A Finitely Convergent Disjunctive Cutting Plane Algorithm for Bilinear Programming

Hamed Rahimian* and Sanjay Mehrotra†

1Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston IL 60208, USA

February 13, 2020

Abstract

We develop a finitely convergent disjunctive cutting plane algorithm to obtain an $\epsilon$-optimal solution or detect infeasibility of a general nonconvex continuous bilinear program. While our cutting planes are obtained in a similar manner to Saxena et al. (2010); Fampa and Lee (2018), a feature of our algorithm that guarantees the finite convergence is exploring near-optimal extreme point solutions to a current relaxation at each iteration. In this sense, our proposed algorithm extends the idea analyzed in 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:

$$\min_{x,y} f_0^T x + g_0^T y + x^T A_0 y $$

s.t. $ f_i^T x + g_i^T y + x^T A_i y + b_i \leq 0, \quad i \in [p]$  \hspace{1cm} (1)

$$ x \leq \underline{x}, \quad y \leq \underline{y} \leq \overline{y}, $$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $A_i$, $i \in 0 \cup [p]$, are $n \times m$ matrices, $f_i, g_i \in \mathbb{R}^n$, $i \in 0 \cup [p]$, $\underline{x}, \overline{x} \in \mathbb{R}^n$, $\underline{y}, \overline{y} \in \mathbb{R}^m$, and $b_i \in \mathbb{R}, i \in [p]$. We do not consider any structure on the matrices $A_i$, $i \in 0 \cup [p]$. A 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),

*1hamed.rahimian@northwestern.edu
†mehrotra@northwestern.edu
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 (Mathiesien, 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 $w = xy$ and Al-Khayyal and Falk (1983) show that this relaxation in fact describes the convex hull of the bilinear set.

Discretizing a subset of continuous variables obtains 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.

The reformulation-linearization technique (RLT) (Sherali and Adams, 2013) has been applied to continuous BLPS. For the case that there is no interaction between the continuous variables $x$ and $y$, 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.

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 proposed 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 propose a finitely-convergent cutting plane algorithm to obtain a solution differing in objective value from the global optimal value by no more than $\epsilon$. 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 $x$ and $y$ 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. To close this gap, we modify the approach in Saxena et al. (2010); Fampa and Lee (2018) and propose a disjunctive programming-based cutting plane approach.

To the best of our knowledge, this is the first 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 (\(x^*, y^*)\), with \(x^*\) and \(y^*\) 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 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 propose 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 improvement gained from a partial implementation of our proposed cutting plane algorithms. 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 \(\text{Proj}_x(\mathcal{B})\) denote the projection of \(\mathcal{B}\) onto the \(x\)-space. Let \(e_i\) be the \(i\)-th unit vector in \(\mathbb{R}^d\). Consider two sets \(\mathcal{B}_1^2 \subseteq \mathcal{B}_1 \subseteq \mathbb{R}^d\). The Hausdorff distance between \(\mathcal{B}_1^t\) and \(\mathcal{B}_2\) is denoted by \(d_H(\mathcal{B}_1^t, \mathcal{B}_2)\) and is defined as \(d_H(\mathcal{B}_1^t, \mathcal{B}_2) := \sup_{b^1 \in \mathcal{B}_1^t} \inf_{b^2 \in \mathcal{B}_2} \|b^1 - b^2\|\). A sequence of sets \(\{\mathcal{B}^t\}\) is called a decreasing sequence of nested sets if \(\mathcal{B}^{t+1} \subseteq \mathcal{B}^t\), for \(t \geq 0\). We say that a decreasing sequence of nested closed sets \(\{\mathcal{B}^t\}\) of \(\mathbb{R}^d\) converges to a closed set \(\mathcal{B} \subseteq \mathbb{R}^d\) in Hausdorff distance, and denote it by \(\lim_{t \to \infty} \mathcal{B}^t = \mathcal{B}\), if \(d_H(\mathcal{B}^t, \mathcal{B}) \to 0\) 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 \(t \geq \hat{t}\) or for any \(\delta > 0\), there exists \(\hat{t} \geq 0\) such that for all \(t \geq \hat{t}\), we have \(\inf_{b \in \mathcal{B}} \|b - b^t\| \leq \delta\) for all \(b^t \in \mathcal{B}^t\). We say that a sequence of sets \(\{\mathcal{B}^t\}\) of \(\mathbb{R}^d\) converges to \(\mathcal{B} \subseteq \mathbb{R}^d\) in the sense of Kuratowski, and denote it by \(\mathcal{B}^t \xrightarrow{K} \mathcal{B}\) as \(t \to \infty\), if \(\limsup_{t \to \infty} \mathcal{B}^t = \liminf_{t \to \infty} \mathcal{B}^t = \mathcal{B}\). For two matrices \(A\) and \(B\), \(A \bullet B = Tr(A^\top B)\) 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:

$$\min_{(x,y,W) \in \mathcal{K}} f_0^\top x + g_0^\top y + A_0 \cdot W$$

s.t. $W = xy^\top$, \hspace{1cm} (BLP)

where

$$\mathcal{K} := \left\{ (x, y, W) \left| f_i^\top x + g_i^\top y + A_i \cdot W + b_i \leq 0, i \in [p], \right. \frac{x}{x} \leq x \leq \bar{x}, \frac{y}{y} \leq y \leq \bar{y} \right\}.$$ \hspace{1cm} (2)

Set

$$\mathcal{F} := \left\{ (x, y, W) \in \mathcal{K} \left| W = xy^\top \right. \right\}$$ \hspace{1cm} (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 $x$, $y$, and $W$, and $\mathcal{K}$ is a convex set. On the other hand, $W = xy^\top$ induces a nonconvex region.

In this paper, we are interested in disjunctive programming procedures in the space of $(x, y, W)$. A disjunctive programming procedure to treat the bilinear terms is studied in Fampa and Lee (2018), by applying McCormick convexification of $W = xy^\top$ 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 $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^m$, any feasible solution to (BLP) satisfies

$$u^\top W v = (u^\top x)(v^\top y).$$ \hspace{1cm} (4)

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

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

$$-u^\top W v + \left(\frac{u^\top x + v^\top y}{2}\right)^2 - \left(\frac{u^\top x - v^\top y}{2}\right)^2 \leq 0.$$ \hspace{1cm} (5) \hspace{1cm} (6)

Observe that the concave terms $-\left(\frac{u^\top x + v^\top y}{2}\right)^2$ and $-\left(\frac{u^\top x - v^\top y}{2}\right)^2$, 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 disjunctive programming to derive valid disjunctive cuts for $\text{conv}(\mathcal{F})$ Saxena et al. (2010). More precisely, constraints (5) and (6) give rise to the
following disjunction, which is satisfied by any feasible solution \((x, y, W)\) to (BLP):

\[
\bigvee_{r=1}^{2} \bigvee_{s=1}^{2} \tilde{s}_{rs}(c, \tilde{k}, \beta),
\]

where \(\tilde{k}\) is a (bounded) convex relaxation of \(F\) (e.g., \(K\), \(c = \text{vec}(uv^\top)\), and

\[
\tilde{s}_{rs}(c, \tilde{k}, \beta) := \begin{cases} 
\beta_{1,r} \leq \frac{u^\top x + v^\top y}{2} \leq \beta_{1,r+1}, & \beta_{2,s} \leq \frac{u^\top x - v^\top y}{2} \leq \beta_{2,s+1}, \\
\frac{u^\top Wv - \left(\frac{u^\top x + v^\top y}{2}\right) \left(\beta_{1,r} + \beta_{1,r+1}\right)}{\beta_{2,r} \beta_{1,r+1}} + \left(\frac{u^\top x - v^\top y}{2}\right)^2 & \leq 0, \\
-\frac{u^\top Wv - \left(\frac{u^\top x - v^\top y}{2}\right) \left(\beta_{2,s} + \beta_{2,s+1}\right)}{\beta_{2,s} \beta_{2,s+1}} + \left(\frac{u^\top x + v^\top y}{2}\right)^2 & \leq 0
\end{cases}
\]

for \(r, s = 1, 2\). Disjunction (7) is obtained by simultaneously splitting the range \([\beta_{1,1}, \beta_{1,3}]\) of function \(\frac{u^\top x + v^\top y}{2}\) over \(\tilde{k}\) 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 \(\frac{u^\top x - v^\top y}{2}\) over \(\tilde{k}\) 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 \(-\left(\frac{u^\top x + v^\top y}{2}\right)^2\) and \(-\left(\frac{u^\top x - v^\top y}{2}\right)^2\) 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 \(\frac{u^\top x + v^\top y}{2}\), 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 \(k \in \{1, \ldots, 4\}\). Hence, hereafter, we denote \(\tilde{s}_{rs}(c, \tilde{k}, \beta)\) as \(\tilde{s}_k(c, \tilde{k}, \beta)\). Let \(r(k)\) and \(s(k)\) denote the \(r\) and \(s\) component of the index \(k\). For the ease of exposition, for \(k \in \{1, \ldots, 4\}\), \(\beta_{1,k}\) 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 \(F\) by \(\tilde{k}\) throughout the paper.

### 2.1 Disjunctive Programming

Given a solution \((\hat{x}, \hat{y}, \hat{W})\) to the current relaxation, Fampa and Lee (2018) analyze the singular value decomposition (SVD) of \(\hat{W} - \hat{\beta}\hat{y}^\top\) to find suitable vectors \(u\) and \(v\), corresponding to a nonzero singular value \(\sigma\), i.e., \(u^\top(\hat{W} - \hat{\beta}\hat{y}^\top)v = \sigma \neq 0\), in order to form the disjunction, and subsequently, derive disjunctive cuts for \(\text{conv}(F)\) using a cut-generation linear program (CGLP). For completeness, we comment on a CGLP to obtain disjunctive cuts based on disjunction (7) in this section.

Having a \((2 \times 2)\)-way disjunction over a polyhedral relaxation \(\tilde{k}\), Fampa and Lee (2018) form a CGLP for the current relaxation solution using linearization of the constraints in (8), where the convex quadratic terms are replaced by their outer approximations, obtained at the current relaxation solution. Moreover, \(\tilde{k}\) includes the constraints of McCormick convexification of \(W = \).
\(\mathbf{x}\mathbf{y}^\top\), using the box constraints on \(\mathbf{x}\) and \(\mathbf{y}\), and all the previously added disjunctive cuts. To generate disjunctive cuts, we follow the procedure of Fampa and Lee (2018). Below, we present an abstract form of this CGLP to generate a valid inequality in order to cut off the current relaxation solution \((\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{W}})\).

**Lemma 1.** Consider a polyhedral bounded relaxation \(\tilde{K}\) of \(\mathcal{F}\), a fixed \(\mathbf{c} = \text{vec}(\mathbf{uv}^\top)\), and a choice of breakpoints \(\beta\) for a \((2 \times 2)\)-way disjunction (7). Let \(A_k\mathbf{x} + B_k\mathbf{y} + C_{-k} \cdot \mathbf{W} \geq \mathbf{d}_k\) represent the set of constraints in \(\tilde{S}_k(\mathbf{c}, \tilde{K}, \beta)\), after linearization of the quadratic terms, where \(C_{-k} \in \mathbb{R}^{n \times m}, \ k = 1, \ldots, 4\). \((\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{W}}) \in \text{conv}(\bigvee_{k=1}^{4} A_k\mathbf{x} + B_k\mathbf{y} + C_{-k} \cdot \mathbf{W} \geq \mathbf{d}_k)\) if the optimal value of the following CGLP is nonnegative

\[
\begin{align*}
\underset{\mathbf{x}, \mathbf{y}, \mathbf{W}}{\text{min}} & \quad \alpha^\top \hat{\mathbf{x}} + \theta^\top \hat{\mathbf{y}} + \Gamma \cdot \hat{\mathbf{W}} - \rho \\
\text{s.t.} & \quad A_k^\top \pi_k = \alpha, \quad B_k^\top \pi_k = \theta, \quad C_{-k}^\top \pi_k = \Gamma, \quad d_k^\top \pi_k \geq \rho, \quad \pi_k \geq 0, \forall k.
\end{align*}
\]

If the optimal value of CGLP (9) is negative and \((\alpha, \theta, \Gamma, \pi_1, \ldots, \pi_4)\) is an optimal solution to (9), then \(\alpha^\top \hat{\mathbf{x}} + \theta^\top \hat{\mathbf{y}} + \Gamma \cdot \hat{\mathbf{W}} \geq \rho\) is a valid inequality for \(\text{conv}(\bigvee_{k=1}^{4} A_k\mathbf{x} + B_k\mathbf{y} + C_{-k} \cdot \mathbf{W} \geq \mathbf{d}_k)\), which cuts off \((\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{W}})\).

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

Note that because of the linearization of the quadratic terms, the above CGLP does not necessarily cut off the current relaxation solution \((\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{W}})\) from \(\text{conv}\left(\bigvee_{k=1}^{4} \tilde{S}_k(\mathbf{c}, \tilde{K}, \beta)\right)\).

### 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 \(\mathbf{x} \in \mathbb{R}^2\), \(\mathbf{y} \in \mathbb{R}^2\), and there is only one linear constraint connecting \(\mathbf{x}\), \(\mathbf{y}\), and \(\mathbf{W}\) as follows:

\[
\begin{align*}
\mathbf{x} &= [0, 0]^\top, \quad \bar{\mathbf{x}} = [2, 4]^\top, \quad \mathbf{f}_0 = [1, 2]^\top, \quad \mathbf{f}_1 = [2, 0.5]^\top, \\
\mathbf{y} &= [0, 0]^\top, \quad \bar{\mathbf{y}} = [1, 2]^\top, \quad \mathbf{g}_0 = [1, 1]^\top, \quad \mathbf{g}_1 = [2, 1]^\top, \\
A_0 &= [-1, -2.5; -1, -3], \quad A_1 = [1, 1; 1, 1], \quad b = 3.
\end{align*}
\]

At each iteration, given a solution \((\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{W}})\) to the current relaxation \(\hat{K}\), we obtain the left- and right-singular vectors \(\mathbf{u}\) and \(\mathbf{v}\), respectively, corresponding to the largest singular value of \(\hat{\mathbf{W}} - \hat{\mathbf{x}}\hat{\mathbf{y}}\). Then, using \(\text{vec}(\mathbf{uv}^\top)\), we form a \((2 \times 2)\)-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: \(\beta_{1,2} = \frac{\mathbf{u}^\top \hat{\mathbf{x}} + \mathbf{v}^\top \hat{\mathbf{y}}}{2}\) and \(\beta_{2,2} = \frac{\mathbf{u}^\top \hat{\mathbf{x}} - \mathbf{v}^\top \hat{\mathbf{y}}}{2}\). The other breakpoints are due to the lower and upper bounds of the corresponding functions over \(\hat{K}\). We refer to this algorithm as “SVF” (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 $-0.5$. 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 $\text{vec}(e_i e_j)$, we form a $(2 \times 2)$-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 $\tilde{K}$. We refer to this algorithm as “STD” (standing for the standard basis), and compare the results with “SVD”. The results are summarized in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>STD</th>
<th>SVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bound</td>
<td>-0.7328</td>
<td>-0.5956</td>
</tr>
<tr>
<td>Gap (%)</td>
<td>46.56</td>
<td>19.12</td>
</tr>
</tbody>
</table>

Table 1: Comparison of STD and SVD of $W - xy^\top$ with a $(2 \times 2)$-way disjunction.

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. However, even SVD leaves an optimality gap of 19.12%.

Motivated by Example 1, in Section 4, we propose a finitely-convergent algorithm to reach an $\epsilon$-optimal solution or detect infeasibility of (BLP) using any finite collection of bases $\text{vec}(uv^\top)$ for $\mathbb{R}^n$ and $\mathbb{R}^m$, including the standard bases. We then extend our algorithm to the case that the bases are found through the singular value decomposition of the residual matrix $W - xy^\top$, where a finite collection of bases are generated sequentially.
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 our finitely-convergent algorithms. In Section 3.2, we present valid disjunctions. In Section 3.3, we describe valid separating inequalities and the corresponding minimum distance problem.

3.1 Problem Reformulation

As explained in (4), we have

\[ \mathcal{F} = \left\{ (x, y, W) \in \mathcal{K} \mid u^\top W v = (u^\top x)(v^\top y), \forall u \in \mathbb{R}^n, v \in \mathbb{R}^m \right\}. \]

A set closely related to this set is

\[ \mathcal{F}' := \left\{ (x, y, W) \in \mathcal{K} \mid u_i^\top W v_j - (u_i^\top x)(v_j^\top y) \leq \epsilon, \forall u \in \mathbb{R}^n, v \in \mathbb{R}^m \right\}, \]

which is the set of \( \epsilon \)-feasible solutions to \( \mathcal{F} \) for \( \epsilon > 0 \). A key observation for our proposed algorithms is to represent \( \mathcal{F} \) (and \( \mathcal{F}' \)) with a finite number of nonlinear constraints.

**Proposition 1.** Let \( \{u_1, \ldots, u_n\} \) denote a set of mutually orthonormal vectors in \( \mathbb{R}^n \), and \( \{v_1, \ldots, v_m\} \) denote a set of mutually orthonormal vectors in \( \mathbb{R}^m \). Then, (BLP) can be equivalently written as

\[
\begin{align*}
\min_{(x, y, W) \in \mathcal{K}} & \quad f_0^\top x + g_0^\top y + A_0 \cdot W \\
\text{s.t.} & \quad u_i^\top W v_j = (u_i^\top x)(v_j^\top y), \ i \in [n], \ j \in [m].
\end{align*}
\]

**(BLP)**

Proof. Observe that \( W = xy^\top \Rightarrow u^\top W v = (u^\top x)(v^\top y) \ \forall u \in \mathbb{R}^n, v \in \mathbb{R}^m \), including \( u_i, v_j, i \in [n], j \in [m] \). We show that if \( u_i^\top W v_j = (u_i^\top x)(v_j^\top y), \ \forall i \in [n], \ j \in [m] \Rightarrow W = xy^\top \). Because \( \{u_1, \ldots, u_n\} \) is orthonormal, any \( u \in \mathbb{R}^n \) can be written as \( u = \sum_{i=1}^n \lambda_i u_i \) for some \( \lambda \in \mathbb{R}^n \). Similarly, any \( v \in \mathbb{R}^m \) can be written as \( v = \sum_{j=1}^m \mu_j v_j \) for some \( \mu \in \mathbb{R}^m \). Thus, we have

\[
\begin{align*}
\sum_{i=1}^n \lambda_i u_i^\top W v_j &= (u_i^\top x)(v_j^\top y), \forall i \in [n], \ j \in [m] \\
\Rightarrow \sum_{i=1}^n \lambda_i u_i^\top W v_j &= (u_i^\top x) \left( \sum_{j=1}^m \mu_j v_j \right) = (u_i^\top x)(v_j^\top y), \forall \lambda \in \mathbb{R}^n, \ j \in [m],
\end{align*}
\]

\[
\Rightarrow \sum_{i=1}^n \lambda_i u_i^\top W \left( \sum_{j=1}^m \mu_j v_j \right) = (u_i^\top x)(v_j^\top y), \forall \lambda \in \mathbb{R}^n, \ \mu \in \mathbb{R}^m
\]

\[
\Rightarrow u^\top W v = (u^\top x)(v^\top y), \forall u \in \mathbb{R}^n, v \in \mathbb{R}^m
\]

\[
\Rightarrow W = xy^\top,
\]

by taking \( u \) and \( v \) be the bases vectors. Consequently, the result follows. \( \square \)
Similar to Proposition 1, we can write \( \mathcal{F}^\varepsilon \) as follows:

\[
\mathcal{F}^\varepsilon = \left\{ (x, y, W) \in \mathcal{K} \left| u_i^\top W v_j - (u_i^\top x)(v_j^\top y) \leq \varepsilon, i \in [n], j \in [m] \right\}
\]

for \( \varepsilon > 0 \). Observe that if \( \mathcal{F}^\varepsilon \) is an empty set, then, \( \mathcal{F} \) is an empty set as well.

Throughout this section, we assume that \( \{u_1, \ldots, u_n\} \) is a set of mutually orthonormal vectors in \( \mathbb{R}^n \), and \( \{v_1, \ldots, v_m\} \) 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 \( \text{vec}(u_av_j^\top) \) by \( e_a \).

### 3.2 Valid Single-Vector Disjunction

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

\[
\mathcal{P}_a(\hat{\mathcal{K}}, \beta) := \text{conv}\left( \bigvee_{k=1}^4 \hat{S}_k(e_a, \hat{\mathcal{K}}, \beta) \right).
\]

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

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

\[
\begin{align*}
\beta_{1,1} &= \min \left\{ \frac{u_a^\top \hat{x} + v_j^\top \hat{y}}{2} \right\} (x, y, W) \in \tilde{\mathcal{K}} \right\}, \\
\beta_{1,2} &= \min \left\{ \frac{u_a^\top \hat{x} - v_j^\top \hat{y}}{2} \right\} (x, y, W) \in \tilde{\mathcal{K}} \right\}, \\
\beta_{1,3} &= \max \left\{ \frac{u_a^\top \hat{x} + v_j^\top \hat{y}}{2} \right\} (x, y, W) \in \tilde{\mathcal{K}} \right\}, \\
\beta_{2,1} &= \min \left\{ \frac{u_a^\top \hat{x} - v_j^\top \hat{y}}{2} \right\} (x, y, W) \in \tilde{\mathcal{K}} \right\}, \\
\beta_{2,2} &= \frac{u_a^\top \hat{x} - v_j^\top \hat{y}}, \\
\beta_{2,3} &= \max \left\{ \frac{u_a^\top \hat{x} - v_j^\top \hat{y}}{2} \right\} (x, y, W) \in \tilde{\mathcal{K}} \right\}.
\end{align*}
\]

Lemma 2 shows that the choice of breakpoints \( \beta \) based on Construction 1 leads to a valid single-vector disjunction.

**Lemma 2.** Consider a bounded convex relaxation \( \tilde{\mathcal{K}} \) of \( \mathcal{F} \). Given \( \varepsilon > 0 \), let \((\hat{x}, \hat{y}, \hat{W}) \in \tilde{\mathcal{K}} \) be such that \( |u_a^\top W v_a - (u_a^\top \hat{x})(v_a^\top \hat{y})| > \varepsilon \) for some \( a \in [nm] \). Let \( \bigvee_{k=1}^4 \hat{S}_k(e_a, \tilde{\mathcal{K}}, \beta) \) be a \((2 \times 2)\)-way disjunction, where the breakpoints are chosen as in Construction 1 and using \((\hat{x}, \hat{y}, \hat{W}) \). Then, for all \( k \in \{1, 2, 3, 4\} \), one of the secant inequalities in \( \hat{S}_k(e_a, \tilde{\mathcal{K}}, \beta) \), defined in (8), is violated by \((\hat{x}, \hat{y}, \hat{W})\) and the amount of violation is greater than \( \varepsilon \).

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

$$u_a^\top \hat{W} v_a - \left( \frac{u_a^\top \hat{x} + v_a^\top \hat{y}}{2} \right) = u_a^\top \hat{W} v_a - \left( \frac{u_a^\top \hat{x} + v_a^\top \hat{y}}{2} \right)^2 - \left( \frac{u_a^\top \hat{x} - v_a^\top \hat{y}}{2} \right)^2 \leq 0. \tag{11}$$

Note that if $	ilde{S}_k(c, \hat{K}, \beta)$ is such that $\beta_{1,2} \leq \frac{u_a^\top x + v_a^\top y}{2} \leq \beta_{1,3}$, we would still get a similar conclusion as in (11). So, the definition of $\tilde{S}_k(c, \hat{K}, \beta)$ as above is without loss of generality. With a similar argument, we conclude $-u_a^\top \hat{W} v_a + \left( \frac{u_a^\top \hat{x} + v_a^\top \hat{y}}{2} \right)^2 - \left( \frac{u_a^\top \hat{x} - v_a^\top \hat{y}}{2} \right)^2 \leq 0$. Thus, the above two inequalities imply that $u_a^\top \hat{W} v_a = (u_a^\top \hat{x})(v_a^\top \hat{y})$, yielding a contradiction.

Now, we show that for all $k \in \{1, 2, 3, 4\}$, one of the secant inequalities in $\tilde{S}_k(c, \hat{K}, \beta)$ is violated by $(\hat{x}, \hat{y}, \hat{W})$ and the amount of violation is greater than $\epsilon$. First, suppose that $u_a^\top \hat{W} v_a - (u_a^\top \hat{x})(v_a^\top \hat{y}) > \epsilon$. Using the equality $(u_a^\top \hat{x})(v_a^\top \hat{y}) = \left( \frac{u_a^\top \hat{x} + v_a^\top \hat{y}}{2} \right)^2 - \left( \frac{u_a^\top \hat{x} - v_a^\top \hat{y}}{2} \right)^2$, we have $u_a^\top \hat{W} v_a - \left( \frac{u_a^\top \hat{x} + v_a^\top \hat{y}}{2} \right)^2 + \left( \frac{u_a^\top \hat{x} - v_a^\top \hat{y}}{2} \right)^2 > \epsilon$. The left-hand side of this inequality is the left-hand side of (11), implying that $(\hat{x}, \hat{y}, \hat{W})$ violates the first secant inequality of $\tilde{S}_k(c, \hat{K}, \beta)$ for all $k \in \{1, 2, 3, 4\}$, and the amount of violation is greater than $\epsilon$.

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

### 3.3 Valid Disjunctive Cut and Minimum Distance Problem

So far, we established our construction to choose the breakpoints $\beta$. We now show that the choice of breakpoints $\beta$ based on Construction 1 leads to a valid disjunctive cut. To obtain a valid cut for $\text{conv}(\mathcal{F})$ using a $(2 \times 2)$-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 minimum distance problem that minimizes the distance, measured by some norm, from $(\hat{x}, \hat{y}, \hat{W})$ to a point in $\mathcal{P}_a(\hat{K}, \beta)$.

**Proposition 2.** Consider a bounded convex relaxation $\hat{K}$ of $\mathcal{F}$. Let $(\hat{x}, \hat{y}, \hat{W})$ be an optimal extreme point solution of $\min_{(x, y, W) \in \hat{K}} f_0^\top x + g_0^\top y + A_0 \bullet W$. Suppose that $u_a^\top \hat{W} v_a \neq (u_a^\top \hat{x})(v_a^\top \hat{y})$ for some $a \in [nm]$. Moreover, let $\bigvee_{k=1}^4 \tilde{S}_k(c_a, \hat{K}, \beta)$ be a $(2 \times 2)$-way disjunction, where the breakpoints are chosen as in Construction 1 and using $(\hat{x}, \hat{y}, \hat{W})$. Furthermore, suppose that $\mathcal{P}_a(\hat{K}, \beta)$, defined in (10), is nonempty. Then, the following minimum distance problem

$$\min_{(x, y, W) \in \mathcal{P}_a(\hat{K}, \beta)} \| (x, y, W) - (\hat{x}, \hat{y}, \hat{W}) \| \tag{12}$$

10
Because (\hat{x}, y, W), then return \text{viol}=\text{TRUE}, \alpha, \theta, \Gamma, and \rho. Otherwise, return \text{viol}=\text{FALSE}, \alpha = 0, \theta = 0, 
\Gamma = 0, and \rho = 0.

Algorithm 1: SepCuts(\hat{x}, y, W; \tilde{K}, u, v)

1: \textbf{Input}: \( (\hat{x}, y, W) \) and \( \text{vec}(uW^\top) \).
2: \textbf{Output}: \( \text{viol}, \alpha, \theta, \Gamma, \rho \). If a valid inequality \( \alpha^\top x + \theta^\top y + \Gamma \bullet W \geq \rho \) is found that is violated by 
(\hat{x}, y, W), then return \text{viol}=\text{TRUE}, \alpha, \theta, \Gamma, and \rho. Otherwise, return \text{viol}=\text{FALSE}, \alpha = 0, \theta = 0, 
\Gamma = 0, and \rho = 0.

3: Let \( \alpha = 0, \theta = 0, \Gamma = 0, \) and \( \rho = 0. \)
4: Let \( c = \text{vec}(uW^\top) \) and \( \beta \) is chosen as 
\[ \beta_{1,1} = \min \left\{ \frac{u^\top x + v^\top y}{2} \mid (x, y, W) \in \tilde{K} \right\}, \]
\[ \beta_{1,2} = \max \left\{ \frac{u^\top x + v^\top y}{2} \mid (x, y, W) \in \tilde{K} \right\}, \]
\[ \beta_{2,1} = \min \left\{ \frac{u^\top x - v^\top y}{2} \mid (x, y, W) \in \tilde{K} \right\}, \]
\[ \beta_{2,2} = \max \left\{ \frac{u^\top x - v^\top y}{2} \mid (x, y, W) \in \tilde{K} \right\}. \]
5: if \( S_k(c, \hat{K}, \beta) = \emptyset \) for all \( k \in \{1, 2, 3, 4\} \) then 
6: \text{viol} \leftarrow \text{FALSE}.
7: \textbf{else} 
8: \text{Let} \( (x^*, y^*, W^*) \) be an optimal solution to \( \min_{(x,y,W)\in P_k(\hat{K},\beta)} \| (x,y,W) - (\hat{x},\hat{y},\hat{W}) \| \).
9: \text{Let} \( \alpha, \theta, \Gamma \) be partial subgradients of \( \| (x,y,W) - (\hat{x},\hat{y},\hat{W}) \| \) at \( (x^*, y^*, W^*) \) with respect to 
\( x, y, \) and \( W, \) respectively. \( \rho = \alpha^\top x^* + \theta^\top y^* + \Gamma \bullet W^*. \)
10: \text{viol} \leftarrow \text{TRUE}.
11: \textbf{end if}

has a strictly positive and finite optimal value, where \( \| \cdot \| \) denotes the \( \ell_2 \)-norm.

Proof. By Lemma 2, \( (\hat{x}, \hat{y}, \hat{W}) \notin S_k(c, \hat{K}, \beta), k \in \{1, 2, 3, 4\} \). Thus, \( (\hat{x}, \hat{y}, \hat{W}) \notin \bigvee_{k=1}^4 S_k(c, \hat{K}, \beta) \).
Because \( (\hat{x}, \hat{y}, \hat{W}) \) is an extreme point of \( \hat{K} \), it cannot be written as a convex combination of points in \( \hat{K} \), including the points in \( \bigvee_{k=1}^4 S_k(c, \hat{K}, \beta) \). Thus, \( (\hat{x}, \hat{y}, \hat{W}) \notin P_a(\hat{K}, \beta) \).
Because, \( P_a(\hat{K}, \beta) \neq \emptyset \), then, the optimal value to (12) is finite. Moreover, by Ruszczyński (2006, Theorem 2.14),
(\hat{x}, \hat{y}, \hat{W}) can be strongly separated from \( P_a(\hat{K}, \beta) \). By Rockafellar (1970, Theorem 11.4),
the strong separation holds if and only if the optimal value to (12) is strictly positive.

The implication of Proposition 2 is that by choosing the breakpoints \( \beta \) according to Construction
1, one can separate \((\hat{x}, \hat{y}, \hat{W})\) from \( P_a(\hat{K}, \beta) \), and consequently, from \( \text{conv}(F) \).
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 \( \epsilon \)-optimal solution to (BLP) or to detect infeasibility
in a finite number of iterations.

One may ask how to describe \( P_a(\hat{K}, \beta) \) in the minimum distance problem (12). We devote the
rest of this section to answer this question.

Assume that the convex quadratic-representable set \( S_k(c, \hat{K}, \beta) \) is written as
\[ S_k(c, \hat{K}, \beta) = \left\{ (x, y, W) \in \hat{K} \mid F_{a,k}(x, y, W, \beta) \leq 0 \right\}, \]
for \( k = 1, \ldots, 4 \). Let us define a binary variable \( z_k \), where \( z_k = 0 \) implies that \( F_{a,k}(x, y, W, \beta) \leq 0 \).
We define the convex set

\[
Z_a(\tilde{K}, \beta) := \left\{ (x, y, W, z) \in \tilde{K}, \sum_{k=1}^{4} (1 - z_k) \geq 1, \ F_{a,k}(x, y, W, \beta) - z_k M_{a,k} \leq 0, \ 0 \leq z_k \leq 1, \ k = 1, \ldots, 4 \right\},
\]

(13)

where \(M_{a,k}\) is a sufficiently large vector to ensure that when \(z_k = 1\), constraints \(F_{a,k}(x, y, W, \beta) \leq 0\) are not active. Such a vector exists because \(\|u_a\| = \|v_a\| = 1\). Also, we define the mixed-binary convex set

\[
Z_a(\tilde{K}, \beta) := \left\{ (x, y, W, z) \in Z_a(\tilde{K}, \beta) \mid z_k \in \{0, 1\}, \ k = 1, \ldots, 4 \right\}.
\]

(14)

**Proposition 3.** Consider a relaxation \(\tilde{K}\) and \(c_a, a \in [nm]\). Let \(Z_a(\tilde{K}, \beta)\) and \(Z_a(\tilde{K}, \beta)\) be defined as in (13) and (14), respectively. For a binary variable \(z_{a,k}, k = 1, \ldots, 4\), define \(Z'_a(\tilde{K}, \beta) := \left\{ (x, y, W, z) \in Z_a(\tilde{K}, \beta) \mid z_k = 0 \right\}\) and \(Z'_a(\tilde{K}, \beta) := \left\{ (x, y, W, z) \in Z_a(\tilde{K}, \beta) \mid z_k = 1 \right\}\). Let \(M_k(Z_a(\tilde{K}, \beta)) := \left\{ (x, y, W, z, b_0, b_1, \lambda_0, \lambda_1) \mid (x, y, W, z) = \lambda_0 b_0 + \lambda_1 b_1, \ \lambda_0 + \lambda_1 = 1, \ \lambda_0 \geq 0, \ \lambda_1 \geq 0, \ b_0 \in Z'_a(\tilde{K}, \beta), \ b_1 \in Z'_a(\tilde{K}, \beta) \right\}\) .

Let \(\pi_k(Z_a(\tilde{K}, \beta))\) be the projection of \(M_k(Z_a(\tilde{K}, \beta))\) onto the \((x, y, W, z)\)-space. Then, \(\text{conv}(Z_a(\tilde{K}, \beta)) = \pi_4\left(\pi_3\left(\pi_2\left(\pi_1(Z_a(\tilde{K}, \beta))\right)\right)\right)\). Moreover, we have \(P_a(K, \beta) = \text{Proj}_{(x,y,W)}(Z_a(\tilde{K}, \beta))\), where \(P_a(K, \beta)\) is defined in (10).

To prove Proposition 3, we first provide a result on the sequential convexification of a mixed-binary convex set.

**Lemma 3.** Stubbs and Mehrotra (1999, Theorem 1) Let \(Z\) be a compact mixed-binary set as follows

\[
Z := \{ z \in Z \mid z_i \in \{0, 1\}, \ i \in [\kappa]\},
\]

where

\[
Z := \left\{ z \in \mathbb{R}^\varepsilon \mid h_\varepsilon(z) \leq 0, \ \varepsilon \in [\ell], \ 0 \leq z_\varepsilon \leq 1, \ i \in [\kappa] \right\},
\]

is a compact, convex, continuous relaxation of \(Z\). For a binary variable \(z_i, i \in [\kappa]\), define \(Z_i^0 := \left\{ z \in Z \mid z_i = 0 \right\}\), and \(Z_i^1 := \left\{ z \in Z \mid z_i = 1 \right\}\). Let

\[
\mathcal{M}_i(Z) := \left\{ (z, b_0, b_1, \lambda_0, \lambda_1) \mid z = \lambda_0 b_0 + \lambda_1 b_1, \ b_0 \in Z_i^0, \ b_1 \in Z_i^1, \ \lambda_0 + \lambda_1 = 1, \ \lambda_0 \geq 0, \ \lambda_1 \geq 0 \right\}.
\]

Let \(\pi_i(Z)\) be the projection of \(\mathcal{M}_i(Z)\) onto the \(z\)-space. Then,

\[
\text{conv}(Z) = \pi_1\left(\pi_2\left(\ldots\left(\pi_{\kappa}(Z)\right)\ldots\right)\right).
\]
of Proposition 3. By a direct application of Lemma 3, we can obtain $Z_a(\tilde{K}, \beta)$. We now show that $P_a(\tilde{K}, \beta)$ is derived by projecting $Z_a(\tilde{K}, \beta)$ onto the $(x, y, W)$-space. By construction, $P_a(\tilde{K}, \beta) \subseteq \text{Proj}_{(x,y,W)}(Z_a(\tilde{K}, \beta))$. To prove the reverse inclusion, consider a feasible point $(x, y, W, z) \in Z_a(\tilde{K}, \beta)$. Note that $(x, y, W) \in \tilde{K}$. Because $z_{a,k} \in \{0,1\}$, there exists $k$ such that $z_{a,k} = 0$; hence, $F_{a,k}(x, y, W, \beta) \leq 0$. Thus, in conjunction with $(x, y, W, z) \in \tilde{K}$, we have $(x, y, W) \in \tilde{S}_k(c_a, \tilde{K}, \beta)$. Consequently, $(x, y, W) \in P_a(\tilde{K}, \beta)$.

Remark 1. Note that in Lemma 3, the equality constraints defining $z$ in $\mathcal{M}_i(Z)$ are nonlinear. By a transformation of variables using perspective envelopes, it is shown in Stubbs and Mehrotra (1999, Theorem 2) that the constraints defining $\mathcal{M}_i(Z)$ can be represented by a set of convex constraints as follows:

$$\tilde{\mathcal{M}}_i(Z) := \left\{ (z, d_0, d_1, \lambda_0, \lambda_1) \mid z = d_0 + d_1, (d_0, \lambda_0) \in \tilde{Z}_i^0, (d_1, \lambda_1) \in \tilde{Z}_i^1, \lambda_0 + \lambda_1 = 1, \lambda_0 \geq 0, \lambda_1 \geq 0 \right\},$$

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

$$\tilde{Z} := \{ (\tilde{z}, \lambda) \mid \psi_\zeta(\tilde{z}, \lambda) \leq 0, \zeta \in \{\zeta\}, 0 \leq \tilde{z}_i \leq \lambda, i \in \kappa, 0 \leq \lambda \leq 1 \},$$

where

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

To describe $P_a(\tilde{K}, \beta)$ in the minimum distance problem (12), we use the results in Proposition 3 and Remark 1. Consider the binary variables $z_{a,k}$, $k = 1, \ldots, 4$. 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 3, but by simultaneously considering all different possibilities of all the binary variables $z_{a,k}$, $k = 1, \ldots, 4$. Associated with this set, one can form set $\tilde{\mathcal{M}}_{1,2,3,4}(Z)$ in a similar fashion to Remark 1. Then, it is straightforward to verify that an optimal solution to (12) is given by the $(x, y, W)$ components of an optimal solution to the convex program:

$$\min_{(x, y, W, z, \lambda_{d(0,0,0,0)}, \ldots, \lambda_{d(1,1,1,1)}, \lambda_{(0,0,0,0)}, \ldots, \lambda_{(1,1,1,1)} \in \mathcal{M}_{1,2,3,4}(Z_a(\tilde{K}, \beta))} \| (x, y, W) - (\hat{x}, \hat{y}, \hat{W}) \|$$

(15)

Let $(x^*, y^*, W^*)$ be the $(x, y, W)$ components of an optimal solution to (15). We can generate the valid inequality $\alpha^\top x + \theta^\top y + \Gamma \cdot W \geq \rho$, where $\alpha$, $\theta$, and $\Gamma$ are partial subgradients of $\| (x, y, W) - (\hat{x}, \hat{y}, \hat{W}) \|$ at $(x^*, y^*, W^*)$ with respect to $x$, $y$, and $W$, respectively. Moreover, $\rho = \alpha^\top x^* + \theta^\top y^* + \Gamma \cdot W^*$. 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, we propose finitely-convergent cutting plane algorithms to obtain an \( \epsilon \)-optimal solution to (BLP) or detect infeasibility. Our proposed 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 \( \hat{W} - \hat{x}\hat{y}^\top \), where \((\hat{x}, \hat{y}, \hat{W})\) 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, we propose to generate inequalities at multiple extreme point solutions of the current relaxation. These extreme points are generated by exploring a set of near-optimal solutions to the current relaxation. Our algorithms require two input parameters \( \gamma > 0 \) and \( \epsilon > 0 \). The parameter \( \gamma \) determines the neighboring set of the current solution that the algorithm explores at each iteration. The parameter \( \epsilon \) determines the optimality and feasibility tolerance. This idea was first analyzed in Owen and Mehrotra (2001) for solving mixed-integer linear programs. We extend this idea to the general bilinear programs.

In this section, we describe modified cutting plane algorithms that generate cuts at multiple vertices of the current relaxation. In Section 4.2, we propose a finitely-convergent algorithm 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 Definition, Assumption, and Technical Results

To prove the main results in this section we give a definition of an \( \epsilon \)-optimal solution to (BLP), state a regularity assumption about \( \mathcal{F} \), and present some technical results.

**Definition 1.** Let \( \{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \) denote a set of mutually orthonormal vectors in \( \mathbb{R}^n \), and \( \{\mathbf{v}_1, \ldots, \mathbf{v}_m\} \) denote a set of mutually orthonormal vectors in \( \mathbb{R}^m \). For the optimization problem (BLP) with \( z^* = \min_{(x,y,\mathbf{W}) \in \mathcal{F}} \mathbf{f}_0^\top x + \mathbf{g}_0^\top y + A_0 \bullet \mathbf{W} \), we say that a point \((\hat{x}, \hat{y}, \hat{W}) \in \mathcal{K}\) is an \( \epsilon \)-optimal solution if \( |\mathbf{u}_a^\top \hat{W} \mathbf{v}_a - (\mathbf{u}_a^\top \hat{x})(\mathbf{v}_a^\top \hat{y})| \leq \epsilon \) for \( a \in [nm] \), and \( \mathbf{f}_0^\top \hat{x} + \mathbf{g}_0^\top \hat{y} + A_0 \bullet \hat{W} \leq z^* + \epsilon \).

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

**Assumption 1. (Problem Stability)** If \( \mathcal{F} \) is an empty set, then, \( \mathcal{F}^\eta \) 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 4.** Let \( \{B^t_1\}, \{B^t_2\}, \ldots, \{B^t_\kappa\} \) be convergent, decreasing sequences of nested nonempty compact connected \(^1\) sets of a finite-dimensional Euclidean space. If \( \lim_{t \to \infty} B^t_i = \bar{B}_i \), where \( \bar{B}_i \) is nonempty, for \( i \in [\kappa] \), then,

\[
\lim_{t \to \infty} \text{conv} \left( \cup_{i=1}^{\kappa} B^t_i \right) = \text{conv} \left( \bigcup_{i=1}^{\kappa} \lim_{t \to \infty} B^t_i \right) = \text{conv} \left( \bigcup_{i=1}^{\kappa} \bar{B}_i \right).
\]

**Proof.** We show that for any \( \delta > 0 \), there exists \( \hat{t} \geq 0 \) such that for all \( t \geq \hat{t} \), we have \( \min_{b \in \text{conv} \left( \bigcup_{i=1}^{\kappa} B^t_i \right)} \|b - b^t\| \leq \delta \) for all \( b^t \in \text{conv} \left( \bigcup_{i=1}^{\kappa} B^t_i \right) \).

First, note that because \( \lim_{t \to \infty} B^t_i = \bar{B}_i \), then, for any \( \delta > 0 \), there exists \( \hat{t}_i \geq 0 \) such that for all \( t \geq \hat{t}_i \), we have \( \min_{b \in \bar{B}_i} \|b - b^t_i\| \leq \delta \), for all \( b^t_i \in B^t_i \). Moreover, for any \( t \geq 0 \), \( b^t \in \text{conv} \left( \bigcup_{i=1}^{\kappa} B^t_i \right) \) can be written as \( b^t = \sum_{\lambda_i = 1}^{\kappa} \lambda_i b^t_i \) for some \( b^t_i \in B^t_i \) and \( \lambda_i \in [0,1] \), \( i \in [\kappa] \), such that \( \sum_{i=1}^{\kappa} \lambda_i = 1 \). Therefore,

\[
\min_{b \in \text{conv} \left( \bigcup_{i=1}^{\kappa} B^t_i \right)} \|b - b^t\| = \min_{b \in \text{conv} \left( \bigcup_{i=1}^{\kappa} \bar{B}_i \right)} \left\| \sum_{i=1}^{\kappa} \lambda_i (b - b^t_i) \right\| \\
\leq \min_{b_i \in \bar{B}_i, i \in [\kappa]} \left\| \sum_{i=1}^{\kappa} \lambda_i (b - b^t_i) \right\| \\
= \sum_{i=1}^{\kappa} \min_{b_i \in \bar{B}_i} \|b_i - b^t_i\|,
\]

where the first inequality follows because \( \{b = \sum_{i=1}^{\kappa} \lambda_i b_i \mid b_i \in \bar{B}_i, i \in [\kappa] \} \subseteq \text{conv} \left( \bigcup_{i=1}^{\kappa} \bar{B}_i \right) \). By choosing \( \hat{t} := \max_{i=1}^{\kappa} \hat{t}_i \), the result follows.

**Lemma 5 (Cantor’s Intersection Theorem).** 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 6.** Salinetti and Wets (1979, Proposition 2) Suppose that \( \{B^t\} \) is a decreasing sequence of nested closed sets of a finite-dimensional Euclidean space. Then, \( \{B^t\} \) converges to \( \bigcap_{t=1}^{\infty} B^t \) in the sense of Kuratowski, as \( t \to \infty \).

**Lemma 7.** Salinetti and Wets (1979, Corollary 3A) Suppose that \( \{B^t\} \) is a sequence of nonempty compact connected sets of a finite-dimensional Euclidean space. Then, \( \lim_{t \to \infty} B^t = \bar{B} \) if and only if \( B^t \xrightarrow{K} \bar{B} \) as \( t \to \infty \), i.e., the Hausdorff convergence implies the Kuratowski convergence and vice versa, and the limits are equal.

\(^1\)Set \( B \) is not connected if there are two disjoint open sets \( U \) and \( V \) such that \( B \subset U \cup V \), \( B \cap U \neq \emptyset \), and \( B \cap V \neq \emptyset \).
Lemma 8. Owen and Mehrotra (2001, Lemma 2) Let \( \{B^t\} \) be a convergent sequence of bounded convex sets such that \( B^{t+1} \subseteq B^t \) for all \( t \geq 0 \) and \( \lim_{t \to \infty} B^t = \bar{B} \). For each \( \tilde{b} \in \text{ext}(\bar{B}) \), there exists some sequence \( \{b^t\} \) of points in \( \text{ext}(B^t) \) with a subsequence converging to \( \tilde{b} \).

4.2 General Basis

In Algorithm 2, we propose a finitely-convergent algorithm 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 a suitable regularity assumption (explained shortly), Algorithm 2 either generates an \( \epsilon \)-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 \( \tilde{z}^t = \min_{(x,y,W) \in S^t} f^0_0 x + g^0_0 y + A_0 \cdot W \). Given \( \gamma > 0 \), we define the set of extreme point solutions whose objectives are \( \gamma \)-away from \( \tilde{z}^t \) as \( \Omega^t := \{(x,y,W) \in \text{ext}(S^t) | f^0_0 x + g^0_0 y + A_0 \cdot W - \tilde{z}^t \leq \gamma \} \). Given \( \epsilon > 0 \), we define the subset \( \epsilon \)-feasible solutions of \( \Omega^t \) as \( \Omega^t_\epsilon := \{(x,y,W) \in \Omega^t | |u^t_0 W v_a - (u^t_a x)(v^t_a y)| \leq \epsilon, a \in [nm]\} \).

Before we state and prove the main result of this section, we present some intermediate results.

Lemma 9. Consider two points \( (x,y,W), (\tilde{x}, \tilde{y}, \bar{W}) \in \bar{K} \). For \( \epsilon > 0 \) and some \( a \in [nm] \), suppose that we have \( |u^t_a (x - \tilde{x})| \leq \frac{\epsilon}{4 \max\{u^t_a y, v^t_a y\}} \) and \( |v^t_a (y - \tilde{y})| \leq \frac{\epsilon}{4 \max\{u^t_a x, u^t_a \tilde{x}\}} \). Then, \( |(u^t_a x)(v^t_a y) - (u^t_a \tilde{x})(v^t_a \tilde{y})| \leq \frac{\epsilon}{2} \).

Proof. Note that \( |(u^t_a x)(v^t_a y) - (u^t_a \tilde{x})(v^t_a \tilde{y})| \leq |u^t_a x||v^t_a (y - \tilde{y})| + |v^t_a y||u^t_a (x - \tilde{x})| \). Because \( |u^t_a \tilde{x}| \leq \max\{u^t_a x, u^t_a \tilde{x}\} \), we have \( |u^t_a \tilde{x}| |v^t_a (y - \tilde{y})| \leq \frac{\epsilon}{4} \). Similarly, we have \( |v^t_a y||u^t_a (x - \tilde{x})| \leq \frac{\epsilon}{4} \). Thus, \( |(u^t_a x)(v^t_a y) - (u^t_a \tilde{x})(v^t_a \tilde{y})| \leq \frac{\epsilon}{2} \). \( \square \)

Corollary 1 below is due to Lemma 8 and establishes the existence of a convergent subsequence of points \( \{(x^t, y^t, W^t)\}_{t \in \mathcal{T}} \) in \( \text{ext}(S^t) \).

Corollary 1. Let \( \{S^t\} \) be a sequence of sets generated by Algorithm 2. Suppose that \( \{S^t\} \) converges to a nonempty set \( \bar{S} \). Furthermore, let \( (\tilde{x}, \tilde{y}, \bar{W}) \) be an optimal extreme point solution of \( \min_{(x,y,W) \in \bar{S}} f^0_0 x + g^0_0 y + A_0 \cdot W \). Then, there exists a convergent subsequence \( \{(x^t, y^t, W^t)\}_{t \in \mathcal{T}} \) of points in \( \text{ext}(S^t) \) such that \( \{(x^t, y^t, W^t)\}_{t \in \mathcal{T}} \to (\tilde{x}, \tilde{y}, \bar{W}) \).

Corollaries 2–4 establish some properties of the convergent subsequence of points \( \{(x^t, y^t, W^t)\}_{t \in \mathcal{T}} \) in \( \text{ext}(S^t) \), introduced in Corollary 1.

Corollary 2. Consider the assumptions in Corollary 1 and let \( \{(x^t, y^t, W^t)\}_{t \in \mathcal{T}} \) be the convergent subsequence of points in \( \text{ext}(S^t) \) such that \( \{(x^t, y^t, W^t)\}_{t \in \mathcal{T}} \to (\tilde{x}, \tilde{y}, \bar{W}) \). Then, for any \( \epsilon > 0 \), there exists a sufficiently large \( t \in \mathcal{T} \) such that

(i) \( f^0_0 x^t + g^0_0 y^t + A_0 \cdot W^t - (f^0_0 \tilde{x} + g^0_0 \tilde{y} + A_0 \cdot \bar{W}) \leq \epsilon \).

(ii) \( f^0_0 x^t + g^0_0 y^t + A_0 \cdot \bar{W} - z^t \leq \epsilon \).
Proof. The result in the first part follows from the fact that \( \{(x^t, y^t, W^t)\}_{t \in \mathcal{T}} \to (\bar{x}, \bar{y}, \bar{W}) \). Moreover, because \( \lim_{t \to \infty} S^t = \tilde{S} \), we have \( \{\tilde{z}^t\} \to f_0^t \tilde{x} + g_0^t \tilde{y} + A_0 \cdot \tilde{W} \). Hence, the result in the second part follows.

**Corollary 3.** Consider the assumptions and notation in Corollary 2. Then, there exists a sufficiently large \( t \in \mathcal{T} \) such that \( (x^t, y^t, W^t) \in \Omega^t \).

**Proof.** First, note that \( f_0^t x^t + g_0^t y^t + A_0 \cdot W^t - \tilde{z}^t = f_0^t x^t + g_0^t y^t + A_0 \cdot W^t - (f_0^t \tilde{x} + g_0^t \tilde{y} + A_0 \cdot \tilde{W}) + (f_0^t x^t + g_0^t y^t + A_0 \cdot W^t - (f_0^t \tilde{x} + g_0^t \tilde{y} + A_0 \cdot \tilde{W})) - \tilde{z}^t \). By Corollary 2(i), there exists \( t_1 \) such that for \( t \geq t_1 \), \( t \in \mathcal{T} \), we have \( f_0^t x^t + g_0^t y^t + A_0 \cdot W^t - (f_0^t \tilde{x} + g_0^t \tilde{y} + A_0 \cdot \tilde{W}) \leq \frac{\gamma}{2} \). Moreover, by Corollary 2(ii), there exists \( t_2 \) such that for \( t \geq t_2 \), \( t \in \mathcal{T} \), we have \( f_0^t x^t + g_0^t y^t + A_0 \cdot W^t - \tilde{z}^t \leq \frac{\gamma}{2} \). Consequently, for \( t \geq \max\{t_1, t_2\}, t \in \mathcal{T} \), we have \( f_0^t x^t + g_0^t y^t + A_0 \cdot W^t - \tilde{z}^t \leq \gamma \), which implies that \( (x^t, y^t, W^t) \in \Omega^t \).

**Corollary 4.** Consider the assumptions and notation in Corollary 2.

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

(ii) If \( |u_a^T \tilde{W} v_a - (u_a^T \tilde{x})(v_a^T \tilde{y})| \leq \epsilon \) for all \( a \in [nm] \), then, there exists a sufficiently large \( t \in \mathcal{T} \) such that \( |u_a^T W^t v_a - (u_a^T x^t)(v_a^T y^t)| \leq 2\epsilon \) for all \( a \in [nm] \).

**Proof.** Because the norms are equivalent on a finite-dimensional vector space, there exists a positive constant \( C \) such that \( C(|u_a^T (W^t - \tilde{W}) v_a + u_a^T (x^t - \tilde{x}) + v_a^T (y^t - \tilde{y})|) \leq C(|u_a^T (W^t - \tilde{W}) v_a| + |u_a^T (x^t - \tilde{x})| + |v_a^T (y^t - \tilde{y})|) \leq \|(x^t, y^t, W^t) - (\tilde{x}, \tilde{y}, \tilde{W})\| \). Because \( \{(x^t, y^t, W^t)\}_{t \in \mathcal{T}} \to (\tilde{x}, \tilde{y}, \tilde{W}) \), there exists \( t_1 \) such that for \( t \geq t_1 \), \( t \in \mathcal{T} \), we have \( \|(x^t, y^t, W^t) - (\tilde{x}, \tilde{y}, \tilde{W})\| \leq C \frac{\epsilon}{4} \). Consequently, for \( t \geq t_1 \), \( t \in \mathcal{T} \), we have \( |u_a^T(W^t - \tilde{W})v_a| \leq \frac{\epsilon}{4} \). Similarly, there exist \( t_2 \) and \( t_3 \) such that for \( t \geq t_2 \) and \( t \geq t_3 \), \( t \in \mathcal{T} \), we have \( |u_a^T(x^t - \tilde{x})| \leq \frac{\epsilon}{8 \max\{u_a^T \tilde{x}, v_a^T \tilde{y}\}} \) and \( |v_a^T(y^t - \tilde{y})| \leq \frac{\epsilon}{8 \max\{u_a^T \tilde{x}, v_a^T \tilde{y}\}} \). For the first part, by the reverse triangle inequality and Lemma 9, we have

\[
|u_a^T W^t v_a - (u_a^T x^t)(v_a^T y^t)| \geq |u_a^T \tilde{W} v_a - (u_a^T \tilde{x})(v_a^T \tilde{y})| - |u_a^T (W^t - \tilde{W}) v_a|
\]

\[
- \|(u_a^T x^t + (v_a^T y^t) - (u_a^T \tilde{x})(v_a^T \tilde{y}))| > \frac{\epsilon}{2}
\]

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

**Lemma 10.** Consider the assumptions in Corollary 4(i). Let the breakpoints \( \beta' \) be chosen as in line 4 of Algorithm 1, for \( S^t \) and point \( (x^t, y^t, W^t) \). 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 \( \tilde{S}_k(c_a, S^t, \beta') \), defined in (8), is violated by \((\tilde{x}, \tilde{y}, \tilde{W})\), and the amount of violation is greater than \( \frac{\epsilon}{2} \).

17
Proof. Suppose that \( u_a^T \tilde{W} v_a - (u_a^T \bar{x})(v_a^T \bar{y}) > \epsilon \). We first analyze the violation of the first secant inequality in \( \tilde{S}_k(c_s, S^t, \beta') \) by \( (\bar{x}, \bar{y}, \tilde{W}) \) for all \( k, \, k \in \{1, 2, 3, 4\} \). Let us begin with those \( \tilde{S}_k(c_s, S^t, \beta') \), for which \( \beta_{1,1}' \) is the lower bound on \( \frac{u_a^T \tilde{x} + v_a^T \bar{y}}{2} \). With some algebra, we have

\[
\begin{align*}
  u_a^T \tilde{W} v_a - & \left( \frac{u_a^T \tilde{x} + v_a^T \bar{y}}{2} \right) \left( \beta_{1,1}' + \frac{u_a^T \tilde{x}' + v_a^T \bar{y}'}{2} \right) + \beta_{1,1}' \frac{u_a^T \tilde{x}' + v_a^T \bar{y}'}{2} \\
  & + \left( \frac{u_a^T \tilde{x} - v_a^T \bar{y}}{2} \right)^2 \\
  & + \left( \beta_{1,1} - \frac{u_a^T \tilde{x} + v_a^T \bar{y}}{2} \right) \left( \frac{u_a^T (x' - \tilde{x}) + v_a^T (y' - \bar{y})}{2} \right) \\
  & > \epsilon + \left( \beta_{1,1} - \frac{u_a^T \tilde{x} + v_a^T \bar{y}}{2} \right) \left( \frac{u_a^T (x' - \tilde{x}) + v_a^T (y' - \bar{y})}{2} \right) .
\end{align*}
\] (16)

Observe that for \( t \geq 0 \), we have \( \beta_{1,1}' - \frac{u_a^T \tilde{x} + v_a^T \bar{y}}{2} \leq 0 \) because \( (\bar{x}, \bar{y}, \tilde{W}) \) is suboptimal to \( \beta_{1,1}' = \min \left\{ \frac{u_a^T \tilde{x} + v_a^T \bar{y}}{2} \mid (x, y, W) \in S^t \right\} \). On the other hand, by a similar argument as in the proof of Corollary 4, we have \( \frac{u_a^T (x'-\tilde{x}) + v_a^T (y'-\bar{y})}{2} > -C\|(x', y', W') - (\bar{x}, \bar{y}, \tilde{W})\| \), for some positive constant \( C \). Hence, there exists \( t_1 \) such that for \( t \geq t_1, t \in T \), we have \( -C \left( \beta_{1,1}' - \frac{u_a^T \tilde{x} + v_a^T \bar{y}}{2} \right) \|(x', y', W') - (\bar{x}, \bar{y}, \tilde{W})\| > -\frac{\epsilon}{2} \). Thus, the violation of the inequality (16) is greater than \( \frac{\epsilon}{2} \). With a similar argument, we can conclude that for those \( \tilde{S}_k(c_s, S^t, \beta') \), for which \( \beta_{1,3}' \) is the upper bound on \( \frac{u_a^T \tilde{x} + v_a^T \bar{y}}{2} \), the violation of the first secant inequality is greater than \( \frac{\epsilon}{2} \).

For the case that \( -u_a^T \tilde{W} v_a + (u_a^T \tilde{x})(v_a^T \bar{y}) > \epsilon \), we can similarly conclude that the violation of the second secant inequality is greater than \( \frac{\epsilon}{2} \). Consequently, the result follows.

Proposition 4 shows that there exists a sufficiently large \( t \) such that the valid inequality generated at \( (x', y', W') \) have a sufficiently large depth and subsequently, can cut off point \( (\bar{x}, \bar{y}, \tilde{W}) \) from \( S^t \).

**Proposition 4.** Consider the assumptions in Lemma 10, and let \( \hat{t} \) be the sufficiently large \( t \) for which the result in Lemma 10 holds. Then,

(i) There exists \( \delta > 0 \), where \( \delta \) is defined as

\[
\delta = \min_{(x, y, W) \in P_a(S^t, \beta')} \|(\bar{x}, \bar{y}, \tilde{W}) - (x, y, W)\| .
\]

(ii) The sequence \( \{S^t\}, t \geq \hat{t}, t \in T \), has a subsequence for which the optimal value of the following minimum distance problem

\[
\delta' = \min_{(x, y, W) \in P_a(S^t, \beta')} \|(\bar{x}, \bar{y}, \tilde{W}) - (x, y, W)\|,
\]

is bounded below by \( \delta \)
(iii) There exists a sufficiently large $t \in \mathcal{T}$ such that

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

**Proof.** The first part follows from Lemma 10, in conjunction with an application of Proposition 2. Now, we prove the second part. With a similar argument as in the proof of Lemma 10, 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 $\tilde{S}_k(c_a, S^t, \beta^t)$ is violated by $(\tilde{x}, \tilde{y}, \tilde{W})$ and the amount of violation is greater than $\{\epsilon - \frac{\epsilon}{4}, \epsilon - \frac{\epsilon}{8}, \ldots\}$. Hence, this subsequence $\mathcal{T}'$ yields a subsequence $\{S^t\}_{t \in \mathcal{T}'}$, for which $\forall_{k=1}^{\delta} S_k(c_a, S^t, \beta^t)$ is more restricted than $\forall_{k=1}^{\delta} S_k(c_a, S^t, \beta^t)$. This is due the fact that $\{S^t\}$ is a decreasing sequence of nested sets and the amount of violation of one of the secant inequalities in $\tilde{S}_k(c_a, S^t, \beta^t)$, $t \in \mathcal{T}'$, increases for all $k \in \{1, 2, 3, 4\}$. Consequently, for the subsequence $\mathcal{T}'$, $\mathcal{P}_a(S^t, \beta^t)$ is more restricted than $\mathcal{P}_a(S^t, \beta^t)$, implying $\delta^t \geq \delta$.

To prove the third part, note that $\| (x^t, y^t, W^t) - (x, y, W) \| \geq \| (\tilde{x}, \tilde{y}, \tilde{W}) - (x, y, W) \| - \| (\tilde{x}, \tilde{y}, \tilde{W}) - (x^t, y^t, W^t) \|$. Thus, by part 2 and Corollary 1, we have $\min_{(x, y, W) \in \mathcal{P}_a(S^t, \beta^t)} \| (x^t, y^t, W^t) - (x, y, W) \| > \min_{(x, y, W) \in \mathcal{P}_a(S^t, \beta^t)} \| (\tilde{x}, \tilde{y}, \tilde{W}) - (x, y, W) \| - \frac{\delta}{2} \geq \delta - \frac{\delta}{2}$, for a sufficiently large $t \in \mathcal{T}$.

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 $\{e_1, \ldots, e_n\}$ and $\{e_1, \ldots, e_m\}$.

**Theorem 1.** Let $\{u_1, \ldots, u_n\}$ denote a set of mutually orthonormal vectors in $\mathbb{R}^n$, and $\{v_1, \ldots, v_m\}$ denote a set of mutually orthonormal vectors in $\mathbb{R}^m$. Assume that the stability Assumption 1 holds for $\eta \geq \epsilon$, i.e., $\mathcal{F}^0 = \emptyset$. Then, Algorithm 2 either generates an $\epsilon$-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 $\text{conv} (\mathcal{F})$ (recall $\text{conv} (\mathcal{F}) \subseteq \mathcal{P}_a(\mathcal{K}, \beta)$ by Proposition 2). 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 6. 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 Lemma 5. Hence, the algorithm terminates at line 26 of Algorithm 2 after detecting infeasibility of (BLP).

Case 2. $\tilde{S} \neq \emptyset$. Let $(\tilde{x}, \tilde{y}, \tilde{W})$ be an optimal extreme point solution of $\min_{(x, y, W) \in \tilde{S}} \mathbf{f}_0^\top x + \mathbf{g}_0^\top y + A_0 \cdot W$. By Corollary 1, there exists a convergent subsequence $\{(x^t, y^t, W^t)\}_{t \in \mathcal{T}}$ of points in $\text{ext}(S^t)$ such that $\{(x^t, y^t, W^t)\}_{t \in \mathcal{T}} \rightarrow (\tilde{x}, \tilde{y}, \tilde{W})$. We examine the cases that $(\tilde{x}, \tilde{y}, \tilde{W}) \notin \mathcal{F}^{\tilde{S}}$ and $(\tilde{x}, \tilde{y}, \tilde{W}) \in \mathcal{F}^{\tilde{S}}$ separately.

Case 2.1. $(\tilde{x}, \tilde{y}, \tilde{W}) \notin \mathcal{F}^{\tilde{S}}$. In this case, there exists some $a, a \in [nm]$, such that $|u_a^\top W v_a$
Algorithm 2 Disjunctive cutting plane for (BLP) using general bases

1: Input: $\mathcal{K}$, $\{u_i\}_{i=1}^n$, $\{v_j\}_{j=1}^m$, $\gamma$, and $\epsilon$.
2: Output: An $\epsilon$-optimal solution.
3: Set $t \leftarrow 0$ and $S^0 = \mathcal{K}$.
4: while $S^t \neq \emptyset$ do
5:     Let $\hat{z}$ be the optimal value of $\min_{(x, y, W) \in S^t} f_0^\top x + g_0^\top y + A_0 \cdot W$.
6:     Let $\Omega^t := \left\{ (x, y, W) \in \text{ext}(S^t) \left| f_0^\top x + g_0^\top y + A_0 \cdot W - \hat{z} \leq \gamma \right. \right\}$.
7:     Let $\Omega := \left\{ (x, y, W) \in \Omega^t \left| |u_0^\top W v_a - (u_0^\top x)(v_0^\top y)| \leq \epsilon, a \in [nm] \right. \right\}$.
8:     if $\Omega \neq \emptyset$ then
9:         for each $(\hat{x}, \hat{y}, \hat{W}) \in \Omega$ do
10:             if $f_0^\top \hat{x} + g_0^\top \hat{y} + A_0 \cdot \hat{W} - \hat{z} \leq \epsilon$ then
11:                 STOP and output $(\hat{x}, \hat{y}, \hat{W})$ as an $\epsilon$-optimal solution.
12:         end if
13:     end for
14:     end if
15:     $S^{t+1} := S^t$.
16:     for each $(\hat{x}, \hat{y}, \hat{W}) \in \Omega^t$ and $a \in [nm]$ such that $|u_0^\top \hat{W} v_a - (u_0^\top \hat{x})(v_0^\top \hat{y})| > \frac{\epsilon}{4}$ do
17:         Call the procedure SepCuts$(\hat{x}, \hat{y}, \hat{W}; S^t, u_a, v_a)$ to obtain $(\text{viol}, \alpha, \theta, \Gamma, \rho)$.
18:         if viol=FALSE then
19:             STOP.
20:         else
21:             Let $S^{t+1} := \left\{ (x, y, W) \in S^{t+1} \left| \alpha^\top x + \theta^\top y + \Gamma \cdot W \geq \rho \right. \right\}$.
22:         end if
23:     end for
24:     Set $t \leftarrow t + 1$.
25: end while
26: STOP.

$(u_0^\top \hat{x})(v_0^\top \hat{y}) \geq \frac{\epsilon}{4}$. Let us choose $\beta$ as in line 4 of Algorithm 1, for $\tilde{S}$ and $(\hat{x}, \hat{y}, \hat{W})$. Let us similarly define $\beta^t$ for $S^t$ and $(x^t, y^t, W^t)$.

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

1. $\| (x^t, y^t, W^t) - (\hat{x}, \hat{y}, \hat{W}) \| < \frac{\delta}{2}$ (Corollary 1),
2. $(x^t, y^t, W^t) \in \Omega^t$ (Corollary 3),
3. $|u_0^\top W^t v_a - (u_0^\top x^t)(v_0^\top y^t)| > \frac{\epsilon}{4}$ (Corollary 4(i)),
4. $\delta = \min_{(x, y, W) \in \mathcal{P}} \|(x, y, W) - (x, y, W)\| > 0$ (Proposition 4(i)),
5. $\min_{(x, y, W) \in \mathcal{P}} \|(x^t, y^t, W^t) - (x, y, W)\| > \frac{\delta}{2}$ (Proposition 4(iii)).

Hence, in iteration $t$, the algorithm generates a valid inequality (line 21 of Algorithm 2) that is violated by $(\hat{x}, \hat{y}, \hat{W})$ (this can be seen from parts 1 and 5 of the above claim, and using the triangle inequality). Thus, $(\hat{x}, \hat{y}, \hat{W}) \not\in S^{t+1}$, contradicting $(\hat{x}, \hat{y}, \hat{W}) \in \tilde{S} \subseteq S^{t+1}$. So, the case that $(\hat{x}, \hat{y}, \hat{W}) \not\in \mathcal{F}^{\frac{T}{2}}$ will not happen. If $\mathcal{F} \neq \emptyset$, then, this contradiction implies that we must have $(\hat{x}, \hat{y}, \hat{W}) \in \mathcal{F}^{\frac{T}{2}}$. On the other hand, if $\mathcal{F} = \emptyset$ (and $\mathcal{F}^\epsilon = \emptyset$ by assumption), then, this contradiction implies that the case $\tilde{S} \neq \emptyset$ is not relevant.
Case 2.2. \((\tilde{x}, \tilde{y}, \tilde{W}) \in \mathcal{F}_2^\ast\). Now, consider the case that \((\tilde{x}, \tilde{y}, \tilde{W}) \in \mathcal{F}_2^\ast\).

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

1. \((x^t, y^t, W^t) \in \Omega^t\) (Corollary 3),
2. \(|u_a^T W^t v_a - (u_a^T x^t)(v_a^T y^t)| \leq \epsilon\) for \(a \in [nm]\), implying \((x^t, y^t, W^t) \in \Omega_\epsilon\) (Corollary 4(ii)).

For a sufficiently large \(t, t \in \mathcal{T}\), that satisfies the above claim, we have \(z^t \leq \min_{(x,y,W) \in \Omega_\epsilon} f_0^T x + g_0^T y + A_0 \cdot W \leq \gamma + z^t\). Note that \(\Omega_\epsilon \neq \emptyset\) because \((x^t, y^t, W^t) \in \Omega_\epsilon\). Also, note that the optimal solution \((\tilde{x}, \tilde{y}, \tilde{W}) \in \Omega^t\) to the above optimization problem is such that \(f_0^T \tilde{x} + g_0^T \tilde{y} + A_0 \cdot \tilde{W} \leq \gamma\). Hence, regardless that \(\gamma \leq \epsilon\) or \(\gamma > \epsilon\), the algorithm terminates in the \(t\)-th iteration and yields an \(\epsilon\)-optimal solution \((\tilde{x}, \tilde{y}, \tilde{W})\). Thus, we have an \(\epsilon\)-optimal solution \((\tilde{x}, \tilde{y}, \tilde{W})\) to the feasible (BLP).

\(\blacksquare\)

### 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 \((\tilde{x}, \tilde{y}, \tilde{W})\) to the minimization over the relaxation \(S^t\), the left- and right-singular vectors \(u_t\) and \(v_t\), respectively, corresponding to the largest singular value of \(\tilde{W} - \tilde{x}\tilde{y}^T\) are obtained. If \(\text{vec}(u_t v_t^T)\) is not in the span of previously generated vectors \(V^t\), \(\text{vec}(u_t v_t^T)\) is added to the set of bases. While the bases are generated, relaxation \(S^t\) is also refined by adding disjunctive cuts through procedure \(\text{SepCuts}(\tilde{x}, \tilde{y}, \tilde{W}; S^t, u_t, v_t)\). This procedure of generating bases is continued until two sets of bases for \(\mathbb{R}^n\) and \(\mathbb{R}^m\) are available, i.e., \(|V_t| \geq \max\{n, m\}\). 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 \(\gamma\)-optimal extreme point solutions. Thus, the practical performance of our proposed algorithms depends on the choice of \(\gamma\). A larger choice of \(\gamma\) results in a fewer iterations to converge while the computational time per iteration might increase. This is due to exploring a larger set of \(\gamma\)-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 \(\gamma\) might result in a slower convergence. The choice of \(\gamma\) is problem dependent and should be tuned to trade off the computational time and improvements obtained from the cuts.

On a related note, in theory, the set of \(\gamma\)-optimal extreme point solutions can be generated using Simplex pivots of the current extreme point solution. One practical implementation is to 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.
In our numerical experiments, we explored a few “promising” near-optimal extreme point solutions using random objective function coefficients.

5 Numerical Illustration

In Section 4, we proposed finitely-convergent algorithms that are based on generating cuts at all \( \gamma \)-optimal extreme point solutions of the current relaxation. For the computational results in this section, we implemented a practical version of the algorithm proposed 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 modified algorithm was repeated until a time limit was reached or the algorithm could not find a violated cut.

To obtain valid cuts, we use procedure \( \text{SepCutsCGLP}(\mathbf{x}, \mathbf{y}, \tilde{W}; S^t, u^t, v^t) \). This procedure proceeds similar to \( \text{SepCuts}(\mathbf{x}, \mathbf{y}, \tilde{W}; S^t, u^t, v^t) \), outlined in Algorithm 1, except for it obtains valid cuts through the CGLP, introduced in Lemma 1. This CGLP is based on a \((2 \times 2)\)-way disjunction, where the disjunction is formed according to \( \text{vec}(u^tv^T) \), corresponding to the largest singular value of \( \tilde{W} - \tilde{x}\tilde{y}^T \). \( \text{SepCutsCGLP}(\mathbf{x}, \mathbf{y}, \tilde{W}; S^t, u^t, v^t) \) is different from \( \text{SepCuts}(\mathbf{x}, \mathbf{y}, \tilde{W}; S^t, u^t, v^t) \), 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 \( (\mathbf{x}, \mathbf{y}, \tilde{W}) \) from the convex hull \( P_c(S^t, \beta) \), where \( c = \text{vec}(u^tv^T) \), 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 \( f^+x + g^+y + A_0 \cdot W - z^T \leq \gamma \) to the current relaxation \( S^t \) and replaced the objective function with a randomly generated objective function. We tested our modified 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 \( \ell_1 \)-norm from the optimal solution to the current relaxation problem. We let the code start exploring near-optimal extreme points solutions after the 10-th 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++\textbackslash 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 \( \tilde{W} - \tilde{x}\tilde{y}^T \) is performed using Eigen 3.3.7 library (Guennebaud et al., 2010).

We evaluated the quality of the lower bounds (“lb”) obtained from the algorithm studied in Fampa and Lee (2018) (denoted as “SVD”) and our modified algorithm (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 \( x \) and \( y \) 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.

<table>
<thead>
<tr>
<th>SVD</th>
<th>Modified SVD1</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.5</td>
<td>-0.5956</td>
</tr>
</tbody>
</table>

(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

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

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:

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

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

<table>
<thead>
<tr>
<th>(n, m)</th>
<th>No.</th>
<th>Opt. Val.</th>
<th>SVD Gap (%)</th>
<th># It.</th>
<th>Time (s)</th>
<th>Modified SVD1 Gap (%)</th>
<th># It.</th>
<th>Time (s)</th>
<th>Gap Closed (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10, 2)</td>
<td>1</td>
<td>-0.75</td>
<td>56.67</td>
<td>556</td>
<td>148.09</td>
<td>56.54</td>
<td>421</td>
<td>301.97†</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-1.00</td>
<td>39.57</td>
<td>378</td>
<td>51.58</td>
<td>29.25</td>
<td>420</td>
<td>300.40†</td>
<td>26.08</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-1.03</td>
<td>1.84</td>
<td>282</td>
<td>33.43</td>
<td>1.56</td>
<td>323</td>
<td>144.87</td>
<td>15.18</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-1.13</td>
<td>10.43</td>
<td>172</td>
<td>13.16</td>
<td>9.66</td>
<td>387</td>
<td>262.40</td>
<td>7.39</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-1.02</td>
<td>6.15</td>
<td>285</td>
<td>26.16</td>
<td>3.89</td>
<td>256</td>
<td>45.10</td>
<td>36.83</td>
</tr>
<tr>
<td>(10, 3)</td>
<td>1</td>
<td>-1.50</td>
<td>21.63</td>
<td>243</td>
<td>32.73</td>
<td>14.50</td>
<td>437</td>
<td>401.70</td>
<td>32.93</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-1.50</td>
<td>20.11</td>
<td>574</td>
<td>193.05</td>
<td>16.51</td>
<td>369</td>
<td>322.40</td>
<td>17.90</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-1.00</td>
<td>35.22</td>
<td>260</td>
<td>46.03</td>
<td>22.57</td>
<td>323</td>
<td>252.24</td>
<td>4.02</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-0.66</td>
<td>46.95</td>
<td>357</td>
<td>109.74</td>
<td>40.54</td>
<td>335</td>
<td>305.74</td>
<td>13.65</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-1.02</td>
<td>28.13</td>
<td>199</td>
<td>21.13</td>
<td>22.59</td>
<td>475</td>
<td>601.83†</td>
<td>19.72</td>
</tr>
<tr>
<td>(10, 4)</td>
<td>1</td>
<td>-1.02</td>
<td>16.80</td>
<td>278</td>
<td>83.24</td>
<td>11.65</td>
<td>448</td>
<td>739.01</td>
<td>30.67</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-0.53</td>
<td>52.47</td>
<td>230</td>
<td>44.13</td>
<td>46.25</td>
<td>512</td>
<td>583.71</td>
<td>11.86</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-1.37</td>
<td>38.54</td>
<td>549</td>
<td>288.43</td>
<td>35.05</td>
<td>224</td>
<td>170.58</td>
<td>9.07</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-0.28</td>
<td>262.08</td>
<td>333</td>
<td>100.23</td>
<td>250.04</td>
<td>425</td>
<td>575.67</td>
<td>4.59</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-0.66</td>
<td>36.79</td>
<td>578</td>
<td>319.11</td>
<td>33.32</td>
<td>408</td>
<td>661.91</td>
<td>9.43</td>
</tr>
<tr>
<td>(10, 5)</td>
<td>1</td>
<td>-1.02</td>
<td>67.49</td>
<td>541</td>
<td>437.98</td>
<td>61.72</td>
<td>477</td>
<td>1204.94†</td>
<td>8.55</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-1.58</td>
<td>34.75</td>
<td>530</td>
<td>314.81</td>
<td>27.5</td>
<td>480</td>
<td>1204.02†</td>
<td>20.87</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-2.11</td>
<td>12.81</td>
<td>484</td>
<td>319.69</td>
<td>6.67</td>
<td>525</td>
<td>958.00</td>
<td>47.92</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-0.99</td>
<td>51.58</td>
<td>564</td>
<td>646.04</td>
<td>46.98</td>
<td>462</td>
<td>447.93</td>
<td>8.91</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-1.02</td>
<td>136.4</td>
<td>917</td>
<td>1200.51†</td>
<td>136.23</td>
<td>486</td>
<td>1203.92†</td>
<td>0.12</td>
</tr>
<tr>
<td>(10, 10)</td>
<td>1</td>
<td>-2.11</td>
<td>18.74</td>
<td>463</td>
<td>771.55</td>
<td>16.21</td>
<td>553</td>
<td>3601.00†</td>
<td>15.50</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-1.03</td>
<td>97.39</td>
<td>964</td>
<td>3604.52†</td>
<td>92.46</td>
<td>664</td>
<td>3604.53†</td>
<td>5.06</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-1.50</td>
<td>54.85</td>
<td>779</td>
<td>2885.94</td>
<td>46.70</td>
<td>564</td>
<td>3611.08†</td>
<td>14.85</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-1.50</td>
<td>55.98</td>
<td>783</td>
<td>3017.33†</td>
<td>48.52</td>
<td>494</td>
<td>3607.58†</td>
<td>13.33</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-1.50</td>
<td>65.09</td>
<td>900</td>
<td>3605.45†</td>
<td>55.92</td>
<td>599</td>
<td>3605.07†</td>
<td>14.09</td>
</tr>
</tbody>
</table>

Average 50.27 45.31 15.47

1 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.

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 the results. 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 $\gamma$, 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. The only 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 $\gamma$ 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 proposed a finitely-convergent cutting plane algorithm to obtain a global $\epsilon$-optimal solution of a bilinear program. Our algorithm generates disjunctive cutting planes. A feature of our algorithm that guarantees global $\epsilon$-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 $\epsilon$-optimality for a bilinear program.

Since exploring all near-optimal extreme point solutions is computationally expensive, for our numerical experiments, we 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 improves 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).

Acknowledgements

The authors gratefully acknowledge the support of the Office of Naval Research through grant N00014-18-1-2097-P00001.

References


