of orthogonal disjunctions. Finally, preliminary computational results were provided to demonstrate that the developed relaxations close a significant gap for bilinear covering sets. Future work will concentrate on applying these results to other classes of problems, and on incorporating the findings in relaxation constructors within a branch-and-bound algorithm for global optimization.

# References


[1] C. S. Adjiman, I. P. Androulakis, and C. A. Floudas. A global optimization method, $\alpha$BB, for general twice-differentiable constrained NLPs–II. Implementation and computational results. *Computers & Chemical Engineering*, 22:1159–1179, 1998.

[2] F. A. Al-Khayyal and J. E. Falk. Jointly constrained biconvex programming. *Mathematics of Operations Research*, 8:273–286, 1983.

[3] A. Atamturk. Strong formulations of robust mixed 0-1 programming. *Mathematical Programming*, pages 235–250, 2006.

[4] E. Balas. Disjunctive programming: Cutting planes from logical conditions. In O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, editors, *Nonlinear Programming*, pages 279–312. Academic Press, NY, 1975.

[5] E. Balas. Disjunctive programming: Properties of the convex hull of feasible points. *Discrete Applied Mathematics*, 89(1-3):3–44, 1998. Original manuscript was published as a technical report in 1974.

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

[7] E. Balas and M. Perregaard. A precise correspondence between lift-and-project cuts, simple disjunctive cuts, and mixed-integer gomory cuts for 0-1 programming. *Mathematical Programming*, 94:221–245, 2003.

[8] P. Belotti, J. Lee, L. Liberti, F. Margot, and A. Waechter. Branching and bounds tightening techniques for non-convex MINLP. Available at `http://www.optimization-online.org/DB_HTML/2008/08/2059.html`, 2008.

[9] C. Bliek, C. Jermann, and A. Neumaier, editors. *Global Optimization and Constraint Satisfaction*, Fifth Annual Workshop on Global Constraint Optimization and Constraint Satisfaction, COCOS, 2002. Springer.

[10] S. Ceria and J. Soares. Convex programming for disjunctive convex optimization. *Mathematical Programming*, 86A:595–614, 1999.

[11] G. Cornuéjols and C. Lemaréchal. A convex-analysis perspective on disjunctive cuts. *Mathematical Programming*, 106:567–586, 2006.

[12] GAMS Corporation. *GAMS – A User's Guide*, 2008. Available at `http://www.gams.com/docs/gams/GAMSUsersGuide.pdf`.

[13] J. E. Falk and R. M. Soland. An algorithm for separable nonconvex programming problems. *Management Science*, 15:550–569, 1969.

[14] I. Harjunkoski, T. Westerlund, R. Porn, and H. Skrifvars. Different transformations for solving non-convex trim-loss problems by MINLP. *European Journal of Operational Research*, 105: 594–603, 1998.

[15] R. Horst and H. Tuy. *Global Optimization: Deterministic Approaches*. Springer Verlag, Berlin, Third edition, 1996.

[16] ILOG. *CPLEX 11.1 User's Manual*. ILOG Inc., Mountain View, CA, 2007.

[17] LINDO Systems Inc. LINGO 11.0 optimization modeling software for linear, nonlinear, and integer programming. Available at `http://www.lindo.com`, 2008.

[18] G. P. McCormick. Computability of global solutions to factorable nonconvex programs: Part I - convex underestimating problems. *Mathematical Programming*, 10:147–175, 1976.

[19] A. Neumaier. Complete search in continuous global optimization and constraint satisfaction. *Acta Numerica*, 13:271–369, 2004.

[20] J.-P. P. Richard and M. Tawarmalani. Lifting inequalities: A framework for generating strong cuts for nonlinear programs. *Mathematical Programming*, pages 1–43, 2008. doi: 10.1007s/10107-008-0226-9. Online first: http://www.springerlink.com/content/59q56533601g1532.

[21] A. D. Rikun. A convex envelope formula for multilinear functions. *Journal of Global Optimization*, 10:425–437, 1997.

[22] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[23] H. S. Ryoo and N. V. Sahinidis. A branch-and-reduce approach to global optimization. *Journal of Global Optimization*, 8:107–139, 1996.

[24] N. V. Sahinidis and M. Tawarmalani. *BARON*. The Optimization Firm, LLC, Urbana-Champaign, IL, 2005. Available at `http://www.gams.com/dd/docs/solvers/baron.pdf`.

[25] N. W. Sawaya and I. E. Grossmann. A cutting plane method for solving linear generalized disjunctive programming problems. *Computers and Chemical Engineering*, 29:1891–1913, 2005.

[26] S. Sherali and S. Sen. Cuts from combinatorial disjunctions. *Operations Research*, pages 928–933, 1985.

[27] R. Stubbs and S. Mehrotra. A branch and cut method for 0-1 mixed convex programming. *Mathematical Programming*, 86:515–532, 1999.

[28] M. Tawarmalani and N. V. Sahinidis. *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications*. Kluwer, Dordrecht, The Netherlands, 2002.

[29] M. Tawarmalani and N. V. Sahinidis. Convex extensions and envelopes of lower semi-continuous functions. *Mathematical Programming*, 93:247–263, 2002.

[30] M. Tawarmalani and N. V. Sahinidis. Semidefinite relaxations of fractional programs via novel techniques for constructing convex envelopes of nonlinear functions. *Journal of Global Optimization*, 20:137–158, 2001.

[31] M. Tawarmalani and N. V. Sahinidis. A polyhedral branch and cut approach to global optimization. *Mathematical Programming*, 103:225–249, 2005.

[32] R. A. Waltz and T. D. Plantenga. *KNITRO 5.1 User's Manual*. Zienna Optimization Inc., Evanston, IL, 2007. Available at `http://www.ziena.com/docs/Knitro51_UserManual.pdf`.

[33] G. M. Ziegler. *Lectures on Polytopes*. Springer, NY, 1998.