Sequential Convexification of a Bilinear Set

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

April 15, 2021

Abstract

We present a sequential convexification procedure to derive, in the limit, a set arbitrary close to the convex hull of ϵ-feasible solutions to a general nonconvex continuous bilinear set. Recognizing that bilinear terms can be represented with a finite number of nonlinear nonconvex constraints in the lifted matrix space, our procedure performs a sequential convexification with respect to all nonlinear nonconvex constraints. Moreover, our approach relies on generating lift-and-project cuts using simple 0-1 disjunctions, where cuts are generated at all fractional extreme point solutions of the current relaxation. An implication of our convexification procedure is that the constraints describing the convex hull can be used in a cutting plane algorithm to solve a linear optimization problem over the bilinear set to ϵ-optimality.

Keywords: Bilinear programming, Nonconvex programming, Disjunctive programming, Lift-and-project, Convexification

1 Introduction

In this paper, we study a general nonconvex continuous bilinear set (BLS) defined as follows:

$$\left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^n \times \mathbb{R}^m \;\middle|\; \begin{array}{l} \boldsymbol{f}_\iota^\top \boldsymbol{x} + \boldsymbol{g}_\iota^\top \boldsymbol{y} + \boldsymbol{x}^\top \boldsymbol{A}_\iota \boldsymbol{y} + b_\iota \leq 0, \quad \iota \in [p], \\ \underline{\boldsymbol{x}} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}},\ \underline{\boldsymbol{y}} \leq \boldsymbol{y} \leq \overline{\boldsymbol{y}} \end{array} \right\}, \tag{1}$$

where $\boldsymbol{A}_\iota$, $\iota \in [p]$, are $n \times m$ matrices, $\boldsymbol{f}_\iota \in \mathbb{R}^n$, $\boldsymbol{g}_\iota \in \mathbb{R}^m$, $\iota \in [p]$, $\underline{\boldsymbol{x}}, \overline{\boldsymbol{x}} \in \mathbb{R}^n$, $\underline{\boldsymbol{y}}, \overline{\boldsymbol{y}} \in \mathbb{R}^m$, and $b_\iota \in \mathbb{R}$, $\iota \in [p]$. We do not consider any structure on the matrices $\boldsymbol{A}_\iota$, $\iota \in [p]$. Optimization problems over a bilinear set of the form (1) arise in various applications in engineering and management, e.g., (Adams and Sherali, 1993; Davarnia et al., 2017; Harjunkoski et al., 1999; Locatelli and Raber, 2002; Misener and Floudas, 2009; Rodríguez and Vecchietti, 2008).

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

To derive a linear programming relaxation of a bilinear function xy, a common approach is to define a new variable w and then relax the constraint $w = xy$. When variables x and y belong to a hyperrectangle, McCormick (1976) generates a polyhedral relaxation for the bilinear set $w = xy$. Al-Khayyal and Falk (1983) show that this relaxation is equivalent to the convex hull of the bilinear set. A common approach to construct a polyhedral relaxation to (1) is to use McCormick envelopes of each bilinear term. Stronger relaxations can be obtained by using convex envelopes of the entire bilinear function, see, e.g., Tawarmalani et al. (2010); Tawarmalani and Sahinidis (2002).

Other common approaches to derive a relaxation of a bilinear set are the reformulation-linearization technique (RLT) (Sherali and Adams, 2013), second-order cone programming (SOCP), and semidefinite programming (SDP). RLT has been investigated for continuous BLSs (Sherali and Alameddine, 1992) and mixed 0-1 BLSs (Adams et al., 2004; Adams and Sherali, 1990, 1993). Although a description of the convex hull can generally be obtained using RLT, this approach introduces additional variables and constraints. Dey et al. (2019) study a bilinear set 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). Relaxation techniques that utilize SDP have also been studied (Bao et al., 2011). A combination of SDP and RLT is also investigated in Anstreicher (2009, 2012); Burer and Saxena (2012); Burer and Yang (2015); Jiang and Li (2019).

The lift-and-project methodology and disjunctive programming (Balas, 1985, 1998) are also studied to derive a relaxation to the bilinear set. Davarnia et al. (2017) propose a constructive procedure to derive the convex hull of the graphs of bilinear functions in the space of original variables, where one set of variables belongs to a general polytope and the other set of variables belongs to a simplex. Saxena et al. (2010) obtain valid disjunctive cuts for mixed-integer quadratically constrained quadratic sets by investigating the eigenvalue decomposition of the quadratic violation matrix. Fampa and Lee (2021) extend the procedure in Saxena et al. (2010) to continuous BLSs over a rectangle, using the singular value decomposition of the bilinear violation matrix (we review this approach in Section 2).

In this paper, we focus on using the lift-and-project methodology and disjunctive programming to convexify a bilinear set of the form (1). This framework takes a holistic view of the bilinear set and simultaneously considers convex and nonconvex constraints, see, e.g., the discussions in Davarnia et al. (2017); Saxena et al. (2010, 2011).

Motivated by the disjunctive programming approach in Saxena et al. (2010), we present an infinitely-convergent sequential convexification procedure to derive a set arbitrary close to the convex hull of ϵ-feasible solutions. A key observation based on which we develop the results in this paper is that bilinear terms can be represented with a finite number of nonlinear nonconvex constraints in the lifted matrix space. As we shall see in Section 3, each nonlinear nonconvex constraint is fully characterized with one vector.Our procedure performs a sequential convexification with respect to all such vectors. Moreover, our procedure borrows ideas from the infinitely-convergent sequential convexification procedure for a mixed-integer linear set, proposed in Owen and Mehrotra (2001), as well as sequential convexification procedure for a mixed-binary convex set, proposed in Stubbs and Mehrotra (1999). Although a sequential convexification procedure has been investigated for a quadratically constrained sets in Saxena et al. (2010), our procedure is the first of this type for (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, i.e., the approach in Fampa and Lee (2021). In Section 3, we demonstrate a procedure to derive a set arbitrary close to the convex hull of solutions that are ϵ-feasible with respect to a single vector. In Section 4, we extend the single-vector convexification procedure into a sequential procedure to derive a set arbitrary close to the convex hull of ϵ-feasible solutions to (1) in the limit. We end with discussions and directions of future research in Section 5.

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 regular 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. The η-ball around $\mathcal{B}$ is defined as $\mathcal{N}_\eta(\mathcal{B}) := \left\{\boldsymbol{b} \in \mathbb{R}^d \,\middle|\, \inf_{\bar{\boldsymbol{b}}\in\mathcal{B}} \|\boldsymbol{b} - \bar{\boldsymbol{b}}\| \leq \eta\right\}$, where $\|\cdot\|$ is a norm on $\mathbb{R}^d$. 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 $\bar{\mathcal{B}}$ and $\mathcal{B}^t$ are empty for all $t \geq \bar{t}$ or for any $\eta > 0$, there exists $\hat{t} \geq 0$ such that for all $t \geq \hat{t}$, we have $\inf_{\boldsymbol{b}\in\bar{\mathcal{B}}} \|\boldsymbol{b} - \boldsymbol{b}^t\| \leq \eta$ for all $\boldsymbol{b}^t \in \mathcal{B}^t$, i.e., $\mathcal{B}^t \subseteq \mathcal{N}_\eta(\bar{\mathcal{B}})$. 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 $\limsup_{t\to\infty} \mathcal{B}^t = \liminf_{t\to\infty} \mathcal{B}^t = \bar{\mathcal{B}}$, where $\limsup_{t\to\infty} \mathcal{B}^t = \bigcap_{t=1}^{\infty} \bigcup_{s=t}^{\infty} \mathcal{B}^s$ and $\liminf_{t\to\infty} \mathcal{B}^t = \bigcup_{t=1}^{\infty} \bigcap_{s=t}^{\infty} \mathcal{B}^s$. 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 (2021)

By defining additional variables $W_{ij} = x_i y_j$, $i \in [n]$, $j \in [m]$, set (1) can be equivalently written as a projection of a nonlinear set in the lifted space as follows:

$$\mathcal{F} := \left\{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \mathcal{K} \,\middle|\, \boldsymbol{W} = \boldsymbol{x}\boldsymbol{y}^\top\right\}, \tag{2}$$

where

$$\mathcal{K} := \left\{ (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{n \times m} \,\middle|\, \begin{array}{l} \boldsymbol{f}_\iota^\top \boldsymbol{x} + \boldsymbol{g}_\iota^\top \boldsymbol{y} + \boldsymbol{A}_\iota \bullet \boldsymbol{W} + b_\iota \le 0,\ \iota \in [p], \\ \underline{\boldsymbol{x}} \le \boldsymbol{x} \le \overline{\boldsymbol{x}},\ \underline{\boldsymbol{y}} \le \boldsymbol{y} \le \overline{\boldsymbol{y}} \end{array} \right\}. \tag{3}$$

Set $\mathcal{K}$ is a relaxation of (2). All the constraints in $\mathcal{K}$ are linear in $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{W}$; hence, $\mathcal{K}$ is a convex set. On the other hand, $\mathcal{F}$ is a nonconvex set because of the constraints $\boldsymbol{W} = \boldsymbol{x}\boldsymbol{y}^\top$.

In this paper, we are interested in disjunctive programming procedures to treat the bilinear terms $\boldsymbol{W} = \boldsymbol{x}\boldsymbol{y}^\top$. A disjunctive programming procedure to handle the bilinear terms in the space of $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W})$ is studied in Fampa and Lee (2021). The authors of that paper apply McCormick convexification of $\boldsymbol{W} = \boldsymbol{x}\boldsymbol{y}^\top$ and extend the ideas in Saxena et al. (2010) for symmetric quadratic terms to bilinear terms. We build our disjunctive programming approach based on the procedure in Saxena et al. (2010); Fampa and Lee (2021). Let us first recall this procedure.

For any $\boldsymbol{u} \in \mathbb{R}^n$ and $\boldsymbol{v} \in \mathbb{R}^m$, any feasible solution to (2) satisfies

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

By using the equality $(\boldsymbol{u}^\top \boldsymbol{x})(\boldsymbol{v}^\top \boldsymbol{y}) = \left(\frac{\boldsymbol{u}^\top \boldsymbol{x} + \boldsymbol{v}^\top \boldsymbol{y}}{2}\right)^2 - \left(\frac{\boldsymbol{u}^\top \boldsymbol{x} - \boldsymbol{v}^\top \boldsymbol{y}}{2}\right)^2$, we can equivalently write (4) as the following two inequalities

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

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

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, induce a nonconvex set. To treat this nonconvexity, one can approximate the concave terms with their secant inequalities and use disjunctive programming in order to obtain valid disjunctive cuts for $\operatorname{conv}(\mathcal{F})$. To be precise, constraints (5) and (6) lead to the following disjunction, which is satisfied by any feasible solution $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W})$ to (2):

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

where C is a (bounded) convex relaxation of $\mathcal{F}$ (e.g., $\mathcal{K}$ as defined in (3)), $\boldsymbol{c} = (\boldsymbol{u}^\top, \boldsymbol{v}^\top)^\top$, and

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

for $r, s \in [q]$. The $(q \times q)$-way disjunction (7) is formed by simultaneously splitting the range $[\beta_{1,1}, \beta_{1,q+1}]$ of function $\frac{\boldsymbol{u}^\top \boldsymbol{x} + \boldsymbol{v}^\top \boldsymbol{y}}{2}$ over C into q intervals, and by splitting the range $[\beta_{2,1}, \beta_{2,q+1}]$ of function $\frac{\boldsymbol{u}^\top \boldsymbol{x} - \boldsymbol{v}^\top \boldsymbol{y}}{2}$ over C into q intervals. That is, the range $[\beta_{1,1}, \beta_{1,q+1}]$ is split into q intervals $[\beta_{1,1}, \beta_{1,2}]$, $[\beta_{1,2}, \beta_{1,3}]$, ..., $[\beta_{1,q}, \beta_{1,q+1}]$, and the range $[\beta_{2,1}, \beta_{2,q+1}]$ is split into q intervals $[\beta_{2,1}, \beta_{2,2}]$, $[\beta_{2,2}, \beta_{2,3}]$, ..., $[\beta_{2,q}, \beta_{2,q+1}]$. Furthermore, the disjunction simultaneously utilizes 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. Lower and upper bounds of some intervals might coincide. However, as long as the breakpoints $\{\beta_{1,1}, \ldots, \beta_{1,q+1}\}$ are in the range of function $\frac{\boldsymbol{u}^\top \boldsymbol{x} + \boldsymbol{v}^\top \boldsymbol{y}}{2}$, we have a valid disjunction of the form (7). A similar situation might happen for the breakpoints $\{\beta_{2,1}, \ldots, \beta_{2,q+1}\}$.

Fampa and Lee (2021) construct a (2×2)-way disjunction (i.e., $q = 2$), where the breakpoints are chosen such that $\beta_{1,2} \in [\beta_{1,1}, \beta_{1,3}]$ and $\beta_{2,2} \in [\beta_{2,1}, \beta_{2,3}]$. Given a solution $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{W}})$ to the current relaxation, Fampa and Lee (2021) analyze the singular value decomposition (SVD) of the violation matrix $\hat{\boldsymbol{W}} - \hat{\boldsymbol{x}}\hat{\boldsymbol{y}}^\top$ to find suitable vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, corresponding to a nonzero singular value σ, i.e., $\boldsymbol{u}^\top(\hat{\boldsymbol{W}} - \hat{\boldsymbol{x}}\hat{\boldsymbol{y}}^\top)\boldsymbol{v} = \sigma \neq 0$, in order to form the disjunction, and subsequently, derive disjunctive cuts for $\operatorname{conv}(\mathcal{F})$ using a cut-generation linear program. Note that although Fampa and Lee (2021) study a (2×2)-way disjunction, we stated a $(q \times q)$-way disjunction for future references in this paper.

Throughout the paper, we let the index k represent the (r, s)-pair, where $k \in [q^2]$. Consequently, we denote $\mathcal{S}_{rs}(\boldsymbol{c}, C, \boldsymbol{\beta})$ as $\mathcal{S}_k(\boldsymbol{c}, C, \boldsymbol{\beta})$. Moreover, let $r(k)$ and $s(k)$ denote the r and s component of the index k. For the ease of exposition, for $k \in [q^2]$, $\beta_{1,k}$ and $\beta_{2,k}$ should be interpreted as $\beta_{1,r(k)}$ and $\beta_{2,s(k)}$, respectively. Similarly, $\beta_{1,k+1}$ and $\beta_{2,k+1}$ should be interpreted as $\beta_{1,r(k)+1}$ and $\beta_{2,s(k)+1}$, respectively. We also denote a (bounded) convex relaxation of $\mathcal{F}$ by C for the rest of the paper.

3 Convexification Procedure

As mentioned in (4), we have

$$\mathcal{F} := \left\{ (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \mathcal{K} \;\middle|\; \begin{array}{l} \boldsymbol{u}^\top \boldsymbol{W} \boldsymbol{v} = (\boldsymbol{u}^\top \boldsymbol{x})(\boldsymbol{v}^\top \boldsymbol{y}), \\ \forall \boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{v} \in \mathbb{R}^m, \text{ with } \|\boldsymbol{u}\| = 1, \ \|\boldsymbol{v}\| = 1 \end{array} \right\}, \tag{9}$$

A relaxed set is

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

which is the set of ϵ-feasible solutions to $\mathcal{F}$ for $\epsilon > 0$. Note that if set $\bar{\mathcal{F}}^\epsilon$ is empty, then set $\mathcal{F}$ is empty as well. In this paper, we develop a procedure for constructing a set arbitrary close to the convex hull of ϵ-feasible solutions, for $\epsilon > 0$. 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.

In this section, we discuss our procedure for constructing a set arbitrary close to $\text{conv}\left(\left\{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in C \,\middle|\, |\boldsymbol{u}^\top\boldsymbol{W}\boldsymbol{v}-(\boldsymbol{u}^\top\boldsymbol{x})(\boldsymbol{v}^\top\boldsymbol{y})|\le\epsilon\right\}\right)$, for a pair of vectors $\boldsymbol{u}\in\mathbb{R}^n$ and $\boldsymbol{v}\in\mathbb{R}^m$. In Section 3.3, we show how to derive such a set using a $(q\times q)$-way disjunction, for a sufficiently large q. In Section 3.4, we demonstrate our procedure to construct the desired set by generating lift-and-project cuts from simple 0-1 disjunctions. As we shall see in Section 4, the single-vector convexification procedure can be extended into a sequential procedure for constructing a set arbitrary close to the convex hull of ϵ-feasible solutions.

3.1 Set Reformulation

A key observation based on which the convexification procedure is proposed is to represent $\mathcal{F}$ with a finite number of nonlinear constraints, and to present an inner approximation to $\bar{\mathcal{F}}^\epsilon$ with a finite number of nonlinear constraints.

Proposition 1. *Rahimian and Mehrotra (2020, Proposition 1) Let $\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n\}$ denote an orthonormal set of vectors in $\mathbb{R}^n$ and $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_m\}$ denote an orthonormal set of vectors in $\mathbb{R}^m$. Then, $\mathcal{F}$ can be equivalently written as*

$$\mathcal{F}=\left\{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\mathcal{K}\,\middle|\,\boldsymbol{u}_i^\top\boldsymbol{W}\boldsymbol{v}_j=(\boldsymbol{u}_i^\top\boldsymbol{x})(\boldsymbol{v}_j^\top\boldsymbol{y}),\ i\in[n],\ j\in[m]\right\}.$$

Proposition 2. *Rahimian and Mehrotra (2020, Proposition 2) Let $\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n\}$ denote an orthonormal set of vectors in $\mathbb{R}^n$ and $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_m\}$ denote an orthonormal set of 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*

$$\mathcal{F}^\epsilon:=\left\{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\mathcal{K}\,\middle|\,|\boldsymbol{u}_i^\top\boldsymbol{W}\boldsymbol{v}_j-(\boldsymbol{u}_i^\top\boldsymbol{x})(\boldsymbol{v}_j^\top\boldsymbol{y})|\le\epsilon,\ i\in[n],j\in[m]\right\}. \tag{11}$$

An implication of Proposition 2 is that if we obtain a set that is contained in $\text{conv}(\mathcal{F}^\epsilon)$, that set will be contained in $\text{conv}\left(\bar{\mathcal{F}}^\epsilon\right)$ as well (see Theorem 3).

3.2 Disjunctive Programming

Before we proceed, let us establish some conventions for the vectors $\boldsymbol{u}\in\mathbb{R}^n$ and $\boldsymbol{v}\in\mathbb{R}^m$ and the choice of $\boldsymbol{\beta}$ in the definition of disjunction (7). We assume that $\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_n\}$ is an orthonormal set of vectors in $\mathbb{R}^n$, and $\{\boldsymbol{v}_1,\ldots,\boldsymbol{v}_m\}$ is an orthonormal set of vectors in $\mathbb{R}^m$. For the ease of exposition, we let the index a represent the (i,j)-pair, where $a\in[nm]$. We also denote $(\boldsymbol{u}^\top,\boldsymbol{v}^\top)^\top$ by $\boldsymbol{c}_a$, $a\in[nm]$.

Furthermore, unless otherwise stated, we assume that given a relaxation C, the breakpoints $\boldsymbol{\beta}$ are equally spaced for the corresponding functions $\frac{\boldsymbol{u}_a^\top\boldsymbol{x}+\boldsymbol{v}_a^\top\boldsymbol{y}}{2}$ and $\frac{\boldsymbol{u}_a^\top\boldsymbol{x}-\boldsymbol{v}_a^\top\boldsymbol{y}}{2}$. Throughout this section, we either choose breakpoints $\boldsymbol{\beta}$ based on relaxation C of $\mathcal{F}$ or based on the specific relaxation $\mathcal{K}$, defined in (3).

Construction 1. *Consider a bounded convex relaxation C of $\mathcal{F}$. We form a $(q \times q)$-way disjunction of the form (7) on C, based on $\boldsymbol{c}_a$ and using the following choice of the breakpoints:*

$$\beta_{1,1} = \min\left\{ \frac{\boldsymbol{u}_a^\top \boldsymbol{x} + \boldsymbol{v}_a^\top \boldsymbol{y}}{2} \,\middle|\, (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in C \right\}, \quad \beta_{1,q+1} = \max\left\{ \frac{\boldsymbol{u}_a^\top \boldsymbol{x} + \boldsymbol{v}_a^\top \boldsymbol{y}}{2} \,\middle|\, (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in C \right\},$$
$$\beta_{2,1} = \min\left\{ \frac{\boldsymbol{u}_a^\top \boldsymbol{x} - \boldsymbol{v}_a^\top \boldsymbol{y}}{2} \,\middle|\, (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in C \right\}, \quad \beta_{2,q+1} = \max\left\{ \frac{\boldsymbol{u}_a^\top \boldsymbol{x} - \boldsymbol{v}_a^\top \boldsymbol{y}}{2} \,\middle|\, (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in C \right\},$$

$\beta_{1,r} = \beta_{1,1} + (r-1)\frac{\beta_{1,q+1} - \beta_{1,1}}{q}$, *for* $r = 2, \ldots, q$, *and* $\beta_{2,s} = \beta_{2,1} + (s-1)\frac{\beta_{2,q+1} - \beta_{2,1}}{q}$, *for* $s = 2, \ldots, q$. *If the breakpoints are chosen based on $\mathcal{K}$, defined in (3) (i.e., computing the range over $\mathcal{K}$ instead of C in the construction above), we denote them collectively by $\boldsymbol{\theta}$ and refer to them as predetermined breakpoints. Also, we use notation $\mathcal{S}_k(\boldsymbol{c}, C, \boldsymbol{\theta})$ to emphasize that the disjunction is constructed over C but using the predetermined breakpoints $\boldsymbol{\theta}$.*

3.3 Single-Vector Convexification: Existence

Consider a (bounded) convex relaxation C of $\mathcal{F}$ and $\boldsymbol{c}_a$, $a \in [nm]$. For $\epsilon > 0$, in this section, we show that there exists a sufficiently large $q \in \mathbb{N}$ such that the $(q \times q)$-way disjunction gives rise to a set arbitrary close to

$$\text{conv}\left(\left\{ (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in C \,\middle|\, |\boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \boldsymbol{x})(\boldsymbol{v}_a^\top \boldsymbol{y})| \leq \epsilon \right\} \right).$$

We also show that the desired set contains

$$\text{conv}\left(\left\{ (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in C \,\middle|\, \boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a = (\boldsymbol{u}_a^\top \boldsymbol{x})(\boldsymbol{v}_a^\top \boldsymbol{y}) \right\} \right).$$

To state our main result on the existence of such a single-vector convexification, let us define the following sets:

$$\mathcal{P}_a^q(C, \boldsymbol{\beta}) := \text{conv}\left(\bigcup_{k=1}^{q^2} \mathcal{S}_k(\boldsymbol{c}_a, C, \boldsymbol{\beta}) \right), \tag{12}$$

$$\mathcal{P}_a^q(C, \boldsymbol{\theta}) := \text{conv}\left(\bigcup_{k=1}^{q^2} \mathcal{S}_k(\boldsymbol{c}_a, C, \boldsymbol{\theta}) \right), \tag{13}$$

$$\mathcal{C}_a(C) := \text{conv}\left(\left\{ (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in C \,\middle|\, \boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a = (\boldsymbol{u}_a^\top \boldsymbol{x})(\boldsymbol{v}_a^\top \boldsymbol{y}) \right\} \right), \tag{14}$$

and

$$\mathcal{C}_a^\epsilon(C) := \text{conv}\left(\left\{ (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in C \,\middle|\, |\boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \boldsymbol{x})(\boldsymbol{v}_a^\top \boldsymbol{y})| \leq \epsilon \right\} \right), \tag{15}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are chosen as explained in Construction 1.

Theorem 1. *Consider a convex relaxation C and $\boldsymbol{c}_a$, $a \in [nm]$. Let $\mathcal{P}_a^q(C, \boldsymbol{\beta})$, $\mathcal{P}_a^q(C, \boldsymbol{\theta})$, $\mathcal{C}_a(C)$, and $\mathcal{C}_a^\epsilon(C)$ be defined as in (12), (13), (14), and (15), respectively. Then, for $\epsilon > 0$, there exists*

$\tilde{q}_1 \in \mathbb{N}$ *such that for all* $q \geq \tilde{q}_1$*, we have*

$$\mathcal{C}_a(C) \subseteq \mathcal{P}_a^q(C, \boldsymbol{\beta}) \subseteq \mathcal{C}_a^\epsilon(C).$$

Moreover, there exists $\tilde{q}_2 \in \mathbb{N}$ *such that for all* $q \geq \tilde{q}_2$*, we have*

$$\mathcal{C}_a(C) \subseteq \mathcal{P}_a^q(C, \boldsymbol{\theta}) \subseteq \mathcal{C}_a^\epsilon(C).$$

Theorem 1 states that one can construct a relaxed set for $\mathcal{C}_a(C)$. Moreover, for a sufficiently large q, this set is not too "conservative" and is contained in $\mathcal{C}_a^\epsilon(C)$. Another interesting point is that this set can be constructed with respect to the predetermined breakpoints $\boldsymbol{\theta}$, or any other breakpoints, as long as they cover the entire range of the corresponding functions $\frac{\boldsymbol{u}_a^\top \boldsymbol{x}+\boldsymbol{v}_a^\top \boldsymbol{y}}{2}$ and $\frac{\boldsymbol{u}_a^\top \boldsymbol{x}-\boldsymbol{v}_a^\top \boldsymbol{y}}{2}$. Before we prove Theorem 1, we state an auxiliary result.

Lemma 1. *Let* $\gamma : \mathbb{R} \to \mathbb{R}$ *such that* $\gamma(z) = -z^2$*, and let* $\rho(z) = -z(a+b) + ab$ *represent the secant approximation of* $\gamma(z)$ *in the* $[a, b]$ $(a, b \in \mathbb{R})$*. Then, (i)* $\gamma(z) - \rho(z) = (z-a)(b-z)$ *and (ii)* $\max_{z\in[a,b]} \gamma(z) - \rho(z) = \frac{(a-b)^2}{4}$.

Proof. of Theorem 1 Consider $q \in \mathbb{N}$. Observe that

$$\left\{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in C \,\middle|\, \boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a = (\boldsymbol{u}_a^\top \boldsymbol{x})(\boldsymbol{v}_a^\top \boldsymbol{y})\right\} = \bigcup_{k=1}^{q^2} \mathcal{G}_k(\boldsymbol{c}_a, C, \boldsymbol{\beta}),$$

where

$$\mathcal{G}_k(\boldsymbol{c}_a, C, \boldsymbol{\beta}) := \left\{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in C \,\middle|\, \begin{array}{l} \beta_{1,k} \leq \frac{\boldsymbol{u}_a^\top \boldsymbol{x}+\boldsymbol{v}_a^\top \boldsymbol{y}}{2} \leq \beta_{1,k+1},\ \beta_{2,k} \leq \frac{\boldsymbol{u}_a^\top \boldsymbol{x}-\boldsymbol{v}_a^\top \boldsymbol{y}}{2} \leq \beta_{2,k+1}, \\ \boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a - \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x}+\boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)^2 + \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x}-\boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)^2 \leq 0, \\ -\boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a + \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x}+\boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)^2 - \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x}-\boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)^2 \leq 0 \end{array}\right\},$$

for $k \in [q^2]$ and $\boldsymbol{\beta}$ is chosen as in Construction 1 for $\vee_{k=1}^{q^2} \mathcal{S}_k(\boldsymbol{c}_a, C, \boldsymbol{\beta})$. Moreover,

$$\left\{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in C \,\middle|\, |\boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a - (\boldsymbol{u}_a^\top \boldsymbol{x})(\boldsymbol{v}_a^\top \boldsymbol{y})| \leq \epsilon\right\} = \bigcup_{k=1}^{q^2} \mathcal{G}_k^\epsilon(\boldsymbol{c}_a, C, \boldsymbol{\beta}),$$

where

$$\mathcal{G}_k^\epsilon(\boldsymbol{c}_a, C, \boldsymbol{\beta}) := \left\{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in C \,\middle|\, \begin{array}{l} \beta_{1,k} \leq \frac{\boldsymbol{u}_a^\top \boldsymbol{x}+\boldsymbol{v}_a^\top \boldsymbol{y}}{2} \leq \beta_{1,k+1},\ \beta_{2,k} \leq \frac{\boldsymbol{u}_a^\top \boldsymbol{x}-\boldsymbol{v}_a^\top \boldsymbol{y}}{2} \leq \beta_{2,k+1}, \\ \boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a - \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x}+\boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)^2 + \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x}-\boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)^2 \leq \epsilon, \\ -\boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a + \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x}+\boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)^2 - \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x}-\boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)^2 \leq \epsilon \end{array}\right\},$$

for $k \in [q^2]$ and again, $\boldsymbol{\beta}$ is chosen as in Construction 1 for $\vee_{k=1}^{q^2} \mathcal{S}_k(\boldsymbol{c}_a, C, \boldsymbol{\beta})$.

Observe that $\mathcal{G}_k(\boldsymbol{c}_a, C, \boldsymbol{\beta}) \subseteq \mathcal{S}_k(\boldsymbol{c}_a, C, \boldsymbol{\beta})$, $k \in [q^2]$. Thus, the first inclusion in the result follows. Now, we show the second inclusion. Note that by using the first part in Lemma 1, we can equivalently write

$$\mathcal{S}_k(\boldsymbol{c}, C, \boldsymbol{\beta}) = \left\{ (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in C \;\middle|\; \begin{array}{l} \beta_{1,k} \leq \frac{\boldsymbol{u}_a^\top \boldsymbol{x} + \boldsymbol{v}_a^\top \boldsymbol{y}}{2} \leq \beta_{1,k+1},\ \beta_{2,k} \leq \frac{\boldsymbol{u}_a^\top \boldsymbol{x} - \boldsymbol{v}_a^\top \boldsymbol{y}}{2} \leq \beta_{2,k+1}, \\ \boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a - \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x} + \boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)^2 + \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x} - \boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)^2 \leq \\ \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x} + \boldsymbol{v}_a^\top \boldsymbol{y}}{2} - \beta_{1,k}\right)\left(\beta_{1,k+1} - \frac{\boldsymbol{u}_a^\top \boldsymbol{x} + \boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right), \\ -\boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a + \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x} + \boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)^2 - \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x} - \boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)^2 \leq \\ \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x} - \boldsymbol{v}_a^\top \boldsymbol{y}}{2} - \beta_{2,k}\right)\left(\beta_{2,k+1} - \frac{\boldsymbol{u}_a^\top \boldsymbol{x} - \boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right) \end{array} \right\}.$$

Now, using the second part in Lemma 1, we conclude that the right-hand sides of the last two constraints in the expression of $\mathcal{S}_k(\boldsymbol{c}_a, C, \boldsymbol{\beta})$ are bounded above by $\frac{(\beta_{1,k+1}-\beta_{1,k})^2}{4}$ and $\frac{(\beta_{2,k+1}-\beta_{2,k})^2}{4}$, respectively. Note that by construction, $\{\beta_{1,1}, \ldots, \beta_{1,q+1}\}$ and $\{\beta_{2,1}, \ldots, \beta_{2,q+1}\}$ are equally spaced. Let $\hat{\beta}_1(q)$ and $\hat{\beta}_2(q)$ denote the length of the interval between two consecutive $\beta_{1,\cdot}$ and $\beta_{2,\cdot}$, respectively, for a $(q \times q)$-way disjunction. Given $\epsilon > 0$, we can choose $\tilde{q}_1$ as the smallest q such that $\hat{\beta}_1(q) \leq 2\epsilon^{\frac{1}{2}}$ and $\hat{\beta}_2(q) \leq 2\epsilon^{\frac{1}{2}}$. Thus, for $q = \tilde{q}_1$, $\mathcal{S}_k(\boldsymbol{c}_a, C, \boldsymbol{\beta}) \subseteq \mathcal{G}_k^\epsilon(\boldsymbol{c}_a, C, \boldsymbol{\beta})$ for all $k \in [q^2]$. Because, for all $q \geq \tilde{q}_1$, the length of the interval between two consecutive $\beta_{1,\cdot}$ and $\beta_{2,\cdot}$ decreases, the second inclusion follows. Similarly, we can use the predetermined breakpoints $\boldsymbol{\theta}$ instead of $\boldsymbol{\beta}$ and conclude the second part of the theorem. □

Remark 1. *It can be seen from the proof of Theorem 1 that the statement could be strengthened by writing the inclusion of the sets without the convex hull operation.* □

3.4 Single-Vector Convexification: Construction

In Section 3.3, we established that for a sufficiently large q, $\mathcal{P}_a^q(C, \boldsymbol{\beta})$ lies between $\mathcal{C}_a(C)$ and $\mathcal{C}_a^\epsilon(C)$ for an arbitrary $\epsilon > 0$. In this section, we describe a procedure to construct such a desired set $\mathcal{P}_a^q(C, \boldsymbol{\beta})$.

Before we proceed, we define some notation. Let us assume that the convex quadratic-representable set $\mathcal{S}_k(\boldsymbol{c}_a, C, \boldsymbol{\beta})$ is written in the following form

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

$k \in [q^2]$. Define a binary variable z_k, where if $z_k = 0$, then we have $\boldsymbol{F}_{a,k}^q(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}, \boldsymbol{\beta}) \leq \boldsymbol{0}$. Let us define the following convex set

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

where $\boldsymbol{M}_{a,k}$ is chosen to be a sufficiently large vector such that when $z_k = 1$, constraints $\boldsymbol{F}_{a,k}^q(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}, \boldsymbol{\beta})$

$\leq \mathbf{0}$ are not active. Such a vector $\boldsymbol{M}_{a,k}$ exists because vectors $\|\boldsymbol{u}_a\|$ and $\|\boldsymbol{v}_a\|$ are normalized, i.e., $\|\boldsymbol{u}_a\| = \|\boldsymbol{v}_a\| = 1$. Moreover, let us define the following mixed-binary convex set

$$\mathcal{Z}_a^q(C,\boldsymbol{\beta}) := \left\{ (\boldsymbol{x},\boldsymbol{y},\boldsymbol{W},\boldsymbol{z}) \in Z_a^q(C,\boldsymbol{\beta}) \middle| z_k \in \{0,1\},\ k \in [q^2] \right\}. \tag{17}$$

Observe that if there exists $\boldsymbol{z} \in \{0,1\}^{q^2}$ such that $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W},\boldsymbol{z}) \in \mathcal{Z}_a^q(C,\boldsymbol{\beta})$, then we have $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W}) \in \bigcup_{k=1}^{q^2} \mathcal{S}_k(\boldsymbol{c}_a, C, \boldsymbol{\beta})$, and vice versa. In other words, $\bigcup_{k=1}^{q^2} \mathcal{S}_k(\boldsymbol{c}_a, C, \boldsymbol{\beta}) = \text{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}(\mathcal{Z}_a^q(C,\boldsymbol{\beta}))$. Our main goal for the rest of this section is to describe the construction of $\mathcal{P}_a^q(C,\boldsymbol{\beta})$, defined in (12). Given that $\mathcal{Z}_a^q(C,\boldsymbol{\beta})$ is a compact mixed-binary convex set, we simplify the exposition using a generic compact mixed-binary convex set $\mathcal{Z}$. We first present some results on the convexification of such a set in Lemma 2, followed by a technical result in Lemma 3. Then, we propose a cutting plane procedure to construct $\text{conv}(\mathcal{Z})$ in Algorithm 1. Using these results, we translate back to $\mathcal{Z}_a^q(C,\boldsymbol{\beta})$ and describe the construction of $\mathcal{P}_a^q(C,\boldsymbol{\beta})$.

Lemma 2. *Stubbs and Mehrotra (1999, Proposition 1 and Theorem 1) Let us define a compact mixed-binary convex set*

$$\mathcal{Z} := \left\{ \boldsymbol{z} \in Z \middle|\ z_\iota \in \{0,1\},\ \iota \in [\kappa]\ \right\}, \tag{18}$$

where $\boldsymbol{z} \in \mathbb{R}^\vartheta$, and Z 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

$$\mathcal{M}_\iota(Z) := \left\{ (\boldsymbol{z}, \boldsymbol{b}_0, \boldsymbol{b}_1, \lambda_0, \lambda_1) \middle| \begin{array}{l} \boldsymbol{z} = \lambda_0 \boldsymbol{b}_0 + \lambda_1 \boldsymbol{b}_1, \\ \lambda_0 + \lambda_1 = 1,\ \lambda_0 \geq 0,\ \lambda_1 \geq 0, \\ \boldsymbol{b}_0 \in Z_\iota^0,\ \boldsymbol{b}_1 \in Z_\iota^1 \end{array} \right\}. \tag{19}$$

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

$$\pi_\iota(Z) = conv\left(Z_\iota^0 \cup Z_\iota^1\right) = conv\left(Z \cap z_\iota \in \{0,1\}\right).$$

Moreover,

$$conv(\mathcal{Z}) = \pi_\kappa\big(\pi_{\kappa-1}\big(\ldots\big(\pi_1(Z)\big)\ldots\big)\big).$$

Lemma 3. *Let $\mathcal{C}$ be a compact convex set in $\mathbb{R}^{\nu_1}$ defined as $\mathcal{C} := Proj_{\mathbf{a}}(\mathcal{U})$, 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 $(\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}) \in \text{conv}(\mathcal{U})$, we have $(\hat{\boldsymbol{a}}, \hat{\boldsymbol{b}}) \notin \text{ext}(\text{conv}(\mathcal{U}))$. 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 $i \in [\nu_1+\nu_2+1]$, $(\boldsymbol{a}^i, \boldsymbol{b}^i) \in \text{ext}(\text{conv}(\mathcal{U}))$, $\lambda^i \geq 0$, and $\sum_{i=1}^{\nu_1+\nu_2+1} \lambda^i = 1$. Note that for $i \in [\nu_1+\nu_2+1]$, we have $\boldsymbol{a}^i \notin \text{ext}(\mathcal{C})$, 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 $\boldsymbol{a}^i \notin \text{ext}(\mathcal{C})$. Because $\hat{\boldsymbol{a}} = \sum_{i=1}^{\nu_1+\nu_2+1} \lambda^i \boldsymbol{a}^i$, we conclude that $\hat{\boldsymbol{a}} \notin \text{ext}(\mathcal{C})$, which contradicts the hypothesis of the lemma. $\square$

To construct $\text{conv}(\mathcal{Z})$, as defined in Lemma 2, we propose a lift-and-project cutting plane procedure. A cutting plane is generated based on a single binary variable. Given the current continuous relaxation $\hat{Z}$, let $\hat{\boldsymbol{z}}$ be a fractional solution that needs to be cut off by a valid linear inequality. Generating a valid inequality accounts for finding a separating hyperplane that separates $\hat{\boldsymbol{z}}$ from $\pi_\iota(\hat{Z})$, where $\hat{z}_\iota \notin \{0,1\}$ for some $\iota \in [\kappa]$. To obtain a valid cut for $\operatorname{conv}(\mathcal{Z})$, one can solve a projection problem that minimizes the distance from $\hat{\boldsymbol{z}}$ to $\pi_\iota(\hat{Z})$.

Proposition 3. *Consider the notation defined in Lemma 2 and let $\hat{Z}$ be the current continuous relaxation. Consider an extreme point solution $\hat{\boldsymbol{z}} \in \hat{Z}$, where $\hat{z}_\iota \notin \{0,1\}$ for some $\iota \in [\kappa]$. Suppose that $\pi_\iota(\hat{Z})$ is nonempty. Then, the following projection problem*

$$\min_{\boldsymbol{z} \in \pi_\iota(\hat{Z})} \quad \|\boldsymbol{z} - \hat{\boldsymbol{z}}\| \tag{20}$$

has a strictly positive and finite optimal value, where $\|\cdot\|$ denotes a norm. Moreover, a solution to (20) is given by the $\boldsymbol{z}$ component of a solution to

$$\min_{(\boldsymbol{z}, \boldsymbol{b}_0, \boldsymbol{b}_1, \lambda_0, \lambda_1) \in \mathcal{M}_\iota(\hat{Z})} \quad \|\boldsymbol{z} - \hat{\boldsymbol{z}}\|. \tag{21}$$

Proof. Because $\pi_\iota(\hat{Z})$ is nonempty and compact, (20) has a solution and the optimal value is finite. Moreover, by Ruszczyński (2006, Theorem 2.14), $\hat{\boldsymbol{z}}$ can be strongly separated from $\pi_\iota(\hat{Z})$. By Rockafellar (1970, Theorem 11.4), the strong separation holds if and only if the optimal value to (20) is strictly positive. The second part of the results follows from the definition of projection. □

Lemma 4. *Stubbs and Mehrotra (1999, Theorem 3) Consider the notation defined in Proposition 3. Let $\hat{\boldsymbol{z}} \notin \pi_\iota(\hat{Z})$ and $\boldsymbol{z}^*$ be an optimal solution to (20). Then, there exists $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}^\top(\boldsymbol{z} - \boldsymbol{z}^*) \geq 0$ is a valid inequality in $\boldsymbol{z}$ which cuts off $\hat{\boldsymbol{z}}$. Moreover, $\boldsymbol{\alpha}$ is a subgradient of $\|\boldsymbol{z} - \hat{\boldsymbol{z}}\|$ at $\boldsymbol{z}^*$.*

Remark 2. *The equality constraints defining $\boldsymbol{z}$ in $\mathcal{M}_\iota(Z)$, defined in (19), have a nonlinear expression. Stubbs and Mehrotra (1999, Theorem 2) show that the constraints defining $\mathcal{M}_\iota(Z)$ can be represented by a set of convex constraints through a transformation of variables using perspective envelopes. Let us assume that the compact convex set Z is defined as follows*

$$Z := \left\{ \boldsymbol{z} \in \mathbb{R}^\vartheta \,\middle|\, \begin{array}{l} h_l(\boldsymbol{z}) \leq 0, \; l \in [\varsigma], \\ 0 \leq z_\iota \leq 1, \; \iota \in [\kappa] \end{array} \right\}.$$

For completeness, this reformulation is given as follows:

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

Algorithm 1: Lift-and-project cutting plane procedure to obtain $\operatorname{conv}(\mathcal{Z})$

Input: A convex continuous relaxation Z of a compact mixed-binary convex set $\mathcal{Z}$.
Output: $\operatorname{conv}(\mathcal{Z})$.

Set $t \leftarrow 0$ and $Z^0 = Z$.
while $Z^t \neq \emptyset$ **do**
  $Z^{t+1} = Z^t$.
  for ***each*** $\hat{\boldsymbol{z}} \in ext(Z^t)$ *and* $\iota \in [\kappa]$ *such that* $\hat{z}_\iota \notin \{0, 1\}$ **do**
    if $\left\{\boldsymbol{z} \in Z^t \,\middle|\, z_\iota = 0\right\} = \left\{\boldsymbol{z} \in Z^t \,\middle|\, z_\iota = 1\right\} = \emptyset$ **then**
      STOP.
    else
      Let $\boldsymbol{z}^*$ be an optimal solution to $\min_{\boldsymbol{z} \in \pi_\iota(Z^t)} \|\boldsymbol{z} - \hat{\boldsymbol{z}}\|$.
      Let $\boldsymbol{\alpha}$ be a subgradient of $\|\boldsymbol{z} - \hat{\boldsymbol{z}}\|$ at $\boldsymbol{z}^*$ with respect to $\boldsymbol{z}$.
      Let $Z^{t+1} := \left\{\boldsymbol{z} \in Z^{t+1} \,\middle|\, \boldsymbol{\alpha}^\top(\boldsymbol{z} - \boldsymbol{z}^*) \geq 0\right\}$.
    end
  end
  Set $t \leftarrow t + 1$.
end
STOP.

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

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

where

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

This reformulation allows to replace the projection problem (21) *with an equivalent convex optimization problem of the form*

$$\min_{(\boldsymbol{z}, \boldsymbol{b}_0, \boldsymbol{b}_1, \lambda_0, \lambda_1) \in \tilde{\mathcal{M}}_\iota(Z)} \|\boldsymbol{z} - \hat{\boldsymbol{z}}\|. \tag{22}$$

As an alternative to the approach in Stubbs and Mehrotra (1999) to handle the nonlinearity in $\mathcal{M}_\iota(Z)$, *Kılınç et al. (2017) propose to solve a sequence of cut-generation linear programs that relies on polyhedral outer approximations of the convex sets* Z^0_ι *and* Z^1_ι, *defined in Lemma 2. They show that, in the limit, the valid inequality obtained by this procedure is as strong as the valid inequality described in Lemma 4, obtained by solving the projection problem* (22). □

We are now ready to describe a cutting plane procedure to construct $\operatorname{conv}(\mathcal{Z})$. This pricedure repeatedly adds valid inequalities of the form described in Lemma 4 at all extreme point solutions of the current relaxation.

Theorem 2. *Let $\mathcal{Z}$ be defined as in (18). Algorithm 1 generates conv$(\mathcal{Z})$ in the limit, i.e.,* $\lim_{t\to\infty} Z^t = conv(\mathcal{Z})$, *where Z^t is the current relaxation at iteration t.*

To prove Theorem 2 we first state some technical results.

Lemma 5. Rahimian and Mehrotra (2020, *Lemma 4) Let $\{\mathcal{B}_1^t\}, \{\mathcal{B}_2^t\}, \ldots, \{\mathcal{B}_\varsigma^t\}$ be convergent, decreasing sequences of nested nonempty compact connected* [1] *sets of a finite-dimensional Euclidean space. If $\lim_{t\to\infty} \mathcal{B}_l^t = \bar{\mathcal{B}}_l$, where $\bar{\mathcal{B}}_l$ is nonempty, for $l \in [\varsigma]$, then*

$$\lim_{t\to\infty} conv\left(\cup_{l=1}^{\varsigma} \mathcal{B}_l^t\right) = conv\left(\cup_{l=1}^{\varsigma} \lim_{t\to\infty} \mathcal{B}_l^t\right) = conv\left(\cup_{l=1}^{\varsigma} \bar{\mathcal{B}}_l\right).$$

Lemma 6. *(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 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 $\lim_{t\to\infty} \mathcal{B}^t = \bar{\mathcal{B}}$. For each $\bar{\boldsymbol{b}} \in ext(\bar{\mathcal{B}})$, there exists some sequence $\{\boldsymbol{b}^t\}$ of points in $ext(\mathcal{B}^t)$ with a subsequence converging to $\bar{\boldsymbol{b}}$.*

Proof. of Theorem 2 Consider notation defined in the description of Algorithm 1. Suppose by contradiction that the procedure does not converge to conv$(\mathcal{Z})$. Because $Z^0 (= Z)$ is bounded and $Z^{t+1} \subseteq Z^t$ for all $t \geq 0$, then the sequence of closed sets $\{Z^t\}$ converges in the sense of Kuratowski to a set $\bar{Z}$ by Lemma 7. We examine the cases that $\bar{Z} = \emptyset$ and $\bar{Z} \neq \emptyset$.

Case 1. $\bar{Z} = \emptyset$. We conclude that $\mathcal{Z}$ is empty because $\mathcal{Z} \subseteq$ conv$(\mathcal{Z}) \subseteq \bar{Z}$. On the other hand, because $\{Z^t\}$ converges in the sense of Kuratowski to the empty set $\bar{Z}$, then there exists a finite $t \geq 0$ such that $Z^t = \emptyset$. Otherwise, $\bar{Z} = \cap_{t=1}^{\infty} Z^t \neq \emptyset$ by Lemma 6 and $\lim_{t\to\infty} Z^t = \bar{Z}$ by Lemma 8. Hence, the procedure terminates at line 15 of Algorithm 1 after detecting emptiness of $\mathcal{Z}$.

Case 2. $\bar{Z} \neq \emptyset$. We show that for all $\bar{\boldsymbol{z}} \in$ ext$(\bar{Z})$, we have $\bar{z}_\iota \in \{0, 1\}$ for all $\iota \in [\kappa]$. By contradiction, suppose that there exists $\bar{\boldsymbol{z}} \in$ ext$(\bar{Z})$ such that $\bar{z}_\iota \notin \{0, 1\}$ for some $\iota \in [\kappa]$. First, note that by Lemma 9, there exists a convergent subsequence $\{\boldsymbol{z}^t\}_{t\in\mathcal{T}}$ of points in ext(Z^t) such that $\{\boldsymbol{z}^t\}_{t\in\mathcal{T}} \to \bar{\boldsymbol{z}}$. Moreover, because $\bar{\boldsymbol{z}} \in$ ext(Z^t), by Proposition 3, there exists $\delta > 0$, where $\delta = \min_{\boldsymbol{z}\in\pi_\iota(\bar{Z})} \|\boldsymbol{z} - \bar{\boldsymbol{z}}\|$.

[1] A 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 $\mathcal{B} \cap \mathcal{V} \neq \emptyset$.

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

1. $\boldsymbol{z}^t \in ext(Z^t)$ *(Lemma 9),*
2. $z_\iota^t \notin \{0,1\}$ *(Lemma 9),*
3. $\|\boldsymbol{z}^t - \bar{\boldsymbol{z}}\| < \frac{\delta}{2}$ *(Lemma 9),*
4. $-\varepsilon + \min_{\boldsymbol{z}\in\pi_\iota(\bar{Z})} \|\boldsymbol{z} - \boldsymbol{z}^t\| > -\varepsilon + 3\frac{\delta}{4}$ *for* $\varepsilon > 0$*,*
5. $\min_{\boldsymbol{z}\in\pi_\iota(Z^t)} \|\boldsymbol{z} - \boldsymbol{z}^t\| > \frac{\delta}{2}$.

To see part (4) of the above claim, note that for any $\varepsilon > 0$, we have

$$\begin{aligned}-\varepsilon + \min_{\boldsymbol{z}\in\pi_\iota(\bar{Z})} \|\boldsymbol{z} - \boldsymbol{z}^t\| &\geq -\varepsilon - \|\boldsymbol{z}^t - \bar{\boldsymbol{z}}\| + \min_{\boldsymbol{z}\in\pi_\iota(\bar{Z})} \|\boldsymbol{z} - \bar{\boldsymbol{z}}\| \\ &\geq -\varepsilon - \|\boldsymbol{z}^t - \bar{\boldsymbol{z}}\| + \delta.\end{aligned}$$

Thus, there exists a finite $t \in \mathcal{T}$ such that part (4) holds. Now, to see part (5), first note that $\lim_{t\to\infty} \pi_\iota(Z^t) = \pi_\iota(\bar{Z})$ by using Lemma 5. Thus, for $\varepsilon = \frac{\delta}{4}$, there exists a finite $t \in \mathcal{T}$ such that $\min_{\boldsymbol{z}\in\pi_\iota(Z^t)} \|\boldsymbol{z} - \boldsymbol{z}^t\| \geq -\frac{\delta}{4} + \min_{\boldsymbol{z}\in\pi_\iota(\bar{Z})} \|\boldsymbol{z} - \boldsymbol{z}^t\| > \frac{\delta}{2}$.

Hence, in iteration t, the procedure generates a valid inequality (line 10 of Algorithm 1) that is violated by $\bar{\boldsymbol{z}}$) (this can be seen from parts 3 and 5 of the above claim). Thus, $\bar{\boldsymbol{z}} \notin Z^{t+1}$, contradicting $\bar{\boldsymbol{z}} \in \bar{Z} \subseteq Z^{t+1}$. So, the case that $\bar{z}_\iota \notin \{0,1\}$ for some $\iota \in [\kappa]$ will not happen. If $\mathcal{Z} \neq \emptyset$, then this contradiction implies that we must have $\bar{z}_\iota \in \{0,1\}$ for all $\iota \in [\kappa]$. On the other hand, if $\mathcal{Z} = \emptyset$, then this contradiction implies that the case $\bar{Z} \neq \emptyset$ is not relevant. □

Remark 3. *By the definition of convergence in Hausdorff distance, Theorem 2 implies that for any* $\eta > 0$*, there exists* $\hat{t} \geq 0$ *such that for all* $t \geq \hat{t}$*, we have* $\inf_{\boldsymbol{z}\in conv(\mathcal{Z})} \|\boldsymbol{z} - \boldsymbol{z}^t\| \leq \eta$ *for all* $\boldsymbol{z}^t \in Z^t$*, where* Z^t *is the current relaxation at iteration* t *of Algorithm 1. In other words, for any precision* η*, there exists a finite* $\hat{t}$ *such that* $Z^{\hat{t}}$ *is sufficiently close to* $conv(\mathcal{Z})$*.* □

Now, let us translate back to $\mathcal{Z}_a^q(C,\boldsymbol{\beta})$, defined in (17), and describe how to construct $\mathcal{P}_a^q(C,\boldsymbol{\beta})$, defined in (12).

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

$$\mathcal{M}_k\big(Z_a^q(C,\boldsymbol{\beta})\big) := \left\{ (\boldsymbol{x},\boldsymbol{y},\boldsymbol{W},\boldsymbol{z},\boldsymbol{b}_0,\boldsymbol{b}_1,\lambda_0,\lambda_1) \,\middle|\, \begin{array}{l} (\boldsymbol{x},\boldsymbol{y},\boldsymbol{W},\boldsymbol{z}) = \lambda_0\boldsymbol{b}_0 + \lambda_1\boldsymbol{b}_1, \\ \lambda_0 + \lambda_1 = 1,\ \lambda_0 \geq 0,\ \lambda_1 \geq 0, \\ \boldsymbol{b}_0 \in Z_{a,k}^{q,0}(C,\boldsymbol{\beta}),\ \boldsymbol{b}_1 \in Z_{a,k}^{q,1}(C,\boldsymbol{\beta}) \end{array} \right\}.$$

Let $\pi_k\big(Z_a^q(C,\boldsymbol{\beta})\big)$ *be the projection of* $\mathcal{M}_k\big(Z_a^q(C,\boldsymbol{\beta})\big)$ *onto the* $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W},\boldsymbol{z})$*-space. Then,*

$$\pi_k\big(Z_a^q(C,\boldsymbol{\beta})\big) = conv\left(Z_{a,k}^{q,0}(C,\boldsymbol{\beta}) \cup Z_{a,k}^{q,1}(C,\boldsymbol{\beta})\right) = conv\left(Z_a^q(C,\boldsymbol{\beta}) \cap z_k \in \{0,1\}\right).$$

Moreover,

$$\mathit{conv}\left(\mathcal{Z}_a^q(C,\boldsymbol{\beta})\right) = \pi_{q^2}\Big(\pi_{q^2-1}\Big(\dots\Big(\pi_1\big(Z_a^q(C,\boldsymbol{\beta})\big)\Big)\dots\Big)\Big)$$

and $\mathcal{P}_a^q(C,\boldsymbol{\beta}) = \mathit{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\Big(\mathit{conv}\left(\mathcal{Z}_a^q(C,\boldsymbol{\beta})\right)\Big)$, *where* $\mathcal{P}_a^q(C,\boldsymbol{\beta})$ *is defined in* (12). *That is,* $\mathit{conv}\Big(\mathit{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\big(\mathcal{Z}_a^q(C,\boldsymbol{\beta})\big)\Big) = \mathit{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\Big(\mathit{conv}\left(\mathcal{Z}_a^q(C,\boldsymbol{\beta})\right)\Big)$.

Proof. One can derive $\pi_k\big(Z_a^q(C,\boldsymbol{\beta})\big)$ and $\operatorname{conv}\left(\mathcal{Z}_a^q(C,\boldsymbol{\beta})\right)$ by a direct application of Lemma 2. To complete the proof, we show that $\mathcal{P}_a^q(C,\boldsymbol{\beta})$ is obtained by projecting $\operatorname{conv}\left(\mathcal{Z}_a^q(C,\boldsymbol{\beta})\right)$ onto the $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})$-space. Let us define $\mathcal{A} := \mathcal{P}_a^q(C,\boldsymbol{\beta}) = \operatorname{conv}\Big(\operatorname{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\big(\mathcal{Z}_a^q(C,\boldsymbol{\beta})\big)\Big)$ and $\mathcal{B} := \operatorname{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\Big(\operatorname{conv}\left(\mathcal{Z}_a^q(C,\boldsymbol{\beta})\right)\Big)$. 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. We need to show $\operatorname{ext}(\mathcal{A}) = \operatorname{ext}(\mathcal{B})$.

"$\Longrightarrow$": Let $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}}) \in \operatorname{ext}(\mathcal{A})$. Then, 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 $(\hat{\boldsymbol{x}}^i,\hat{\boldsymbol{y}}^i,\hat{\boldsymbol{W}}^i) \in \operatorname{ext}\Big(\operatorname{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\big(\mathcal{Z}_a^q(C,\boldsymbol{\beta})\big)\Big)$, $\lambda_i \geq 0$, and $\sum_i \lambda_i = 1$. Using Lemma 3, there exists $\hat{\boldsymbol{z}}^i$ such that $(\hat{\boldsymbol{x}}^i,\hat{\boldsymbol{y}}^i,\hat{\boldsymbol{W}}^i,\hat{\boldsymbol{z}}^i) \in \operatorname{ext}\Big(\operatorname{conv}\left(\mathcal{Z}_a^q(C,\boldsymbol{\beta})\right)\Big)$. Thus, we can construct a point $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}},\hat{\boldsymbol{z}}) \in \operatorname{conv}\left(\mathcal{Z}_a^q(C,\boldsymbol{\beta})\right)$ as $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}},\hat{\boldsymbol{z}}) = \sum_i \lambda_i(\hat{\boldsymbol{x}}^i,\hat{\boldsymbol{y}}^i,\hat{\boldsymbol{W}}^i,\hat{\boldsymbol{z}}^i)$. Consequently, $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}}) \in \operatorname{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\Big(\operatorname{conv}\left(\mathcal{Z}_a^q(C,\boldsymbol{\beta})\right)\Big) = \mathcal{B}$.

"$\Longleftarrow$": Let $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}}) \in \operatorname{ext}(\mathcal{B})$. Then, by Lemma 3, there exists $\hat{\boldsymbol{z}}$ such that $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}},\hat{\boldsymbol{z}}) \in \operatorname{ext}\Big(\operatorname{conv}\left(\mathcal{Z}_a^q(C,\boldsymbol{\beta})\right)\Big)$. Thus, $\hat{z} \in \{0,1\}$, $k \in [q^2]$ and $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}},\hat{\boldsymbol{z}}) \in \mathcal{Z}_a^q(C,\boldsymbol{\beta})$. Because $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}},\hat{\boldsymbol{z}}) \in \mathcal{Z}_a^q(C,\boldsymbol{\beta})$, we have $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}}) \in \operatorname{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\Big(\mathcal{Z}_a^q(C,\boldsymbol{\beta})\Big)$. Consequently, $(\hat{\boldsymbol{x}},\hat{\boldsymbol{y}},\hat{\boldsymbol{W}}) \in \operatorname{conv}\Big(\operatorname{Proj}_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})}\big(\mathcal{Z}_a^q(C,\boldsymbol{\beta})\big)\Big) = \mathcal{A}$. □

Remark 4. *Proposition 4 implies that by applying Algorithm 1 to the continuous relaxation* $Z_a^q(C,\boldsymbol{\beta})$ *of the compact mixed-binary convex set* $\mathcal{Z}_a^q(C,\boldsymbol{\beta})$*, defined respectively in* (16) *and* (17)*, one can construct a set arbitrary close to* $\mathcal{P}_a^q(C,\boldsymbol{\beta})$ *through lift-and-project cuts.* □

4 A Sequential ∞-Convergent Convexification Procedure

In Section 3, we described a procedure for constructing a set arbitrary close to the convex hull of ϵ-feasible solutions with respect to a single vector $\boldsymbol{c}_a$, $a \in [nm]$ (recall Theorem 1 and Remark 4). In this section, we describe a sequential procedure (over all vectors $\boldsymbol{c}_a$, $a \in [nm]$) for generating a set arbitrary close to $\operatorname{conv}\left(\mathcal{F}^\epsilon\right)$. This procedure is based on the single-vector convexification, presented in Theorem 1, and is infinitely-convergent.

To state the main result in this section, let us construct

$$K^t = \mathcal{P}_{nm}^q\left(\mathcal{P}_{nm-1}^q\left(\cdots\left(\mathcal{P}_1^q(K^{t-1})\right)\dots\right)\right), \tag{23}$$

for $t = 1, 2, \dots$, with $K^0 = \mathcal{K}$. Note that all the constructions and notation in this section that are related to the breakpoints, including K^t, are based on the predetermined breakpoints $\boldsymbol{\theta}$, explained in Construction 1, but we omit $\boldsymbol{\theta}$ from notation for brevity. Observe that in (23), K^{t-1} is sequentially convexified over all vectors $\boldsymbol{c}_a$, $a \in [nm]$. We are now ready to state our infinity-convergent sequential convexification procedure.

Theorem 3. *Let K^t, $t = 1, 2, \ldots$, be defined as in (23). For any $\epsilon > 0$, there exists $\tilde{q} \in \mathbb{N}$ such that for all $q \geq \tilde{q}$, we have*

$$\mathit{conv}\,(\mathcal{F}) \subseteq \lim_{t \to \infty} K^t \subseteq \mathit{conv}\,(\mathcal{F}^{\epsilon})\,.$$

Theorem 3 describes a sequential procedure for generating a relaxed set for conv $(\mathcal{F})$, while this set is not too conservative and is contained in conv $(\mathcal{F}^{\epsilon})$ in the limit. One implication of this theorem is that it helps detect infeasibility of conv $(\mathcal{F})$. That is, if K^t is infeasible, then conv $(\mathcal{F})$ is infeasible as well. On the other hand, the constraints from the description of K^t can be used as constraints for describing conv $(\mathcal{F})$.

Proof. of Theorem 3 In order to prove the theorem, let us simultaneously split the range of functions $\frac{\boldsymbol{u}_a^\top \boldsymbol{x} + \boldsymbol{v}_a^\top \boldsymbol{y}}{2}$ and $\frac{\boldsymbol{u}_a^\top \boldsymbol{x} - \boldsymbol{v}_a^\top \boldsymbol{y}}{2}$ for all vectors $\boldsymbol{c}_a$, $a \in [nm]$. Any feasible solution $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W})$ to (1) satisfies the following disjunction:

$$\vee_{k=1}^{q^{2^{mn}}} \mathcal{H}_k(\mathcal{K}), \tag{24}$$

where

$$\mathcal{H}_k(\mathcal{K}) := \left\{ (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \mathcal{K} \;\middle|\; \begin{array}{l} \theta_{1,k}^a \leq \frac{\boldsymbol{u}_a^\top \boldsymbol{x} + \boldsymbol{v}_a^\top \boldsymbol{y}}{2} \leq \theta_{1,k+1}^a,\ \theta_{2,k}^a \leq \frac{\boldsymbol{u}_a^\top \boldsymbol{x} - \boldsymbol{v}_a^\top \boldsymbol{y}}{2} \leq \theta_{2,k+1}^a, \\ \boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a - \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x} + \boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)(\theta_{1,k}^a + \theta_{1,k+1}^a) + \theta_{1,k}^a \theta_{1,k+1}^a \\ + \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x} - \boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)^2 \leq 0, \\ -\boldsymbol{u}_a^\top \boldsymbol{W} \boldsymbol{v}_a - \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x} - \boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)(\theta_{2,k}^a + \theta_{2,k+1}^a) + \theta_{2,k}^a \theta_{2,k+1}^a \\ + \left(\frac{\boldsymbol{u}_a^\top \boldsymbol{x} + \boldsymbol{v}_a^\top \boldsymbol{y}}{2}\right)^2 \leq 0,\ a \in [nm] \end{array} \right\}. \tag{25}$$

for $k \in [q^{2^{mn}}]$ and $\boldsymbol{\theta}_a$ denotes the predetermined breakpoints corresponding to $\boldsymbol{c}_a$, $a \in [nm]$. Note the difference between $\mathcal{S}_k(\boldsymbol{c}_a, \mathcal{K})$, as defined in (8) for only vector $\boldsymbol{c}_a$, and $\mathcal{H}_k(\mathcal{K})$, as defined in (25) for all vectors $\boldsymbol{c}_a$, $a \in [nm]$.

Let us define $C^q(\mathcal{K}) := \text{conv}\left(\bigcup_{k=1}^{q^{2^{mn}}} \mathcal{H}_k(\mathcal{K})\right)$. By a similar argument as in the proof of Theorem 1, we can show that conv $(\mathcal{F}) \subseteq C^q(\mathcal{K})$ for any $q \in \mathbb{N}$. Similarly, we can show that there exists $\tilde{q}_1 \in \mathbb{N}$ such that for all $q \geq \tilde{q}_1$, we have $C^q(\mathcal{K}) \subseteq$ conv $(\mathcal{F}^{\epsilon})$. On the other hand, by construction, $\{K^t\}$ is a decreasing sequence of nested compact convex sets such that $C^q(\mathcal{K}) \subseteq K^{t+1} \subseteq K^t$ for all $t \geq 0$. Thus, by Lemma 7, the limit of $\{K^t\}$ exists. Let us denote the limit by $\bar{K} \supseteq C^q(\mathcal{K})$. If $\bar{K} = \emptyset$, then by Lemma 6, there exists a finite $t \geq 0$ such that $K^t = \emptyset$. Hence, the theorem holds trivially. Otherwise, if $\bar{K} \neq \emptyset$, then $\{K^t\}$ is a sequence of nonempty sets. Then, by Lemma 7, the Kuratowski limit of $\{K^t\}$ exists, and by Lemma 8, it is equal to the Hausdorff limit, $\lim_{t \to \infty} K^t$. By Salinetti and Wets (1979), $\bar{K}$ is a convex set. We are only left with showing $\bar{K} \subseteq$ conv $(\mathcal{F}^{\epsilon})$ to complete the proof. By construction, we have

$$K^{t+1} \subseteq \mathcal{P}_a^q(K^t) = \operatorname{conv}\left(\bigcup_{k=1}^{q^2} \mathcal{S}_k(\boldsymbol{c}_a, K^t)\right)$$

for all values of a, $a \in [nm]$, and $q \in \mathbb{N}$. By taking the limit in both sides, we have

$$\bar{K} = \lim_{t\to\infty} K^{t+1} \subseteq \lim_{t\to\infty} \operatorname{conv}\left(\bigcup_{k=1}^{q^2} \mathcal{S}_k(\boldsymbol{c}_a, K^t)\right) = \operatorname{conv}\left(\bigcup_{k=1}^{q^2} \mathcal{S}_k(\boldsymbol{c}_a, \bar{K})\right),$$

where the equality follows from an application of Lemma 5, because $\lim_{t\to\infty} \mathcal{S}_k(\boldsymbol{c}_a, K^t) = \mathcal{S}_k(\boldsymbol{c}_a, \bar{K})$ by Lemma 8, $k \in [q^2]$. Moreover, by Theorem 1, there exists $\tilde{q}_2 \in \mathbb{N}$ such that for all $q \geq \tilde{q}_2$, we have $\operatorname{conv}\left(\bigcup_{k=1}^{q^2} \mathcal{S}_k(\boldsymbol{c}_a, \bar{K})\right) \subseteq \mathcal{C}_a^\epsilon(\bar{K})$. We now claim that the inclusion $\bar{K} \subseteq \operatorname{conv}\left(\bigcup_{k=1}^{q^2} \mathcal{S}_k(\boldsymbol{c}_a, \bar{K})\right)$ for $q \geq \tilde{q} := \max\{\tilde{q}_1, \tilde{q}_2\}$ implies that $\bar{K} \subseteq \operatorname{conv}(\mathcal{F}^\epsilon)$. Suppose by contradiction that there exists $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \in \bar{K}$ and $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \notin \operatorname{conv}(\mathcal{F}^\epsilon)$. Hence, there exists $\boldsymbol{c}_{\hat{a}}$, $\hat{a} \in [nm]$, such that $|\boldsymbol{u}_{\hat{a}}^\top \boldsymbol{W} \boldsymbol{v}_{\hat{a}} - (\boldsymbol{u}_{\hat{a}}^\top \boldsymbol{x})(\boldsymbol{v}_{\hat{a}}^\top \boldsymbol{y})| > \epsilon$. Thus, $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W})$ can be separated from $\mathcal{C}_a^\epsilon(\bar{K})$. This, in turns, implies that for $q \geq \tilde{q}$, $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W})$ can be separated from $\operatorname{conv}\left(\bigcup_{k=1}^{q^2} \mathcal{S}_k(\boldsymbol{c}_a, \bar{K})\right)$, i.e., $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W}) \notin \operatorname{conv}\left(\bigcup_{k=1}^{q^2} \mathcal{S}_k(\boldsymbol{c}_a, \bar{K})\right)$. This is a contradiction with the above inclusion $\bar{K} \subseteq \operatorname{conv}\left(\bigcup_{k=1}^{q^2} \mathcal{S}_k(\boldsymbol{c}_a, \bar{K})\right)$. Thus, $\bar{K} \subseteq \operatorname{conv}(\mathcal{F}^\epsilon)$ and this completes the proof for the case that $\{K^t\}$ is a nonempty sequence of sets. □

Corollary 1. *Let K^t, $t = 1, 2, \ldots$, be defined as in (23). For $\epsilon > 0$ and $\eta > 0$, there exists $\tilde{q} \in \mathbb{N}$ and $\tilde{t} \in \mathbb{N}$ such that for all $q \geq \tilde{q}$ and $t \geq \tilde{t}$, we have*

$$\mathit{conv}(\mathcal{F}) \subseteq K^t \subseteq \mathcal{N}_\eta\big(\mathit{conv}(\mathcal{F}^\epsilon)\big).$$

Proof. By Theorem 3, we know that for a sufficiently large q, we have $\operatorname{conv}(\mathcal{F}) \subseteq K^t$ for $t \geq 0$. Let $\bar{K} := \lim_{t\to\infty} K^t$. Thus, by definition, for a sufficiently large t, we have $K^t \subseteq \mathcal{N}_\eta(\bar{K})$. Because $\bar{K} \subseteq \operatorname{conv}(\mathcal{F}^\epsilon)$ by Theorem 3, we conclude that for a sufficiently large t and q, we have $K^t \subseteq \mathcal{N}_\eta\big(\operatorname{conv}(\mathcal{F}^\epsilon)\big)$. This completes the proof. □

Remark 5. *To construct K^t, one can sequentially use Algorithm 1 to derive $\mathcal{P}_a(\cdot)$, for all $a \in [nm]$, in the limit. Depending on the precision $\eta > 0$ in Corollary 1, we can control the error to construct $\mathcal{P}_a(\cdot)$. Hence, we can stop Algorithm 1 at the desired level of precision (recall Remark 3).* □

Remark 6. *As mentioned before, because $\mathit{conv}(\mathcal{F}^\epsilon) \subseteq \mathit{conv}\big(\bar{\mathcal{F}}^\epsilon\big)$, both Theorem 3 and Corollary 1 can be stated with $\mathit{conv}\big(\bar{\mathcal{F}}^\epsilon\big)$.* □

The infinite-convergence of the sequential convexification procedure in this section is in contrast to the finite sequential convexification procedure described in Saxena et al. (2010) for a quadratically constrained set. Recall Proposition 1. Observe that a quadratically constrained set in $\mathbb{R}^n$ can be equivalently written with n nonlinear constraints. Given an orthonormal set of vectors for

$\mathbb{R}^n$, Saxena et al. (2010, Theorem 2) show that sequential convexification procedure is finite and constructs the convex hull in n steps. However, as we stated in Theorem 3, a sequential convexification procedure for the bilinear set (1), over all vectors $\boldsymbol{c}_a$, $a \in [nm]$, constructs a set arbitrary close the convex hull of $\mathcal{F}^\epsilon$ in the limit.

This difference may be reminiscent of the difference between the convexification of a mixed-binary set and a mixed-integer set. A mixed-binary set (with n binary variables) can be sequentially convexified in n steps (Sherali and Adams, 1990; Stubbs and Mehrotra, 1999). However, for a mixed-integer set, a sequential convexification procedure, over all integer variables, constructs the convex hull in the limit (Owen and Mehrotra, 2001, Theorem 3). For a mixed-integer set, of course, one may replace integer variables with their binary expansions (Owen and Mehrotra, 2002). Hence, it is possible to sequentially convexify the resulting mixed-binary set in a lifted space and in a finite number of steps (Stubbs and Mehrotra, 1999). Instead, Owen and Mehrotra (2001) propose an infinitely-convergent sequential convexification procedure that neither introduce any new binary variables in the problem nor maintain a disjunctive convex hull tree (Chen et al., 2011).

In a similar spirit, for a bilinear set of the form (2), one may derive $\operatorname{conv}(\mathcal{F})$ with a finite sequential convexification of exponentially many (in fact, $q^{2^{mn}}$) binary variables in a lifted space (this goal can be reached by generalizing Proposition 4). Instead, in Theorem 3, we described a sequential convexification procedure in the space of $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{W})$ that derives a set arbitrary close to $\operatorname{conv}(\mathcal{F}^\epsilon)$ in the limit. Nevertheless, we used the construction underlying that finite sequential convexification procedure (i.e., $\vee_{k=1}^{q^{2^{mn}}} \mathcal{H}_k(\mathcal{K})$, defined in (24)) as a tool to prove the validity of our proposed ∞-convergent sequential convexification procedure.

We conclude this section by establishing the relationship between K^t and the following sets:

$$Q^t = \mathcal{C}_{nm}\left(\mathcal{C}_{nm-1}\left(\cdots\left(\mathcal{C}_1(Q^{t-1})\right)\ldots\right)\right), \tag{26}$$

for $t = 1, 2, \ldots$, with $Q^0 = \mathcal{K}$, and

$$E^t = \mathcal{C}^\epsilon_{nm}\left(\mathcal{C}^\epsilon_{nm-1}\left(\cdots\left(\mathcal{C}^\epsilon_1(E^{t-1})\right)\ldots\right)\right), \tag{27}$$

for $t = 1, 2, \ldots$, with $E^0 = \mathcal{K}$, where the operators $\mathcal{C}_a(\cdot)$ and $\mathcal{C}^\epsilon_a(\cdot)$, $a \in [mn]$, are defined as in (14) and (15), respectively.

Theorem 4. *For $\epsilon > 0$, let K^t, Q^t, and E^t, $t = 1, 2, \ldots$, be defined as in (23), (26), and (27), respectively. There exists $\tilde{q} \in \mathbb{N}$ such that for all $q \geq \tilde{q}$, we have*

$$\mathit{conv}(\mathcal{F}) \subseteq Q^t \subseteq K^t \subseteq E^t$$

for any $t \geq 0$.

Proof. Consider a relaxation C of $\mathcal{F}$. Note that for any $a \in [nm]$, we have $\mathcal{C}_a(C) \subseteq \mathcal{P}^q_a(C)$ for all values $q \in \mathbb{N}$. Thus, by the sequential construction of Q^t and K^t, we have $Q^t \subseteq K^t$ for any $t \geq 0$. On the other hand, by construction, $\{Q^t\}$ is a sequence of nested compact convex sets such that conv $(\mathcal{F}) \subseteq Q^t$ for all $t \geq 0$. Now, we show that $K^t \subseteq E^t$ for a sufficiently large q and any $t \geq 0$. By Theorem 1, we have $\mathcal{P}_a^q(C) \subseteq \mathcal{C}_a^\epsilon(C)$ for a sufficiently large q. Thus, by a sequential application of Theorem 1, we conclude that there exists $\tilde{q} \in \mathbb{N}$ such that for all $q \geq \tilde{q}$, we have $K^t \subseteq E^t$. □

5 Discussions and Future Research

In this paper, we studied a general nonconvex bilinear program with continuous variables. We presented a sequential convexification procedure to construct a set arbitrary close to the convex hull of ϵ-feasible solutions in the limit. We showed that the single-vector convexification based on a pair of basis for the space of $(\boldsymbol{x}, \boldsymbol{y})$ can be extended to a sequential procedure over all pairs of basis vectors. One may obtain the bases through the singular value decomposition of the residual bilinear matrix, proposed in Saxena et al. (2010) and Fampa and Lee (2021).

Constraints from the description of conv $(\mathcal{F})$ can be utilized as cutting planes in an algorithmic fashion. Although these constraints are not immediately available, one can consider using the constraints obtained from the single-vector convexification, for a vector $\boldsymbol{c}_a$, a relaxation C, a fixed $q \in \mathbb{N}$, and a proper choice of breakpoints $\boldsymbol{\beta}$. In our companion paper (Rahimian and Mehrotra, 2020), we show that to find an ϵ-optimal solution to $\min_{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{W})\in\mathcal{F}} \ \boldsymbol{f}_0^\top \boldsymbol{x} + \boldsymbol{g}_0^\top \boldsymbol{y} + \boldsymbol{A}_0 \bullet \boldsymbol{W}$ through cutting planes, a (2×2)-way disjunction with a proper choice of breakpoints $\boldsymbol{\beta}$ suffices. Such a cutting plane algorithm generates cuts at all near-optimal extreme point solutions of the current relaxation. We refer the reader to Rahimian and Mehrotra (2020) for a more detailed discussion.

Acknowledgements

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

References

Adams, W. P., R. J. Forrester, and F. W. Glover (2004). Comparisons and enhancement strategies for linearizing mixed 0-1 quadratic programs. *Discrete Optim. 1*(2), 99–120.

Adams, W. P. and H. D. Sherali (1990). Linearization strategies for a class of zero-one mixed integer programming problems. *Oper. Res. 38*(2), 217–226.

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.

Anstreicher, K. M. (2009). Semidefinite programming versus the reformulation-linearization technique for nonconvex quadratically constrained quadratic programming. *J. Global Optim. 43*(2-3), 471–484.

Anstreicher, K. M. (2012). On convex relaxations for quadratically constrained quadratic programming. *Math. Program. 136*(2), 233–251.

Balas, E. (1985). Disjunctive programming and a hierarchy of relaxations for discrete optimization problems. *SIAM J. Algebr. Discrete Meth. 6*(3), 466–486.

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.

Burer, S. and A. Saxena (2012). The milp road to miqcp. In J. Lee and S. Leyffer (Eds.), *Mixed integer nonlinear programming*, pp. 373–405. Springer.

Burer, S. and B. Yang (2015). The trust region subproblem with non-intersecting linear constraints. *Math. Program. 149*(1-2), 253–264.

Chen, B., S. Küçükyavuz, and S. Sen (2011). Finite disjunctive programming characterizations for general mixed-integer linear programs. *Oper. Res. 59*(1), 202–210.

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 (2021). Convexification of bilinear forms through non-symmetric lifting. *Journal of Global Optimization*.

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.

Jiang, R. and D. Li (2019). Second order cone constrained convex relaxations for nonconvex quadratically constrained quadratic programming. *J. Global Optim. 75*(2), 461–494.

Kılınç, M. R., J. Linderoth, and J. Luedtke (2017). Lift-and-project cuts for convex mixed integer nonlinear programs. *Math. Prog. Comput. 9*(4), 499–526.

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.

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.

Owen, J. H. and S. Mehrotra (2002). On the value of binary expansions for general mixed-integer linear programs. *Oper. Res. 50*(5), 810–819.

Rahimian, H. and S. Mehrotra (2020). A finitely convergent disjunctive cutting plane algorithm for bilinear programming. Optimization Online http://www.optimization-online.org/DB_HTML/2020/01/7594.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 (1990). A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM J. Discrete Math. 3*(3), 411–430.

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.

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

Tawarmalani, M., J.-P. P. Richard, and K. Chung (2010). Strong valid inequalities for orthogonal disjunctions and bilinear covering sets. *Math. Program. 124*(1-2), 481–512.

Tawarmalani, M. and N. Sahinidis (2002). *Convexification and global optimization in continuous and mixed-integer nonlinear programming: theory, algorithms, software, and applications*, Volume 65. Springer Science & Business Media.