A Finitely Convergent Disjunctive Cutting Plane Algorithm for Bilinear Programming

Hamed Rahimian*[1] and Sanjay Mehrotra†[2]

[1]Department of Industrial Engineering, Clemson University, Clemson SC 29634, USA
[2]Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston IL 60208, USA

Abstract

In this paper we present and analyze a finitely-convergent disjunctive cutting plane algorithm to obtain an ϵ-optimal solution or detect infeasibility of a general nonconvex continuous bilinear program. While the cutting planes are obtained in a manner similar to Saxena et al. (2010); Fampa and Lee (2018), a feature of the algorithm that guarantees finite convergence is exploring near-optimal extreme point solutions to a current relaxation at each iteration. In this sense, the presented algorithm and its analysis extends the work of Owen and Mehrotra (2001) for solving mixed-integer linear programs to the general bilinear programs.

Keywords: Bilinear programming, Nonconvex programming, Disjunctive programming, Global optimization, Cutting planes

1 Introduction

In this paper we study a general nonconvex continuous bilinear program (BLP) defined as follows:

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{y}} \ & \boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{A}_0 \boldsymbol{y} \\ \text{s.t.} \ & \boldsymbol{f}_\iota^\top \boldsymbol{x} + \boldsymbol{g}_\iota^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{A}_\iota \boldsymbol{y} + b_\iota \leq 0, \quad \iota \in [p] \\ & \underline{\boldsymbol{x}} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}},\ \underline{\boldsymbol{y}} \leq \boldsymbol{y} \leq \overline{\boldsymbol{y}}, \end{aligned} \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{y} \in \mathbb{R}^m$, $\boldsymbol{A}_\iota$, $\iota \in 0 \cup [p]$, are $n \times m$ matrices, $\boldsymbol{f}_\iota \in \mathbb{R}^n$, $\boldsymbol{g}_\iota \in \mathbb{R}^m$, $\iota \in 0 \cup [p]$, $\underline{\boldsymbol{x}}, \overline{\boldsymbol{x}} \in \mathbb{R}^n$, $\underline{\boldsymbol{y}}, \overline{\boldsymbol{y}} \in \mathbb{R}^m$, and $b_\iota \in \mathbb{R}$, $\iota \in [p]$. We do not consider any structure on the matrices $\boldsymbol{A}_\iota$, $\iota \in 0 \cup [p]$. A

*hrahimi@clemson.edu
†mehrotra@northwestern.edu

bilinear program of the form (1) finds various applications in production, location-allocation, and product distribution situations (Adams and Sherali, 1993), pooling (Misener and Floudas, 2009), trim-loss and cutting stock (Harjunkoski et al., 1999; Rodríguez and Vecchietti, 2008), packing (Locatelli and Raber, 2002), network interdiction (Davarnia et al., 2017), and economic equilibrium (Mathiesen, 1985).

The problem of generating relaxations of a bilinear program has been investigated in the literature. A common method to obtain a linear programming relaxation of a bilinear function xy is to introduce a new variable w and then relax the constraint $w = xy$. For the case that variables x and y are restricted to a box, McCormick (1976) constructs a polyhedral relaxation for the bilinear set defined by $w = xy$. Al-Khayyal and Falk (1983) show that this relaxation describes the convex hull of the bilinear set.

Discretizing a subset of continuous variables gives a mixed-integer BLP that approximates the original BLP, see, e.g., Pham et al. (2009); Glover (1975). Gupte et al. (2013) obtain an exact mixed-integer linear programming reformulation of a mixed-integer BLP using the binary expansion of integer variables. By studying the polyhedral structure of the set arising from McCormick envelopes for an individual bilinear term, Gupte et al. (2013) obtain the convex hull of these reformulated individual bilinear sets and use them in a branch-and-bound algorithm to solve the reformulated mixed-integer linear program.

Other relaxations based on the reformulation-linearization technique (RLT) (Sherali and Adams, 2013), second-order cone programming (SOCP), see, e.g., Dey et al. (2019); Santana and Dey (2020), and semidefinite programming (SDP), see, e.g., Bao et al. (2011), have been applied to continuous BLPs. For the case that there is no interaction between the continuous variables $\boldsymbol{x}$ and $\boldsymbol{y}$, except for in the bilinear objective function, Sherali and Alameddine (1992) develop a RLT-based relaxation that theoretically dominates the McCormick relaxation. Using this RLT-based relaxation, they propose a finitely-convergent branch-and-bound algorithm. Dey et al. (2019) study a bilinear program of the form (1), where the variables can be partitioned into two sets such that fixing the variables in any of the sets results in a linear program. They show that the convex hull of the set induced by a single constraint is SOC representable in the extended space (see also Santana and Dey (2020) for results on a more general quadratic equation). The intersection of such sets gives a relaxation which is stronger than the standard SDP relaxation intersected with the boolean quadratic polytope Dey et al. (2019).

In this paper, we focus on using the lift-and-project methodology and disjunctive programming (Balas, 1998). Our motivation to use this framework is that it simultaneously takes into account convex and nonconvex constraints, see, e.g., Davarnia et al. (2017); Saxena et al. (2010, 2011). An infinitely-convergent disjunctive sequential convexification procedure for a continuous bilinear set is studied in a companion paper Rahimian and Mehrotra (2020).

Treating bilinear terms in the context of global optimization has also been studied in the literature (Konno, 1976; Vaish and Shetty, 1977; Sherali and Shetty, 1980; Fampa and Lee, 2018). Konno (1976) proposes an infinitely-convergent cutting plane procedure to obtain a solution differing in objective value from the global optimal value of the studied BLP by no more than a predetermined quantity $\epsilon > 0$. Vaish and Shetty (1977) propose an infinitely-convergent cutting plane procedure to obtain a global optimal solution to the studied BLP. They also porpose a finitely-convergent cutting plane algorithm to obtain a solution differing in objective value from the global optimal value by no more than ϵ. Sherali and Shetty (1980) propose a finitely-convergent cutting plane algorithm to obtain a global optimal solution by generating polar cuts at an extreme point solution and generating disjunctive cuts at other points. In the studied BLP in Konno (1976); Vaish and Shetty (1977); Sherali and Shetty (1980) it is assumed that variables $\boldsymbol{x}$ and $\boldsymbol{y}$ belong to their own polytopes and there is no nonlinearity in the constraints. That is, the objective function is nonconvex while the feasible region is convex.

For mixed-integer quadratically constrained quadratic programs, Saxena et al. (2010) propose to obtain valid disjunctive cuts using the eigenvalue decomposition of the quadratic violation matrix. For a continuous BLP, with bilinear terms in the objective function and constraints, Fampa and Lee (2018) further extend the approach in Saxena et al. (2010) using the singular value decomposition of the bilinear violation matrix (we shall shortly review this approach in Section 2). They conduct extensive computational experiments to assess the performance of this approach and methods that convert a bilinear program to a quadratic program with a symmetric matrix.

Although Fampa and Lee (2018) are concerned with the global optimization of the studied BLP, they do not provide any theoretical result to guarantee that an optimal solution is found finitely. To close this gap, in this paper, a modification to the approach in Saxena et al. (2010); Fampa and Lee (2018) is analyzed to guarantee a finitely-convergent disjunctive programming-based cutting plane approach. This modification is inspired by the cutting plane approach proposed in Owen and Mehrotra (2001) in the context of solving mixed-integer linear programs with general integer variables. As in Owen and Mehrotra (2001), a fundamental feature of the analyzed algorithm in this paper is to generates valid inequalities at *all* near-optimal extreme point solutions of the current relaxation (to be defined precisely in Section 4). We theoretically analyze that modifying the idea investigated in Fampa and Lee (2018) with this vertex exploration guarantees finite convergence.

Although, in theory, this vertex exploration guarantees finite convergence, there are some computational limitations. On one hand, exploring all near-optimal extreme point solutions is computationally expensive. On the other hand, not all generated cuts through this vertex exploration would necessarily have computational values. Thus, for practical implementations, a judicious (problem-dependent) vertex exploration is necessary. We provide some indications to these limitations in our numerical results in Section 5, where we generate cuts at only *a few* extreme point solutions. We conduct numerical experiments to compare an *implementation* of the idea proposed in Fampa and Lee (2018) (see Section 2.2 for more details on our implementation) and a *practical implementation* of the algorithm analyzed in this paper (see Section 5 for more details). It is worth noting that although we conduct a comparative study, our primary aim in this paper is not to reach a conclusion about the computational efficiency of one over another.

To the best of our knowledge, the analyzed algorithm in this paper is the first pure cutting plane approach that solves (1) to ϵ-optimality (to be defined precisely in Section 4) or detects infeasibility in a finite number of iterations. We emphasize that the feasible region in (1) is nonconvex. This is different from the studied BLP in Konno (1976); Vaish and Shetty (1977); Sherali and Shetty (1980), where an optimal solution is attained at an extreme point $(\boldsymbol{x}^*, \boldsymbol{y}^*)$, with $\boldsymbol{x}^*$ and $\boldsymbol{y}^*$ to be the extreme points of their corresponding polytopes, see, e.g., (Konno, 1976, Theorem 2.1).

This paper is organized as follows. In Section 2, we review the lift-and-project methodology of Saxena et al. (2010) in the context of a BLP, and the basic ideas of disjunctive programming. We also illustrate our motivation to theoretically enhance the procedure studied by Fampa and Lee (2018). In Section 3, we present the cut generation component of our algorithms. In Section 4, we analyze disjunctive cutting plane algorithms that find an ϵ-optimal solution to (1) or detect infeasibility in a finite number of iterations. In Section 5, we demonstrate the optimality gap improvement gained from a practical implementation of the analyzed cutting plane algorithms in this paper. We end with conclusions in Section 6.

Notation and Definitions: Throughout this paper, vectors are denoted by boldface lowercase letters and matrices are denoted by boldface uppercase letters. Sets are denoted by calligraphic or normal uppercase letters. All sets in this paper are subsets of a finite-dimensional Euclidean space $\mathbb{R}^d$. Consider a set $\mathcal{B} \subseteq \mathbb{R}^d$. Let $\text{ext}(\mathcal{B})$ and $\text{conv}(\mathcal{B})$ denote the set of extreme points and convex hull of the set $\mathcal{B}$. Let $\text{Proj}_{\boldsymbol{x}}(\mathcal{B})$ denote the projection of $\mathcal{B}$ onto the $\boldsymbol{x}$-space. Let $\boldsymbol{e}_i$ be the i-th unit vector in $\mathbb{R}^d$. Consider two sets $\mathcal{B}^2 \subseteq \mathcal{B}^1 \subseteq \mathbb{R}^d$. The Hausdorff distance between $\mathcal{B}^1$ and $\mathcal{B}^2$ is denoted by $d_H(\mathcal{B}^1, \mathcal{B}^2)$ and is defined as $d_H(\mathcal{B}^1, \mathcal{B}^2) := \sup_{\boldsymbol{b}^1 \in \mathcal{B}^1} \inf_{\boldsymbol{b}^2 \in \mathcal{B}^2} \|\boldsymbol{b}^1 - \boldsymbol{b}^2\|$. A sequence of sets $\{\mathcal{B}^t\}$ is called a decreasing sequence of nested sets if $\mathcal{B}^{t+1} \subseteq \mathcal{B}^t$, $t \geq 0$. We say that a decreasing sequence of nested closed sets $\{\mathcal{B}^t\}$ of $\mathbb{R}^d$ converges to a closed set $\bar{\mathcal{B}} \subseteq \mathbb{R}^d$ in Hausdorff distance, and denote it by $\lim_{t\to\infty} \mathcal{B}^t = \bar{\mathcal{B}}$, if $d_H(\mathcal{B}^t, \bar{\mathcal{B}}) \to 0$ as $t \to \infty$. According to (Salinetti and Wets, 1979, Lemma 1), it means that either $\mathcal{B}$ and $\mathcal{B}^t$ are empty for all $t \geq \bar{t}$ or for any $\delta > 0$, there exists $\hat{t} \geq 0$ such that for all $t \geq \hat{t}$, we have $\inf_{\boldsymbol{b} \in \bar{\mathcal{B}}} \|\boldsymbol{b} - \boldsymbol{b}^t\| \leq \delta$ for all $\boldsymbol{b}^t \in \mathcal{B}^t$. We say that a sequence of sets $\{\mathcal{B}^t\}$ of $\mathbb{R}^d$ converges to $\bar{\mathcal{B}} \subseteq \mathbb{R}^d$ in the sense of Kuratowski, and denote it by $\mathcal{B}^t \xrightarrow{\text{K}} \bar{\mathcal{B}}$ as t $\to \infty$, if $\limsup_{t\to\infty} \mathcal{B}^t = \liminf_{t\to\infty} \mathcal{B}^t = \bar{\mathcal{B}}$. For two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\boldsymbol{A} \bullet \boldsymbol{B} = Tr(\boldsymbol{A}^\top \boldsymbol{B})$ denotes the Frobenius inner product between matrices. We let $[d]$ denote the index set $\{1, \ldots, d\}$.

2 Lift-and-Project Methodology of Saxena et al. (2010); Fampa and Lee (2018)

By introducing additional variables $W_{ij} = x_i y_j$, $i \in [n]$, $j \in [m]$, problem (1) can be equivalently written as the following nonlinear program in the lifted space:

$$\begin{aligned} \min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\mathcal{K}} \quad & \boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{A}_0 \bullet \boldsymbol{W} \\ \text{s.t.} \quad & \boldsymbol{W} = \boldsymbol{x}\boldsymbol{y}^\top, \end{aligned} \tag{BLP}$$

where

$$\mathcal{K} := \left\{ (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \,\middle|\, \begin{array}{l} \boldsymbol{f}_\iota^\top \boldsymbol{x} + \boldsymbol{g}_\iota^\top \boldsymbol{y} + \boldsymbol{A}_\iota \bullet \boldsymbol{W} + b_\iota \leq 0,\ \iota \in [p], \\ \underline{\boldsymbol{x}} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}},\ \underline{\boldsymbol{y}} \leq \boldsymbol{y} \leq \overline{\boldsymbol{y}} \end{array} \right\}. \tag{2}$$

Set

$$\mathcal{F} := \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \mathcal{K} \,|\, \boldsymbol{W} = \boldsymbol{x}\boldsymbol{y}^\top\} \tag{3}$$

is the feasible region of (BLP), and set $\mathcal{K}$ is the feasible region of a relaxation of (BLP). Note that all the constraints in $\mathcal{K}$ are linear in $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{W}$, and $\mathcal{K}$ is a convex set. On the other hand, $\boldsymbol{W} = \boldsymbol{x}\boldsymbol{y}^\top$ induces a nonconvex region.

In this paper, we are interested in disjunctive programming procedures in the space of $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W})$. A disjunctive programming procedure to treat the bilinear terms is studied in Fampa and Lee (2018), by applying McCormick convexification of $\boldsymbol{W} = \boldsymbol{x}\boldsymbol{y}^\top$ and extending the ideas in Saxena et al. (2010) for symmetric quadratic terms to bilinear terms. Because the approach in Saxena et al. (2010); Fampa and Lee (2018) forms a basis for our work, let us first recall their procedure.

For any $\boldsymbol{u} \in \mathbb{R}^n$ and $\boldsymbol{v} \in \mathbb{R}^m$, any feasible solution to (BLP) satisfies

$$\boldsymbol{u}^\top \boldsymbol{W} \boldsymbol{v} = (\boldsymbol{u}^\top \boldsymbol{x})(\boldsymbol{v}^\top \boldsymbol{y}). \tag{4}$$

Because $(\boldsymbol{u}^\top \boldsymbol{x})(\boldsymbol{v}^\top \boldsymbol{y}) = \left(\frac{\boldsymbol{u}^\top \boldsymbol{x} + \boldsymbol{v}^\top \boldsymbol{y}}{2}\right)^2 - \left(\frac{\boldsymbol{u}^\top \boldsymbol{x} - \boldsymbol{v}^\top \boldsymbol{y}}{2}\right)^2$, (4) is equivalent to the following two inequalities

$$\boldsymbol{u}^\top \boldsymbol{W} \boldsymbol{v} - \left(\frac{\boldsymbol{u}^\top \boldsymbol{x} + \boldsymbol{v}^\top \boldsymbol{y}}{2}\right)^2 + \left(\frac{\boldsymbol{u}^\top \boldsymbol{x} - \boldsymbol{v}^\top \boldsymbol{y}}{2}\right)^2 \leq 0, \tag{5}$$

$$-\boldsymbol{u}^\top \boldsymbol{W} \boldsymbol{v} + \left(\frac{\boldsymbol{u}^\top \boldsymbol{x} + \boldsymbol{v}^\top \boldsymbol{y}}{2}\right)^2 - \left(\frac{\boldsymbol{u}^\top \boldsymbol{x} - \boldsymbol{v}^\top \boldsymbol{y}}{2}\right)^2 \leq 0. \tag{6}$$

Observe that the concave terms $-\left(\frac{\boldsymbol{u}^\top \boldsymbol{x} + \boldsymbol{v}^\top \boldsymbol{y}}{2}\right)^2$ and $-\left(\frac{\boldsymbol{u}^\top \boldsymbol{x} - \boldsymbol{v}^\top \boldsymbol{y}}{2}\right)^2$, in (5) and (6), respectively, result in a nonconvex region. A way to handle this nonconvexity is to approximate the concave terms with their secant inequalities and to utilize disjuncntive programming to derive valid disjunctive cuts for $\operatorname{conv}(\mathcal{F})$ Saxena et al. (2010). More precisely, constraints (5) and (6) give rise to the following disjunction, which is satisfied by any feasible solution $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W})$ to (BLP):

$$\vee_{r=1}^2 \vee_{s=1}^2 \tilde{\mathcal{S}}_{rs}(\boldsymbol{c}, \tilde{\mathcal{K}}, \boldsymbol{\beta}), \tag{7}$$

where $\tilde{\mathcal{K}}$ is a (bounded) convex relaxation of $\mathcal{F}$ (e.g., $\mathcal{K}$), $\boldsymbol{c} = \text{vec}(\boldsymbol{u}\boldsymbol{v}^\top)$, and

$$\tilde{\mathcal{S}}_{rs}(\boldsymbol{c}, \tilde{\mathcal{K}}, \boldsymbol{\beta}) := \left\{ (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \tilde{\mathcal{K}} \;\middle|\; \begin{array}{l} \beta_{1,r} \leq \frac{\boldsymbol{u}^\top\boldsymbol{x}+\boldsymbol{v}^\top\boldsymbol{y}}{2} \leq \beta_{1,r+1},\ \beta_{2,s} \leq \frac{\boldsymbol{u}^\top\boldsymbol{x}-\boldsymbol{v}^\top\boldsymbol{y}}{2} \leq \beta_{2,s+1}, \\ \boldsymbol{u}^\top\boldsymbol{W}\boldsymbol{v} - \left(\frac{\boldsymbol{u}^\top\boldsymbol{x}+\boldsymbol{v}^\top\boldsymbol{y}}{2}\right)(\beta_{1,r}+\beta_{1,r+1}) + \beta_{1,r}\beta_{1,r+1} \\ \quad + \left(\frac{\boldsymbol{u}^\top\boldsymbol{x}-\boldsymbol{v}^\top\boldsymbol{y}}{2}\right)^2 \leq 0, \\ -\boldsymbol{u}^\top\boldsymbol{W}\boldsymbol{v} - \left(\frac{\boldsymbol{u}^\top\boldsymbol{x}-\boldsymbol{v}^\top\boldsymbol{y}}{2}\right)(\beta_{2,s}+\beta_{2,s+1}) + \beta_{2,s}\beta_{2,s+1} \\ \quad + \left(\frac{\boldsymbol{u}^\top\boldsymbol{x}+\boldsymbol{v}^\top\boldsymbol{y}}{2}\right)^2 \leq 0 \end{array} \right\}, \tag{8}$$

for $r, s = 1, 2$. Disjunction (7) is obtained by simultaneously splitting the range $[\beta_{1,1}, \beta_{1,3}]$ of function $\frac{\boldsymbol{u}^\top\boldsymbol{x}+\boldsymbol{v}^\top\boldsymbol{y}}{2}$ over $\tilde{\mathcal{K}}$ into two intervals $[\beta_{1,1}, \beta_{1,2}]$ and $[\beta_{1,2}, \beta_{1,3}]$, and by splitting the range $[\beta_{2,1}, \beta_{2,3}]$ of function $\frac{\boldsymbol{u}^\top\boldsymbol{x}-\boldsymbol{v}^\top\boldsymbol{y}}{2}$ over $\tilde{\mathcal{K}}$ into two intervals $[\beta_{2,1}, \beta_{2,2}]$ and $[\beta_{2,2}, \beta_{2,3}]$. Moreover, the disjunction simultaneously constructs secant inequalities of functions $-\left(\frac{\boldsymbol{u}^\top\boldsymbol{x}+\boldsymbol{v}^\top\boldsymbol{y}}{2}\right)^2$ and $-\left(\frac{\boldsymbol{u}^\top\boldsymbol{x}-\boldsymbol{v}^\top\boldsymbol{y}}{2}\right)^2$ in each corresponding interval. The breakpoints $\{\beta_{1,1}, \beta_{1,2}, \beta_{1,3}\}$ might have overlaps. However, as long as these breakpoints are in the range of function $\frac{\boldsymbol{u}^\top\boldsymbol{x}+\boldsymbol{v}^\top\boldsymbol{y}}{2}$, a disjunction of the form (7) is valid. A similar situation might happen for the breakpoints $\{\beta_{2,1}, \beta_{2,2}, \beta_{2,3}\}$.

For the rest of the paper, we let the index k represent the (r, s)-pair, where $k \in \{1, 2, 3, 4\}$. Hence, hereafter, we denote $\tilde{\mathcal{S}}_{rs}(\boldsymbol{c}, \tilde{\mathcal{K}}, \boldsymbol{\beta})$ as $\tilde{\mathcal{S}}_k(\boldsymbol{c}, \tilde{\mathcal{K}}, \boldsymbol{\beta})$. Let $r(k)$ and $s(k)$ denote the r and s component of the index k. For the ease of exposition, for $k \in \{1, 2, 3, 4\}$, $\beta_{1,k}$ and $\beta_{2,k}$ should be understood as $\beta_{1,r(k)}$ and $\beta_{2,s(k)}$, respectively. Similarly, $\beta_{1,k+1}$ and $\beta_{2,k+1}$ should be understood as $\beta_{1,r(k)+1}$ and $\beta_{2,s(k)+1}$, respectively. We also denote a (bounded) convex relaxation of $\mathcal{F}$ by $\tilde{\mathcal{K}}$ throughout the paper.

2.1 Disjunctive Programming

Given a solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$ to the current relaxation $\tilde{\mathcal{K}}$, Fampa and Lee (2018) analyze the singular value decomposition (SVD) of $\hat{\boldsymbol{W}} - \hat{\boldsymbol{x}}\hat{\boldsymbol{y}}^\top$ to find suitable vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, corresponding to a nonzero singular value σ, i.e., $\boldsymbol{u}^\top(\hat{\boldsymbol{W}} - \hat{\boldsymbol{x}}\hat{\boldsymbol{y}}^\top)\boldsymbol{v} = \sigma \neq 0$. The vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are used in order to form a disjunction of the form (7), and subsequently, to derive disjunctive cuts for $\text{conv}(\mathcal{F})$ through a cut-generation linear program (CGLP). The CGLP used in Fampa and Lee (2018) contains linearization of the constraints in (8), where the convex quadratic terms are replaced by their outer approximations, obtained at the current relaxation solution. Moreover, $\tilde{\mathcal{K}}$ includes the constraints of McCormick convexification of $\boldsymbol{W} = \boldsymbol{x}\boldsymbol{y}^\top$, using the box constraints on $\boldsymbol{x}$ and $\boldsymbol{y}$, and all the previously added disjunctive cuts. Below, we present an abstract form of this CGLP to generate a valid inequality in order to cut off the current solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$.

Lemma 1. *Consider a polyhedral bounded relaxation $\tilde{\mathcal{K}}$ of $\mathcal{F}$, a fixed $\boldsymbol{c} = \text{vec}(\boldsymbol{u}\boldsymbol{v}^\top)$, and a choice of breakpoints $\boldsymbol{\beta}$ for a (2×2)-way disjunction (7). Let $\boldsymbol{A}_k\boldsymbol{x} + \boldsymbol{B}_k\boldsymbol{y} + \boldsymbol{C}_{\cdot k} \bullet \boldsymbol{W} \geq \boldsymbol{d}_k$ represent the set of constraints in $\tilde{\mathcal{S}}_k(\boldsymbol{c}, \tilde{\mathcal{K}}, \boldsymbol{\beta})$, after linearization of the quadratic terms, where $\boldsymbol{C}_{\cdot k} \in \mathbb{R}^{n \times m}$,*

$k = 1,2,3,4$. *Then,* $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \in$ *conv* $\left(\vee_{k=1}^4 \boldsymbol{A}_k \boldsymbol{x} + \boldsymbol{B}_k \boldsymbol{y} + \boldsymbol{C}_{\cdot k} \bullet \boldsymbol{W} \geq \boldsymbol{d}_k\right)$ *if the optimal value of the following CGLP is nonnegative*

$$\min \boldsymbol{\alpha}^\top \hat{\boldsymbol{x}} + \boldsymbol{\theta}^\top \hat{\boldsymbol{y}} + \boldsymbol{\Gamma} \bullet \hat{\boldsymbol{W}} - \rho \tag{9a}$$

$$s.t.\ \boldsymbol{A}_k^\top \boldsymbol{\pi}_k = \boldsymbol{\alpha},\ \boldsymbol{B}_k^\top \boldsymbol{\pi}_k = \boldsymbol{\theta},\ \boldsymbol{C}_{\cdot k}^\top \boldsymbol{\pi}_k = \boldsymbol{\Gamma},\ \boldsymbol{d}_k^\top \boldsymbol{\pi}_k \geq \rho,\ \boldsymbol{\pi}_k \geq \boldsymbol{0},\ \forall k. \tag{9b}$$

If the optimal value of CGLP (9) is negative and $(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\Gamma}, \boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_4)$ *is an optimal solution to (9), then* $\boldsymbol{\alpha}^\top \boldsymbol{x} + \boldsymbol{\theta}^\top \boldsymbol{y} + \boldsymbol{\Gamma} \bullet \boldsymbol{W} \geq \rho$ *is a valid inequality for conv* $\left(\vee_{k=1}^4 \boldsymbol{A}_k \boldsymbol{x} + \boldsymbol{B}_k \boldsymbol{y} + \boldsymbol{C}_{\cdot k} \bullet \boldsymbol{W} \geq \boldsymbol{d}_k\right)$, *which cuts off* $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$.

Proof. The result follows from an application of Balas (1998, Theorem 3.1). □

2.2 Motivating Example

It is illustrated in the numerical experiments of Fampa and Lee (2018) that their proposed procedure is not guaranteed to reach an optimal solution of (BLP). In a simple example, we illustrate this issue.

Example 1. *Consider a problem of the form (BLP), where* $\boldsymbol{x} \in \mathbb{R}^2$, $\boldsymbol{y} \in \mathbb{R}^2$, *and there is only one linear constraint connecting* $\boldsymbol{x}$, $\boldsymbol{y}$, *and* $\boldsymbol{W}$ *as follows:*

$$\begin{aligned}
&\underline{\boldsymbol{x}} = [0,0]^\top,\ \overline{\boldsymbol{x}} = [2,4]^\top,\ \boldsymbol{f}_0 = [1,2]^\top,\ \boldsymbol{f}_1 = [2,0.5]^\top,\\
&\underline{\boldsymbol{y}} = [0,0]^\top,\ \overline{\boldsymbol{y}} = [1,2]^\top,\ \boldsymbol{g}_0 = [1,1]^\top,\ \boldsymbol{g}_1 = [2,1]^\top,\\
&\boldsymbol{A}_0 = [-1,-2.5;-1,-3],\ \boldsymbol{A}_1 = [1,1;1,1],\ b = 3.
\end{aligned}$$

At each iteration, given a solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$ *to the current relaxation* $\tilde{\mathcal{K}}$, *we obtain the left- and right-singular vectors* $\boldsymbol{u}$ *and* $\boldsymbol{v}$, *respectively, corresponding to the largest singular value of* $\hat{\boldsymbol{W}} - \hat{\boldsymbol{x}}\hat{\boldsymbol{y}}^\top$. *Then, using* vec$(\boldsymbol{u}\boldsymbol{v}^\top)$, *we form a* (2×2)*-way disjunction (7) and obtain a disjunctive cut using the CGLP (9), stated in Lemma 1. In order to form the disjunction, we choose the breakpoints* $\beta_{1,2}$ *and* $\beta_{2,2}$ *using the current solution as follows:* $\beta_{1,2} = \frac{\boldsymbol{u}^\top \hat{\boldsymbol{x}} + \boldsymbol{v}^\top \hat{\boldsymbol{y}}}{2}$ *and* $\beta_{2,2} = \frac{\boldsymbol{u}^\top \hat{\boldsymbol{x}} - \boldsymbol{v}^\top \hat{\boldsymbol{y}}}{2}$. *The other breakpoints are due to the lower and upper bounds of the corresponding functions over* $\tilde{\mathcal{K}}$. *We refer to this particular implementation of the methodology described in Fampa and Lee (2018), as "*`SVD`*" (standing for singular value decomposition). When this algorithm terminates after 77 iterations, we obtain a (lower) bound* -0.5956, *while the optimal value to this problem is* -0.5. `SVD` *leaves an optimality gap of* 19.12%. *Figure 1 depicts evolution of the lower bound over the iteration number.*

Alternatively, consider a procedure that generates disjunctive cuts based on standard bases of $\mathbb{R}^2$, *and for all* $i \in \{1,2\}$, $j \in \{1,2\}$ *such that* $\hat{W}_{ij} \neq \hat{x}_i \hat{y}_j$. *Using* vec$(\boldsymbol{e}_i \boldsymbol{e}_j)$, *we form a* (2×2)*-way disjunction, where the breakpoints* $\beta_{1,2}$ *and* $\beta_{2,2}$ *are as follows:* $\beta_{1,2} = \frac{\hat{x}_i + \hat{y}_j}{2}$ *and* $\beta_{2,2} = \frac{\hat{x}_i - \hat{y}_j}{2}$. *The other breakpoints are due to the lower and upper bounds of the corresponding functions over* $\tilde{\mathcal{K}}$. *We refer to this algorithm as "*`STD`*" (standing for the standard basis), and compare the results with "*`SVD`*".* `STD` *leaves an optimality gap of* 46.56%. *The results are summarized in Table 1.*

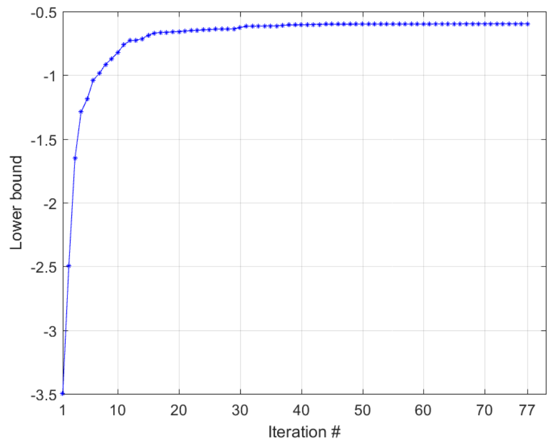


Figure 1: Lower bounds for Example 1 using `SVD`.

Table 1: Comparison of `STD` and `SVD` of $\boldsymbol{W} - \boldsymbol{x}\boldsymbol{y}^\top$ with a (2×2)-way disjunction.

`STD`		`SVD`		
Bound	Gap (%)	Bound	Gap (%)	Opt. Value
-0.7328	46.56	-0.5956	19.12	-0.5

As expected, algorithm `STD` yielded a worse lower bound than `SVD`, because `STD` is focused on only one bilinear term at a time, while `SVD` has a holistic view of all the bilinear terms.

We emphasize that the `SVD` algorithm implemented to get the results in Table 1 is not an exact reproduction of Fampa and Lee (2018) for two reasons. First, at each iteration, the authors in Fampa and Lee (2018) find four vectors $\text{vec}(\boldsymbol{u}\boldsymbol{v}^\top)$, corresponding to the four largest nonzero singular values, and hence, at most 4 disjunctive cuts are added. Second, the authors in Fampa and Lee (2018) do not clearly specify how the breakpoints are chosen, to the best of our knowledge. On the contrary, in our implementation of `SVD` here and all subsequent experiments in Section 5, we use the vector $\text{vec}(\boldsymbol{u}\boldsymbol{v}^\top)$, corresponding to the largest nonzero singular value. It goes without saying that although in Example 1 we have only two singular values, even for larger matrices $\boldsymbol{W}$, our implementation only generates one such disjunctive cut at each iteration. Moreover, again, for Example 1 and all subsequent experiments in Section 5, we use a specific construction to choose the breakpoints (see more details in section 3.2), inspired by the construction in Saxena et al. (2010) for quadratically-constrained quadratic programs.

Nevertheless, motivated by Example 1, in Section 4, we analyze a finitely-convergent algorithm to reach an ϵ-optimal solution or detect infeasibility of (BLP) using any finite collection of bases $\text{vec}(\boldsymbol{u}\boldsymbol{v}^\top)$ for $\mathbb{R}^n$ and $\mathbb{R}^m$, including the standard bases. We then extend the analyzed algorithm to the case that the bases are found through the singular value decomposition of the residual matrix $\boldsymbol{W} - \boldsymbol{x}\boldsymbol{y}^\top$, where a finite collection of bases is generated sequentially. We emphasize that our goal in this paper is not to propose an algorithm that is necessarily superior to the algorithm investigated in Fampa and Lee (2018) in terms of the computational time—partially, due to the reasons laid out above—but to analyze a fundamental modification to that algorithm that guarantees finite convergence.

3 Separating Inequalities and Minimum Distance Problem

A key observation around which this paper is developed is a reformulation of (BLP), discussed in section 3.1. Then, we discuss the cut generation component of the analyzed finitely-convergent algorithms. In section 3.2, we present valid disjunctions. In section 3.3, we describe valid separating inequalities and the corresponding projection problem.

3.1 Problem Reformulation

As explained in (4), we have

$$\mathcal{F} := \left\{ (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \mathcal{K} \;\middle|\; \begin{array}{l} \boldsymbol{u}^\top \boldsymbol{W} \boldsymbol{v} = (\boldsymbol{u}^\top \boldsymbol{x})(\boldsymbol{v}^\top \boldsymbol{y}), \\ \forall \boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{v} \in \mathbb{R}^m, \text{ with } \|\boldsymbol{u}\| = 1, \ \|\boldsymbol{v}\| = 1 \end{array} \right\}. \tag{10}$$

A set closely related set is

$$\bar{\mathcal{F}}^\epsilon := \left\{ (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \mathcal{K} \;\middle|\; \begin{array}{l} |\boldsymbol{u}^\top \boldsymbol{W} \boldsymbol{v} - (\boldsymbol{u}^\top \boldsymbol{x})(\boldsymbol{v}^\top \boldsymbol{y})| \le mn\epsilon, \\ \forall \boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{v} \in \mathbb{R}^m, \text{ with } \|\boldsymbol{u}\| = 1, \ \|\boldsymbol{v}\| = 1 \end{array} \right\}, \tag{11}$$

which is the set of ϵ-feasible solutions to $\mathcal{F}$ for $\epsilon > 0$. A key observation to analyze the algorithms in this paper is that $\mathcal{F}$ can be equivalently reformulated with a *finite* number of nonlinear constraints, corresponding to the bases of $\mathbb{R}^n$ and $\mathbb{R}^m$. We will also show that a conservative approximation to $\bar{\mathcal{F}}^\epsilon$ can be reformulated with a finite number of nonlinear constraints. Observe that if $\bar{\mathcal{F}}^\epsilon$ is an empty set, then, $\mathcal{F}$ is an empty set as well. Also, note that for both sets $\mathcal{F}$ and $\bar{\mathcal{F}}^\epsilon$, we restrict $\boldsymbol{u}$ and $\boldsymbol{v}$ to be unit vectors, without loss of generality. In the remainder of the paper, we may implicitly drop these restrictions from the set definition to simplify the exposition.

Proposition 1. *Let $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ denote a set of mutually orthonormal vectors in $\mathbb{R}^n$, and $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_m\}$ denote a set of mutually orthonormal vectors in $\mathbb{R}^m$. Then,* (BLP) *can be equivalently written as*

$$\begin{aligned} \min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W}) \in \mathcal{K}} \quad & \boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{A}_0 \bullet \boldsymbol{W} \\ s.t. \quad & \boldsymbol{u}_i^\top \boldsymbol{W} \boldsymbol{v}_j = (\boldsymbol{u}_i^\top \boldsymbol{x})(\boldsymbol{v}_j^\top \boldsymbol{y}), \ i \in [n], \ j \in [m]. \end{aligned} \tag{BLP}$$

Proof. Observe that $\boldsymbol{W} = \boldsymbol{x}\boldsymbol{y}^\top \Rightarrow \boldsymbol{u}^\top \boldsymbol{W} \boldsymbol{v} = (\boldsymbol{u}^\top \boldsymbol{x})(\boldsymbol{v}^\top \boldsymbol{y}) \ \forall \boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{v} \in \mathbb{R}^m$, including $\boldsymbol{u}_i$, $\boldsymbol{v}_j$, $i \in [n]$, $j \in [m]$. We show that if $\boldsymbol{u}_i^\top \boldsymbol{W} \boldsymbol{v}_j = (\boldsymbol{u}_i^\top \boldsymbol{x})(\boldsymbol{v}_j^\top \boldsymbol{y})$, $\forall i \in [n]$, $j \in [m] \Rightarrow \boldsymbol{W} = \boldsymbol{x}\boldsymbol{y}^\top$. Because $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ is orthonormal, any $\boldsymbol{u} \in \mathbb{R}^n$ can be written as $\boldsymbol{u} = \sum_{i=1}^n \lambda_i \boldsymbol{u}_i$ for some $\boldsymbol{\lambda} \in \mathbb{R}^n$.

Similarly, any $\boldsymbol{v} \in \mathbb{R}^m$ can be written as $\boldsymbol{v} = \sum_{j=1}^m \mu_j \boldsymbol{v}_j$ for some $\boldsymbol{\mu} \in \mathbb{R}^m$. Thus, we have

$$
\begin{aligned}
&\boldsymbol{u}_i^\top \boldsymbol{W} \boldsymbol{v}_j = (\boldsymbol{u}_i^\top \boldsymbol{x})(\boldsymbol{v}_j^\top \boldsymbol{y}), \ \forall i \in [n], \ j \in [m] \\
\Rightarrow &\left(\sum_{i=1}^n \lambda_i \boldsymbol{u}_i^\top\right) \boldsymbol{W} \boldsymbol{v}_j = \left(\sum_{i=1}^n \lambda_i \boldsymbol{u}_i^\top \boldsymbol{x}\right) \left(\boldsymbol{v}_j^\top \boldsymbol{y}\right), \ \forall \boldsymbol{\lambda} \in \mathbb{R}^n, \ j \in [m], \\
\Rightarrow &\left(\sum_{i=1}^n \lambda_i \boldsymbol{u}_i^\top\right) \boldsymbol{W} \left(\sum_{i=1}^n \mu_j \boldsymbol{v}_j\right) = \left(\sum_{i=1}^n \lambda_i \boldsymbol{u}_i^\top \boldsymbol{x}\right) \left(\sum_{j=1}^m \mu_j \boldsymbol{v}_j^\top \boldsymbol{y}\right), \ \forall \boldsymbol{\lambda} \in \mathbb{R}^n, \ \boldsymbol{\mu} \in \mathbb{R}^m \\
\Rightarrow &\boldsymbol{u}^\top \boldsymbol{W} \boldsymbol{v} = (\boldsymbol{u}^\top \boldsymbol{x})(\boldsymbol{v}^\top \boldsymbol{y}), \ \forall \boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{v} \in \mathbb{R}^m \\
\Rightarrow &\boldsymbol{W} = \boldsymbol{x}\boldsymbol{y}^\top,
\end{aligned}
$$

by taking $\boldsymbol{u}$ and $\boldsymbol{v}$ be the bases vectors. Consequently, the result follows. $\square$

Proposition 2. *Let $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ denote a set of mutually orthonormal vectors in $\mathbb{R}^n$, and $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ denote a set of mutually orthonormal vectors in $\mathbb{R}^m$. Then, if $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \mathcal{F}^\epsilon$, then $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \bar{\mathcal{F}}^\epsilon$, where*

$$
\mathcal{F}^\epsilon := \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \mathcal{K} \,|\, |\boldsymbol{u}_i^\top \boldsymbol{W} \boldsymbol{v}_j - (\boldsymbol{u}_i^\top \boldsymbol{x})(\boldsymbol{v}_j^\top \boldsymbol{y})| \le \epsilon, \ i \in [n], j \in [m]\}. \tag{12}
$$

Proof. Consider unit vectors $\boldsymbol{u} \in \mathbb{R}^n$ and $\boldsymbol{v} \in \mathbb{R}^m$. Let us write $\boldsymbol{u} = \sum_{i=1}^n \lambda_i \boldsymbol{u}_i$ for some $\boldsymbol{\lambda} \in \mathbb{R}^n$, and $\boldsymbol{v} = \sum_{j=1}^m \mu_j \boldsymbol{v}_j$ for some $\boldsymbol{\mu} \in \mathbb{R}^m$. First, note that we have $\boldsymbol{u}^\top \boldsymbol{u} = \boldsymbol{\lambda}^\top \boldsymbol{U}^\top \boldsymbol{U} \boldsymbol{\lambda}$, where $\boldsymbol{U}$ is a matrix whose columns are $\boldsymbol{u}_i$, $i = 1, \ldots, n$. Because $\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix, and $\boldsymbol{u}^\top \boldsymbol{u} = 1$, we have $\boldsymbol{\lambda}^\top \boldsymbol{\lambda} = 1$. Thus, $|\lambda_i| \le 1$ for $i \in [n]$. Similarly, we have $|\mu_j| \le 1$ for $j \in [m]$. Observe that

$$
\begin{aligned}
\boldsymbol{u}^\top \boldsymbol{W} \boldsymbol{v} - (\boldsymbol{u}^\top \boldsymbol{x})(\boldsymbol{v}^\top \boldsymbol{y}) =& \boldsymbol{u}^\top (\boldsymbol{W} - \boldsymbol{x}\boldsymbol{y}^\top) \boldsymbol{v} \\
=& \left(\sum_{i=1}^n \lambda_i \boldsymbol{u}_i\right)^\top (\boldsymbol{W} - \boldsymbol{x}\boldsymbol{y}^\top) \left(\sum_{j=1}^m \mu_j \boldsymbol{v}_j\right) \\
=& \sum_{i=1}^n \sum_{j=1}^m \lambda_i \mu_j \boldsymbol{u}_i^\top (\boldsymbol{W} - \boldsymbol{x}\boldsymbol{y}^\top) \boldsymbol{v}_j.
\end{aligned}
$$

Thus, for $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \mathcal{F}^\epsilon$, we have

$$
\begin{aligned}
\left|\boldsymbol{u}^\top \boldsymbol{W} \boldsymbol{v} - (\boldsymbol{u}^\top \boldsymbol{x})(\boldsymbol{v}^\top \boldsymbol{y})\right| \le& \left|\sum_{i=1}^n \sum_{j=1}^m \lambda_i \mu_j \boldsymbol{u}_i^\top (\boldsymbol{W} - \boldsymbol{x}\boldsymbol{y}^\top) \boldsymbol{v}_j\right| \\
\le& \sum_{i=1}^n \sum_{j=1}^m \left|\lambda_i \mu_j \boldsymbol{u}_i^\top (\boldsymbol{W} - \boldsymbol{x}\boldsymbol{y}^\top) \boldsymbol{v}_j\right| \\
\le& \max_{i=1}^n \max_{j=1}^m |\lambda_i \mu_j| \times \sum_{i=1}^n \sum_{j=1}^m \left|\boldsymbol{u}_i^\top (\boldsymbol{W} - \boldsymbol{x}\boldsymbol{y}^\top) \boldsymbol{v}_j\right|
\end{aligned}
$$

$$\leq \sum_{i=1}^{n}\sum_{j=1}^{m}\left|\boldsymbol{u}_i^\top(\boldsymbol{W}-\boldsymbol{x}\boldsymbol{y}^\top)\boldsymbol{v}_j\right|$$
$$\leq mn\epsilon.$$

This completes the proof. □

Throughout this section, we assume that $\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n\}$ is a set of mutually orthonormal vectors in $\mathbb{R}^n$, and $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_m\}$ is a set of mutually orthonormal vectors in $\mathbb{R}^m$. For the ease of exposition, let the index a represent the (i,j)-pair, where $a\in[nm]$. For the index a, we denote $\text{vec}(\boldsymbol{u}_a\boldsymbol{v}_a^\top)$ by $\boldsymbol{c}_a$.

3.2 Valid Single-Vector Disjunction

Consider a (bounded) convex relaxation $\tilde{\mathcal{K}}$ of $\mathcal{F}$ and $\boldsymbol{c}_a$, $a\in[nm]$. Let us define the following set:

$$\mathcal{P}_a(\tilde{\mathcal{K}},\boldsymbol{\beta}) := \text{conv}\left(\vee_{k=1}^4 \tilde{\mathcal{S}}_k(\boldsymbol{c}_a,\tilde{\mathcal{K}},\boldsymbol{\beta})\right). \tag{13}$$

A cutting plane is generated based on a single-vector disjunction. Given a relaxation $\tilde{\mathcal{K}}$ of $\mathcal{F}$, let $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}})$ be an optimal extreme point solution to the minimization problem over $\tilde{\mathcal{K}}$ that needs to be cut off by a valid linear inequality. In particular, suppose that the solution $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}})$ is not satisfying the constraint $\boldsymbol{u}_a^\top\boldsymbol{W}\boldsymbol{v}_a=(\boldsymbol{u}_a^\top\boldsymbol{x})(\boldsymbol{v}_a^\top\boldsymbol{y})$ for some $a\in[nm]$. Generating a valid inequality accounts for finding a separating hyperplane that separates $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}})$ from $\mathcal{P}_a(\tilde{\mathcal{K}},\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a proper choice of breakpoints. For the rest of the paper, we choose the breakpoints $\boldsymbol{\beta}$ for a (2×2)-way disjunction in a specific manner, detailed in Construction 1.

Construction 1. *Consider a bounded convex relaxation $\tilde{\mathcal{K}}$ of $\mathcal{F}$. Let $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}})\in\tilde{\mathcal{K}}$ be such that $\boldsymbol{u}_a^\top\hat{\boldsymbol{W}}\boldsymbol{v}_a\neq(\boldsymbol{u}_a^\top\hat{\boldsymbol{x}})(\boldsymbol{v}_a^\top\hat{\boldsymbol{y}})$ for some $a\in[nm]$. We form a (2×2)-way disjunction of the form (7) on $\tilde{\mathcal{K}}$, based on $\boldsymbol{c}_a=\text{vec}(\boldsymbol{u}_a\boldsymbol{v}_a^\top)$ and $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}})$, using the following choice of the breakpoints:*

$$\begin{aligned}
&\beta_{1,1}=\min\{\tfrac{\boldsymbol{u}_a^\top\boldsymbol{x}+\boldsymbol{v}_a^\top\boldsymbol{y}}{2}\,|\,(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\tilde{\mathcal{K}}\}, &&\beta_{2,1}=\min\{\tfrac{\boldsymbol{u}_a^\top\boldsymbol{x}-\boldsymbol{v}_a^\top\boldsymbol{y}}{2}\,|\,(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\tilde{\mathcal{K}}\},\\
&\beta_{1,2}=\tfrac{\boldsymbol{u}_a^\top\hat{\boldsymbol{x}}+\boldsymbol{v}_a^\top\hat{\boldsymbol{y}}}{2}, &&\beta_{2,2}=\tfrac{\boldsymbol{u}_a^\top\hat{\boldsymbol{x}}-\boldsymbol{v}_a^\top\hat{\boldsymbol{y}}}{2},\\
&\beta_{1,3}=\max\{\tfrac{\boldsymbol{u}_a^\top\boldsymbol{x}+\boldsymbol{v}_a^\top\boldsymbol{y}}{2}\,|\,(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\tilde{\mathcal{K}}\}, &&\beta_{2,3}=\max\{\tfrac{\boldsymbol{u}_a^\top\boldsymbol{x}-\boldsymbol{v}_a^\top\boldsymbol{y}}{2}\,|\,(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\tilde{\mathcal{K}}\}.
\end{aligned}$$

Lemma 2 shows that the choice of breakpoints $\boldsymbol{\beta}$ based on Construction 1 leads to a single-vector disjunction that can be used to generate a valid disjunctive cut to cut off an extreme point $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}})$, where $\boldsymbol{u}_a^\top\hat{\boldsymbol{W}}\boldsymbol{v}_a\neq(\boldsymbol{u}_a^\top\hat{\boldsymbol{x}})(\boldsymbol{v}_a^\top\hat{\boldsymbol{y}})$ for some $a\in[mn]$.

Lemma 2. *Consider a bounded convex relaxation $\tilde{\mathcal{K}}$ of $\mathcal{F}$. Let $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}})\in\tilde{\mathcal{K}}$ be an extreme point solution such that $|\boldsymbol{u}_a^\top\hat{\boldsymbol{W}}\boldsymbol{v}_a-(\boldsymbol{u}_a^\top\hat{\boldsymbol{x}})(\boldsymbol{v}_a^\top\hat{\boldsymbol{y}})|>\epsilon$ for some $\epsilon>0$ and some $a\in[nm]$. Let $\vee_{k=1}^4\tilde{S}_k(\boldsymbol{c}_a,\tilde{\mathcal{K}},\boldsymbol{\beta})$ be a (2×2)-way disjunction, where the breakpoints are chosen as in Construction 1 and using $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}})$. Then, $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}})\notin\tilde{S}_k(\boldsymbol{c}_a,\tilde{\mathcal{K}},\boldsymbol{\beta})$ for all $k\in\{1,2,3,4\}$. Moreover, for all $k\in\{1,2,3,4\}$, one of the secant inequalities in $\tilde{S}_k(\boldsymbol{c}_a,\tilde{\mathcal{K}},\boldsymbol{\beta})$, defined in (8), is violated with an amount greater than ϵ by $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}})$.*

Proof. We first show that $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \notin \tilde{S}_k(\boldsymbol{c}_a, \tilde{\mathcal{K}}, \boldsymbol{\beta})$ for all $k \in \{1,2,3,4\}$, consequently, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \notin \vee_{k=1}^4 \tilde{S}_k(\boldsymbol{c}_a, \tilde{\mathcal{K}}, \boldsymbol{\beta})$. Suppose by contradiction that $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \in \tilde{S}_k(\boldsymbol{c}_a, \tilde{\mathcal{K}}, \boldsymbol{\beta})$ for some $\hat{k} \in \{1,2,3,4\}$. Without loss of generality, suppose that $\tilde{S}_{\hat{k}}(\boldsymbol{c}_a, \tilde{\mathcal{K}}, \boldsymbol{\beta})$ in (8) is such that $\beta_{1,1}$ and $\beta_{1,2}$ are the lower and upper bounds of $\frac{\boldsymbol{u}_a^\top \boldsymbol{x} + \boldsymbol{v}_a^\top \boldsymbol{y}}{2}$, respectively. Moreover, $\beta_{2,1}$ and $\beta_{2,2}$ are the lower and upper bounds of $\frac{\boldsymbol{u}_a^\top \boldsymbol{x} - \boldsymbol{v}_a^\top \boldsymbol{y}}{2}$. Because $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \in \tilde{\mathcal{S}}_{\hat{k}}(\boldsymbol{c}, \tilde{\mathcal{K}}, \boldsymbol{\beta})$, we have

$$\begin{aligned} &\boldsymbol{u}_a^\top \hat{\boldsymbol{W}} \boldsymbol{v}_a - \left(\frac{\boldsymbol{u}_a^\top \hat{\boldsymbol{x}} + \boldsymbol{v}_a^\top \hat{\boldsymbol{y}}}{2}\right)(\beta_{1,1} + \frac{\boldsymbol{u}_a^\top \hat{\boldsymbol{x}} + \boldsymbol{v}_a^\top \hat{\boldsymbol{y}}}{2}) + \beta_{1,1}\frac{\boldsymbol{u}_a^\top \hat{\boldsymbol{x}} + \boldsymbol{v}_a^\top \hat{\boldsymbol{y}}}{2} + \\ &\left(\frac{\boldsymbol{u}_a^\top \hat{\boldsymbol{x}} - \boldsymbol{v}_a^\top \hat{\boldsymbol{y}}}{2}\right)^2 = \boldsymbol{u}_a^\top \hat{\boldsymbol{W}} \boldsymbol{v}_a - \left(\frac{\boldsymbol{u}_a^\top \hat{\boldsymbol{x}} + \boldsymbol{v}_a^\top \hat{\boldsymbol{y}}}{2}\right)^2 + \left(\frac{\boldsymbol{u}_a^\top \hat{\boldsymbol{x}} - \boldsymbol{v}_a^\top \hat{\boldsymbol{y}}}{2}\right)^2 \le 0. \end{aligned} \tag{14}$$

Note that if $\tilde{\mathcal{S}}_{\hat{k}}(\boldsymbol{c}, \tilde{\mathcal{K}}, \boldsymbol{\beta})$ is such that $\beta_{1,2} \le \frac{\boldsymbol{u}^\top \boldsymbol{x} + \boldsymbol{v}^\top \boldsymbol{y}}{2} \le \beta_{1,3}$, we would still get a similar conclusion as in (14). So, the definition of $\tilde{\mathcal{S}}_{\hat{k}}(\boldsymbol{c}, \tilde{\mathcal{K}}, \boldsymbol{\beta})$ as above is without loss of generality. With a similar argument, we conclude $-\boldsymbol{u}_a^\top \hat{\boldsymbol{W}} \boldsymbol{v}_a + \left(\frac{\boldsymbol{u}_a^\top \hat{\boldsymbol{x}} + \boldsymbol{v}_a^\top \hat{\boldsymbol{y}}}{2}\right)^2 - \left(\frac{\boldsymbol{u}_a^\top \hat{\boldsymbol{x}} - \boldsymbol{v}_a^\top \hat{\boldsymbol{y}}}{2}\right)^2 \le 0$. Thus, the above two inequalities imply that $\boldsymbol{u}_a^\top \hat{\boldsymbol{W}} \boldsymbol{v}_a = (\boldsymbol{u}_a^\top \hat{\boldsymbol{x}})(\boldsymbol{v}_a^\top \hat{\boldsymbol{y}})$, yielding a contradiction.

Now, we show that for all $k \in \{1,2,3,4\}$, one of the secant inequalities in $\tilde{S}_k(\boldsymbol{c}_a, \tilde{\mathcal{K}}, \boldsymbol{\beta})$ is violated by $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$ and the amount of violation is greater than ϵ. First, suppose that $\boldsymbol{u}_a^\top \hat{\boldsymbol{W}} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \hat{\boldsymbol{x}})(\boldsymbol{v}_a^\top \hat{\boldsymbol{y}}) > \epsilon$. Using the equality $(\boldsymbol{u}_a^\top \hat{\boldsymbol{x}})(\boldsymbol{v}_a^\top \hat{\boldsymbol{y}}) = \left(\frac{\boldsymbol{u}_a^\top \hat{\boldsymbol{x}} + \boldsymbol{v}_a^\top \hat{\boldsymbol{y}}}{2}\right)^2 - \left(\frac{\boldsymbol{u}_a^\top \hat{\boldsymbol{x}} - \boldsymbol{v}_a^\top \hat{\boldsymbol{y}}}{2}\right)^2$, we have $\boldsymbol{u}_a^\top \hat{\boldsymbol{W}} \boldsymbol{v}_a - \left(\frac{\boldsymbol{u}_a^\top \hat{\boldsymbol{x}} + \boldsymbol{v}_a^\top \hat{\boldsymbol{y}}}{2}\right)^2 + \left(\frac{\boldsymbol{u}_a^\top \hat{\boldsymbol{x}} - \boldsymbol{v}_a^\top \hat{\boldsymbol{y}}}{2}\right)^2 > \epsilon$. The left-hand side of this inequality is the left-hand side of (14), implying that $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$ violates the first secant inequality of $\tilde{S}_k(\boldsymbol{c}_a, \tilde{\mathcal{K}}, \boldsymbol{\beta})$ for all $k \in \{1,2,3,4\}$, and the amount of violation is greater than ϵ.

Now, suppose that $-\boldsymbol{u}_a^\top \hat{\boldsymbol{W}} \boldsymbol{v}_a + (\boldsymbol{u}_a^\top \hat{\boldsymbol{x}})(\boldsymbol{v}_a^\top \hat{\boldsymbol{y}}) > \epsilon$. Similarly, we conclude that $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$ violates the second secant inequality of $\tilde{S}_k(\boldsymbol{c}_a, \tilde{\mathcal{K}}, \boldsymbol{\beta})$ for all $k \in \{1,2,3,4\}$, and the amount of violation is greater than ϵ. □

3.3 Valid Disjunctive Cut and Projection Problem

So far, we established our construction to choose the breakpoints $\boldsymbol{\beta}$. We now show that the choice of breakpoints $\boldsymbol{\beta}$ based on Construction 1 leads to a valid disjunctive cut. To obtain a valid cut for $\operatorname{conv}(\mathcal{F})$ using a (2×2)-way disjunction, one may solve the corresponding CGLP, introduced in Lemma 1. As mentioned before, this CGLP contains the outer approximation to the convex quadratic terms. Alternatively, one can solve a projection problem that minimizes the distance, measured by some norm, from $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$ to a point in $\mathcal{P}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})$.

Proposition 3. *Consider a bounded convex relaxation $\tilde{\mathcal{K}}$ of $\mathcal{F}$. Let $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$ be an optimal extreme point solution of $\min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W}) \in \tilde{\mathcal{K}}} \ \boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{A}_0 \bullet \boldsymbol{W}$. Suppose that $\boldsymbol{u}_a^\top \hat{\boldsymbol{W}} \boldsymbol{v}_a \neq (\boldsymbol{u}_a^\top \hat{\boldsymbol{x}})(\boldsymbol{v}_a^\top \hat{\boldsymbol{y}})$ for some $a \in [nm]$. Moreover, let $\vee_{k=1}^4 \tilde{S}_k(\boldsymbol{c}_a, \tilde{\mathcal{K}}, \boldsymbol{\beta})$ be a (2×2)-way disjunction, where the breakpoints are chosen as in Construction 1 and using $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$. Furthermore, suppose that $\mathcal{P}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})$, defined*

Algorithm 1 $\texttt{SepCuts}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}; \tilde{\mathcal{K}}, \boldsymbol{u}, \boldsymbol{v})$

Input: $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$, $\tilde{\mathcal{K}}$, and $\text{vec}(\boldsymbol{u}\boldsymbol{v}^\top)$.

Output: $(\texttt{viol}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\Gamma}, \rho)$. If a valid inequality $\boldsymbol{\alpha}^\top \boldsymbol{x} + \boldsymbol{\theta}^\top \boldsymbol{y} + \boldsymbol{\Gamma} \bullet \boldsymbol{W} \geq \rho$ is found that is violated by $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$, then return $\texttt{viol}$=TRUE, $\boldsymbol{\alpha}$, $\boldsymbol{\theta}$, $\boldsymbol{\Gamma}$, and ρ. Otherwise, return $\texttt{viol}$=FALSE, $\boldsymbol{\alpha} = \boldsymbol{0}$, $\boldsymbol{\theta} = \boldsymbol{0}$, $\boldsymbol{\Gamma} = \boldsymbol{0}$, and $\rho = 0$.

Let $\boldsymbol{\alpha} = \boldsymbol{0}$, $\boldsymbol{\theta} = \boldsymbol{0}$, $\boldsymbol{\Gamma} = \boldsymbol{0}$, and $\rho = 0$.

Let $\boldsymbol{c} = \text{vec}(\boldsymbol{u}\boldsymbol{v}^\top)$ and $\boldsymbol{\beta}$ is chosen as $\beta_{1,1} = \min\{\frac{\boldsymbol{u}^\top \boldsymbol{x} + \boldsymbol{v}^\top \boldsymbol{y}}{2} \,|\, (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \tilde{\mathcal{K}}\}$, $\beta_{1,3} = \max\{\frac{\boldsymbol{u}^\top \boldsymbol{x} + \boldsymbol{v}^\top \boldsymbol{y}}{2} \,|\, (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \tilde{\mathcal{K}}\}$, and $\beta_{1,2} = \frac{\boldsymbol{u}^\top \hat{\boldsymbol{x}} + \boldsymbol{v}^\top \hat{\boldsymbol{y}}}{2}$. Also, $\beta_{2,1} = \min\{\frac{\boldsymbol{u}^\top \boldsymbol{x} - \boldsymbol{v}^\top \boldsymbol{y}}{2} \,|\, (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \tilde{\mathcal{K}}\}$, $\beta_{2,3} = \max\{\frac{\boldsymbol{u}^\top \boldsymbol{x} - \boldsymbol{v}^\top \boldsymbol{y}}{2} \,|\, (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \tilde{\mathcal{K}}\}$, and $\beta_{2,2} = \frac{\boldsymbol{u}^\top \hat{\boldsymbol{x}} - \boldsymbol{v}^\top \hat{\boldsymbol{y}}}{2}$.

if $\mathcal{S}_k(\boldsymbol{c}, \tilde{\mathcal{K}}, \boldsymbol{\beta}) = \emptyset$ for all $k \in \{1, 2, 3, 4\}$ **then**

    $\texttt{viol} \leftarrow$ FALSE.

else

    Let $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{W}^*)$ be an optimal solution to $\min_{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \mathcal{P}_{\boldsymbol{c}}(\tilde{\mathcal{K}}, \boldsymbol{\beta})} \|(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) - (\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})\|$.

    Let $\boldsymbol{\alpha}$, $\boldsymbol{\theta}$, and $\boldsymbol{\Gamma}$ be partial subgradients of $\|(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) - (\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})\|$ at $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{W}^*)$ with respect to $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{W}$, respectively. Let $\rho = \boldsymbol{\alpha}^\top \boldsymbol{x}^* + \boldsymbol{\theta}^\top \boldsymbol{y}^* + \boldsymbol{\Gamma} \bullet \boldsymbol{W}^*$.

    $\texttt{viol} \leftarrow$ TRUE.

end if

in (13)*, is nonempty. Then, the following projection problem*

$$\min_{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \mathcal{P}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})} \|(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) - (\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})\| \tag{15}$$

has a strictly positive and finite optimal value.

Proof. By Lemma 2, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \notin \tilde{\mathcal{S}}_k(\boldsymbol{c}_a, \tilde{\mathcal{K}}, \boldsymbol{\beta})$, $k \in \{1, 2, 3, 4\}$. Thus, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \notin \vee_{k=1}^4 \tilde{\mathcal{S}}_k(\boldsymbol{c}_a, \tilde{\mathcal{K}}, \boldsymbol{\beta})$. Because $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$ is an extreme point of $\tilde{\mathcal{K}}$, it cannot be written as a convex combination of points in $\tilde{\mathcal{K}}$, including the points in $\vee_{k=1}^4 \tilde{\mathcal{S}}_k(\boldsymbol{c}_a, \tilde{\mathcal{K}}, \boldsymbol{\beta})$. Thus, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \notin \mathcal{P}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})$. Because, $\mathcal{P}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta}) \neq \emptyset$, then, the optimal value to (15) is finite. Moreover, by Ruszczyński (2006, Theorem 2.14), $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$ can be strongly separated from $\mathcal{P}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})$. By Rockafellar (1970, Theorem 11.4), the strong separation holds if and only if the optimal value to (15) is strictly positive. □

The implication of Proposition 3 is that by choosing the breakpoints $\boldsymbol{\beta}$ according to Construction 1, one can separate $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$ from $\mathcal{P}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})$, and consequently, from $\text{conv}\,(\mathcal{F})$. We summarize the cut generation procedure in Algorithm 1. In Section 4, we describe how these cutting planes can be utilized in an algorithmic fashion to obtain an ϵ-optimal solution to (BLP) or to detect infeasibility in a finite number of iterations.

One may ask how to describe $\mathcal{P}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})$ in the projection problem (15). We devote the rest of this section to answer this question.

Assume that the convex quadratic-representable set $\tilde{\mathcal{S}}_k(\boldsymbol{c}_a, \tilde{\mathcal{K}}, \boldsymbol{\beta})$ is written as

$$\tilde{\mathcal{S}}_k(\boldsymbol{c}_a, \tilde{\mathcal{K}}, \boldsymbol{\beta}) = \left\{ (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \tilde{\mathcal{K}} \,\middle|\, \boldsymbol{F}_{a,k}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}, \boldsymbol{\beta}) \leq \boldsymbol{0} \right\},$$

for $k = 1, 2, 3, 4$. Let us define a binary variable z_k, where $z_k = 0$ implies that $\boldsymbol{F}_{a,k}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}, \boldsymbol{\beta}) \leq \boldsymbol{0}$.

We define the convex set

$$Z_a(\tilde{\mathcal{K}},\boldsymbol{\beta}) := \left\{ (\boldsymbol{x},\boldsymbol{y},\boldsymbol{W},\boldsymbol{z}) \,\middle|\, \begin{array}{l} (\boldsymbol{x},\boldsymbol{y},\boldsymbol{W}) \in \tilde{\mathcal{K}},\ \sum_{k=1}^4 (1-z_k) \geq 1, \\ \boldsymbol{F}_{a,k}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W},\boldsymbol{\beta}) - z_k \boldsymbol{M}_{a,k} \leq \boldsymbol{0}, \\ 0 \leq z_k \leq 1,\ k=1,2,3,4 \end{array} \right\}, \tag{16}$$

where $\boldsymbol{M}_{a,k}$ is a sufficiently large vector to ensure that when $z_k = 1$, constraints $\boldsymbol{F}_{a,k}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W},\boldsymbol{\beta}) \leq \boldsymbol{0}$ are not active. Such a vector exists because $\|\boldsymbol{u}_a\| = \|\boldsymbol{v}_a\| = 1$. Also, we define the mixed-binary convex set

$$\mathcal{Z}_a(\tilde{\mathcal{K}},\boldsymbol{\beta}) := \left\{ (\boldsymbol{x},\boldsymbol{y},\boldsymbol{W},\boldsymbol{z}) \in Z_a(\tilde{\mathcal{K}},\boldsymbol{\beta}) \,\middle|\, z_k \in \{0,1\},\ k=1,2,3,4 \right\}. \tag{17}$$

Proposition 4. *Consider a relaxation $\tilde{\mathcal{K}}$ and $\boldsymbol{c}_a$, $a \in [nm]$. Let $Z_a(\tilde{\mathcal{K}},\boldsymbol{\beta})$ and $\mathcal{Z}_a(\tilde{\mathcal{K}},\boldsymbol{\beta})$ be defined as in (16) and (17), respectively. For a binary variable $z_{a,k}$, $k=1,2,3,4$, define $Z_a^0(\tilde{\mathcal{K}},\boldsymbol{\beta}) := \{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W},\boldsymbol{z}) \in Z_a(\tilde{\mathcal{K}},\boldsymbol{\beta}) \,|\, z_k = 0\}$ and $Z_a^1(\tilde{\mathcal{K}},\boldsymbol{\beta}) := \{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W},\boldsymbol{z}) \in Z_a(\tilde{\mathcal{K}},\boldsymbol{\beta}) \,|\, z_k = 1\}$. Let $\mathcal{M}_k\big(Z_a(\tilde{\mathcal{K}},\boldsymbol{\beta})\big) :=$*

$$\left\{ (\boldsymbol{x},\boldsymbol{y},\boldsymbol{W},\boldsymbol{z},\boldsymbol{b}_0,\boldsymbol{b}_1,\lambda_0,\lambda_1) \,\middle|\, \begin{array}{l} (\boldsymbol{x},\boldsymbol{y},\boldsymbol{W},\boldsymbol{z}) = \lambda_0 \boldsymbol{b}_0 + \lambda_1 \boldsymbol{b}_1, \\ \lambda_0 + \lambda_1 = 1,\ \lambda_0 \geq 0,\ \lambda_1 \geq 0, \\ \boldsymbol{b}_0 \in Z_a^0(\tilde{\mathcal{K}},\boldsymbol{\beta}),\ \boldsymbol{b}_1 \in Z_a^1(\tilde{\mathcal{K}},\boldsymbol{\beta}) \end{array} \right\}.$$

Let $\pi_k\big(Z_a(\tilde{\mathcal{K}},\boldsymbol{\beta})\big)$ be the projection of $\mathcal{M}_k\big(Z_a(\tilde{\mathcal{K}},\boldsymbol{\beta})\big)$ onto the $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W},\boldsymbol{z})$-space. Then, $conv\Big(\mathcal{Z}_a(\tilde{\mathcal{K}},\boldsymbol{\beta})\Big)$ $= \pi_4\Big(\pi_3\Big(\pi_2\Big(\pi_1\big(Z_a(\tilde{\mathcal{K}},\boldsymbol{\beta})\big)\Big)\Big)\Big)$. Moreover, we have $\mathcal{P}_a(\tilde{\mathcal{K}},\boldsymbol{\beta}) = Proj_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\Big(conv\Big(\mathcal{Z}_a(\tilde{\mathcal{K}},\boldsymbol{\beta})\Big)\Big)$, where $\mathcal{P}_a(\tilde{\mathcal{K}},\boldsymbol{\beta})$ is defined in (13). That is,

$$conv\Big(Proj_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\big(\mathcal{Z}_a(\tilde{\mathcal{K}},\boldsymbol{\beta})\big)\Big) = Proj_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\Big(conv\Big(\mathcal{Z}_a(\tilde{\mathcal{K}},\boldsymbol{\beta})\Big)\Big).$$

To prove Proposition 4, we first provide a technical result, followed by a result on the sequential convexification of a mixed-binary convex set.

Lemma 3. *Let $\mathcal{C}$ be a compact convex set in $\mathbb{R}^{\nu_1}$ defined as $\mathcal{C} := Proj_{\boldsymbol{a}}(\mathcal{U})$, where $\mathcal{U} \subseteq \mathbb{R}^{\nu_1} \times \mathbb{R}^{\nu_2}$ is a compact convex set. If $\hat{\boldsymbol{a}} \in ext(\mathcal{C})$, then, there exists $\hat{\boldsymbol{b}}$ such that $(\hat{\boldsymbol{a}},\hat{\boldsymbol{b}}) \in ext(conv(\mathcal{U}))$.*

Proof. Suppose by contradiction that for all $\hat{\boldsymbol{b}}$ such that $(\hat{\boldsymbol{a}},\hat{\boldsymbol{b}}) \in \mathrm{conv}(\mathcal{U})$, we have $(\hat{\boldsymbol{a}},\hat{\boldsymbol{b}}) \notin \mathrm{ext}(\mathrm{conv}(\mathcal{U}))$. Thus, by the Carathéodory theorem, we can write $(\hat{\boldsymbol{a}},\hat{\boldsymbol{b}}) = \sum_{i=1}^{\nu_1+\nu_2+1} \lambda^i (\boldsymbol{a}^i,\boldsymbol{b}^i)$, where for $i \in [\nu_1+\nu_2+1]$, $(\boldsymbol{a}^i,\boldsymbol{b}^i) \in \mathrm{ext}(\mathrm{conv}(\mathcal{U}))$, $\lambda^i \geq 0$, and $\sum_{i=1}^{\nu_1+\nu_2+1} \lambda^i = 1$. Note that for $i \in [\nu_1+\nu_2+1]$, we have $\boldsymbol{a}^i \notin \mathrm{ext}(\mathcal{C})$, because otherwise, this contradicts the hypothesis of the contradiction. For $i \in [\nu_1+\nu_2+1]$, let $\boldsymbol{a}^i \in \mathcal{C}$ but $\boldsymbol{a}^i \notin \mathrm{ext}(\mathcal{C})$. Because $\hat{\boldsymbol{a}} = \sum_{i=1}^{\nu_1+\nu_2+1} \lambda^i \boldsymbol{a}^i$, we conclude that $\hat{\boldsymbol{a}} \notin \mathrm{ext}(\mathcal{C})$, which contradicts the hypothesis of the lemma. □

Lemma 4. Stubbs and Mehrotra (1999, *Proposition 1 and Theorem 1) Let $\mathcal{Z}$ be a compact mixed-binary set as follows*

$$\mathcal{Z} := \{\boldsymbol{z} \in Z \,|\, z_\iota \in \{0,1\},\ \iota \in [\kappa]\},$$

where

$$Z := \left\{ \boldsymbol{z} \in \mathbb{R}^{\vartheta} \,\middle|\, h_{\varsigma}(\boldsymbol{z}) \le 0,\ \varsigma \in [l],\ 0 \le z_{\iota} \le 1,\ \iota \in [\kappa] \right\},$$

is a compact, convex, continuous relaxation of $\mathcal{Z}$. *For a binary variable* z_{ι}, $\iota \in [\kappa]$, *define* $Z_{\iota}^{0} := \{\boldsymbol{z} \in Z \,|\, z_{\iota} = 0\}$, *and* $Z_{\iota}^{1} := \{\boldsymbol{z} \in Z \,|\, z_{\iota} = 1\}$. *Let*

$$\mathcal{M}_{\iota}(Z) := \left\{ (\boldsymbol{z}, \boldsymbol{b}_0, \boldsymbol{b}_1, \lambda_0, \lambda_1) \,\middle|\, \begin{array}{l} \boldsymbol{z} = \lambda_0 \boldsymbol{b}_0 + \lambda_1 \boldsymbol{b}_1,\ \boldsymbol{b}_0 \in Z_{\iota}^{0},\ \boldsymbol{b}_1 \in Z_{\iota}^{1}, \\ \lambda_0 + \lambda_1 = 1,\ \lambda_0 \ge 0,\ \lambda_1 \ge 0. \end{array} \right\}.$$

Let $\pi_{\iota}(Z)$ *be the projection of* $\mathcal{M}_{\iota}(Z)$ *onto the* $\boldsymbol{z}$*-space. Then,*

$$conv(\mathcal{Z}) = \pi_1\big(\pi_2\big(\dots\big(\pi_{\kappa}(Z)\big)\dots\big)\big).$$

Proof of Proposition 4. By a direct application of Lemma 4, we can obtain $\text{conv}\left(\mathcal{Z}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})\right)$. To complete the proof, we show that $\mathcal{P}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})$ is obtained by projecting $\text{conv}\left(\mathcal{Z}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})\right)$ onto the $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W})$-space. Let us define $\mathcal{A} := \mathcal{P}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta}) = \text{conv}\left(\text{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\big(\mathcal{Z}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})\big)\right)$ and $\mathcal{B} := \text{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\left(\text{conv}\left(\mathcal{Z}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})\right)\right)$. By the boundedness assumptions, both $\mathcal{A}$ and $\mathcal{B}$ are compact convex sets. Hence, they are completely determined by the closed convex hull of their extreme points.

"$\Longrightarrow$": Let $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \in \mathcal{A}$. By the Carathéodory theorem, we can write $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) = \sum_i \lambda_i (\hat{\boldsymbol{x}}^i, \hat{\boldsymbol{y}}^i, \hat{\boldsymbol{W}}^i)$, where $(\hat{\boldsymbol{x}}^i, \hat{\boldsymbol{y}}^i, \hat{\boldsymbol{W}}^i) \in \text{ext}\Big(\text{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\big(\mathcal{Z}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})\big)\Big)$, $\lambda_i \ge 0$, and $\sum_i \lambda_i = 1$. Using Lemma 3, there exists $\hat{\boldsymbol{z}}^i$ such that $(\hat{\boldsymbol{x}}^i, \hat{\boldsymbol{y}}^i, \hat{\boldsymbol{W}}^i, \hat{\boldsymbol{z}}^i) \in \text{ext}\Big(\text{conv}\left(\mathcal{Z}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})\right)\Big)$. Thus, we can construct a point $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}, \hat{\boldsymbol{z}}) \in \text{conv}\left(\mathcal{Z}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})\right)$ as $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}, \hat{\boldsymbol{z}}) = \sum_i \lambda_i (\hat{\boldsymbol{x}}^i, \hat{\boldsymbol{y}}^i, \hat{\boldsymbol{W}}^i, \hat{\boldsymbol{z}}^i)$. This implies that $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \in \text{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\left(\text{conv}\left(\mathcal{Z}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})\right)\right) = \mathcal{B}$.

"$\Longleftarrow$": Let $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \in \text{ext}(\mathcal{B})$. By Lemma 3, there exists $\hat{\boldsymbol{z}}$ such that $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}, \hat{\boldsymbol{z}}) \in \text{ext}\Big(\text{conv}\left(\mathcal{Z}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})\right)\Big)$. Thus, $\hat{z} \in \{0,1\}$, $k = 1,2,3,4$, and $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}, \hat{\boldsymbol{z}}) \in \mathcal{Z}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})$. Because $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}, \hat{\boldsymbol{z}}) \in \mathcal{Z}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})$, we have $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \in \text{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\left(\mathcal{Z}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})\right)$. Consequently, $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \in \text{conv}\left(\text{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\big(\mathcal{Z}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})\big)\right) = \mathcal{A}$.

□

Remark 1. *Note that in Lemma 4, the equality constraints defining* $\boldsymbol{z}$ *in* $\mathcal{M}_{\iota}(Z)$ *are nonlinear. By a transformation of variables using perspective envelopes, it is shown in Stubbs and Mehrotra (1999, Theorem 2) that the constraints defining* $\mathcal{M}_{\iota}(Z)$ *can be represented by a set of convex constraints as follows:*

$$\tilde{\mathcal{M}}_{\iota}(Z) := \left\{ (\boldsymbol{z}, \boldsymbol{d}_0, \boldsymbol{d}_1, \lambda_0, \lambda_1) \,\middle|\, \begin{array}{l} \boldsymbol{z} = \boldsymbol{d}_0 + \boldsymbol{d}_1,\ (\boldsymbol{d}_0, \lambda_0) \in \tilde{Z}_{\iota}^{0},\ (\boldsymbol{d}_1, \lambda_1) \in \tilde{Z}_{\iota}^{1}, \\ \lambda_0 + \lambda_1 = 1,\ \lambda_0 \ge 0,\ \lambda_1 \ge 0 \end{array} \right\},$$

where $\tilde{Z}_\iota^0 := \{(\tilde{\boldsymbol{z}}, \lambda) \in \tilde{Z} \mid \tilde{z}_\iota = 0\}$, *and* $\tilde{Z}_\iota^1 := \{(\tilde{\boldsymbol{z}}, \lambda) \in \tilde{Z} \mid \tilde{z}_\iota = \lambda\}$. *Moreover,* $\tilde{Z}$ *is defined as*

$$\tilde{Z} := \left\{ (\tilde{\boldsymbol{z}}, \lambda) \middle| \ \psi_\varsigma(\tilde{\boldsymbol{z}}, \lambda) \leq 0, \ \varsigma \in [l], \ 0 \leq \tilde{z}_\iota \leq \lambda, \ \iota \in [\kappa], \ 0 \leq \lambda \leq 1 \ \right\},$$

where

$$\psi_\varsigma(\tilde{\boldsymbol{z}}, \lambda) = \begin{cases} \lambda h_\varsigma(\frac{\tilde{\boldsymbol{z}}}{\lambda}) & \text{if } \frac{\tilde{\boldsymbol{z}}}{\lambda} \in Z, \ \lambda > 0, \\ 0 & \text{if } \tilde{\boldsymbol{z}} = 0, \ \lambda = 0. \end{cases}$$

To describe $\mathcal{P}_a(\tilde{\mathcal{K}}, \boldsymbol{\beta})$ in the projection problem (15), we use the results in Proposition 4 and Remark 1. Consider the binary variables $z_{a,k}$, $k = 1, 2, 3, 4$. For a convex set Z, let $\mathcal{M}_{1,2,3,4}(Z)$ be defined in a similar fashion to $\mathcal{M}_k(Z)$ in Lemma 4, but by simultaneously considering all different possibilities of all the binary variables $z_{a,k}$, $k = 1, 2, 3, 4$. Associated with this set, one can form set $\bar{\mathcal{M}}_{1,2,3,4}(Z)$ in a similar fashion to Remark 1. Then, it is straightforward to verify that an optimal solution to (15) is given by the $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W})$ components of an optimal solution to the convex program:

$$\min_{\substack{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W},\boldsymbol{z},\boldsymbol{d}_{\{0,0,0,0\}},\ldots,\boldsymbol{d}_{\{1,1,1,1\}}, \\ \lambda_{\{0,0,0,0\}},\ldots,\lambda_{\{1,1,1,1\}})\in\bar{\mathcal{M}}_{1,2,3,4}\left(\mathcal{Z}_a(\tilde{\mathcal{K}},\boldsymbol{\beta})\right)}} \|(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) - (\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})\|. \tag{18}$$

Let $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{W}^*)$ be the $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W})$ components of an optimal solution to (18). We can generate the valid inequality $\boldsymbol{\alpha}^\top \boldsymbol{x} + \boldsymbol{\theta}^\top \boldsymbol{y} + \boldsymbol{\Gamma} \bullet \boldsymbol{W} \geq \rho$, where $\boldsymbol{\alpha}$, $\boldsymbol{\theta}$, and $\boldsymbol{\Gamma}$ are partial subgradients of $\|(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) - (\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})\|$ at $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{W}^*)$ with respect to $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{W}$, respectively. Moreover, $\rho = \boldsymbol{\alpha}^\top \boldsymbol{x}^* + \boldsymbol{\theta}^\top \boldsymbol{y}^* + \boldsymbol{\Gamma} \bullet \boldsymbol{W}^*$. For more details on how to obtain such a valid inequality, we refer to Stubbs and Mehrotra (1999).

4 A Finitely-Convergent Cutting Plane Algorithm

Motivated by Example 1 and numerical experiments in Fampa and Lee (2018), in this section, a finitely-convergent cutting plane algorithms to obtain an ϵ-optimal solution to (BLP) or detect infeasibility is analyzed. The analyzed algorithms utilize cuts obtained from the single-vector disjunction, described in Section 3.

A usual approach for such a cutting plane algorithm is to generate cuts by using one or more disjunctions obtained from *one* optimal solution to the current relaxation. For example, in the numerical experiments in Fampa and Lee (2018), at most four cuts are generated, based on the largest four singular values of $\hat{\boldsymbol{W}} - \hat{\boldsymbol{x}}\hat{\boldsymbol{y}}^\top$, where $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$ is an optimal solution to the current relaxation. In section 2.2, we illustrated that such an approach is not sufficient to get arbitrary close to the convex hull of solutions and obtain a global optimal solution.

Unlike the usual cutting plane approach that generates a valid inequality *only* at the current optimal solution, in this section, generating inequalities at *multiple* extreme point solutions of the current relaxation is analyzed in an algorithmic fashion. These extreme points are generated by exploring a set of near-optimal solutions to the current relaxation. The analyzed algorithms require two input parameters $\gamma > 0$ and $\epsilon > 0$. The parameter γ determines the neighboring set of the current solution that the algorithm explores at each iteration. The parameter ϵ determines the optimality and feasibility tolerance.

In this section, we describe a modification to the cutting plane algorithm proposed in Fampa and Lee (2018) to generate cuts at multiple vertices of the current relaxation. In section 4.2, a finitely-convergent algorithm is analyzed when two sets of bases for $\mathbb{R}^n$ and $\mathbb{R}^m$ are available a priori. In section 4.3, we suppose that such bases are not available a priori and are obtained through SVD.

4.1 Definitions, Assumption, and Technical Results

In this section, we give definitions of an ϵ-feasible solution and ϵ-optimal solution to (BLP), state a regularity assumption about $\mathcal{F}$, and present some technical results.

Definition 1. *(ϵ-**Feasible Solution**) For the optimization problem (BLP) with $z^* = \min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\mathcal{F}} \boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{A}_0 \bullet \boldsymbol{W}$, we say that a point $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \in \mathcal{K}$ is ϵ-feasible if $|\boldsymbol{u}_a^\top \hat{\boldsymbol{W}} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \hat{\boldsymbol{x}})(\boldsymbol{v}_a^\top \hat{\boldsymbol{y}})| \le \epsilon$ for all $a \in [mn]$. That is, $|\boldsymbol{u}^\top \hat{\boldsymbol{W}} \boldsymbol{v} - (\boldsymbol{u}^\top \hat{\boldsymbol{x}})(\boldsymbol{v}^\top \hat{\boldsymbol{y}})| \le mn\epsilon$, i.e., $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$ in $\bar{\mathcal{F}}^\epsilon$, as defined in (11).*

Definition 2. *(ϵ-**Optimal Solution**) For the optimization problem (BLP) with $z^* = \min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\mathcal{F}} \boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{A}_0 \bullet \boldsymbol{W}$, we say that a point $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \in \mathcal{K}$ is an ϵ-optimal solution if $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$ is ϵ-feasible and $\boldsymbol{f}_0^\top \hat{\boldsymbol{x}} + \boldsymbol{g}_0^\top \hat{\boldsymbol{y}} + \boldsymbol{A}_0 \bullet \hat{\boldsymbol{W}} \le z^* + \epsilon$.*

In order to establish finite convergence of the analyzed algorithms, we need to make a regularity assumption about the feasible region of (BLP) and the set $\bar{\mathcal{F}}^\epsilon$, defined in (11).

Assumption 1. *(**Problem Stability**) If $\mathcal{F}$ is an empty set, then, $\bar{\mathcal{F}}^\eta$ is an empty set for some $\eta > 0$.*

The above assumption implies that if (BLP) is infeasible, then, it remains infeasible for a small perturbation in the bilinear constraints. This assumption allows the algorithms in sections 4.2 and 4.3 to detect infeasibility of (BLP).

We now state some technical results.

Lemma 5. *Let $\{\mathcal{B}_1^t\}, \{\mathcal{B}_2^t\}, \ldots, \{\mathcal{B}_\kappa^t\}$ be convergent, decreasing sequences of nested nonempty compact connected [1] sets of a finite-dimensional Euclidean space. If $\lim_{t\to\infty} \mathcal{B}_\iota^t = \bar{\mathcal{B}}_\iota$, where $\bar{\mathcal{B}}_\iota$ is nonempty, for $\iota \in [\kappa]$, then,*

$$\lim_{t\to\infty} \mathrm{conv}\left(\cup_{\iota=1}^{\kappa} \mathcal{B}_\iota^t\right) = \mathrm{conv}\left(\cup_{\iota=1}^{\kappa} \lim_{t\to\infty} \mathcal{B}_\iota^t\right) = \mathrm{conv}\left(\cup_{\iota=1}^{\kappa} \bar{\mathcal{B}}_\iota\right).$$

Proof. We show that for any $\delta > 0$, there exists $\hat{t} \ge 0$ such that for all $t \ge \hat{t}$, we have $\min_{\boldsymbol{b}\in\mathrm{conv}\left(\cup_{\iota=1}^{\kappa}\bar{\mathcal{B}}_\iota\right)} \|\boldsymbol{b} - \boldsymbol{b}^t\| \le \delta$ for all $\boldsymbol{b}^t \in \mathrm{conv}\left(\cup_{\iota=1}^{\kappa} \mathcal{B}_\iota^t\right)$.

First, note that because $\lim_{t\to\infty} \mathcal{B}_\iota^t = \bar{\mathcal{B}}_\iota$, then, for any $\delta > 0$, there exists $\hat{t}_\iota \ge 0$ such that for all $t \ge \hat{t}$, we have $\min_{\boldsymbol{b}_\iota\in\bar{\mathcal{B}}_\iota} \|\boldsymbol{b}_\iota - \boldsymbol{b}_\iota^t\| \le \delta$, for all $\boldsymbol{b}_\iota^t \in \mathcal{B}_\iota^t$. Moreover, for any $t \ge 0$, $\boldsymbol{b}^t \in \mathrm{conv}\left(\cup_{\iota=1}^{\kappa} \mathcal{B}_\iota^t\right)$

[1] Set $\mathcal{B}$ is not connected if there are two disjoint open sets $\mathcal{U}$ and $\mathcal{V}$ such that $\mathcal{B} \subset \mathcal{U} \cup \mathcal{V}$, $\mathcal{B} \cap \mathcal{U} \neq \emptyset$, and $\mathcal{B} \cap \mathcal{V} \neq \emptyset$.

can be written as $\boldsymbol{b}^t = \sum_{\iota=1}^{\kappa} \lambda_\iota^t \boldsymbol{b}_\iota^t$ for some $\boldsymbol{b}_\iota^t \in \mathcal{B}_\iota^t$ and $\lambda_\iota^t \in [0,1]$, $\iota \in [\kappa]$, such that $\sum_{\iota=1}^{\kappa} \lambda_\iota^t = 1$. Therefore,

$$\begin{aligned}
&\min_{\boldsymbol{b} \in \operatorname{conv}\left(\cup_{\iota=1}^{\kappa} \bar{\mathcal{B}}_\iota\right)} \|\boldsymbol{b} - \boldsymbol{b}^t\| = \min_{\boldsymbol{b} \in \operatorname{conv}\left(\cup_{\iota=1}^{\kappa} \bar{\mathcal{B}}_\iota\right)} \Big\| \sum_{\iota=1}^{\kappa} \lambda_\iota^t (\boldsymbol{b} - \boldsymbol{b}_\iota^t) \Big\| \\
&\leq \min_{\boldsymbol{b}_\iota \in \bar{\mathcal{B}}_\iota,\, \iota \in [\kappa]} \Big\| \sum_{\iota=1}^{\kappa} \lambda_\iota^t (\boldsymbol{b}_\iota - \boldsymbol{b}_\iota^t) \Big\| \leq \min_{\boldsymbol{b}_\iota \in \bar{\mathcal{B}}_\iota,\, \iota \in [\kappa]} \sum_{\iota=1}^{\kappa} \lambda_\iota^t \|\boldsymbol{b}_\iota - \boldsymbol{b}_\iota^t\| \\
&= \sum_{\iota=1}^{\kappa} \lambda_\iota^t \min_{\boldsymbol{b}_\iota \in \bar{\mathcal{B}}_\iota} \|\boldsymbol{b}_\iota - \boldsymbol{b}_\iota^t\|,
\end{aligned}$$

where the first inequality follows because $\{\boldsymbol{b} = \sum_{\iota=1}^{\kappa} \lambda_\iota^t \boldsymbol{b}_\iota \,|\, \boldsymbol{b}_\iota \in \bar{\mathcal{B}}_\iota,\ \iota \in [\kappa]\} \subseteq \operatorname{conv}\left(\cup_{\iota=1}^{\kappa} \bar{\mathcal{B}}_\iota\right)$. By choosing $\hat{t} := \max_{\iota=1}^{\kappa} \hat{t}_\iota$, the result follows. □

Lemma 6. *O'Searcoid (2006, Theorem 12.1.3) A decreasing sequence of nonempty, nested, closed sets of a compact metric space has a nonempty compact intersection.*

Lemma 7. *Salinetti and Wets (1979,* Proposition 2*) Suppose that $\{\mathcal{B}^t\}$ is a decreasing sequence of nested closed sets of a finite-dimensional Euclidean space. Then, $\{\mathcal{B}^t\}$ converges to $\cap_{t=1}^{\infty} \mathcal{B}^t$ in the sense of Kuratowski, as $t \to \infty$.*

Lemma 8. *Salinetti and Wets (1979,* Corollary 3A*) Suppose that $\{\mathcal{B}^t\}$ is a sequence of nonempty compact connected sets of a finite-dimensional Euclidean space. Then, $\lim_{t\to\infty} \mathcal{B}^t = \bar{\mathcal{B}}$ if and only if $\mathcal{B}^t \xrightarrow{K} \bar{\mathcal{B}}$ as $t \to \infty$, i.e., the Hausdorff convergence implies the Kuratowski convergence and vice versa, and the limits are equal.*

Lemma 9. *Owen and Mehrotra (2001,* Lemma 2*) Let $\{\mathcal{B}^t\}$ be a convergent sequence of bounded convex sets such that $\mathcal{B}^{t+1} \subseteq \mathcal{B}^t$ for all $t \geq 0$ and $\lim_{t\to\infty} \mathcal{B}^t = \tilde{\mathcal{B}}$. For each $\tilde{b} \in ext(\tilde{\mathcal{B}})$, there exists some sequence $\{b^t\}$ of points in $ext(\mathcal{B}^t)$ with a subsequence converging to $\tilde{b}$.*

4.2 General Basis

In Algorithm 2, a finitely-convergent algorithm is analyzed when two sets of bases for $\mathbb{R}^n$ and $\mathbb{R}^m$ are available. This algorithm relies on the cut-generation procedure, described in Algorithm 1. Under the regularity Assumption 1, Algorithm 2 either generates an ϵ-optimal solution to (BLP) or detects its infeasibility (i.e., $\mathcal{F} = \emptyset$) in a finite number of iterations (see Theorem 1).

Before we proceed, let us introduce the notation we use in Algorithm 2. At each iteration t, we denote the current relaxation by S^t. Let $\underline{z}^t = \min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in S^t} \ \boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{A}_0 \bullet \boldsymbol{W}$. Recall that the input parameter $\gamma > 0$ controls the vertex exploration in the neighborhood of the current solution. We impose the vertex exploration by defining the set of extreme point solutions whose objective values are γ-away from $\underline{z}^t$ as

$$\Omega^t := \{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W}) \in \operatorname{ext}(S^t) \big| \boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{A}_0 \bullet \boldsymbol{W} - \underline{z}^t \leq \gamma\}. \tag{19}$$

Given $\epsilon > 0$, we define a subset of ϵ-feasible solutions in Ω^t as

$$\Omega_\epsilon := \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \Omega^t \big| |\boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \boldsymbol{x})(\boldsymbol{v}_a^\top \boldsymbol{y})| \leq \epsilon,\ a \in [nm]\}. \tag{20}$$

As it can be seen from Algorithm 2, the algorithm proposed in Fampa and Lee (2018) is modified to generate disjunctive cuts at all points $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}, \bar{\boldsymbol{W}}) \in \Omega^t$ and $a \in [nm]$ such that $|\boldsymbol{u}_a^\top \bar{\boldsymbol{W}} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \bar{\boldsymbol{x}})(\boldsymbol{v}_a^\top \bar{\boldsymbol{y}})| > \frac{\epsilon}{4}$. We show that the analyzed vertex exploration, imposed by $\gamma > 0$, enables the algorithm to generate cuts that are deep enough to cut off points that are not in $\operatorname{conv}(\mathcal{F})$, and eventually lead to an ϵ-optimal solution to (BLP) or detects infeasibility. Before we state and prove the main result of this section, we address the building blocks of Algorithm 2 and state some intermediate results. In particular, Lemma 10 shows that Algorithm 2 explores extreme point solutions in Ω^t that will lead to deep-enough cuts. Lemma 13 shows that cuts generated at these extreme point solutions have a sufficiently large depth to cut off points that violate constraints in $\mathcal{F}$.

Lemma 10. *Let $\{S^t\}$ be a sequence of (nonempty) sets generated by Algorithm 2. Suppose that $\{S^t\}$ converges to a nonempty set $\tilde{S}$, and let $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$ be an optimal extreme point solution of $\min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\tilde{S}}\ \boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{A}_0 \bullet \boldsymbol{W}$. Let $\{(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t)\}_{t\in\mathcal{T}}$ be a convergent subsequence of points in ext(S^t) such that $\{(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t)\}_{t\in\mathcal{T}} \to (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$. Then, for any $\gamma > 0$, there exists a sufficiently large $t \in \mathcal{T}$ such that $(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t) \in \Omega^t$, where Ω^t is defined in (19).*

Proof. First, because $S^0 (= \mathcal{K})$ is bounded and $S^{t+1} \subseteq S^t$ for all $t \geq 0$, then the sequence of nonempty closed sets $\{S^t\}$ converges to a set $\tilde{S}$ by Lemma 7. Furthermore, by Lemma 9, we know that there exists a convergent subsequence $\{(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t)\}_{t\in\mathcal{T}}$ of points in ext(S^t) such that $\{(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t)\}_{t\in\mathcal{T}} \to (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$. Now, note that $\boldsymbol{f}_0^\top \boldsymbol{x}^t + \boldsymbol{g}_0^\top \boldsymbol{y}^t + \boldsymbol{A}_0 \bullet \boldsymbol{W}^t - \underline{z}^t = \boldsymbol{f}_0^\top \boldsymbol{x}^t + \boldsymbol{g}_0^\top \boldsymbol{y}^t + \boldsymbol{A}_0 \bullet \boldsymbol{W}^t - (\boldsymbol{f}_0^\top \tilde{\boldsymbol{x}} + \boldsymbol{g}_0^\top \tilde{\boldsymbol{y}} + \boldsymbol{A}_0 \bullet \tilde{\boldsymbol{W}}) + (\boldsymbol{f}_0^\top \tilde{\boldsymbol{x}} + \boldsymbol{g}_0^\top \tilde{\boldsymbol{y}} + \boldsymbol{A}_0 \bullet \tilde{\boldsymbol{W}}) - \underline{z}^t$. By the fact that $\{(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t)\}_{t\in\mathcal{T}} \to (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$, there exists t_1 such that for $t \geq t_1$, $t \in \mathcal{T}$, we have $\boldsymbol{f}_0^\top \boldsymbol{x}^t + \boldsymbol{g}_0^\top \boldsymbol{y}^t + \boldsymbol{A}_0 \bullet \boldsymbol{W}^t - (\boldsymbol{f}_0^\top \tilde{\boldsymbol{x}} + \boldsymbol{g}_0^\top \tilde{\boldsymbol{y}} + \boldsymbol{A}_0 \bullet \tilde{\boldsymbol{W}}) \leq \frac{\gamma}{2}$. Moreover, because $\lim_{t\to\infty} S^t = \tilde{S}$, we have $\{\underline{z}^t\} \to \boldsymbol{f}_0^\top \tilde{\boldsymbol{x}} + \boldsymbol{g}_0^\top \tilde{\boldsymbol{y}} + \boldsymbol{A}_0 \bullet \tilde{\boldsymbol{W}}$. That is, there exists t_2 such that for $t \geq t_2$, $t \in \mathcal{T}$, we have $\boldsymbol{f}_0^\top \tilde{\boldsymbol{x}} + \boldsymbol{g}_0^\top \tilde{\boldsymbol{y}} + \boldsymbol{A}_0 \bullet \tilde{\boldsymbol{W}} - \underline{z}^t \leq \frac{\gamma}{2}$. Consequently, for $t \geq \max\{t_1, t_2\}$, $t \in \mathcal{T}$, we have $\boldsymbol{f}_0^\top \boldsymbol{x}^t + \boldsymbol{g}_0^\top \boldsymbol{y}^t + \boldsymbol{A}_0 \bullet \boldsymbol{W}^t - \underline{z}^t \leq \gamma$, which implies that $(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t) \in \Omega^t$. ☐

The key piece in Lemma 10 is choosing γ to be strictly positive (because of the limiting arguments in the proof). In fact, when $\gamma = 0$, there could be no neighboring set, and hence, no vertex exploration takes place. This choice of $\gamma > 0$, and hence the vertex exploration, is the conceptual modification to the algorithm investigated in Fampa and Lee (2018). We shall shortly see in the proof of Theorem 1 that the vertex exploration plays a role in the finite-convergence analysis of Algorithm 2.

Lemma 11. *Consider the assumptions and notation in Lemma 10.*

(i) If $|\boldsymbol{u}_a^\top \tilde{\boldsymbol{W}} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}})(\boldsymbol{v}_a^\top \tilde{\boldsymbol{y}})| > \epsilon$ for some $a \in [nm]$, then, there exists a sufficiently large $t \in \mathcal{T}$ such that $|\boldsymbol{u}_a^\top \boldsymbol{W}^t \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \boldsymbol{x}^t)(\boldsymbol{v}_a^\top \boldsymbol{y}^t)| > \frac{\epsilon}{2}$.

(ii) If $|\boldsymbol{u}_a^\top \tilde{\boldsymbol{W}} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}})(\boldsymbol{v}_a^\top \tilde{\boldsymbol{y}})| \le \epsilon$ *for all* $a \in [nm]$*, then, there exists a sufficiently large* $t \in \mathcal{T}$ *such that* $|\boldsymbol{u}_a^\top \boldsymbol{W}^t \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \boldsymbol{x}^t)(\boldsymbol{v}_a^\top \boldsymbol{y}^t)| \le 2\epsilon$ *for all* $a \in [nm]$.

To prove Lemma 11, we state and prove a technical lemma first.

Lemma 12. *Consider two points* $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}), (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}}) \in \tilde{\mathcal{K}}$*. For* $\epsilon > 0$ *and some* $a \in [nm]$*, suppose that we have* $|\boldsymbol{u}_a^\top(\boldsymbol{x} - \tilde{\boldsymbol{x}})| \le \frac{\epsilon}{4\max\{\boldsymbol{v}_a^\top \overline{\boldsymbol{y}}, \boldsymbol{v}_a^\top \underline{\boldsymbol{y}}\}}$ *and* $|\boldsymbol{v}_a^\top(\boldsymbol{y} - \tilde{\boldsymbol{y}})| \le \frac{\epsilon}{4\max\{\boldsymbol{u}_a^\top \overline{\boldsymbol{x}}, \boldsymbol{u}_a^\top \underline{\boldsymbol{x}}\}}$*. Then,* $|(\boldsymbol{u}_a^\top \boldsymbol{x})(\boldsymbol{v}_a^\top \boldsymbol{y}) - (\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}})(\boldsymbol{v}_a^\top \tilde{\boldsymbol{y}})| \le \frac{\epsilon}{2}$*.*

Proof. Note that $|(\boldsymbol{u}_a^\top \boldsymbol{x})(\boldsymbol{v}_a^\top \boldsymbol{y}) - (\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}})(\boldsymbol{v}_a^\top \tilde{\boldsymbol{y}})| \le |\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}}||\boldsymbol{v}_a^\top(\boldsymbol{y} - \tilde{\boldsymbol{y}})| + |\boldsymbol{v}_a^\top \boldsymbol{y}||\boldsymbol{u}_a^\top(\boldsymbol{x} - \tilde{\boldsymbol{x}})|$. Because $|\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}}| \le \max\{\boldsymbol{u}_a^\top \overline{\boldsymbol{x}}, \boldsymbol{u}_a^\top \underline{\boldsymbol{x}}\}$, we have $|\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}}||\boldsymbol{v}_a^\top(\boldsymbol{y} - \tilde{\boldsymbol{y}})| \le \frac{\epsilon}{4}$. Similarly, we have $|\boldsymbol{v}_a^\top \boldsymbol{y}||\boldsymbol{u}_a^\top(\boldsymbol{x} - \tilde{\boldsymbol{x}})| \le \frac{\epsilon}{4}$. Thus, $|(\boldsymbol{u}_a^\top \boldsymbol{x})(\boldsymbol{v}_a^\top \boldsymbol{y}) - (\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}})(\boldsymbol{v}_a^\top \tilde{\boldsymbol{y}})| \le \frac{\epsilon}{2}$. □

Proof of Lemma 11. Because the norms are equivalent on a finite-dimensional vector space, there exists a positive constant C such that $C(|\boldsymbol{u}_a^\top(\boldsymbol{W}^t - \tilde{\boldsymbol{W}})\boldsymbol{v}_a + \boldsymbol{u}_a^\top(\boldsymbol{x}^t - \tilde{\boldsymbol{x}}) + \boldsymbol{v}_a^\top(\boldsymbol{y}^t - \tilde{\boldsymbol{y}})|) \le C(|\boldsymbol{u}_a^\top(\boldsymbol{W}^t - \tilde{\boldsymbol{W}})\boldsymbol{v}_a| + |\boldsymbol{u}_a^\top(\boldsymbol{x}^t - \tilde{\boldsymbol{x}})| + |\boldsymbol{v}_a^\top(\boldsymbol{y}^t - \tilde{\boldsymbol{y}})|) \le \|(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t) - (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})\|$. Because $\{(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t)\}_{t\in\mathcal{T}} \to (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$, there exists t_1 such that for $t \ge t_1$, $t \in \mathcal{T}$, we have $\|(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t) - (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})\| \le C\frac{\epsilon}{4}$. Consequently, for $t \ge t_1$, $t \in \mathcal{T}$, we have $|\boldsymbol{u}_a^\top(\boldsymbol{W}^t - \tilde{\boldsymbol{W}})\boldsymbol{v}_a| \le \frac{\epsilon}{4}$. Similarly, there exist t_2 and t_3 such that for $t \ge t_2$ and $t \ge t_3$, $t \in \mathcal{T}$, we have $|\boldsymbol{u}_a^\top(\boldsymbol{x}^t - \tilde{\boldsymbol{x}})| \le \frac{\epsilon}{8\max\{\boldsymbol{v}_a^\top \overline{\boldsymbol{y}}, \boldsymbol{v}_a^\top \underline{\boldsymbol{y}}\}}$ and $|\boldsymbol{v}_a^\top(\boldsymbol{y}^t - \tilde{\boldsymbol{y}})| \le \frac{\epsilon}{8\max\{\boldsymbol{u}_a^\top \overline{\boldsymbol{x}}, \boldsymbol{u}_a^\top \underline{\boldsymbol{x}}\}}$. For the first part, by the reverse triangle inequality and Lemma 12, we have

$$\begin{aligned}|\boldsymbol{u}_a^\top \boldsymbol{W}^t \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \boldsymbol{x}^t)(\boldsymbol{v}_a^\top \boldsymbol{y}^t)| \ge & |\boldsymbol{u}_a^\top \tilde{\boldsymbol{W}} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}})(\boldsymbol{v}_a^\top \tilde{\boldsymbol{y}})| - |\boldsymbol{u}_a^\top(\boldsymbol{W}^t - \tilde{\boldsymbol{W}})\boldsymbol{v}_a| \\ & - |(\boldsymbol{u}_a^\top \boldsymbol{x}^t)(\boldsymbol{v}_a^\top \boldsymbol{y}^t) - (\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}})(\boldsymbol{v}_a^\top \tilde{\boldsymbol{y}})| > \frac{\epsilon}{2}\end{aligned}$$

for $t \ge \max\{t_1, t_2, t_3\}$. By a similar argument, but using the triangle inequality, the result in the second part also follows. □

Lemma 13 shows that there exists a sufficiently large t such that the valid inequality generated at $(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t)$ have a sufficiently large depth and subsequently, can cut off point $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$, with $|\boldsymbol{u}_a^\top \tilde{\boldsymbol{W}} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}})(\boldsymbol{v}_a^\top \tilde{\boldsymbol{y}})| > \epsilon$ for some $a \in [nm]$, from S^t.

Lemma 13. *Consider the assumptions in Lemma 11(i). Let the breakpoints* $\tilde{\boldsymbol{\beta}}$ *be chosen as in line 4 of Algorithm 1, for* $\tilde{S}$ *and point* $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$*. There exists* $\delta > 0$*, where* δ *is defined as*

$$\delta = \min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W}) \in \mathcal{P}_a(\tilde{S}, \tilde{\boldsymbol{\beta}})} \|(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}}) - (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W})\|.$$

Furthermore, there exists a sufficiently large $t \in \mathcal{T}$ *such that*

$$\min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W}) \in \mathcal{P}_a(S^t, \boldsymbol{\beta}^t)} \|(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t) - (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W})\| > \frac{\delta}{2},$$

where the breakpoints $\boldsymbol{\beta}^t$ *are chosen as in line 4 of Algorithm 1, for* S^t *and point* $(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t)$*.*

To prove Lemma 13, we first show in Lemma 14 that there exists a sufficiently large t such that point $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$, introduced in Lemma 10, with $|\boldsymbol{u}_a^\top \tilde{\boldsymbol{W}} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}})(\boldsymbol{v}_a^\top \tilde{\boldsymbol{y}})| > \epsilon$ for some $a \in [nm]$, violates the disjunction formed over S^t and using the breakpoints $\boldsymbol{\beta}^t$, chosen according to line 4 of Algorithm 1 for point $(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t)$.

Lemma 14. *Consider the assumptions in Lemma 11(i). Let the breakpoints $\boldsymbol{\beta}^t$ be chosen as in line 4 of Algorithm 1, for S^t and point $(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t)$. Then, there exists a sufficiently large $t \in \mathcal{T}$ such that for all $k \in \{1, 2, 3, 4\}$, one of the secant inequalities in $\tilde{S}_k(\boldsymbol{c}_a, S^t, \boldsymbol{\beta}^t)$, defined in (8), is violated by $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$, and the amount of violation is greater than $\frac{\epsilon}{2}$.*

Proof. Suppose that $\boldsymbol{u}_a^\top \tilde{\boldsymbol{W}} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}})(\boldsymbol{v}_a^\top \tilde{\boldsymbol{y}}) > \epsilon$. We first analyze the violation of the first secant inequality in $\tilde{S}_k(\boldsymbol{c}_a, S^t, \boldsymbol{\beta}^t)$ by $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$ for all k, $k \in \{1, 2, 3, 4\}$. Let us begin with those $\tilde{S}_k(\boldsymbol{c}_a, S^t, \boldsymbol{\beta}^t)$, for which $\beta^t_{1,1}$ is the lower bound on $\frac{\boldsymbol{u}_a^\top \boldsymbol{x} + \boldsymbol{v}_a^\top \boldsymbol{y}}{2}$. With some algebra, we have

$$
\begin{aligned}
&\boldsymbol{u}_a^\top \tilde{\boldsymbol{W}} \boldsymbol{v}_a - \left(\frac{\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}} + \boldsymbol{v}_a^\top \tilde{\boldsymbol{y}}}{2}\right)\left(\beta^t_{1,1} + \frac{\boldsymbol{u}_a^\top \boldsymbol{x}^t + \boldsymbol{v}_a^\top \boldsymbol{y}^t}{2}\right) + \beta^t_{1,1}\frac{\boldsymbol{u}_a^\top \boldsymbol{x}^t + \boldsymbol{v}_a^\top \boldsymbol{y}^t}{2} \\
&+ \left(\frac{\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}} - \boldsymbol{v}_a^\top \tilde{\boldsymbol{y}}}{2}\right)^2 = \boldsymbol{u}_a^\top \tilde{\boldsymbol{W}} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}})(\boldsymbol{v}_a^\top \tilde{\boldsymbol{y}}) \\
&+ \left(\beta^t_{1,1} - \frac{\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}} + \boldsymbol{v}_a^\top \tilde{\boldsymbol{y}}}{2}\right)\left(\frac{\boldsymbol{u}_a^\top (\boldsymbol{x}^t - \tilde{\boldsymbol{x}}) + \boldsymbol{v}_a^\top (\boldsymbol{y}^t - \tilde{\boldsymbol{y}})}{2}\right) \\
>&\epsilon + \left(\beta^t_{1,1} - \frac{\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}} + \boldsymbol{v}_a^\top \tilde{\boldsymbol{y}}}{2}\right)\left(\frac{\boldsymbol{u}_a^\top (\boldsymbol{x}^t - \tilde{\boldsymbol{x}}) + \boldsymbol{v}_a^\top (\boldsymbol{y}^t - \tilde{\boldsymbol{y}})}{2}\right).
\end{aligned}
\tag{21}
$$

Observe that for $t \geq 0$, we have $\beta^t_{1,1} - \frac{\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}} + \boldsymbol{v}_a^\top \tilde{\boldsymbol{y}}}{2} \leq 0$ because $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$ is suboptimal to $\beta^t_{1,1} = \min\{\frac{\boldsymbol{u}_a^\top \boldsymbol{x} + \boldsymbol{v}_a^\top \boldsymbol{y}}{2} \,|\, (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in S^t\}$. On the other hand, by a similar argument as in the proof of Lemma 11, we have $\frac{\boldsymbol{u}_a^\top (\boldsymbol{x}^t - \tilde{\boldsymbol{x}}) + \boldsymbol{v}_a^\top (\boldsymbol{y}^t - \tilde{\boldsymbol{y}})}{2} > -C\|(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t) - (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})\|$, for some positive constant C. Hence, there exists t_1 such that for $t \geq t_1$, $t \in \mathcal{T}$, we have $-C\left(\beta^t_{1,1} - \frac{\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}} + \boldsymbol{v}_a^\top \tilde{\boldsymbol{y}}}{2}\right)\|(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t) - (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})\| > -\frac{\epsilon}{2}$. Thus, the violation of the inequality (21) is greater than $\frac{\epsilon}{2}$. With a similar argument, we can conclude that for those $\tilde{S}_k(\boldsymbol{c}_a, S^t, \boldsymbol{\beta}^t)$, for which $\beta^t_{1,3}$ is the upper bound on $\frac{\boldsymbol{u}_a^\top \boldsymbol{x} + \boldsymbol{v}_a^\top \boldsymbol{y}}{2}$, the violation of the first secant inequality is greater than $\frac{\epsilon}{2}$.

For the case that $-\boldsymbol{u}_a^\top \tilde{\boldsymbol{W}} \boldsymbol{v}_a + (\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}})(\boldsymbol{v}_a^\top \tilde{\boldsymbol{y}}) > \epsilon$, we can similarly conclude that the violation of the second secant inequality is greater than $\frac{\epsilon}{2}$. Consequently, the result follows. ☐

Proof of Lemma 13. The first part follows from a direct application of Proposition 3. To prove the second part, let $\hat{t}_1$ be the sufficiently large t for which the result in Lemma 14 holds. Note that with a similar argument as in the proof of Lemma 14, we can find a sequence of sufficiently large t, $t \in \mathcal{T}$, namely $\mathcal{T}' = \{\hat{t}_1, \hat{t}_2, \hat{t}_3, \ldots\}$, for which, for all $k \in \{1, 2, 3, 4\}$, one of the secant inequalities in $\tilde{S}_k(\boldsymbol{c}_a, S^t, \boldsymbol{\beta}^t)$ is violated by $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$ and the amount of violation is greater than $\{\epsilon - \frac{\epsilon}{2}, \epsilon - \frac{\epsilon}{4}, \epsilon - \frac{\epsilon}{8}, \ldots\}$. Hence, this subsequence $\mathcal{T}'$ yields a subsequence $\{S^t\}_{t \in \mathcal{T}'}$, for which $\vee_{k=1}^4 \tilde{S}_k(\boldsymbol{c}_a, S^t, \boldsymbol{\beta}^t)$ is a decreasing sequence. This is due to the fact that $\{S^t\}$ is a decreasing sequence of nested sets and the amount of violation of one of the secant inequalities in $\tilde{S}_k(\boldsymbol{c}_a, S^t, \boldsymbol{\beta}^t)$, $t \in \mathcal{T}'$, increases for all $k \in \{1, 2, 3, 4\}$. Consequently, for the subsequence $\mathcal{T}'$, $\mathcal{P}_a(S^t, \boldsymbol{\beta}^t)$ is a

Algorithm 2 Disjunctive cutting plane for (BLP) using general bases

Input: $\mathcal{K}$, $\{\boldsymbol{u}_i\}_{i=1}^n$, $\{\boldsymbol{v}_j\}_{j=1}^m$, γ, and ϵ.
Output: An ϵ-optimal solution.
Set $t \leftarrow 0$ and $S^0 = \mathcal{K}$.
while $S^t \neq \emptyset$ **do**
  Let $\underline{z}^t$ be the optimal value of $\min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in S^t} \ \boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{A}_0 \bullet \boldsymbol{W}$.
  Let $\Omega^t := \{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W}) \in \text{ext}(S^t) \,|\, \boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{A}_0 \bullet \boldsymbol{W} - \underline{z}^t \leq \gamma\}$.
  Let $\Omega_\epsilon := \{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W}) \in \Omega^t \,|\, |\boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \boldsymbol{x})(\boldsymbol{v}_a^\top \boldsymbol{y})| \leq \epsilon,\ a \in [nm]\}$.
  if $\Omega_\epsilon \neq \emptyset$ **then**
    for each $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}}) \in \Omega_\epsilon$ **do**
      if $\boldsymbol{f}_0^\top \hat{\boldsymbol{x}} + \boldsymbol{g}_0^\top \hat{\boldsymbol{y}} + \boldsymbol{A}_0 \bullet \hat{\boldsymbol{W}} - \underline{z}^t \leq \epsilon$ **then**
        STOP and output $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}})$ as an ϵ-optimal solution.
      end if
    end for
  end if
  $S^{t+1} = S^t$.
  for each $(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}},\bar{\boldsymbol{W}}) \in \Omega^t$ and $a \in [nm]$ such that $|\boldsymbol{u}_a^\top \bar{\boldsymbol{W}} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \bar{\boldsymbol{x}})(\boldsymbol{v}_a^\top \bar{\boldsymbol{y}})| > \frac{\epsilon}{4}$ **do**
    Call the procedure `SepCuts`$(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}},\bar{\boldsymbol{W}}; S^t, \boldsymbol{u}_a, \boldsymbol{v}_a)$ to obtain $(\texttt{viol}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\Gamma}, \rho)$.
    if `viol`=FALSE **then**
      STOP.
    else
      Let $S^{t+1} := \{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W}) \in S^{t+1} \,|\, \boldsymbol{\alpha}^\top \boldsymbol{x} + \boldsymbol{\theta}^\top \boldsymbol{y} + \boldsymbol{\Gamma} \bullet \boldsymbol{W} \geq \rho\}$.
    end if
  end for
  Set $t \leftarrow t+1$.
end while
STOP.

decreasing sequence. On the other hand, $\{(\boldsymbol{x}^t,\boldsymbol{y}^t,\boldsymbol{W}^t)\}_{t\in\mathcal{T}} \to (\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}},\tilde{\boldsymbol{W}})$ and $\lim_{t\to\infty} S^t = \tilde{S}$. These imply that $\{\boldsymbol{\beta}^t\}_{t\in\mathcal{T}} \to \tilde{\boldsymbol{\beta}}$. Putting these all together, we conclude that $\{\mathcal{P}_a(S^t,\boldsymbol{\beta}^t)\}_{t\in\mathcal{T}'} \to \mathcal{P}_a(\tilde{S},\tilde{\boldsymbol{\beta}})$ by Lemma 5. Thus, we have $\min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\mathcal{P}_a(S^t,\boldsymbol{\beta}^t)} \|(\boldsymbol{x}^t,\boldsymbol{y}^t,\boldsymbol{W}^t) - (\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\| > -\frac{\delta}{2} + \min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\mathcal{P}_a(\tilde{S},\tilde{\boldsymbol{\beta}})} \|(\boldsymbol{x}^t,\boldsymbol{y}^t,\boldsymbol{W}^t) - (\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\| = \frac{\delta}{2}$ for a sufficiently large $t \in \mathcal{T}'$. This completes the proof. □

We are now ready to state the main result of this section for any general bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, including the standard bases $\{\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n\}$ and $\{\boldsymbol{e}_1,\ldots,\boldsymbol{e}_m\}$.

Theorem 1. *Consider two parameters $\gamma > 0$ and $\epsilon > 0$. Let $\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n\}$ denote a set of mutually orthonormal vectors in $\mathbb{R}^n$, and $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_m\}$ denote a set of mutually orthonormal vectors in $\mathbb{R}^m$. Assume that the stability Assumption 1 holds for $\eta > mn\epsilon$, i.e., $\bar{\mathcal{F}}^\eta = \emptyset$. Then, Algorithm 2 either generates an ϵ-optimal solution to (BLP) or detects infeasibility (i.e., $\mathcal{F} = \emptyset$) in a finite number of iterations.*

Proof. Consider notation defined in the description of Algorithm 2. Note that the algorithm generates cutting planes for conv $(\mathcal{F})$ (recall conv $(\mathcal{F}) \subseteq \mathcal{P}_a(\check{\mathcal{K}},\boldsymbol{\beta})$ by Proposition 3). Suppose by contradiction that the algorithm does not converge in a finite number of iterations. Because $S^0(=\mathcal{K})$ is bounded and $S^{t+1} \subseteq S^t$ for all $t \geq 0$, then the sequence of closed sets $\{S^t\}$ converges to a set $\tilde{S}$ by Lemma 7. We examine the cases that $\tilde{S} = \emptyset$ and $\tilde{S} \neq \emptyset$.

Case 1. $\tilde{S} = \emptyset$. We conclude that $\mathcal{F}$ is empty because $\mathcal{F} \subseteq \text{conv}(\mathcal{F}) \subseteq \tilde{S}$. On the other hand, because $\{S^t\}$ converges to the empty set $\tilde{S}$, then, there exists a finite $t \geq 0$ such that $S^t = \emptyset$. Otherwise, $\tilde{S} = \cap_{t=1}^{\infty} S^t \neq \emptyset$ by the Cantor's intersection theorem, stated in Lemma 6. Hence, the algorithm terminates at line 26 of Algorithm 2 after detecting infeasibility of (BLP).

Case 2. $\tilde{S} \neq \emptyset$. Let $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$ be an optimal extreme point solution of $\min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\tilde{S}} \ \boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{A}_0 \bullet \boldsymbol{W}$. By Lemma 9, there exists a convergent subsequence $\{(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t)\}_{t\in\mathcal{T}}$ of points in $\text{ext}(S^t)$ such that $\{(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t)\}_{t\in\mathcal{T}} \to (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$. We examine the cases that $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}}) \notin \mathcal{F}^{\frac{\epsilon}{2}}$ and $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}}) \in \mathcal{F}^{\frac{\epsilon}{2}}$ separately, where $\mathcal{F}^{\frac{\epsilon}{2}}$ is defined in (12).

Case 2.1. $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}}) \notin \mathcal{F}^{\frac{\epsilon}{2}}$. In this case, there exists some a, $a \in [nm]$, such that $|\boldsymbol{u}_a^\top \tilde{\boldsymbol{W}} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \tilde{\boldsymbol{x}})(\boldsymbol{v}_a^\top \tilde{\boldsymbol{y}})| > \frac{\epsilon}{2}$. Let us choose $\tilde{\boldsymbol{\beta}}$ as in line 4 of Algorithm 1, for $\tilde{S}$ and $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$. Let us similarly define $\boldsymbol{\beta}^t$ for S^t and $(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t)$. By the first part of Lemma 13, there exists $\delta > 0$, where δ is defined as follows:

$$\delta = \min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\mathcal{P}_a(\tilde{S},\tilde{\boldsymbol{\beta}})} \|(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}}) - (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W})\|.$$

Claim 1. *There exists a finite $t \in \mathcal{T}$ such that*

1. $\|(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t) - (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})\| < \frac{\delta}{2}$ *(lemma 9),*
2. $(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t) \in \Omega^t$ *(lemma 10),*
3. $|\boldsymbol{u}_a^\top \boldsymbol{W}^t \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \boldsymbol{x}^t)(\boldsymbol{v}_a^\top \boldsymbol{y}^t)| > \frac{\epsilon}{4}$ *(Lemma 11(i)),*
4. $\min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\mathcal{P}_a(S^t,\boldsymbol{\beta}^t)} \|(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t) - (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W})\| > \frac{\delta}{2}$ *(Lemma 13).*

Hence, in iteration t, the algorithm generates a valid inequality (line 21 of Algorithm 2) that is violated by $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$ (this can be seen from parts 1 and 4 of the above claim). Thus, $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}}) \notin S^{t+1}$, contradicting $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}}) \in \tilde{S} \subseteq S^{t+1}$. So, the case that $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}}) \notin \mathcal{F}^{\frac{\epsilon}{2}}$ will not happen. If $\mathcal{F} \neq \emptyset$, then, this contradiction implies that we must have $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}}) \in \mathcal{F}^{\frac{\epsilon}{2}}$.

Case 2.2. $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}}) \in \mathcal{F}^{\frac{\epsilon}{2}}$. Now, consider the case that $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}}) \in \mathcal{F}^{\frac{\epsilon}{2}}$.

Claim 2. *There exists a finite $t \in \mathcal{T}$ such that*

1. $(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t) \in \Omega^t$ *(Lemma 10),*
2. $|\boldsymbol{u}_a^\top \boldsymbol{W}^t \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \boldsymbol{x}^t)(\boldsymbol{v}_a^\top \boldsymbol{y}^t)| \leq \epsilon$ *for $a \in [nm]$, implying $(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t) \in \Omega_\epsilon$ (Lemma 11(ii)),*
3. $\boldsymbol{f}_0^\top \tilde{\boldsymbol{x}} + \boldsymbol{g}_0^\top \tilde{\boldsymbol{y}} + \boldsymbol{A}_0 \bullet \tilde{\boldsymbol{W}} - \underline{z}^t \leq \gamma$ *(using the fact that $\lim_{t\infty} S^t = \tilde{S}$).*

For a sufficiently large t, $t \in \mathcal{T}$, that satisfies the above claim, we have $\underline{z}^t \leq \min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\Omega_\epsilon} \ \boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{A}_0 \bullet \boldsymbol{W} \leq \boldsymbol{f}_0^\top \tilde{\boldsymbol{x}} + \boldsymbol{g}_0^\top \tilde{\boldsymbol{y}} + \boldsymbol{A}_0 \bullet \tilde{\boldsymbol{W}}$. Note that $\Omega_\epsilon \neq \emptyset$ because $(\boldsymbol{x}^t, \boldsymbol{y}^t, \boldsymbol{W}^t) \in \Omega_\epsilon$. Also, note that the optimal solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}}) \in \Omega^\epsilon$ to the above optimization problem is such that $\boldsymbol{f}_0^\top \hat{\boldsymbol{x}} + \boldsymbol{g}_0^\top \hat{\boldsymbol{y}} + \boldsymbol{A}_0 \bullet \hat{\boldsymbol{W}} - \underline{z}^t \leq \boldsymbol{f}_0^\top \tilde{\boldsymbol{x}} + \boldsymbol{g}_0^\top \tilde{\boldsymbol{y}} + \boldsymbol{A}_0 \bullet \tilde{\boldsymbol{W}} - \underline{z}^t \leq \gamma$. If $0 < \gamma \leq \epsilon$, then, the algorithm terminates in the t-th iteration and yields an ϵ-optimal solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$. Otherwise, if $\gamma > \epsilon$, with a similar argument, we can find a sufficiently large t, $t \in \mathcal{T}$, such that $\boldsymbol{f}_0^\top \tilde{\boldsymbol{x}} + \boldsymbol{g}_0^\top \tilde{\boldsymbol{y}} + \boldsymbol{A}_0 \bullet \tilde{\boldsymbol{W}} - \underline{z}^t \leq \epsilon$, in addition to the first two parts of the above claim. Note that $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}, \tilde{\boldsymbol{W}})$ provides an upper bound to the optimization problem $\min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\Omega_\epsilon} \boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{A}_0 \bullet \boldsymbol{W}$. Thus, $\boldsymbol{f}_0^\top \hat{\boldsymbol{x}} + \boldsymbol{g}_0^\top \hat{\boldsymbol{y}} + \boldsymbol{A}_0 \bullet \hat{\boldsymbol{W}} - \underline{z}^t \leq \boldsymbol{f}_0^\top \tilde{\boldsymbol{x}} + \boldsymbol{g}_0^\top \tilde{\boldsymbol{y}} + \boldsymbol{A}_0 \bullet \tilde{\boldsymbol{W}} - \underline{z}^t \leq \epsilon$. Consequently, the algorithm terminates in the t-th iteration and yields an ϵ-optimal solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$. In any case, we have an ϵ-optimal solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$ to the feasible (BLP). □

4.3 SVD Basis

In Algorithm 2, it is assumed that two sets of bases are available a priori. In this section, we assume that bases are obtained through the course of the algorithm.

Given an optimal solution $(\breve{\boldsymbol{x}}, \breve{\boldsymbol{y}}, \breve{\boldsymbol{W}})$ to the minimization over the relaxation S^t, the left- and right-singular vector $\boldsymbol{u}_t$ and $\boldsymbol{v}_t$, respectively, corresponding to the largest singular value of $\breve{\boldsymbol{W}} - \breve{\boldsymbol{x}}\breve{\boldsymbol{y}}^\top$ are obtained. If $\text{vec}(\boldsymbol{u}_t \boldsymbol{v}_t^\top)$ is not in the span of previously generated vectors $\mathcal{V}^t$, $\text{vec}(\boldsymbol{u}_t \boldsymbol{v}_t^\top)$ is added to the set of bases. While the bases are generated, relaxation S^t is also refined by adding disjunctive cuts through procedure $\texttt{SepCuts}(\breve{\boldsymbol{x}}, \breve{\boldsymbol{y}}, \breve{\boldsymbol{W}}; S^t, \boldsymbol{u}_t, \boldsymbol{v}_t)$. This procedure of generating bases is continued until two sets of bases for $\mathbb{R}^n$ and $\mathbb{R}^m$ are available, i.e., $|\mathcal{V}_t| \geq \max\{n, m\}$. Because these spaces are finite-dimensional, the procedure of generating bases stops after a finite number of iterations. Once the bases are available, the algorithm continues as in Algorithm 2.

4.4 Discussion

The algorithms in Sections 4.2 and 4.3 need the set of γ-optimal extreme point solutions. Thus, the practical performance of the analyzed algorithms depends on the choice of $\gamma > 0$. A larger choice of γ results in a fewer iterations to converge while the computational time per iteration might increase. This is due to exploring a larger set of γ-optimal solutions, solving a relaxation problem with more constraints, and a potentially more demanding cut separation problem. On the other hand, a smaller choice of γ might result in a slower convergence. The choice of γ is problem-dependent and should be tuned to trade off the computational time and improvements obtained from the cuts. In other words, although in theory, the vertex exploration guarantees the finite convergence, not all cuts generated through the vertex exploration may have computational benefits.

On a related note, in theory, the set of γ-optimal extreme point solutions can be generated using Simplex pivots of the current extreme point solution. Alternatively, one may first encode basic feasible solutions of the current relaxation using binary variables and obtain a mixed-binary linear program (see the idea in Lee et al. (2000)). Now, one can use the solution pool feature of a commercial optimization solver (e.g., CPLEX) to enumerate all near-optimal solutions. Recognizing that this limitation to perform a thorough vertex exploration might incur additional computational burdens, in our numerical experiments in Section 5, we explored a few "promising" near-optimal extreme point solutions using random objective function coefficients.

5 Numerical Illustration

In Section 4, finitely-convergent algorithms that are based on generating cuts at *all* γ-optimal extreme point solutions of the current relaxation were analyzed. For the computational results in this section, we implemented a more practical version of the algorithm analyzed in section 4.3, which generates cuts at the current relaxation solution and only *a few* additional near-optimal extreme point solutions. For our numerical experiments, this practical consideration was repeated until a time limit was reached or the algorithm could not find a violated cut.

To obtain valid cuts, we use procedure $\texttt{SepCutsCGLP}(\breve{\boldsymbol{x}}, \breve{\boldsymbol{y}}, \breve{\boldsymbol{W}}; S^t, \boldsymbol{u}_t, \boldsymbol{v}_t)$. This procedure proceeds similar to $\texttt{SepCuts}(\breve{\boldsymbol{x}}, \breve{\boldsymbol{y}}, \breve{\boldsymbol{W}}; S^t, \boldsymbol{u}_t, \boldsymbol{v}_t)$, outlined in Algorithm 1, expect for it obtains valid cuts through the CGLP, introduced in Lemma 1. This CGLP is based on a (2×2)-way disjunction, where the disjunction is formed according to $\text{vec}(\boldsymbol{u}_t \boldsymbol{v}_t^\top)$, corresponding to the largest singular value of $\breve{\boldsymbol{W}} - \breve{\boldsymbol{x}}\breve{\boldsymbol{y}}^\top$. $\texttt{SepCutsCGLP}(\breve{\boldsymbol{x}}, \breve{\boldsymbol{y}}, \breve{\boldsymbol{W}}; S^t, \boldsymbol{u}_t, \boldsymbol{v}_t)$ is different from $\texttt{SepCuts}(\breve{\boldsymbol{x}}, \breve{\boldsymbol{y}}, \breve{\boldsymbol{W}}; S^t, \boldsymbol{u}_t, \boldsymbol{v}_t)$, in the sense that the former utilizes an outer approximation on the convex quadratic terms to form the disjunction. So, it is possible that the minimum distance of a point $(\breve{\boldsymbol{x}}, \breve{\boldsymbol{y}}, \breve{\boldsymbol{W}})$ from the convex hull $\mathcal{P}_{\boldsymbol{c}}(S^t, \boldsymbol{\beta})$, where $\boldsymbol{c} = \text{vec}(\boldsymbol{u}_t \boldsymbol{v}_t^\top)$, is positive, while the corresponding CGLP cannot find a violated disjunctive cut; hence, our modified algorithm would stop.

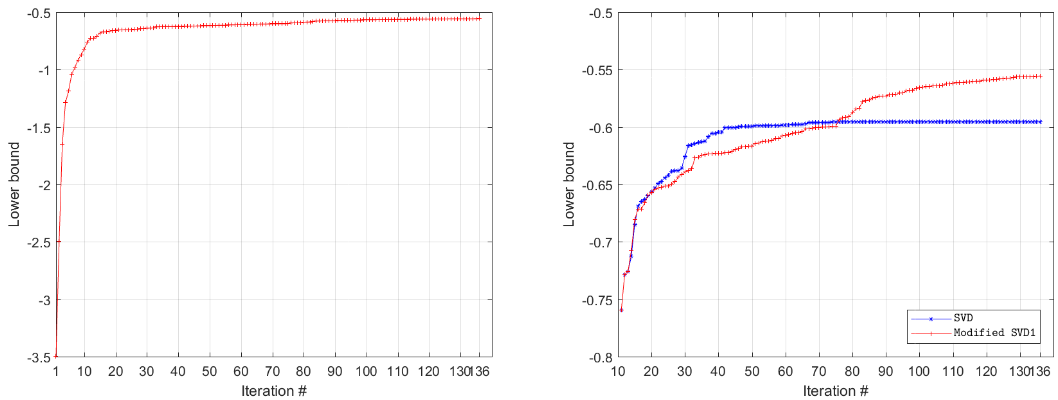
To obtain a near-optimal solution, we added the constraint $\boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{A}_0 \bullet \boldsymbol{W} - \underline{z}^t \leq \gamma$ to the current relaxation S^t and replaced the objective function with a randomly generated objective function. We then tested this practical version of the analyzed algorithm for the case that the cuts are generated at one additional near-optimal extreme point solution. To find this additional point at each iteration, we generated three candidate near-optimal extreme point solutions and chose the point with the highest ℓ_1-norm from the optimal solution to the current relaxation problem. We let the code start exploring near-optimal extreme points solutions after the 10th iteration.

We report the results for Example 1 and a set of randomly generated problems for $n = 10$ and $m \in \{2, 3, 4, 5, 10\}$. For each pair (n, m), we generated 5 instances. For the case that $m = 2$, 3, 4, 5, or 10, we set the time limit to 300, 600, 900, 1200, or 3600 seconds, respectively. All problems were solved on a 64-bit Windows environment using C++\CPLEX 12.7 on a PC with an Intel Core i7-2640M 2.80 GHz processor and 8.00 GB of RAM. The singular value decomposition of the residual matrix $\breve{\boldsymbol{W}} - \breve{\boldsymbol{x}}\breve{\boldsymbol{y}}^\top$ is performed using Eigen 3.3.7 library (Guennebaud et al., 2010).

We evaluated the quality of the lower bounds ("lb") obtained from our implementation of the algorithm studied in Fampa and Lee (2018) (as introduced in Section 2.2 and denoted as "`SVD`") and a practical version of the algorithm analyzed in Section 4.3 (denoted as "`Modified SVDx`") with respect to the optimal value z^* (denoted as "Opt. Val."). In `Modified SVDx`, `x` denotes the number of additional near-optimal extreme point solutions. To obtain the optimal value z^*, we replaced variables $\boldsymbol{x}$ and $\boldsymbol{y}$ with their binary expansions, linearized the individual bilinear terms using McCormick envelopes, and solved the reformulated mixed-binary linear program to optimality

Table 2: Effect of exploring one near-optimal extreme point solution for Example 1.

	`SVD`			`Modified SVD1`			
Opt. Val.	Bound	Gap (%)	# It.	Bound	Gap (%)	# It.	Gap Closed (%)
-0.5	-0.5956	19.12	77	-0.5555	11.10	136	41.95



(a) `Modified SVD1` (b) `SVD` and `Modified SVD1`

Figure 2: Lower bounds for Example 1 using `SVD` and `Modified SVD1`.
Notes: Note the difference in the y-axis scales.

using CPLEX. We computed the optimality gap by

$$\text{Gap} := \frac{z^* - \text{lb}}{|z^*|} \times 100n\%.$$

Also, in order to investigate the effect of exploring near-optimal solutions, we computed the gap closed by `Modified SVDx` relative to `SVD` as follows:

$$\text{Gap Closed} := \frac{\text{lb}(\texttt{SVD}) - \text{lb}(\texttt{Modified SVDx})}{z^* - \text{lb}(\texttt{SVD})} \times 100\%.$$

Table 2 shows that exploring one near-optimal solution resulted in a 41.95 % reduction in the remaining optimality gap for Example 1. Column "# It." of this table reports the number of iterations that took the algorithm to stop. Figure 2 depicts the evolution of the lower bound over the iteration number using `SVD` and `Modified SVD1` for Example 1. `Modified SVD1` terminated after 136 iterations. As shown in Figure 2b, although `Modified SVD` achieved a worse lower bound than that of `SVD` in the first 77 iterations (when `SVD` stopped), `Modified SVD` continued 59 more ietartions and improved the lower bound.

We summarize the results for the randomly generated instances in Table 3. The column "Time" of this table reports the total time spent to achieve the lower bound (in seconds). For `Modified SVDx`, this includes the time spent to explore the near-optimal extreme point solutions. We reported the bounds at the termination of the algorithm. The number of iterations that took the algorithm

Table 3: Effect of exploring one near-optimal extreme point solution.

			SVD			Modified SVD1			
(n, m)	No.	Opt. Val.	Gap (%)	# It.	Time (s)	Gap (%)	# It.	Time (s)	Gap Closed (%)
$(10, 2)$	1	-0.75	56.67	556	148.09	56.54	421	301.97†	0.23
	2	-1.00	39.57	378	51.58	29.25	420	300.40†	26.08
	3	-1.03	1.84	282	33.43	1.56	323	144.87	15.18
	4	-1.13	10.43	172	13.16	9.66	387	262.40	7.39
	5	-1.02	6.15	285	26.16	3.89	256	45.10	36.83
$(10, 3)$	1	-1.50	21.63	243	32.73	14.50	437	401.70	32.93
	2	-1.50	20.11	574	193.05	16.51	369	322.40	17.90
	3	-1.00	23.52	260	46.03	22.57	323	252.24	4.02
	4	-0.66	46.95	357	109.74	40.54	335	305.74	13.65
	5	-1.02	28.13	199	21.13	22.59	475	601.83†	19.72
$(10, 4)$	1	-1.02	16.80	278	83.24	11.65	448	739.01	30.67
	2	-0.53	52.47	230	44.13	46.25	512	583.71	11.86
	3	-1.37	38.54	549	288.43	35.05	224	170.58	9.07
	4	-0.28	262.08	333	100.23	250.04	425	575.67	4.59
	5	-0.66	36.79	578	319.11	33.32	408	661.91	9.43
$(10, 5)$	1	-1.02	67.49	541	437.98	61.72	477	1204.94†	8.55
	2	-1.58	34.75	530	314.81	27.5	480	1204.02†	20.87
	3	-2.11	12.81	484	319.69	6.67	525	958.00	47.92
	4	-0.99	51.58	564	646.04	46.98	462	447.93	8.91
	5	-1.02	136.4	917	1200.51†	136.23	486	1203.92†	0.12
$(10, 10)$	1	-2.11	18.74	463	771.55	16.21	553	3601.00†	13.50
	2	-1.03	97.39	964	3604.52†	92.46	664	3604.53†	5.06
	3	-1.50	54.85	779	2885.94	46.70	564	3611.08†	14.85
	4	-1.50	55.98	783	3017.33†	48.52	494	3607.58†	13.33
	5	-1.50	65.09	900	3605.45†	55.92	599	3605.07†	14.09
Average			**50.27**			**45.31**			**15.47**

† Reached the time limit. For the case that $m = 2$, 3, 4, 5, or 10, we set the time limit to 300, 600, 900, 1200, or 3600 seconds, respectively.

to stop is also reported in column "# It."

Observe from Table 3 that generating valid cuts for one additional near-optimal extreme point solution improved the optimally gap for all instances at the end of the time limit. On average, `Modified SVD1` improved the optimality gap 15.47%. This improvement was generally achieved at the expense of a higher computational time for `Modified SVD1`. As expected, adding more cuts at each iteration increases the size of the relaxation problem, which in turns, leads to a larger CGLP. So, the overall computational time might increase. The only exceptions are the 3rd instance of $(10, 4)$ and the 4th instance of $(10, 5)$, where the computational time decreased and `Modified SVD1` still improved the optimality gap.

We also tested the effect of exploring two near-optimal extreme points solutions for our randomly generated instances. Because the results were not conclusive, we do not report them here. However, the following observations are noteworthy. We observed that similar to `Modified SVD1`, `Modified SVD2` improved the optimality gap relative to `SVD` for all instances. As expected, this algorithm is computationally more demanding than `SVD`, but it still improved the optimality gap at the end of of a 3600 seconds time-limit, on average 11.18%. For our choice of parameter γ, however, we observed that `Modified SVD2` resulted in a higher or similar optimality gap relative to `Modified SVD1` in most instances. Generally speaking, this is due to the fact that exploring two near-optimal extreme point solutions leads to a computationally more demanding algorithm. So, it might take a longer time for `Modified SVD2` to reach the lower bound achieved by `Modified SVD1`. An instance for which we observed a significant improvement in the optimality gap relative to `Modified SVD1` was the 4th instance of problem $(10, 4)$, where the optimality gap reduced to 172.11%, resulting in a 34.33% reduction relative to `SVD` and a 31.17% reduction relative to `Modified SVD1`. A further tuning of parameter γ for other instances might improve the optimality gap relative to `Modified SVD1` at the termination of the algorithm.

6 Conclusion

In this paper, we studied a general nonconvex bilinear program with continuous variables. We analyzed finitely-convergent disjunctive programming-based cutting plane algorithms to obtain a global ϵ-optimal solution of a bilinear program. While the analyzed algorithms heavily rely on the ideas investigated Fampa and Lee (2018); Saxena et al. (2010), a feature of them that guarantees global ϵ-optimality is to explore *all* near-optimal extreme point solutions to a current relaxation. We provided a theoretical foundation to demonstrate that generating cuts at all near-optimal solutions can guarantee global ϵ-optimality for a bilinear program.

Since exploring all near-optimal extreme point solutions is computationally expensive, for our numerical experiments, we implemented a "practical" version of the analyzed algorithms. In fact, we generated valid cuts at the current relaxation solution and only *a few* near-optimal extreme point solutions. The results suggested that this practical implementation improved the optimality gap relative to the optimality gap resulted from the procedure that generates cuts *only* at the optimal extreme point solution using the singular value decomposition Fampa and Lee (2018). We acknowledge that the computational benefits of the analyzed algorithms lie in a judicious vertex exploration to generate cuts. Such considerations are problem-dependent and deserve further computational studies.

References

Adams, W. P. and H. D. Sherali (1993). Mixed-integer bilinear programming problems. *Math. Program. 59*(1-3), 279–305.

Al-Khayyal, F. A. and J. E. Falk (1983). Jointly constrained biconvex programming. *Math. Oper. Res. 8*(2), 273–286.

Balas, E. (1998). Disjunctive programming: Properties of the convex hull of feasible points. *Discrete Appl. Math. 89*(1-3), 3–44.

Bao, X., N. V. Sahinidis, and M. Tawarmalani (2011). Semidefinite relaxations for quadratically constrained quadratic programming: A review and comparisons. *Math. Program. 129*(1), 129.

Davarnia, D., J.-P. P. Richard, and M. Tawarmalani (2017). Simultaneous convexification of bilinear functions over polytopes with application to network interdiction. *SIAM J. Optim. 27*(3), 1801–1833.

Dey, S. S., A. Santana, and Y. Wang (2019). New SOCP relaxation and branching rule for bipartite bilinear programs. *Optim. Eng. 20*(2), 307–336.

Fampa, M. and J. Lee (2018). Efficient treatment of bilinear forms in global optimization. `arXiv preprint arXiv:1803.07625 [math.OC]`.

Glover, F. (1975). Improved linear integer programming formulations of nonlinear integer problems. *Management Sci. 22*(4), 455–460.

Guennebaud, G., B. Jacob, et al. (2010). Eigen v3. http://eigen.tuxfamily.org.

Gupte, A., S. Ahmed, M. S. Cheon, and S. Dey (2013). Solving mixed integer bilinear problems using MILP formulations. *SIAM J. Optim. 23*(2), 721–744.

Harjunkoski, I., T. Westerlund, and R. Pörn (1999). Numerical and environmental considerations on a complex industrial mixed integer non-linear programming (MINLP) problem. *Comput. Chem. Eng. 23*(10), 1545–1561.

Konno, H. (1976). A cutting plane algorithm for solving bilinear programs. *Math. Program. 11*(1), 14–27.

Lee, S., C. Phalakornkule, M. M. Domach, and I. E. Grossmann (2000). Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Comput. Chem. Eng. 24*(2-7), 711–716.

Locatelli, M. and U. Raber (2002). Packing equal circles in a square: a deterministic global optimization approach. *Discrete Appl. Math. 122*(1-3), 139–166.

Mathiesen, L. (1985). Computational experience in solving equilibrium models by a sequence of linear complementarity problems. *Oper. Res. 33*(6), 1225–1250.

McCormick, G. P. (1976). Computability of global solutions to factorable nonconvex programs: Part I – convex underestimating problems. *Math. Program. 10*(1), 147–175.

Misener, R. and C. A. Floudas (2009). Advances for the pooling problem: Modeling, global optimization, and computational studies. *Appl. Comput. Math. 8*(1), 3–22.

O'Searcoid, M. (2006). *Metric spaces.* Springer Science & Business Media.

Owen, J. H. and S. Mehrotra (2001). A disjunctive cutting plane procedure for general mixed-integer linear programs. *Math. Program. 89*(3), 437–448.

Pham, V., C. Laird, and M. El-Halwagi (2009). Convex hull discretization approach to the global optimization of pooling problems. *Ind. Eng. Chem. Res. 48*(4), 1973–1979.

Rahimian, H. and S. Mehrotra (2020). Sequential convexification of a bilinear set. Optimization Online http://www.optimization-online.org/DB_HTML/2020/01/7595.html.

Rockafellar, R. T. (1970). *Convex analysis*, Volume 28. Princeton University Press.

Rodríguez, M. A. and A. Vecchietti (2008). Enterprise optimization for solving an assignment and trim-loss non-convex problem. *Comput. Chem. Eng. 32*(11), 2812–2822.

Ruszczyński, A. (2006). *Nonlinear optimization*, Volume 13. Princeton university press.

Salinetti, G. and R. J.-B. Wets (1979). On the convergence of sequences of convex sets in finite dimensions. *SIAM Review 21*(1), 18–33.

Santana, A. and S. S. Dey (2020). The convex hull of a quadratic constraint over a polytope. *SIAM J. Optim. 30*(4), 2983–2997.

Saxena, A., P. Bonami, and J. Lee (2010). Convex relaxations of non-convex mixed integer quadratically constrained programs: extended formulations. *Math. Program. 124*(1-2), 383–411.

Saxena, A., P. Bonami, and J. Lee (2011). Convex relaxations of non-convex mixed integer quadratically constrained programs: projected formulations. *Math. Program. 130*(2), 359–413.

Sherali, H. D. and W. P. Adams (2013). *A reformulation-linearization technique for solving discrete and continuous nonconvex problems*, Volume 31. Springer Science & Business Media.

Sherali, H. D. and A. Alameddine (1992). A new reformulation-linearization technique for bilinear programming problems. *J. Global Optim. 2*(4), 379–410.

Sherali, H. D. and C. Shetty (1980). A finitely convergent algorithm for bilinear programming problems using polar cuts and disjunctive face cuts. *Math. Program. 19*(1), 14–31.

Stubbs, R. A. and S. Mehrotra (1999). A branch-and-cut method for 0-1 mixed convex programming. *Math. Program. 86*(3), 515–532.

Vaish, H. and C. Shetty (1977). A cutting plane algorithm for the bilinear programming problem. *Nav. Res. Logist. Q. 24*(1), 83–94.