Convexification of Permutation-Invariant Sets and an Application to Sparse PCA

A Preprint

Jinhak Kim  
College of Business  
Northern Illinois University  
jkim23@niu.edu

Mohit Tawarmalani  
Krannert School of Management  
Purdue University  
mtawarma@purdue.edu

Jean-Philippe P. Richard  
Industrial and Systems Engineering  
University of Minnesota  
jrichar@umn.edu

August 10, 2021

ABSTRACT

We develop techniques to convexify a set that is invariant under permutation and/or change of sign of variables and discuss applications of these results. First, we convexify the intersection of the unit ball of a permutation and sign-invariant norm with a cardinality constraint. This gives a nonlinear formulation for the feasible set of sparse principal component analysis (sparse PCA) and an alternative proof of the $K$-support norm. Second, we characterize the convex hull of sets of matrices defined by constraining their singular values. As a consequence, we generalize an earlier result that characterizes the convex hull of rank-constrained matrices whose spectral norm is below a given threshold. Third, we derive convex and concave envelopes of various permutation-invariant nonlinear functions and their level-sets over hypercubes, with congruent bounds on all variables. Finally, we develop new relaxations for the exterior product of sparse vectors. Using these relaxations for sparse PCA, we show that our relaxation closes 98% of the gap left by a classical SDP relaxation for instances where the covariance matrices are of dimension up to $50 \times 50$.

Keywords  Convexification, permutation-invariant sets, majorization, sparse PCA

1 Introduction

Convexification, permutation-invariant sets, and an application to sparse PCA

This paper is concerned with convexification of permutation-invariant sets. A set $S \subseteq \mathbb{R}^n$ is permutation-invariant if $x \in S$ implies that $Px \in S$ for all $n$-dimensional permutation matrices $P$. Permutation-invariant sets appear in a variety of optimization problems. To support this claim and highlight the relevance of our construction, we next provide a variety of example applications where exploiting permutation invariance proves fruitful. Consider first sparse principal component analysis, a problem first introduced in [34], that consists in finding sparse vectors that explain the most variance in a data set. Specifically, the problem of finding the first sparse principal component can be formulated as $\max \{x^\top \Sigma x \mid x \in S\}$ where $S = \{x \in \mathbb{R}^n \mid \text{card}(x) \leq K, \|x\| \leq 1\}$, card$(x)$ is the number of nonzero components of $x$, and $\Sigma$ is the covariance matrix of the given data; see, for example, [8]. The feasible set of this model is permutation-invariant because card$(Px) = \text{card}(x)$ and $\|Px\| = \|x\|$ for any vector $x \in \mathbb{R}^n$ and permutation matrix $P$. The convex hull of $S$ is the unit ball associated with the $K$-support norm [3], a result that yields a tighter approximation of $S$ compared to the elastic net $\{x \in \mathbb{R}^n \mid \|x\|_1 \leq \sqrt{K}, \|x\| \leq 1\}$ where $\| \cdot \|_1$ is the $l_1$-norm. Recognizing the set as permutation-invariant allows for a streamlined derivation of these results. Second, observe that permutation-invariance also arises when studying sets of matrices as the rank of a matrix can be thought of as a permutation-invariant cardinality constraint on the singular values. This equivalent formulation suggests that the applicability of this concept extends to sets of matrices whose singular values belong to a permutation-invariant set. For example, [14], considers the set $\{M \in \mathcal{M}^{m \times n}(\mathbb{R}) \mid \text{rank}(M) \leq K, \|M\|_{sp} \leq r\}$ where $\mathcal{M}^{m \times n}(\mathbb{R})$ is the set of $m$-by-$n$ real matrices and $\|M\|_{sp}$ is the spectral norm of $M$. Since the spectral norm is the largest singular value, it follows easily that the singular values of these matrices belong to a permutation-invariant set. Using a permutation-invariant perspective to study such sets helps to generalize the results of [14] in various ways. Finally, observe that a variety of sets with specific structures have been studied in nonconvex optimization because of their uses in creating
convex relaxations of more general problems. For example, specialized techniques have been used to identify the convex hull of the multilinear monomial $\prod_{i=1}^n x_i$ over $[0,1]^n$ and $[-1,1]^n$ [30, 24]. We will show that these results can be obtained as special cases of our general approach to convexifying permutation-invariant sets. Besides, we will show that we can obtain hitherto unknown polynomial-sized convex hull descriptions for various commonly occurring sets in global optimization.

We now describe why permutation-invariance is a useful property to consider while constructing convex hulls of sets. To see this, let $\Delta_\pi := \{ x \mid x_{\pi(1)} \geq \cdots \geq x_{\pi(n)} \}$, where $\pi$ is an $n$-dimensional permutation and denote by $\text{conv}(S)$, the convex hull of a given set $S$. Since $S$ can be expressed as a union of $n!$ sets of the form $S \cap \Delta_\pi$, if there is a (possibly lifted) polynomial-sized representation of $\text{conv}(S \cap \Delta_\pi)$ for each $\pi$, disjunctive programming [4] can be used to represent $\text{conv}(S)$ in a higher-dimensional space. Unfortunately, this representation is exponentially-sized in $n$, because the construction creates copies, one for each $\pi$, for each of the variables $(x_1, \ldots, x_n)$. What is remarkable about the permutation-invariance is that, by exploiting this property, we will show that a much more compact formulation can be derived. To appreciate the significance of this construction, we first argue that, if $S$ is not permutation-invariant, such a compact formulation for $\text{conv}(S)$ may not even exist. To illustrate this point, consider the set, $B$, which is a face of the boolean-quadratic polytope, and is defined as $B := \{ xx^T \mid x \in \{0,1\}^n, x_1 = 1 \}$. If we order all the $\binom{n(n+1)}{2}$ products, these imply an order for all $x_j$ variables since $x_1 x_j = x_j$ for $j = 2, \ldots, n$.

Moreover, if $x_j \leq x_i$, it follows that $x_j x_i = x_j$. Therefore, the ordering of the $\binom{n(n+1)}{2}$ bilinear terms reduces to that of all $x_j$ variables, possibly with some equalities. In other words, it suffices to consider $B \cap \Delta_\pi$, where $\Delta_\pi = \{ xx^T \mid x_{\pi(r)} x_{\pi(i)} \geq x_{\pi(r)} x_{\pi(j)} \}$ whenever $\max(r, i) \leq \max(k, j)$ is a lifting of $\Delta_\pi$ and $\pi(1) = 1$. Since $x \in \{0, 1\}^n \cap \Delta_\pi$ implies that $x_{\pi(r)} x_{\pi(i)} = x_{\pi(r)} x_{\pi(i)}$, we can write $\text{conv}(B \cap \Delta_\pi) = \{ X \in \{0, 1\}^{n \times n} \mid X_{ij} = X_{1, \max(i,j)}, X_{11} = 1 \}$. However, it is known that $\text{conv}(B)$ does not have a polynomial-sized formulation [9]. In contrast, treating permutation-invariant sets $S$ in this way provides a significant advantage since the sets $S \cap \Delta_\pi$ are congruent to one another. Exploiting this fact, we show that it is possible to construct a polynomial-sized extended formulation for $S$ whenever a polynomial-sized formulation exists for $\text{conv}(S \cap \Delta_\pi)$. Our construction makes use of well-known extended formulations of a permutahedron along-side the convex hull of $S \cap \Delta_\pi$. The outline of the construction is as follows: first, we consider a permutation-invariant set $S$ and assume that its convex hull over $\Delta^n := \{ x \in \mathbb{R}^n \mid x_1 \geq \cdots \geq x_n \}$, has a polynomial description. Then, the convex hull is simply the union of permutahedral where each permutahedron is generated by a point in $\text{conv}(S \cap \Delta^n)$. Each permutahedron can then be modeled using a polynomial number of linear equalities and inequalities to provide an extended formulation for $\text{conv}(S)$. The techniques involved apply to other settings. For example, they can be used to obtain convex hulls of sign-invariant sets using convex hull representations of $S \cap \{ x \mid x \geq 0 \}$.

The remainder of the paper is organized as follows. We present basic convexification results for permutation- and/or sign-invariant sets in Section 2. We then explore various applications of the results in the ensuing sections. In Section 3, we derive the convex hull of the intersection of a unit ball associated with a permutation-invariant norm and a cardinality constraint. The resulting convex hull defines another norm for which we give an explicit formula. As a result, we show that it is simple to determine whether an arbitrary point belongs to the convex hull, and to construct a separating hyperplane when it does not. We study the connection between permutation-invariant sets and sets of matrices characterized by their singular values. Furthermore, we investigate the semidefinite-representability of rank-constrained sets of matrices.

In Section 4, we develop convex and concave envelope characterizations of various permutation-invariant functions and sets described using such functions. For example, we derive the convex hull of the lower level-set of a Schur-concave function, which is convex when all but one variable are fixed, a lifted representation of the convex hull of the graph of $\prod_{i=1}^n x_i$ over $[a, b]^n$, where $a, b \in \mathbb{R}$, and the convex hull of $\prod_{i=1}^n \mu_i \geq \prod_{i=1}^n x_{ij}$ over $[c, d]^m \times [a, b]^n$ with $a, c \geq 0$ and $\alpha, \beta > 0$. We show numerically that, for general $a$ and $b$, the convex hull of $\prod_{i=1}^n x_i$ over $[a, b]^n$ is much tighter than the widely-used recursive McCormick relaxation, in contrast to the well-known fact that the recursive McCormick procedure yields the convex hull of $\prod_{i=1}^n x_i$ when $a = -1$ and $b = 1$ [24].

In Section 5, we study the set of rank-one matrices whose generating vectors lie in a permutation-invariant set. We construct semidefinite programming relaxations of the convex hull by proposing various valid inequalities derived from the rank-one condition of the matrix and the fact that every extreme point of a permutahedron generated by a vector is a permutation of the generating vector. Finally, we perform computational experiments with our relaxation for sparse PCA on several instances taken from the literature and other randomly generated instances. We compare our results to the relaxations proposed by [8] and [7].

To increase the readability of the paper, we include a table of some frequently used notations in the appendix.
2 Convex hull of permutation-invariant and sign-invariant sets

In this section, we show that the convex hulls of permutation-invariant and sign-invariant sets can be readily constructed if their convex hulls over a fundamental sub-domain are known. Given a set \( S \), we use the notation \( \operatorname{int}(S) \) to represent its interior, \( \operatorname{vert}(S) \) to denote the set of its extreme points, and \( \operatorname{conv}(S) \) to represent its convex hull.

For a positive integer \( k \), we denote the set of \( k \)-by-\( k \) permutation matrices by \( \mathcal{P}_k \). Given a positive integer \( n \) and a nonnegative integer \( p \), a set \( S \subseteq \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^p\} \) is called permutation-invariant with respect to \( x \) if \( (P^j x, z) \in S \) for all permutation matrices \( P \in \mathcal{P}_n \). When \( S \subseteq \{x \in \mathbb{R}^n\} \) is permutation-invariant with respect to \( x \), we simply say that \( S \) is permutation-invariant. A real-valued function \( (x, z) \mapsto f(x, z) \) with \( (x, z) \in \mathbb{R}^n \times \mathbb{R}^p \) is called permutation-invariant with respect to \( x \) if \( f(x, z) = f(P^j x, z) \) for all permutation matrices \( P \in \mathcal{P}_n \). When \( f : x \mapsto f(x) \) is permutation-invariant with respect to \( x \), we say that \( f \) is permutation-invariant.

Moreover, any permutation-invariant set \( S \) can be written as the lower-level set \( S = \{(x, z) \mid f(x, z) \leq 0\} \) of a permutation-invariant function \( f \) by choosing this function to be the indicator function of the \( S \), which takes value 0 on the set and \( \infty \) otherwise.

A set \( S \subseteq \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^p\} \) where \( n \) is a positive integer and \( p \) is a nonnegative integer is called sign-invariant with respect to \( x \) if \( (x, z) \in S \) implies that \( (\tilde{x}, z) \in S \) for all \( \tilde{x} \) that satisfy \( |\tilde{x}| = |x| \).

Lemma 2.1 describes an important property of the convex hull of sets that are closed under certain linear transformations of the coordinates of their elements.

**Lemma 2.1.** Let \( T \in \mathcal{M}^{n \times n}(\mathbb{R}) \) and let \( S \subseteq \mathbb{R}^n \) be such that for each \( x \in S \), \( T x \in S \) as well. Then, if \( x \in \operatorname{conv}(S) \), \( T x \in \operatorname{conv}(S) \).

**Proof.** Proof. The result follows because \( T S \subseteq S \) implies that \( T \operatorname{conv}(S) = \operatorname{conv}(T S) \subseteq \operatorname{conv}(S) \).

It follows from Lemma 2.1 that if \( S \) is permutation-invariant (resp. sign-invariant) then \( \operatorname{conv}(S) \) is also permutation-invariant (resp. sign-invariant).

For each \( x \in \mathbb{R}^n \), we denote the \( i \)th largest component of \( x \) by \( x[i] \) for \( i = 1, \ldots, n \).

**Definition 2.1.** Given two vectors \( x, y \in \mathbb{R}^n \), we say that \( x \) majorizes \( y \), a property we denote by \( x \succeq_m y \), if \( \sum_{i=1}^j x[i] \geq \sum_{i=1}^j y[i] \) for \( j = 1, \ldots, n-1 \), and \( \sum_{i=1}^n x[i] = \sum_{i=1}^n y[i] \). We say that \( y \) is weakly submajorized by \( x \), and denote this relation as \( x \preceq_{\text{wm}} y \) if \( \sum_{i=1}^j x[i] \geq \sum_{i=1}^j y[i] \), \( \forall j = 1, \ldots, n \). For simplicity of notation, we shall refer to this relation as weak majorization of \( y \) by \( x \).

The result of Lemma 2.2 relates majorization and permutation. Its proof follows from combining Hardy, Littlewood, and Pólya’s theorem with Birckhoff’s theorem; see 2.B.2 and 2.A.2 in [25].

**Lemma 2.2** ([29], Corollary 2.B.3 of [25]). For \( x, y \in \mathbb{R}^n \), \( x \succeq_m y \) if and only if \( y \) is a convex combination of \( x \) and its permutations.

An extension of majorization, which is known as \( G \)-majorization, introduced by Rado [29], in the context of a group of transformations, is defined using the property in Lemma 2.2 as the set of all doubly stochastic matrices from a semigroup; see 14.C in [25] for more detail and references about \( G \)-majorization.

**Lemma 2.3.** Let \( K \) be a convex subset of \( \mathbb{R}^n \times \mathbb{R}^p \). Then

\[
Y := \left\{ (x, u, z) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^p \mid \begin{aligned}
(u, z) &\in K, \\
u &\geq_m x, \\
u_1 &\geq \cdots \geq u_n
\end{aligned} \right\}
\]

is convex.

**Proof.** Proof. First, observe that \( \sum_{i=1}^j u[i] = \sum_{i=1}^j u_i \) since \( u_1 \geq u_2 \geq \cdots \geq u_n \). Further, \( \sum_{i=1}^j x[i] \) is a convex function being the maximum of all possible sums of \( j \) elements chosen from \( x \). Next, \( \sum_{i=1}^j x[i] = \sum_{i=1}^n x_i \) and is, therefore, linear. Therefore, \( Y \) has the following convex representation:

\[
Y = \left\{ (x, u, z) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^p \mid \begin{aligned}
(u, z) &\in K, \\
\sum_{i=1}^j u_i &\geq \sum_{i=1}^j x[i], \text{ for } j = 1, \ldots, n-1, \\
u_1 &\geq \cdots \geq u_n
\end{aligned} \right\}.
\]
We next present Theorem 2.1, which gives an explicit description of the convex hull of a permutation-invariant set when an explicit description of the convex hull of its intersection with the cone \( x_1 \geq \cdots \geq x_n \) is available. This description only requires the introduction of a copy \( u \) of the variables \( x \) together with majorization constraints.

**Theorem 2.1.** Suppose \( S \subseteq \{(x, z) \mid \mathbb{R}^n \times \mathbb{R}^p\} \) is permutation-invariant with respect to \( x \in \mathbb{R}^n \). Then,

\[
\text{conv}(S) = X := \{ (x, z) \mid (u, z) \in \text{conv}(S_0), \ u \geq_m x \},
\]

where \( S_0 = S \cap \{(u, z) \mid u_1 \geq \cdots \geq u_n \} \).

**Proof.** To prove that \( X \) is convex using Lemma 3, it suffices to show that \((u, z) \in \text{conv}(S_0)\) implies \( u \in \Delta^n \). This is clear because \( \text{conv}(S_0) \subseteq \text{conv}(S) \cap \{(u, z) \mid u_1 \geq \cdots \geq u_n \} \).

We now show that \( S \subseteq X \). As \( X \) is convex, this will also show that \( \text{conv}(S) \subseteq X \). Consider an arbitrary \((x, z) \in S\) and define \( u \) as \( u_i = x_{\iota} \) for \( i = 1, \ldots, n \). Then, \((u, z) \in S_0\) because \( S \) is permutation-invariant and \( u \) is in descending order. Since \( u \geq_m x \), \((x, z) \in X\).

We next prove that \( X \subseteq \text{conv}(S) \). Let \((x, z) \in X\). We show that this point can be expressed as a convex combination of points in \( S \). Since \((x, z) \in X\), there exists \( u \) such that \((u, z) \in \text{conv}(S_0) \subseteq \text{conv}(S) \) and \( u \geq_m x \). It follows from the permutation-invariance of \( S \) with respect to \( x \) and Lemma 2.1 that \( \text{conv}(S) \) is permutation-invariant with respect to \( x \). By Lemma 2, \( x \) can be written as \( x = \sum \lambda_i(P_i u) \), where \( P_i \) is a permutation matrix, \( \lambda_i \geq 0 \), and \( \sum \lambda_i = 1 \). Therefore, \((x, z) = (\sum \lambda_i(P_i u), z) = \sum \lambda_i(P_i u, z)\), concluding the proof.

We next present classical results that allow for a linear formulation of the majorization constraints; see Section 3.3.4 of [27] for a more thorough discussion. To model \( \sum_{i=1}^n x_{\iota[i]} \), we express it as the value function of the following optimization problem, where \( x_1, \ldots, x_n \in \mathbb{R} \) and \( j \in \{1, \ldots, n-1\} \):

\[
\max \sum_{i=1}^n x_{\iota[i]} s_i \\
\text{s.t.} \quad \sum_{i=1}^n s_i = j, \quad 0 \leq s_i \leq 1, \quad i = 1, \ldots, n.
\]

To model majorization constraints, we need to enforce that for all feasible \( s_i \), \( \sum_{i=1}^n x_{\iota[i]} s_i \) does not exceed \( \sum_{i=1}^n u_j \). By taking the dual of (3), we exchange the quantifier “for-all” to “there-exists”, and obtain the following formulation which is amenable to direct inclusion in the model:

\[
\text{LS}(j) : \min \sum_{i=1}^n t_i \\
\text{s.t.} \quad x_i \leq t_i + r, \quad i = 1, \ldots, n, \\
\quad t_i \geq 0, \quad i = 1, \ldots, n
\]

where the dual variables \( r \) and \( t \) correspond to the primal constraints \( \sum_{i=1}^n s_i = j \) and \( s \leq 1 \), respectively. Since (3) is feasible, (4) exhibits no duality gap. The constraint \( \sum_{i=1}^n u_i \geq \sum_{i=1}^n x_{\iota[i]} \) is then modeled through the existence of an \((r, t)\) that is feasible to (4) such that \( \sum_{i=1}^n u_i \geq \sum_{j=1}^n t_j \).

**Theorem 2.2.** Suppose \( S \subseteq \{(x, z) \mid \mathbb{R}^n \times \mathbb{R}^p\} \) is permutation-invariant with respect to \( x \). Then,

\[
\text{conv}(S) = \left\{ (x, z) \mid \begin{align*}
(u, z) & \in \text{conv}(S_0) \\
u_1 \geq \cdots \geq u_n \\
\sum_{i=1}^n u_i & = \sum_{i=1}^n x_i \\
\sum_{i=1}^n u_i & \geq j r^j + \sum_{i=1}^n t_i^j, \quad j = 1, \ldots, n-1, \\
x_i & \leq t_i^j + r^j, \quad j = 1, \ldots, n-1, \quad i = 1, \ldots, n, \\
t_i^j & \geq 0, \quad j = 1, \ldots, n-1, \quad i = 1, \ldots, n,
\end{align*} \right\},
\]

where \( S_0 = S \cap \{(u, z) \mid u_1 \geq \cdots \geq u_n \} \).

**Remark 2.1.** In fact, Theorems 2.1 and 2.2 remain valid for any choice of \( S_0 \) that satisfies \( \text{conv}(S) \cap \{(u, z) \mid u_1 \geq \cdots \geq u_n \} \supseteq S_0 \supseteq S \cap \{(u, z) \mid u_1 \geq \cdots \geq u_n \} \).

The formulation given in (5) expresses \( \text{conv}(S) \) as the projection of a convex set with \( n^2 + n + p \) variables. This formulation is much smaller than that which would have been obtained using a classical application of disjunctive programming [see 4, Chapter 2]. An even smaller representation is possible using a more compact formulation of the permutahedron. Goemans [10] proposed such an extended formulation for the permutahedron, the convex hull
of all possible permutations of a fixed vector \( u \in \mathbb{R}^n \) using a sorting network where the numbers of variables and inequalities of the extended formulation depend on the number of comparators of the associated sorting network. When the Ajtai–Komlós–Szemerédi sorting network \([1]\) is used, the extended formulation for the permutahedron has \( \Theta(n \log n) \) variables and inequalities. As imposing the majorization constraint \( u \geq_m x \) is equivalent to requiring that \( x \) belongs to the permutahedron generated by \( u \), alternative extended formulations of \( \text{conv}(S) \) can be obtained by replacing \( u \geq_m x \) in (2) with such extended formulations. This results in a formulation of \( \text{conv}(S) \) that is more compact than (5). For relaxations of sparse principal component analysis, we show in Section 5.2 that this smaller formulation provides some computational benefit on larger instances in our test set.

The ideas underlying the proof of Theorem 2.2 can also be applied when sets are invariants with respect to collections of linear transformations that are not permutation matrices. In particular, we describe next a related convexification result for sign-invariant sets.

**Theorem 2.3.** Suppose \( S \subseteq \{(x, z) \in \mathbb{R}^n \times \mathbb{R}^p \} \) is sign-invariant with respect to \( x \). Then,

\[
\text{conv}(S) = X := \{(x, z) \mid (u, z) \in \text{conv}(S_0), u \geq |x|\}
\]

where \( S_0 = S \cap (\mathbb{R}_+^n \times \mathbb{R}^p) \).

**Proof.** Set \( X \) is convex because it is the projection of an intersection of two convex sets. We now show that \( S \subseteq X \). For an arbitrary \( (x, z) \in S \), define \( u = |x| \). By sign-invariance of \( S \), \( (u, z) \in S \) and hence \( (u, z) \in S_0 \subseteq \text{conv}(S_0) \). By definition, \( u \) satisfies \( u \geq |x| \). This shows that \( \text{conv}(S) \subseteq X \).

We next show that \( X \subseteq \text{conv}(S) \). Let \( (x, z) \in X \). There exists \( u \in \mathbb{R}^n \) such that \( (u, z) \in \text{conv}(S_0) \subseteq \text{conv}(S) \) and \( u \geq |x| \). Since \( \text{conv}(S) \) is sign-invariant by Lemma 2.1, it follows that \( \{(x, z) \mid x_i \in \{u_i, -u_i\}\} \subseteq \text{conv}(S) \).

Therefore, \( (x, z) \in \{(x, z) \mid |x_i| \leq u_i\} \subseteq \text{conv}(S) \), where the containment follows from the convexity of \( \text{conv}(S) \).

Next, we consider the case where the target set \( S \) is both sign-invariant and permutation-invariant. While two separate representations (2) and (6) for the convex hull already exist, we can derive a unified representation. Inequality \( u \geq_m |x| \) would not be a useful description because \( \sum_{i=1}^n u_i = \sum_{i=1}^n |x_i| \) is a non-convex constraint. In fact, such a set would not even be convex, as can be seen from \( S = \{-1, 1\} \). Instead, we prove that the convex hull can be represented using weak majorization. More precisely, suppose that \( S \subseteq \{(x, z) \mid \mathbb{R}^n \times \mathbb{R}^p \} \) is sign- and permutation-invariant with respect to \( x \in \mathbb{R}^n \). Then, \( \text{conv}(S) = X := \{(x, z) \mid (u, z) \in \text{conv}(S_0), u \geq_{wm} |x|\} \), where \( S_0 = S \cap \{(u, z) \mid u_1 \geq \cdots \geq u_n \geq 0\} \). We first prove that \( S \subseteq X \). Consider an arbitrary \( (x, z) \in S \), and define \( u \) as \( u_i = |x_i| \) for \( i = 1, \ldots, n \). Then, \( (u, z) \in S_0 \subseteq \text{conv}(S_0) \) and \( u \geq_{wm} |x| \), showing the inclusion. We next prove \( X \subseteq \text{conv}(S) \). Consider an arbitrary \( (x, z) \in X \). Then, there exists \( u \) such that \( (u, z) \in \text{conv}(S_0) \subseteq \text{conv}(S) \) and \( u \geq_{wm} |x| \). It follows that there exists \( u' \) such that \( u \geq_m u' \) and \( u' \geq |x| \); see 5.2.9. of [25], for example. Then, it can be shown that \( (u', z) \in \text{conv}(S) \) using the same arguments that were used in the proof of Theorem 2.1. Since \( \text{conv}(S) \) is sign-invariant, this implies that all sign variants of \( (u', z) \in \text{conv}(S) \). Then, by the last part of the proof of Theorem 2.3, this implies that \( (x, z) \in \text{conv}(S) \). (This convex hull description can also be derived by first constructing \( \text{conv}(S \cap (\mathbb{R}_+^n \times \mathbb{R}^p)) \)) using Theorem 2.1. Thus, \( \text{conv}(S) \) is obtained by Theorem 2.3. The variables \( u' \) introduced by Theorem 2.3 can finally be projected using 5.A.9. from [25].

The above convexification results can be easily extended to the sets which are permutation-invariant or/and sign-invariant with respect to multiple subsets of independent variables.

**Theorem 2.4.** Let \( S \subseteq \{(x^1, \ldots, x^m, z) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m} \times \mathbb{R}^p \} \).

1. Suppose \( S \) is a permutation-invariant set with respect to \( x^k \) for \( k = 1, \ldots, m \). Then,

\[
\text{conv}(S) = \left\{(x^1, \ldots, x^m, z) \mid (u^1, \ldots, u^m, z) \in \text{conv}(S_0), \right. \\
\left. u^k \geq_m x^k, k = 1, \ldots, m \right\}
\]

where \( S_0 = S \cap (\Delta^{n_1} \times \cdots \times \Delta^{n_m} \times \mathbb{R}^p) \).

2. Suppose \( S \) is sign-invariant with respect to \( x^k \) for \( k = 1, \ldots, m \). Then,

\[
\text{conv}(S) = \left\{(x^1, \ldots, x^m, z) \mid (u^1, \ldots, u^m, z) \in \text{conv}(S_0), \right. \\
\left. u^k \geq |x^k|, k = 1, \ldots, m \right\}
\]

where \( S_0 = S \cap (\mathbb{R}_+^{n_1} \times \cdots \times \mathbb{R}_+^{n_m} \times \mathbb{R}^p) \).
3. Suppose $S$ is permutation-invariant and sign-invariant with respect to $x^k$ for $k = 1, \ldots, m$. Then,

$$\text{conv}(S) = \left\{ (x^1, \ldots, x^m, z) \mid (u^1, \ldots, u^m, z) \in \text{conv}(S_0), \quad u^k \geq \omega_m |x^k|, \quad k = 1, \ldots, m \right\}$$

(9)

where

$$S_0 = S \cap \left\{ (u^1, \ldots, u^m, z) \mid u^1 \geq \cdots \geq u^k \geq 0, \quad k = 1, \ldots, m \right\}.$$

We remark that our convexification result on sign-invariant sets can also be used to convexify reflections on a hyperplane [17]. In particular, consider a set $S$ and assume that the set is closed under reflection on $\langle a, \cdot \rangle = 0$ for some $a$ with $\|a\|_2 = 1$. Assume further that $S_0 = S \cap \{ x \mid \langle a, x \rangle \geq 0 \}$ is available. Consider any orthogonal matrix $U$ that aligns $a$ along the first principal direction, so that $Ua = e_1$. Observe that $US$ is sign-invariant with respect to variable $x_1$. Then, $\text{conv}(US) = \{ z \mid (t, z_2, \ldots, z_n) \in U \text{conv}(S_0), t \geq |z_1| \}$ and so, $\text{conv}(S) = \{ x \mid x + (t - z_1)a \in \text{conv}(S_0), t \geq |z_1| \}$, where we have used the fact that $U^{'e_1} = a$.

3 Sparsity theorem

In this section, we first study the convex hull of the set

$$N^K_{\| \cdot \|_s} = \{ x \in \mathbb{R}^n \mid \|x\|_s \leq 1, \text{card}(x) \leq K \},$$

(10)

where $\| \cdot \|_s$ is a sign- and permutation-invariant norm (also known as a symmetric gauge function) and introduce the Sparsity Theorem (Theorem 3.2), which shows that the optimal value of

$$\min_{u} \|u\|_s \quad \text{s.t.} \quad u_1 \geq \cdots \geq u_K \geq 0,$$

$$u_{K+1} = \cdots = u_n = 0,$$

$$u \geq \omega_m |x|$$

(11)

is no more than 1 if and only if $x \in \text{conv}(N^K_{\| \cdot \|_s})$ and that an optimal solution to (11) can be obtained in closed-form.

In other words, given $x \in \mathbb{R}^n$, Theorem 3.2 gives a closed-form expression for a sparse vector $u$, in terms of $x$, in the representative disjunct $\Delta^n \cap \mathbb{R}^n_+$ that weakly majorizes $|x|$ and minimizes $\|u\|_s$. In fact, this $u$ majorizes $|x|$. This is not surprising. If the optimal $u$ to (11) did not majorize $|x|$, there exists $u'$ such that $u \geq u' \geq_m |x|$; see 5.A.9. in [25]. Then, since $u'$ is in the convex hull of $u$ and its sign-variants, it follows by sign-invariance of $\| \cdot \|_s$ that $\|u'\|_s \leq \|u\|_s$, and so $u'$ is also optimal and majorizes $|x|$.

We denote the set $\{ x \in \mathbb{R}^n \mid \|x\|_s \leq r \}$ by $B_s(r)$. When $K = 1$, the convex hull is an $\ell_1$-norm ball. The set is trivial when $K = n$. Therefore, we assume $1 < K < n$. When the associated norm $\| \cdot \|_s$ is the $\ell_2$-norm, $N^K_{\| \cdot \|_s}$ is the feasible set of the sparse principal component analysis problem (sparse PCA); see [8].

We define

$$\Delta^n_+ := \Delta^n \cap \mathbb{R}^n_+$$

(12)

and, for any vector $x \in \mathbb{R}^n$, define

$$(x_{\Delta^n})_i = x_{[i]} \quad \text{and} \quad ((x_{\Delta^n})_i = |x_{[i]}|, \quad \text{for } i = 1, \ldots, n.$$

When the dimension $n$ of the set and the associated vector is clear in the context, we use the simpler notations $\Delta, \Delta_+, x_{\Delta}$, and $x_{\Delta_+}$.

By sign- and permutation-invariance of the norm $\| \cdot \|_s$ and that of the cardinality requirement, $N^K_{\| \cdot \|_s}$ is sign- and permutation-invariant and hence we can apply Theorem 2.4 to obtain its convex hull as a projection of a higher dimensional set

$$\text{conv} \left( N^K_{\| \cdot \|_s} \right) = \left\{ x \in \mathbb{R}^n \mid u \in N^K_{\| \cdot \|_s} \cap \Delta_+, \quad u \geq \omega_m |x| \right\} = \left\{ x \in \mathbb{R}^n \mid \|u\|_s \leq 1, \quad u_1 \geq \cdots \geq u_K \geq 0, \quad u_{K+1} = \cdots = u_n = 0, \quad u \geq \omega_m |x| \right\}.$$

(13)

The extended formulation (13) can be written in closed-form with $O(nK)$ additional variables and constraints based on the modeling technique described in Section 2. Other extended formulations are proposed in [21] and [22] for the
case where $\| \cdot \|_s$ is an $\ell_p$-norm. In these papers, the formulations are obtained either through dynamic programming concepts or Goemans’ extended formulation of the permutahedron using a sorting network [10].

In this section, we describe the convex hull as a norm ball in the original variable space. The induced norm is easily calculable if the associated norm $\| \cdot \|_s$ is calculable. Moreover, given an arbitrary point in $\mathbb{R}^n$ not in the convex hull, we devise an algorithm to construct a separating hyperplane.

We first present the following lemma introduced in [20].

**Lemma 3.1.** Suppose $x \geq_m y$. Then, for any permutation-invariant norm $f(\cdot)$, $f(x) \geq f(y)$.

A set in $\mathbb{R}^n$ is called a convex body if it is a compact convex set with non-empty interior. In the next proposition, we show that $\text{conv}(N^K_{\| \cdot \|})$ is a convex body.

**Proposition 3.1.** The set $\text{conv}(N^K_{\| \cdot \|})$ is a convex body.

**Proof.** Since $N^K_{\| \cdot \|}$ is a compact set, it follows that $\text{conv}(N^K_{\| \cdot \|})$ is a compact convex set [5, Corollary I.2.4]. To see that $\text{conv}(N^K_{\| \cdot \|})$ has a non-empty interior, observe that there exists $\epsilon > 0$ such that $B_1(\epsilon) \subseteq B_s(1)$ where $B_1(\epsilon)$ represents the $\ell_1$-norm ball with radius $\epsilon$. This follows from the equivalence of norms in a finite vector space. Notice that $\text{vert}(B_1(\epsilon)) = \{ \pm \epsilon e_i \mid i = 1, \ldots, n \}$ where $e_i$ is the $i$th canonical vector. Since, for any $x \in \text{vert}(B_1(\epsilon))$, $\text{card}(x) = 1$, it follows that $\text{vert}(B_1(\epsilon)) \subseteq N^K_{\| \cdot \|}$, and, so $B_1(\epsilon) \subseteq \text{conv}(N^K_{\| \cdot \|})$. The result follows because $0 \in \text{int}(B_1(\epsilon))$.

It is well-known that there exists a one-to-one correspondence between norms in $\mathbb{R}^n$ and convex bodies symmetric about 0 and containing 0 in their interior; see Section 14.4 of [26] for instance. Given an arbitrary norm $\| \cdot \|$, we can construct its unit ball $\{ x \mid \| x \| \leq 1 \}$, which is a convex body of the desired type. Conversely, given any compact convex body $C$ that is symmetric about 0 and contains 0 in its interior, we can define the function

$$f_C(x) := \min \left\{ t > 0 \mid \frac{x}{t} \in C \right\} \quad (14)$$

for $x \in \mathbb{R}^n$. It is known that the function $f_C$ satisfies the properties of norms; see Section 14.4 of [26], for example. Further, the convex body $C$ is a lower-level set of this norm, that is, $C = \{ x \mid f_C(x) \leq 1 \}$.

Since $\text{conv}(N^K_{\| \cdot \|})$ is a compact convex body that is symmetric about 0 and contains 0 in its interior, a norm associated with $\text{conv}(N^K_{\| \cdot \|})$ can be defined as in (14). We denote the corresponding norm by $\| \cdot \|_c$. Since $\| \cdot \|_c$ is sign- and permutation-invariant, the following result holds.

**Proposition 3.2.** The set $\text{conv}(N^K_{\| \cdot \|})$ is the unit ball associated with a sign- and permutation-invariant norm, that is, $\text{conv}(N^K_{\| \cdot \|}) = B_c(1)$.

We next show that the values of the norms $\| \cdot \|_c$ and $\| \cdot \|_s$ are the same for vectors that satisfy the cardinality constraint.

**Proposition 3.3.** If $\text{card}(x) \leq K$, $\| x \|_c = \| x \|_s$.

**Proof.** Proof. We first show that $\| x \|_c \geq \| x \|_s$. Since $N^K_{\| \cdot \|} \subseteq B_s(1)$, it follows that $B_c(1) = \text{conv}(N^K_{\| \cdot \|}) \subseteq B_s(1)$. This implies that $B_c(r) \subseteq B_s(r)$ for any $r$. In particular, when $r = \| x \|_c$, we have $x \in B_c(r) \subseteq B_s(r)$, which implies that $\| x \|_c \geq \| x \|_s$. We now show $\| x \|_c \leq \| x \|_s$ when $\text{card}(x) \leq K$. Let $r = \| x \|_s$ and observe that $\frac{x}{r} \in N^K_{\| \cdot \|} \subseteq B_c(1)$. Therefore, $x \in B_c(r)$ or $\| x \|_c \leq \| x \|_s$.

We present an explicit formula to evaluate $\| \cdot \|_c$. For an arbitrary $x \in \mathbb{R}^n$, define $s(x) \in \mathbb{R}^{K+2}$ as

$$s(x)_i = \sum_{j=1}^{n} \frac{|x_j|}{K - i + 1}, \quad i = 1, \ldots, K; \quad s(x)_0 = s(x)_{K+1} = \infty.$$ 

Let $i_x$ be the minimum among those indices that minimize $s(x)_i$, and let $\delta(x) = s(x)_{i_x}$. Now, define $u(x) \in \mathbb{R}^n$ as

$$u(x)_i = \begin{cases} |x|_i, & i \in \{1, \ldots, i_x - 1\} \\ \delta(x), & i \in \{i_x, \ldots, K\} \\ 0, & \text{otherwise}. \end{cases} \quad (15)$$

In the following proposition, we show that, for arbitrary $x \in \mathbb{R}^n$, we can construct a vector $u(x) \in \Delta_+$ that satisfies the cardinality constraint and majorizes $x$. 

7
Proposition 3.4. Let \( s(x), i_x, \delta(x) \), and \( u(x) \) be defined as above. Then,

1. \( s(x)_{i+1} - s(x)_i = \frac{1}{K-i} (s(x)_{i+1} - |x|_{i+1}) = \frac{1}{K-i} (s(x)_i - |x|_i) \) for \( i = 1, \ldots, K-1 \)
2. \( s(x)_1 \geq \cdots \geq s(x)_{i_x} \) and \( s(x)_{i_x} \leq \cdots \leq s(x)_K \)
3. \( u(x) \geq_{\text{m}} |x| \)
4. \( u(x)_1 = \max\{ |x|_1, s(x)_1 \} \)

Proof. By definition of \( s(x) \),

\[
(K - i) s(x)_{i+1} - (K - i + 1) s(x)_i = \sum_{k=i+1}^n |x|_j - (K - i + 1) \sum_{j=i}^n |x|_j
\]

By adding \( s(x)_{i+1} \) on both sides, we have \( (K - i + 1) s(x)_{i+1} - s(x)_i = s(x)_{i+1} - |x|_{i+1} \), implying the first equality of Part 1. The remainder of the part can be shown similarly by adding \( s(x)_i \) on both sides.

To see Part 2, if there is no index \( i \) in \( \{1, \ldots, K-1\} \) so that \( s(x)_i < s(x)_{i+1} \), the result holds trivially. Assume now that \( i' \) is the smallest such index. Then, \( s(x)_{i'+1} > s(x)_{i'} \geq |x|_{i'} \), where the second inequality follows because \( s(x)_{i'+1} - s(x)_i > 0 \) implies \( s(x)_{i'} > |x|_{i'} > 0 \) from Part 1. This in turn shows, using Part 1 that \( s(x)_{i'+2} > s(x)_{i'+1} \), as long as \( i' < K - 1 \). By induction, \( s(x)_{i'} < \cdots < s(x)_K \) and, by the definition of \( i' \), \( s(x)_1 \geq \cdots \geq s(x)_{i'} \). Then, Part 2 follows by defining \( i_x \) as the first index such that \( s(x)_{i_x} = s(x)_i \).

We next prove Part 3. We first show that \( u(x) \) is nonincreasing. If \( i_x = 1 \) then all components of \( u(x) \) equal \( \delta(x) \), and hence it is nonincreasing. Now, assume that \( i_x \geq 2 \). By definition of \( u(x) \), it suffices to show that \( |x|_{i_x} \geq \delta(x) \). Then, \( s(x)_{i_x} - |x|_{i_x-1} = (K - i_x)(s(x)_{i_x} - |x|_{i_x-1}) \leq 0 \) by plugging in \( i = i_x - 1 \) in the first equality of Part 1. Therefore, \( |x|_{i_x} \geq s(x)_{i_x} = \delta(x) \), completing the proof that \( u(x) \) is nonincreasing. Thus, \( u(x)_{i} = u(x)_{i_x} = 0 \) for all \( i = 1, \ldots, n \). Next, observe that \( \sum_{i=i_x}^n u(x)_{i} = \sum_{i=i_x}^n |x|_{i} \) by definition of \( \delta(x) \). This implies in turn that \( \sum_{i=i_x}^n u(x)_{i} = \sum_{i=i_x}^n |x|_{i} \). We next show that \( \sum_{i=i_x}^n u(x)_{i} \geq \sum_{i=i_x}^n |x|_{i} \) for all \( j = 1, \ldots, n - 1 \). When \( j = 1, \ldots, i_x - 1 \), the inequality holds by definition of \( u(x) \). We next consider the case \( j \geq i_x \). If \( i_x = K \), the inequality holds because \( \sum_{i=i_x}^n u(x)_{i} = \sum_{i=i_x}^n |x|_{i} \) where the first and the second equalities follow because \( j \geq K \) and \( u(x)_{K+1} = \cdots = u(x)_n = 0 \). Now assume that \( i_x < K \). Since \( s(x)_{i_x+1} \geq s(x)_{i_x} \) and \( s(x)_{i_x+1} - s(x)_{i_x} = \frac{1}{K-i_x} (s(x)_{i_x} - |x|_{i_x}) \), we have \( s(x)_{i_x} \geq |x|_{i_x} \) for all \( i \geq i_x \). Therefore, \( \sum_{i=i_x}^j u(x)_{i} - \sum_{i=i_x}^j |x|_{i} = \sum_{i=i_x}^j u(x)_{i} - \sum_{i=i_x}^j |x|_{i} = \sum_{i=i_x}^j (\delta(x) - |x|_{i}) \geq 0 \).

For Part 4, first assume \( i_x = 1 \). Then, \( u(x)_1 = s(x)_1 \). By Part 1, \( s(x)_1 \geq |x|_1 \) and hence \( u(x)_1 = \max\{ |x|_1, s(x)_1 \} \). Next, assume that \( i_x \geq 2 \). Then, \( u(x)_1 = |x|_1 \). By Part 2, \( s(x)_2 \leq s(x)_1 \) and hence, by Part 1, \( s(x)_1 \leq |x|_1 \). Therefore, \( u(x)_1 = \max\{ |x|_1, s(x)_1 \} \).

We next show in the following theorem that \( u(x) \) can be used to compute \( \|x\|_c \) if \( \| \cdot \|_a \) is calculable. To this end, we introduce the following notations. Let \( \beta \) be an optimal solution to

\[
\max\{ \beta^T u(x) \mid \|\beta\|_{a^*} \leq 1 \},
\]

where \( \| \cdot \|_{a^*} \) is the dual norm of \( \| \cdot \|_a \). We now define \( \theta \) and \( \chi \) in \( \mathbb{R}^n \) as follows:

\[
\theta_i = \begin{cases} \frac{\beta_i}{\sum_{k=1}^K \beta_k}, & i = 1, \ldots, i_x - 1 \\ 0, & \text{otherwise} \end{cases}, \quad \chi_i = \begin{cases} \frac{\beta_i}{\sum_{k=i_x}^K \beta_k}, & i = 1, \ldots, i_x - 1 \\ \sum_{k=i_x}^K \beta_k, & \text{otherwise} \end{cases}
\]

Theorem 3.1. For an arbitrary \( x \in \mathbb{R}^n \), let \( u(x) \) be defined as in (15). Then, \( \|x\|_c = \|u(x)\|_a \).

Proof. Proof. Recall the definitions of \( s(x) \), \( i_x \), and \( \delta(x) \). We have \( \|x\|_c \leq \|u(x)\|_c = \|u(x)\|_a \), where the first inequality is because of Part 3 of Proposition 3.4 and Lemma 3.1, while the equality is due to Proposition 3.3. We next show \( \|x\|_c \geq \|u(x)\|_a \). Let \( r = \|u(x)\|_a \) so that \( u(x) \in B_r(r) \). By the definition of \( \| \cdot \|_a \) and \( \tilde{\beta} \), we have \( \tilde{\beta}^T u(x) = r \). Moreover, \( \tilde{\beta}^T u \leq \|u\|_a \leq r \) for all \( u \in B_r(r) \), where the first inequality is because \( \tilde{\beta} \in B_{\|u\|_a} \) and the
second inequality is because \( u \in B_2(r) \). Because \( u(x) \in \Delta_+ \), it follows from rearrangement inequality that, without loss of generality, we may assume that \( \beta \in \Delta_+ \).

We next show that \( \theta^T u \leq r \) for all \( u \in B_2(r) \) where \( \theta \) is as defined in (17). Define \( \beta := (\hat{\beta}_1, \ldots, \hat{\beta}_K, 0, \ldots, 0) \) and \( \beta := (\hat{\beta}_1, \ldots, \hat{\beta}_K, -\hat{\beta}_{K+1}, \ldots, -\hat{\beta}_n) \) and observe that \( \beta = \frac{\beta + \beta}{2} \). By sign-invariance of \( \| \cdot \|_\ast \) and thus of \( \| \cdot \|_\ast \), \( \beta \in B_2 \ast (1) \), and, so \( \beta \in B_2 \ast (1) \). However, since \( \beta \geq_m \theta \), Lemma 3.1 shows that \( \theta \in B_2 \ast (1) \). This in turn shows that \( \theta^T u \leq r \) is valid for \( B_2(r) \).

We next claim that \( \chi^T u \leq 1 \) is valid for \( N_\| \cdot \|_\ast \), where \( \chi \) as is defined in (17). Assume that, on the contrary, there exists \( \hat{u} \in N_\| \cdot \|_\ast \) such that \( \chi^T \hat{u} > 1 \). Because of the rearrangement inequality and the fact that \( \chi \in \Delta_+ \), we may assume that \( \hat{u} \in \Delta_+ \). This yields the contradiction \( 1 \geq \theta^T \hat{u} = \chi^T \hat{u} > 1 \), where the first inequality is by the validity of \( \theta^T u \leq 1 \) for \( B_2(1) \) which outer-approximates \( N_\| \cdot \|_\ast \), and the equality can be arranged by choosing \( \hat{u} \) with support at most \( K \). It follows that \( \chi^T u \leq r \) is valid for \( B_c(r) \) or, in other words, that \( \chi \in B_c \ast (1) \), where \( \| \cdot \|_\ast \) is the dual norm of \( \| \cdot \|_\ast \).

Therefore,

\[
\|x\|_c = \|x_{\Delta_+}\|_c = \max\{\beta^T x_{\Delta_+} : \|\beta\|_\ast \leq 1\} \geq \chi^T x_{\Delta_+},
\]

where the inequality is because \( \|\chi\|_\ast \leq 1 \). However, the following calculation shows that

\[
\chi^T x_{\Delta_+} = \sum_{i=1}^{c-1} \frac{\hat{\beta}_i}{K} |x_{[i]} + \sum_{i=1}^{n} |x_{[i]}| = \sum_{i=1}^{c-1} \hat{\beta}_i |x_{[i]}| + \sum_{i=1}^{K} \frac{1}{K} \sum_{i=1}^{n} |x_{[i]}| = \hat{\beta}^T u(x) = r.
\]

Combining (18) and (19), we conclude that \( \|x\|_c \geq \|u(x)\|_s \), where we have used that \( r \) was defined to be \( \|u(x)\|_s \).

**Corollary 3.1.** For a fixed \( x \in \mathbb{R}^n \), let \( \chi \) be as defined in (17). Then, \( \chi^T u \leq 1 \) is valid for \( N_\| \cdot \|_\ast \). In particular, \( \chi^T u \leq 1 \) separates \( x_{\Delta_+} \), if \( x \notin \text{conv}(N_\| \cdot \|_\ast) \).

**Proof.** Proof. It was already shown in the proof of Theorem 3.1 that \( \chi^T u \leq 1 \) is valid for \( N_\| \cdot \|_\ast \). Let \( u(x) \) be as defined in (15). Then, \( \chi^T x_{\Delta_+} = \|u(x)\|_s = \|x\|_c \), where the first equality is from (19) and the second equality is from Theorem 3.1. Therefore, \( \|x_{\Delta_+}\|_c = \chi^T x_{\Delta_+} \leq \|\chi\|_\ast \|x_{\Delta_+}\|_\ast \leq \|x_{\Delta_+}\|_c \), where the first inequality is due to Cauchy-Schwarz inequality and the second inequality is because the proof of Theorem 3.1 shows that \( \|\chi\|_\ast \leq 1 \). Therefore, equality holds throughout and, in particular, \( \|\chi\|_\ast \leq 1 \). If \( x \notin \text{conv}(N_\| \cdot \|_\ast) \), then \( \chi^T x_{\Delta_+} = \|x\|_c > 1 \), where the inequality is because \( x \notin B_c(1) \).

**Remark 3.1.** In the proof of Corollary 3.1, let \( T \) be the transformation (a composition of sign-conversions and permutations) that maps \( x \) to \( x_{\Delta_+} \). Then, the hyperplane that separates \( x \) and \( N_\| \cdot \|_\ast \) is \( T^{-1}(\chi)^T u \leq 1 \).

**Theorem 3.2** (Sparsity Theorem). For an arbitrary \( x \in \mathbb{R}^n \), \( u(x) \) as defined in (15) is an optimal solution to

\[
\min \|u\|_s \quad \text{s.t.} \quad u_1 \geq \cdots \geq u_K \geq 0, \quad u_{K+1} = \cdots = u_n = 0, \quad u \geq_{\text{min}} |x|.
\]

**Proof.** Proof. First, \( u(x) \) is feasible because of its definition and Part 3 of Proposition 3.4. Now, it suffices to show that \( \|u(x)\|_s \leq \|u\|_s \) for any feasible solution \( u \). Since \( u \geq_{\text{min}} |x| \), there exists \( \hat{u} \) such that \( u \geq_m \hat{u} \) and \( \hat{u} \geq |x| \); see 5.A.9. of [25], for example. Then, \( \|u\|_s = \|u\|_c \geq \|\hat{u}\|_c \geq \|x\|_c = \|u(x)\|_c \), where the first inequality follows from Proposition 3.3, the first inequality follows from Lemma 3.1, and the last equality follows from Theorem 3.1. Finally, to prove the second inequality, observe that \( |x| \) can be written as a convex combination of \( u' \) and its sign-invariants because \( u' \geq |x| \). By the triangle inequality, positive homogeneity, and the sign-invariance of \( \| \cdot \|_\ast \) and \( \| \cdot \|_\ast \),

**Example 3.1.** Consider the set \( N_\| \cdot \|_\ast \), where \( n = 6 \). Let \( x := (\frac{27}{28}, \frac{27}{28}, \frac{27}{28}, \frac{27}{28}, \frac{27}{28}, \frac{27}{28}) \). Note that \( \|x\|_2 = 1 \) and \( x \in \Delta_+ \). For illustration, we establish that \( x \notin \text{conv}(N_\| \cdot \|_\ast) \) by constructing an explicit separating hyperplane described in (16) and (17). First, we construct the vector \( s(x) \in \mathbb{R}^3 \) as follows:

\[
\begin{align*}
s(x)_1 &= \frac{\sum_{j=1}^{6} x_j}{3+3+1} = \frac{1}{12} \left( 27 + \frac{27}{28} + \frac{27}{28} + \frac{27}{28} + \frac{27}{28} + \frac{27}{28} \right) = \frac{28}{56}, \\
s(x)_2 &= \frac{\sum_{j=2}^{6} x_j}{3+3+1} = \frac{1}{12} \left( \frac{27}{28} + \frac{27}{28} + \frac{27}{28} + \frac{27}{28} + \frac{27}{28} \right) = \frac{28}{56}, \\
s(x)_3 &= \frac{\sum_{j=3}^{6} x_j}{3+3+1} = \frac{1}{12} \left( \frac{27}{28} + \frac{27}{28} + \frac{27}{28} + \frac{27}{28} \right) = \frac{20}{56}.
\end{align*}
\]
Observe that \( s(x)_2 = \min \{ s(x)_1, s(x)_2, s(x)_3 \} \). Next, we compute that
\[
u(1) = x_1 = 27 \frac{27}{28}, \quad u(2) = u(x)_3 = s(x)_2 = 15 \frac{56}{56}, \quad u(4) = u(x)_5 = u(x)_6 = 0.
\]
Since \( \|u(x)\|_2 = 1.036 \cdots > 1 \), we conclude from Theorem 3.1 that \( x \notin \text{conv}(N^3_{\| \cdot \|_2}) \). We now derive the separating hyperplane. We first separate \( u(x) \) from \( B_2(1) \). Since \( \| \cdot \|_2 \) is self-dual, \( \hat{\beta} = u(x)/\|u(x)\|_2 \) is an optimal solution to (16). Then, the inequality \( \hat{\beta}^\top u \leq 1 \) (or equivalently \( u(x)^\top u \leq \|u(x)\|_2^2 \)) is valid for \( B_2(1) \). Furthermore, it separates \( u(x) \) because \( u(x)^\top u(x) = \|u(x)\|_2^2 > \|u(x)\|_2 \). We next construct a hyperplane that separates \( x \) from \( N^3_{\| \cdot \|_2} \). Define \( \theta \) and \( \chi \) as follows:
\[
\theta_1 = \beta_1 = \frac{1}{\|u(x)\|_2} u(x)_1 = \frac{1}{\|u(x)\|_2} \left( 27 \frac{27}{28}, 56 \frac{56}{56}, 56 \right), \quad \theta_2 = \beta_3 = \frac{1}{\|u(x)\|_2} \left( \beta_2 + \beta_3 \right) = \frac{1}{\|u(x)\|_2} \left( 15 \frac{56}{56}, 15 \frac{56}{56}, 15 \frac{56}{56} \right), \quad \theta_4 = \cdots = \theta_6 = 0
\]
\[
\chi_1 = \beta_1 = \frac{1}{\|u(x)\|_2} u(x)_1 = \frac{1}{\|u(x)\|_2} \left( 27 \frac{27}{28}, 56 \frac{56}{56}, 56 \right), \quad \chi_2 = \cdots = \chi_6 = \frac{1}{\|u(x)\|_2} \left( \beta_2 + \beta_3 \right) = \frac{1}{\|u(x)\|_2} \left( 15 \frac{56}{56}, 15 \frac{56}{56}, 15 \frac{56}{56} \right).
\]
Observe that \( \|\theta\|_2 = \|\beta\|_2 = 1 \), but \( \|\chi\|_2 > 1 \). Now consider the inequality \( \chi^\top y \leq 1 \). It is valid for \( N^3_{\| \cdot \|_2} \) because for any \( u \in N^3_{\| \cdot \|_2} \), \( \chi^\top u \leq \chi^\top u_{\Delta^+} = \theta^\top u_{\Delta^+} \leq \|\theta\|_2 \|u_{\Delta^+}\|_2 \leq 1 \). Moreover, it separates \( x \) because \( \chi^\top x = \frac{1}{\|u(x)\|_2} \left( 27 \frac{27}{28}, 56 \frac{56}{56}, 56 \right) \left( 56 \frac{56}{56}, 4 \frac{4}{28}, 3 \frac{3}{28}, 2 \frac{2}{28}, 1 \frac{1}{28} \right) = 1.036 \cdots > 1 \).

Next, we consider some special cases of the set \( N^K_{\| \cdot \|_2} \) defined in (10) and provide explicit convex hull descriptions.

**Proposition 3.5.** Let \( S = \{ x \in \mathbb{R}^q : |\text{card}(x)_1 \leq K, \|x\|_\infty \leq r, \} \) where \( \|x\|_\infty = |x|_1 \). Then,
\[
\text{conv}(S) = \{ x \in \mathbb{R}^q : \|x\|_1 \leq rK, \|x\|_\infty \leq r \}.
\]

**Proof.** Proof. Observe that \( S/r = \{ x \in \mathbb{R}^q : |\text{card}(x)_1 \leq K, \|x\|_\infty \leq 1 \} \). Then,
\[
\text{conv}(S) = r \text{conv}(S/r) = \{ y \in \mathbb{R}^q : \|y\|_\infty \leq r \} = \{ y \in \mathbb{R}^q : \max\{|y|_1, s(y)_1 \} \leq r \},
\]
where the second equality follows from Proposition 3.2 and Theorem 3.1 and the third equality from Part 4 of Proposition 3.4. Since \( s(y)_1 = k \sum_{j=1}^{q} |y|_j \) by definition of \( s(y) \), the result follows.

When \( \| \cdot \|_\infty \) is the \( \ell_\infty \)-norm, the norm \( \| \cdot \|_\infty \) associated with \( \text{conv}(N^K_{\| \cdot \|_\infty}) \) is known to be the \( K \)-support norm (or \( K \)-overlap norm). An explicit formula for this norm is introduced in [3]. We next provide an alternate derivation of this formula using our arguments. For consistency with literature, we denote the \( K \)-support norm by \( \| \cdot \|_{K}^p \).

**Lemma 3.2.** The unique integer \( r \in \{ 0, \ldots, K-1 \} \) that satisfies
\[
|x|_{K-r-1} > s(x)_{K-r} \geq |x|_{K-r},
\]
is \( r = K - i_x \) where \( |x|_{[0]} = \infty \) by convention.

**Proof.** Proof. This result follows from Proposition 3.4 and we refer to its parts directly in the proof. Let \( |x|_{i} \leq s(x)_i \) for some \( i \in \{ 0, \ldots, K \} \). We show that \( i = i_x \). First, we show that \( |x|_{i} \leq s(x)_i \) for all \( j \geq i \). Let \( j+1 \) be the first index no less than \( i \) so that \( |x|_{j+1} > s(x)_{j+1} \) and observe that, in fact, \( j+1 > i \). Then, we obtain the contradiction \( |x|_{j+1} \leq |x|_j \leq s(x)_j \leq s(x)_{j+1} < |x|_{j+1} \), where the third inequality is by Part 1. Therefore, \( |x|_{j} \leq s(x)_{j} \) for \( j \geq i \) which implies by Part 1 that \( s(x)_i \leq \cdots \leq s(x)_{K} \) and by Part 2 that \( i \geq i_x \). Since \( |x|_{i-1} > s(x)_i \), it follows by Part 1 that either \( i = 1 \) or \( s(x)_{i-1} > s(x)_i \). In either case, it follows that \( i \leq i_x \). Therefore, \( i = i_x = K - r \).

**Proposition 3.6** (Proposition 2.1 of [3]).
\[
\|x\|_{K}^p = \left( \sum_{i=1}^{K-1} |x|_i^2 \right) + \frac{1}{r+1} \left( \sum_{i=K-r}^{n} |x|_{i}^2 \right) \right) \frac{1}{2}.
\]
where \( r \) is the unique integer in \( \{ 0, \ldots, K-1 \} \) satisfying (21).

**Proof.** Proof. By Theorem 3.1,
\[
\|u(x)\|_2 = \left( \sum_{i=1}^{K-1} |x|_i^2 \right) + \frac{1}{K - i_x + 1} \left( \sum_{i=i_x}^{n} |x|_{i}^2 \right) \right) \frac{1}{2}.
\]
The result then follows since Lemma 3.2 establishes that \( r = K - i_x \).
3.1 Convexification of sets of matrices characterized by their singular values

Let $\mathcal{M}^{m,n}(\mathbb{R})$ be the set of $m \times n$ real-valued matrices. For $M \in \mathcal{M}^{m,n}(\mathbb{R})$, let $\sigma_1(M) \geq \cdots \geq \sigma_{q}(M)$ denote the singular values of $M$ where $q = \min\{m,n\}$ and let $\sigma : \mathcal{M}^{m,n}(\mathbb{R}) \to \mathbb{R}^q$ be defined as $\sigma(M) = (\sigma_1(M), \ldots, \sigma_q(M))$. Let $\|M\|_p = \sigma_1(M)$ and $\|M\|_c = \sum_{i=1}^{q} \sigma_i(M)$ be the spectral norm and the nuclear norm of $M$, respectively. In this subsection, we consider sets of matrices that are characterized by their singular values. More specifically, we are interested in sets of the form $S = \{ M \in \mathcal{M}^{m,n}(\mathbb{R}) \mid f_i(\sigma(M)) \leq 1, i = 1, \ldots, r \}$ and their convex hulls where each $f_i$ is a sign- and permutation-invariant function. Define $\bar{S} = \{ x \in \mathbb{R}^q \mid f_i(x) \leq 1, i = 1, \ldots, r \}$ where $q = \min\{m,n\}$. It is clear that $\bar{S} \subseteq S$ if and only if $\sigma(M) \in S$. As we show next, Theorem 2.4 implies that convex hulls of sets of the form $\bar{S}$ can be obtained by studying $\bar{S}$ instead. Similar results, although dealing with closed convex hulls, can be derived using conjugacy results of [19]. We include a direct proof based on Theorem 2.4.

**Theorem 3.3.** For $p \in \mathbb{Z}_{++}$ and $q \in \mathbb{Z}_+$ and each $i \in \{1, \ldots, r\}$, let $f_i : (x, z) \to \mathbb{R}$, where $(x, z) \in \mathbb{R}^p \times \mathbb{R}^q$, be sign- and permutation-invariant functions with respect to $x \in \mathbb{R}^p$. Let $m, n \in \mathbb{N}$, such that $\min\{m,n\} = p$, and define $\bar{S} = \{ (M, z) \in \mathcal{M}^{m,n}(\mathbb{R}) \times \mathbb{R}^q \mid f_i(\sigma(M), z) \leq 1, i = 1, \ldots, r \}$. Further, define $S = \{ (x, z) \in \mathbb{R}^p \times \mathbb{R}^q \mid f_i(x, z) \leq 1, i = 1, \ldots, r \}$. Then, $\text{conv}(\bar{S}) = X := \left\{ (M, z) \in \mathcal{M}^{m,n}(\mathbb{R}) \times \mathbb{R}^q \mid (\sigma(M), z) \in \text{conv}(S) \right\}$.

**Proof.** We first show that $\text{conv}(\bar{S}) \subseteq X$. Consider $(M, z) \in \text{conv}(\bar{S})$, so that $(M, z) = \sum_j \gamma_j(M_j, z')$, where $(M_j, z') \in \bar{S}$ and $\gamma_j$ are convex multipliers. For $k \in \{1, \ldots, p\}$ and any $Y \in \mathcal{M}^{m,n}(\mathbb{R})$, define $s_k(Y) := \sum_{i=1}^{k} \sigma(Y[i])$ to be the $k$th Ky Fan norm. Then, it follows by sublinearity and positive-homogeneity of norms that $s_k'(M) \leq \sum_j \gamma_j s_k'(M_j)$. In other words, $\sum_j \gamma_j \sigma(M_j) \geq \sum_{i=m}^{\infty} \sigma(M)$. Let $\sigma^i = \sigma(M_j)$. Since $(M_j, z') \in \bar{S}$, it follows that $f_i(\sigma^i, z') \leq 1$. Therefore, $\left(\sum_j \gamma_j \sigma^i, z\right) \in \text{conv}(S_0)$ where $S_0 = S \cap \{(\sigma, z) \mid \sigma_1 \geq \cdots \geq \sigma_p \geq 0\}$. Then, it follows by Part 3 of Theorem 2.4 that $(\sigma(M), z) \in \text{conv}(S)$ and $(M, z) \in X$.

We now show that $\text{conv}(\bar{S}) \supseteq X$. Let $(M, z) \in X$ and let $U \text{ diag}(\sigma)V^T$ be the singular value decomposition of $M$, where diag$(\sigma) \in \mathcal{M}^{m,n}(\mathbb{R})$ is the diagonal matrix, whose diagonal is the vector $\sigma$ and $\sigma_1 \geq \cdots \geq \sigma_p \geq 0$. Since $(\sigma, z) \in \text{conv}(S)$, it follows by Part 3 of Theorem 2.4 that there exist $\sigma' \in \mathbb{R}^p$, $(\sigma', z') \in S_0$ and convex multipliers $\gamma_j$ so that $(\sigma', z) = \sum j \gamma_j(\sigma', z')$ and $\sigma' \geq \sum_{i=m}^{\infty} \sigma(M)$. Now, if $\theta^i$ is obtained by permuting $\sigma^i$ or changing the sign of some of its entries, it follows readily that $(U \text{ diag}(\theta^i)V^T, z') \in \bar{S}$, because these operations do not alter the singular values of the matrix. Since $\sigma' \geq \sum_{i=m}^{\infty} \sigma(M)$, $\sigma = \sum \chi_k T_k \sigma^i$, where $\chi_k$ are convex multipliers and each $T_k$ permutes the entries of $\sigma'$ and possibly changes the sign of a few of the entries. Then, it follows that $(\sigma, z) = \sum_k \chi_k(T_k \sigma^i, z) = \sum \chi_k \gamma_j(T_k \sigma^i, z)$. Since $U \text{ diag}(\theta)V^T$ is a linear operator of $\theta$, $\sum \chi_k \gamma_j = 1$, and we have already shown that $(U \text{ diag}(T_k \sigma^i)V^T, z') \in \bar{S}$, it follows that $(M, z) \in \text{conv}(S)$.

In the following, we denote the set of $p \times p$ real symmetric matrices as $\mathcal{S}^p$ and, for any $M \in \mathcal{S}^p$, we denote the eigenvalues as $\lambda(M)$. In this context, a similar result can be shown using eigenvalues instead of singular values.

**Theorem 3.4.** For $p \in \mathbb{Z}_{++}$ and $q \in \mathbb{Z}_+$ and each $i \in \{1, \ldots, r\}$, let $f_i : (x, z) \to \mathbb{R}$, where $(x, z) \in \mathbb{R}^p \times \mathbb{R}^q$ be permutation-invariant functions with respect to $x \in \mathbb{R}^p$. Define $\bar{S} = \{ (M, z) \in \mathcal{S}^p \times \mathbb{R}^q \mid f_i(\lambda(M), z) \leq 1, i = 1, \ldots, r \}$. Further, define $S = \{ (x, z) \in \mathbb{R}^p \times \mathbb{R}^q \mid f_i(x, z) \leq 1, i = 1, \ldots, r \}$. Then, $\text{conv}(\bar{S}) = X := \{ (M, z) \in \mathcal{S}^p \times \mathbb{R}^q \mid (\lambda(M), z) \in \text{conv}(S) \}$.

**Proof.** We only provide a proof sketch since the proof is similar to that of Theorem 3.3. To show that $\text{conv}(\bar{S}) \subseteq X$, we consider $(M, z) = \sum \gamma_j(M_j, z') \in \text{conv}(\bar{S})$ and use the fact that $\sum \gamma_j \lambda(M_j) \geq \lambda(M)$ [15, Theorem 4.3.27]. Then, the result follows from Part 1 of Theorem 2.4. To show that $\text{conv}(S) \supseteq X$, we consider $(M, z) \in X$ and express $\sum \gamma_j(\lambda^i, z') \geq \lambda(M), z)$ for some $(\lambda^i, z') \in S$. Then, we observe that this implies that for any orthogonal matrix $U$ and a permutation $\pi$, $(U \text{ diag}(\pi(\lambda^i))U^T, z) \in \bar{S}$. The result is then derived in a manner similar to that in the proof of Theorem 2.4 except that instead of using the singular value decomposition $U \text{ diag}(\sigma)V^T$ of $M$, we use the eigenvalue decomposition $M = U\lambda(M)U^T$.

The rank of a matrix can be represented as the cardinality of the vector of singular values. Since cardinality is a sign- and permutation-invariant function, we obtain the following result as a special case of Theorem 3.3.

**Corollary 3.2.** Let $\bar{S} = \{ M \in \mathcal{M}^{m,n}(\mathbb{R}) \mid \text{rank}(M) \leq K, \|\sigma(M)\|_c \leq r \}$. Then,

$\text{conv}(\bar{S}) = \{ M \in \mathcal{M}^{m,n}(\mathbb{R}) \mid \|\sigma(M)\|_c \leq r \}$. 

11
Consider $\bar{S}$ in Corollary 3.2. Recall that determining if an arbitrary matrix $M \in \mathcal{M}^{m,n}(\mathbb{R})$ is in the convex hull $\text{conv}(\bar{S})$ can be easily done when the norm $\| \cdot \|$ is calculable. In particular, when $\| \cdot \|$ is the Euclidean norm, a given matrix $M$ is in $\text{conv}(\bar{S})$ if $\|\sigma(M)\|_2^2 \leq r$; see (22) for an explicit formula for $\| \cdot \|_K^2$. Semidefinite representability of the convex hull will be discussed in Section 3.2.

Next, we consider the special case where $\| \cdot \|$ is the $l_\infty$ norm. Proposition 3.5 and Theorem 3.3 together give an alternative proof for the following result.

**Proposition 3.7 (Theorem 1 of [14]).** Let $\bar{S} = \{ M \in \mathcal{M}^{m,n}(\mathbb{R}) \mid \text{rank}(M) \leq K, \|M\|_{sp} \leq r \}$. Then, $\text{conv}(\bar{S}) = \{ M \in \mathcal{M}^{m,n}(\mathbb{R}) \mid \|M\|_\infty \leq rK, \|M\|_{sp} \leq r \}$.

**3.2 Semidefinite-representability of sets of matrices characterized by their singular values**

We presented in Corollary 3.2 a convex hull result for a set of matrices $\bar{S}$ that is described using their singular values. The resulting convex hull is written in a norm $\| \cdot \|_e$ induced by the defining norm $\| \cdot \|_s$ of $\bar{S}$. In this subsection, we discuss the representability of this set as the feasible set of a semidefinite programming (SDP) problem. A set is called semidefinite-representable if it is a projection of a set expressed by a linear matrix inequality. We remark the following well-known results about semidefinite-representability; see Section 4.2 of [6].

**Lemma 3.3.** The following sets are semidefinite-representable:

1. The epigraph of the sum of $p$ largest singular values of a rectangular matrix.
2. The epigraph of the sum of $p$ largest eigenvalues of a symmetric matrix.
3. The graph of the sum of all eigenvalues of a symmetric matrix.
4. The set $A \cap B$, where $A$ and $B$ are semidefinite-representable.

In particular, we consider the set: $S = \{ M \in \mathcal{M}^{m,n}(\mathbb{R}) \mid \text{rank}(M) \leq K, f(\sigma(M)) \leq r \}$ where $q = \min\{m, n\}$ and $f: \Delta_+ \to \mathbb{R}$ is a quasiconvex function. A function $f$ is said to be quasiconvex on $\Delta_+$ if, for $\lambda \in [0, 1]$ and $x, y \in \Delta_+$, we have $f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}$. We assume this function has semidefinite-representable lower-level sets. Then, we show that the convex hull of $S$, the set of rank-constrained matrices whose singular values belong to a lower-level set of $f$ are semidefinite-representable.

**Theorem 3.5.** Let $q = \min\{m, n\}$ and $\bar{S} = \{ M \in \mathcal{M}^{m,n}(\mathbb{R}) \mid \text{rank}(M) \leq K, f(\sigma(M)) \leq r \}$, where $f: \Delta_+ \to \mathbb{R}$ has semidefinite-representable lower-level sets. Then, $\text{conv}(\bar{S})$ is semidefinite-representable.

**Proof.** Define $g: \mathbb{R}^q \to \mathbb{R}$ so that $g(x) = f(\sigma_{\Delta_+}(x))$. Observe further that $g$ is sign- and permutation-invariant. Then, let $S = \{ x \in \mathbb{R}^q \mid \text{card}(x) \leq K, g(x) \leq r \}$. By sign- and permutation-invariance of $S$ and Theorem 2.4,

$$\text{conv}(S) = \left\{ x \in \mathbb{R}^q \mid f(u) \leq r, \begin{array}{l} u_1 \geq \cdots \geq u_K \geq 0, \\ u_{K+1} = \cdots = u_n = 0, \\ u \geq_{\text{perm}} |x| \end{array} \right\}.$$ 

Therefore, by Theorem 3.3,

$$\text{conv}(\bar{S}) = \left\{ M \in \mathcal{M}^{m,n}(\mathbb{R}) \mid f(u) \leq r, \begin{array}{l} u_1 \geq \cdots \geq u_K \geq 0, \\ u_{K+1} = \cdots = u_n = 0, \\ u \geq_{\text{perm}} \sigma(M) \end{array} \right\}.$$ 

By the definition of weak majorization, the convex hull has the following representation:

$$\begin{cases} f(u) \leq r, \\ u_1 \geq \cdots \geq u_K \geq 0, \\ u_{K+1} = \cdots = u_n = 0, \\ \sum_{i=1}^j u_i \geq \sum_{i=1}^j \sigma_j(M), \quad j = 1, \ldots, K. \end{cases}$$

The semidefinite-representability of (23) follows from Lemma 3.3 and the semidefinite-representability of the level set $\{ u \mid f(u) \leq r \}$ and the introduced linear inequalities.

Although Theorem 3.5 is similar to Proposition 4.2.2 in [6], we discuss next how these results differ. First, we introduce a rank constraint and thus treat a nonconvex set. Second, we discuss the representation of the convex hull
rather than the set itself. Third, we do not require monotonicity of $f(x)$ and require semidefinite representability only over $\Delta_+ (= \Delta_+^q)$ instead of $\mathbb{R}_+^q$. We briefly describe why the added assumptions are required in Proposition 4.2.2 in [6], but not in our result. This is because, when $f(x)$ is not monotone but is quasiconvex over $\Delta_+$, its extension to $\mathbb{R}_+^q$ defined using $g(x) := f(x_{\Delta_+})$ is not necessarily quasiconvex. As such, the lower-level sets of $g(x)$ are not always semidefinite representable. To see this, consider $f(x) = 1 - x$, where $x \in \mathbb{R}$. Then, $\{ x \mid g(x) \leq 0 \}$ is not a convex set because $g(1) = g(-1) = 0$, while $g(0) = 1$. On the contrary, consider an $f(x)$ that is monotone, permutation-invariant and quasiconvex over $\mathbb{R}_+^q$. Let $C = \{ x \mid g(x) \leq r \}$. We argue that $C$ is convex and can be expressed as $X = \{ x \mid f(u) \leq r, u \in \Delta_+, u \geq \text{em} \mid |x| \}$. First, observe that Theorem 2.4 shows that $X = \text{conv}(C) \supseteq C$. Now, we argue that $X \subseteq C$. To see this, assume $x \in X$. Then, $x_{\Delta_+} \in X$ because $x$, being $\text{conv}(C)$ inherits the sign and permutation-invariance of $C$. Then, there exist $u$ and $u'$ such that $u \geq u' \geq x_{\Delta_+}$ and $f(u) \leq r$. Therefore, we have $g(x) = f(x_{\Delta_+}) \leq f(u') \leq f(u) \leq r$, where the first inequality is from monotonicity of $f$. The second inequality is because $u$ and $u'$ are non-negative, $u'$ is in the convex hull of $u$ and its permutation-variants, and $f$ is quasiconvex and permutation-invariant. The third inequality follows by definition of $u$. Therefore, it follows that $x \in C$.

**Corollary 3.3.** Let $S = \{ M \in M^{m,n}(\mathbb{R}) \mid \text{rank}(M) \leq K, \| \sigma(M) \|_s \leq r \}$ where $\| \cdot \|_s$ is a permutation-invariant monotone norm. Then, $\text{conv}(S)$ is semidefinite-representable. In particular, when $\| \sigma(\cdot) \|_s$ is a Ky Fan $p$-norm (the sum of $p$ largest singular values) for some $p = 1, \ldots, \min\{m, n\}$, the convex hull is semidefinite-representable.

Similarly, we can prove the following result, where $S^q_+ \in M^{q \times q}(\mathbb{R})$ is the set of positive semidefinite symmetric matrices.

**Theorem 3.6.** Let $\bar{S} = \{ M \in S^q_+ \mid \text{rank}(M) \leq K, f(\lambda(M)) \leq r \}$, where $f : \Delta_+ \to \mathbb{R}$ has semidefinite-representable lower-level sets. Then, $\text{conv}(\bar{S})$ is semidefinite-representable.

**Proof.** Define $g : \mathbb{R}_+^q \to \mathbb{R}$ so that $g(x) = f(x_{\Delta_+})$. Then, let $S = \{ x \in \mathbb{R}_+^q \mid \text{card}(x) \leq K, g(x) \leq r \}$. By permutation-invariance of $S$ and Theorem 2.4,

$$\text{conv}(S) = \left\{ x \in \mathbb{R}_+^q \mid \begin{array}{l}
f(u) \leq r, \\
u_1 \geq \cdots \geq u_K \geq 0, \\
u_{K+1} = \cdots = u_n = 0, \\
u \geq_m x
eq \end{array} \right\}.$$

Therefore, by Theorem 3.4,

$$\text{conv}(\bar{S}) = \left\{ M \in S^q \mid \begin{array}{l}
f(u) \leq r, \\
u_1 \geq \cdots \geq u_K \geq 0, \\
u_{K+1} = \cdots = u_n = 0, \\
u \geq_m \lambda(M)
eq \end{array} \right\}.$$

By the definition of majorization, the convex hull has the following representation:

$$\left\{\begin{array}{l}
f(u) \leq r, \\
u_1 \geq \cdots \geq u_K \geq 0, \\
u_{K+1} = \cdots = u_n = 0, \\
\sum_{i=1}^j u_i \geq \sum_{i=1}^j \lambda_j(M), \quad j = 1, \ldots, K-1, \\
\sum_{i=1}^K u_j = \sum_{i=1}^K \lambda_j(M).
\end{array}\right\} \tag{24}$$

The semidefinite-representability of (24) follows from Lemma 3.3 and the semidefinite-representability of the level set $\{ u \mid f(u) \leq r \}$ and the introduced linear inequalities. \hfill \Box

The ideas in the above proof can be extended, using disjunctive programming techniques, to symmetric matrices, that are not necessarily positive semidefinite. Since the eigenvalues are no longer non-negative, we cannot impose the restriction that $u \geq 0$ and, thus, assume that $u_{K+1}, \ldots, u_n = 0$. Instead, we express the rank constraint as a union of sets each of which satisfies $u_{a+1} = \cdots = u_{a+n-K} = 0$ for some $a \in \{0, \ldots, K\}$. Then, we obtain $\text{conv}(S)$ as the convex hull of a union of semidefinite representable sets. Using the disjunctive programming argument of Proposition 3.3.5 in [6], this yields a lifted representation of a set that outer-approximates $\text{conv}(S)$ and is contained in $\text{cl} \text{conv}(S)$.

### 4 Convex envelopes of nonlinear functions

The problem of finding convex envelopes of nonlinear functions is central to the global solution of factorable problems through branch-and-bound. When the domain over which the envelope is constructed is a polytope $P$, it is often the
case that the envelope is completely determined by the values that the function takes on a subset of the faces of \( P \), or more generally, a subset of its feasible points. If the above property holds, disjunctive programming techniques can often be employed to provide an explicit (although typically large) description of the envelope, through the introduction of new variables for each of the important subsets of \( P \). In this section, using Theorem 2.4 as a foundation, we show that for certain functions defined over permutation-invariant polytopes (i) envelopes can be built without recourse to disjunctive programming (Proposition 4.1), and (ii) polynomially-size disjunctive programming formulations can be constructed even when the number of faces of \( P \) important in the construction of the envelope is exponential (Theorem 4.1). These results yield compact envelopes descriptions for specific families of functions (Propositions 4.2 and 4.3). We also provide numerical evidence that the use of these techniques produces relaxations of multilinear functions over permutation-invariant hypercubes that are significantly stronger than those obtained using a recursive application of McCormick’s procedure. The techniques can be extended to handle epigraphs of singular values/eigenvalues of matrices using the ideas presented in Sections 3.1 and 3.2.

**Definition 4.1.** A function \( \phi : C \to \mathbb{R} \) is said to be Schur-concave on \( C \), if for every \( x, y \in C \), \( x \geq_{m} y \) implies that \( \phi(x) \leq \phi(y) \).

Various functions have been shown to be Schur-concave including the Shannon entropy \( \sum_{i=1}^{n} x_i \log\left(\frac{x_i}{\bar{x}}\right) \) and elementary symmetric functions \( \sum_{j \subseteq \{1, \ldots, n\} : |j| = k} \prod_{i \in j} x_i \). Symmetric concave functions are also Schur-concave. More complex Schur-concave functions can be constructed from known Schur-concave functions using some compositions or operations that preserve Schur-concavity; see Chapter 3 of [25] for further detail.

In this section, for any function \( \phi : C \to \mathbb{R} \) we denote the convex envelope of \( \phi \) over \( C \) by \( \text{conv}_C(\phi) \). A common tool in the construction of convex envelopes is to restrict the domain of the function to a smaller subset. We say that a function \( \phi : C \to \mathbb{R} \) can be restricted to \( X \), where \( X \subseteq C \), for the purpose of obtaining \( \text{conv}_C(\phi) \) if \( \text{conv}_C(\phi|_X) = \text{conv}_X(\phi) \), where \( \phi|_X(x) \) is defined as \( \phi(x) \) for any \( x \in X \) and \( +\infty \) otherwise.

First, we establish in Lemma 4.1 that, when deriving the envelope of a Schur-concave function over a permutation-invariant polytope, it is sufficient to restrict attention to those points in the domain that are not majorized by other feasible points. When coupled with a simple domain structure, this result permits a description of convexification results without the use of majorization inequalities, in a smaller dimensional space.

**Lemma 4.1.** Let \( \phi : P \to \mathbb{R} \) be a Schur-concave function, where \( P \subseteq \mathbb{R}^n \) is a permutation-invariant polytope. Let \( M := \{x \in P \mid \exists u \in P \text{ and } u \geq_{m} x \text{ and } u_{\Delta} \neq x_{\Delta} \} \). Let \( S := \{(x, \phi) \mid \phi(x) \leq \phi \leq \alpha, x \in P \} \) and \( X := \{(x, \phi) \mid \phi(x) \leq \phi \leq \alpha, x \in M \} \). Then, \( \text{conv}(S) = \text{conv}(X) \).

**Proof.** Since \( M \subseteq P \) it follows that \( X \subseteq S \) and, therefore, \( \text{conv}(X) \subseteq \text{conv}(S) \). Now, consider \( (x', \phi') \in S \setminus X \). Therefore, \( \phi(x') \leq \phi \leq \alpha \) and \( x' \in P \setminus M \). Let \( x''_i = \frac{1}{n} \sum_{i=1}^{n} x'_i \) for all \( i \in \{1, \ldots, n\} \). That is, all components of \( x'' \) are identical. Let \( u' := \arg \max \{\|u - x''\| \mid u \geq_{m} x', u \in P \} \). The maximum is achieved because the feasible set is compact and because the objective is upper-semicontinuous. Assume by contradiction that there exists \( u' \in P \) such that \( y' \geq_{m} u' \) and \( y'_{\Delta} \neq u'_{\Delta} \). Since \( u' \) can be written as a convex combination of at least two permutations of \( y' \) and the objective of the problem defining \( u' \) is permutation-invariant and strictly convex, it follows that \( \|y' - x''\| > \|u' - x''\| \) violating the optimality of \( u' \). Therefore, there does not exist \( u' \in P \) such that \( y' \geq_{m} u' \) and \( y'_{\Delta} \neq u'_{\Delta} \). In other words, \( u' \in M \). That is, for any permutation matrix \( Q \in \Sigma_n \), \( Q u' \in M \). Since \( \phi \) is Schur-concave, then \( \phi(Q u') = \phi(u') \leq \phi(x') \leq \phi \leq \alpha \). Therefore, \( (Q u', \phi') \in X \). Finally, since \( x' \leq_{m} u' \), \( x' \) can be written as a convex combination of permutations of \( u' \). Therefore, \( (x', \phi') \in \text{conv}(X) \). We conclude that \( S \subseteq \text{conv}(X) \).

Lemma 4.1 requires the identification of the set \( M \), which consists of points in \( P \) which are not expressible as a convex combination of another point in \( P \) and its permutations. In Lemma 4.2, we characterize such points as those which have a supporting hyperplane of a specific form. Later, we discuss how these results can be combined to obtain, in closed-form, convex envelopes of various Schur-concave functions.

**Lemma 4.2.** Let \( x' \in P \), where \( P \) is a permutation-invariant polytope. Let \( \pi \) be a permutation of \( \{1, \ldots, n\} \) such that for each \( i \in \{1, \ldots, n - 1\} \), \( x'_{\pi(i)} \geq x'_{\pi(i+1)} \). Then, there exists \( u' \in P \) with \( u' \geq_{m} x' \) and \( u'_{\Delta} \neq x'_{\Delta} \) if and only if there does not exist \( a \in \mathbb{R}^n \) such that \( a_{\pi(i)} > a_{\pi(i+1)} \) for all \( i \in \{1, \ldots, n - 1\} \) and \( \sum_{i=1}^{n} a_{i}(x_i - x'_i) \leq 0 \) is valid for \( P \).

**Proof.** We first show that if such \( a \) exists, there does not exist \( u' \in P \) such that \( u' \geq_{m} x' \) and \( u'_{\Delta} \neq x'_{\Delta} \). Assume by contradiction that such a \( u' \) exists. Because \( P \) is permutation-invariant we may assume that \( u'_{\pi(i)} \geq u'_{\pi(i+1)} \) for all \( i \in \{1, \ldots, n - 1\} \) by sorting \( u' \) if necessary. Since \( u' \geq_{m} x' \) and \( u' \neq x' \), there exists \( y' \neq 0 \),
\( k \in \{1, \ldots, n - 1\} \), and \( r \in \{1, \ldots, n - k\} \) such that \( y' \geq m' \), \( y' \geq m' x' \), \( y'_\pi(k) = x'_\pi(k) + \theta \), and \( y'_\pi(k+r) = x'_\pi(k+r) - \theta \) and \( y'_\pi(x) = x'_\pi(x) \) otherwise; see Lemma 2.1 in [25]. Since \( u' \in \mathcal{P} \), \( \mathcal{P} \) is convex and permutation-invariant, and since \( y' \) can be written as a convex combination of \( u' \) and its permutations, it is clear that \( y' \in \mathcal{P} \). Therefore, \( \sum_{i=1}^{n} a_i (y'_i - x'_i) \leq 0 \) or \( a_{\pi(k)} - a_{\pi(k+r)} \leq 0 \). This is a contradiction to the assumed ordering of \( a \).

Now, we show that if there does not exist \( u' \in \mathcal{P} \) such that \( u' \geq m' x' \) and \( u'_\Delta \neq x'_\Delta \), we can construct such a vector \( a \).

Define a polyhedral cone \( K := \{ \sum_{i=1}^{n} \alpha_i (e_\pi(i) - e_{\pi(i+1)}) \mid \alpha \geq 0 \} \) where \( e_i \) represents the \( i \)-th standard basis. Observe that \( \nu \in K \) (equivalently, \( \nu \geq K \)) if and only if \( \sum_{i=1}^{k} \nu_i \geq 0 \), \( \nu_i \geq 0 \) for all \( k \). More specifically, given \( \nu \in \mathbb{R}_+^n \), let \( S_k^\nu \) be the partial sums so that \( S_k^\nu \geq K \) if and only if \( S_k^\nu \) is non-negative for every \( k \) and \( S_k^\nu \geq 0 \).

To verify that the latter set is contained in \( K \), consider any \( \nu \) in \( K \) such that \( \nu - \sum_{i=1}^{n} \nu_i = \sum_{i=1}^{n} \nu_i \geq 0 \), \( \nu \in K \) implies that \( \nu \geq 0 \). We now construct the polyhedron \( C := K - \{ x' \} \), where the difference is the Minkowski difference. Since \( x' \in \mathcal{P} \), \( \mathcal{P} \) is component-wise non-convex. This follows because there exists a point in \( \nu \in \mathcal{P} \) so that \( \nu \geq K \) such that there is a \( k \) for which \( S_k^\nu \geq 0 \). This suffices because, if there exists \( u = u - v \in \mathcal{P} \), where \( u \in \mathcal{P} \) and \( v \in \mathcal{P} \). Then, we have \( u \geq K \) \( v \geq K \), which in turn proves the existence of \( u \) that majorizes \( x' \). Furthermore, by showing \( S_k^\nu \) is non-negative for some \( k \), we have \( u'_\Delta \neq x'_\Delta \), which contradicts the assumption that such \( u \) does not exist.

Assume that the inequality is not strict for some \( k \in \{1, \ldots, n - 1\} \) so that \( \langle a', x - x' \rangle \leq 0 \). Then, for \( \epsilon > 0 \), we define \( w_\epsilon = \epsilon (e(x) - e(x') \in K \) and show that, for a sufficiently small \( \epsilon \), \( w_\epsilon \in \mathcal{P} \). Since \( w_\epsilon \in \mathcal{P} \), \( \mathcal{P} \) is the relative interior of \( F \), it follows that there exists an \( \epsilon \) such that \( w_\epsilon \in \mathcal{P} \). Moreover, \( S_k^\nu \epsilon \in \mathcal{P} \) if \( \epsilon > 0 \). As we argued above, the existence of such \( w_\epsilon \) contradicts our assumption and, so, we conclude that \( a'_{\pi(k+1)} - a'_{\pi(k)} \leq 0 \) for all \( i \in \{1, \ldots, n - 1\} \). 

Lemmas 4.1 and 4.2 can be combined to develop convex envelopes of Schur-concave functions. This is because, taken together, they prove that it suffices to restrict attention to a subset of \( M \in \mathcal{P} \) in order to construct the convex envelope. To better understand the structure of \( M \) and to illustrate potential applications, we derive in Proposition 4.1 the closed-form description of the convex envelope of a Schur-concave function over \([a, b]^n\). In this case, \( M \) is contained in the one-dimensional faces of \([a, b]^n\), the key insight that allows for the derivation of the closed-form. Although we provide a self-contained argument for this fact in the proof of Proposition 4.1, this inclusion can be seen as a special case of Lemma 4.2, a visualization which serves to illustrate the use of Lemma 4.2 in characterizing \( M \). To see this special case, observe that \( x' \in M \cap \Delta^n \) only if there is an inequality \( \langle \beta, x - x' \rangle \leq 0 \) valid for \([a, b]^n\) such that \( \beta_1 > \cdots > \beta_n \). Since such an inequality is tight at \( x' \), it can be derived as a conic combination of facet-defining inequalities tight at \( x' \). The facet-defining inequalities of \([a, b]^n\) are \( -x_i \leq -a \) and \( x_i - b \leq 0 \) for \( i = 1, \ldots, n \). Let \( F(x') \) be the set of facet-defining inequalities tight at \( x' \), and for any facet-defining inequality in this set, say \( \langle a, x - x' \rangle \leq 0 \), let \( L_a = \{ (t, t+1) \mid a_i > a_{i+1} \} \). It is easy to see that, for \( x' \in [a, b]^n \), \( L_a \leq 1 \) for each \( a \in F(x') \), i.e. each \( L_a \) contains at most one point. Then, for \( \beta \) to be derived as a conic combination of inequalities in \( F(x') \), it must be that \( \bigcup_{a \in F(x')} L_a \leq 1 \). Since \( L_a \leq 1 \), it follows that \( |F(x')| \geq n - 1 \). Therefore, \( M \) is a subset of one-dimensional faces of the hypercube. Moreover, a similar argument shows that if there exists a \( k \in \{1, \ldots, n - 1\} \) such that \( |L_a| \leq k \), then \( |F(x')| \geq \left[ \frac{n-1}{k} \right] \), and, consequently, \( M \) is a subset of \( n - \left[ \frac{n-1}{k} \right] \) faces of \( P \).

**Proposition 4.1.** Consider a function \( \phi(x) : \mathbb{R}^n \to \mathbb{R} \) that is Schur-concave over \([a, b]^n\) and let \( S' := \{ (x, \phi) \mid \phi(x) \leq \phi, x \in [a, b]^n \} \). For any \( x \in [a, b]^n \), define \( S(x) = \sum_{i=1}^{n} (x_i - a_i) \). For any \( s \in [0, n(b-a)] \), let \( i^s = \max \{ i \mid i(b-a) < s \} \) and \( u^s = \begin{cases} b & \text{if } i \leq i^s \\ a + s - (b-a)i^s & \text{if } i = i^s + 1 \\ a & \text{otherwise.} \end{cases} \) (25)

Let \( \Theta^s := \{ (x, \phi) \mid \phi(u^s(x)) \leq \phi, x \in [a, b]^n \} \). Then, \( \text{conv}(S') = \text{conv}(\Theta^s) \). Moreover, if \( \phi \) is component-wise convex then \( \Theta^s \) is convex.

**Proof.** We first show that \( u^s(x') \geq m \) \( x' \) for each \( x' \in [a, b]^n \). This follows because \( u^s(x') \) simultaneously maximizes the continuous knapsack problems \( \max \{ \sum_{i=1}^{n} x_i \mid \sum_{i=1}^{n} x_i = S(x') + na, x \in [a, b]^n \} \) for all \( j \) because the ratio of objective and knapsack coefficient of \( x_i \) reduces with increasing \( i \), and \( x' \) is a feasible solution to these knapsack problems.
We now show that $S^\alpha \subseteq \Theta^\alpha$. Let $(x', \phi') \in S^\alpha$. Therefore, $\phi(u^{S(x')}) \leq \phi(x') \leq \phi \leq \alpha$, where the first inequality follows from Schur-concavity of $\phi$ and $u^{S(x')} \geq_m x'$, and the remaining inequalities follow because $(x', \phi')$ is feasible to $S^\alpha$. Therefore, $(x', \phi') \in \Theta^\alpha$.

Now, we show that $\Theta^\alpha \subseteq \text{conv}(S^\alpha)$. Let $(x', \phi') \in \Theta^\alpha$. Since $u^{S(x')} \in [a, b]^n$ and $\phi(u^{S(x')}) \leq \phi' \leq \alpha$, we conclude that $(u^{S(x')}, \phi') \in S^\alpha$. Then it follows that $(x', \phi') \in \text{conv}(S^\alpha)$ since $u^{S(x')} \geq_m x'$ implies that $x'$ can be written as a convex combination of permutations of $u^{S(x')}$ and $S^\alpha$ is permutation-invariant in $x$.

To show that $\Theta^\alpha$ is convex when $\phi$ is component-wise convex, we write $\Theta^\alpha$ as $\text{proj}_{(x, \phi)} \Xi^\alpha$, where $\Xi^\alpha = \{(x, \phi) \mid \phi(s) \leq \alpha, x \in [a, b]^n, s = \sum_{i=1}^n (x_i - a)\}$ and $\phi(s) = \phi(u^s)$. The result follows if $\phi(s)$ is convex over $[0, n(b - a)]$ since $\Theta^\alpha$ is expressed as the projection of the convex set $\Xi^\alpha$. First, observe that, for $s \in (i(b - a), (i + 1)(b - a))$, the convexity of $\phi(s)$ follows from the assumed convexity of $\phi(u^s)$ since, in this interval, $u^s$ varies only along the $i$th coordinate. Now, choose $k \in \{0, \ldots, n - 1\}$ and let $\bar{s} = k(b - a)$. To prove the result, it suffices to verify that the left derivative of $\phi(s)$ at $\bar{s}$ is no more than the corresponding right derivative. For any $\epsilon > 0$, observe that $u^s + \epsilon e_k \geq_m u^s + \epsilon e_{k+1}$. This follows because $(1 - \lambda)(u_{k-1}^s + \epsilon, u_{k+2}^s) + \lambda(u_{k+1}^s, u_{k+1}^s + \epsilon) = (u_k^s, u_{k+1}^s + \epsilon)$, where $0 < \lambda = \frac{\epsilon}{v_{k-1}^s - u_{k+1}^s + \epsilon} \leq 1$ showing that $u^s + \epsilon e_{k+1}$ can be expressed as a convex combination of $u^s + \epsilon e_k$ and its permutations. Since $\phi(\cdot)$ is Schur-concave, it follows that $\phi(u^s + \epsilon e_k) \leq \phi(u^s + \epsilon e_{k+1}) = \varphi(\bar{s} + \epsilon)$. Then, the following chain of inequalities follows

$$
\lim_{\epsilon \downarrow 0} \frac{\varphi(\bar{s}) - \varphi(\bar{s} - \epsilon)}{\epsilon} = \lim_{\epsilon \downarrow 0} \frac{\phi(u^s) - \phi(u_{\bar{s}} - \epsilon)}{\epsilon} \leq \lim_{\epsilon \downarrow 0} \frac{\phi(u^s + \epsilon e_k) - \phi(u^s)}{\epsilon} \leq \lim_{\epsilon \downarrow 0} \frac{\varphi(\bar{s} + \epsilon) - \varphi(\bar{s})}{\epsilon},
$$

where the first equality is by the definition of $\varphi(\cdot)$ and $u^s$, the first inequality is from the assumed convexity of $\phi(\cdot)$ when the argument is perturbated only along the $k$th coordinate, and the second inequality is because $\phi(u^s + \epsilon e_k) \leq \varphi(\bar{s} + \epsilon)$ and $\phi(u^s) = \varphi(\bar{s})$. \hfill \Box

In essence, Proposition 4.1 shows that we can reduce our attention to the edges of the hypercube belonging to $\Delta^n$ in our construction of $\text{conv}(S^\alpha)$; see Figure 1 for an illustration when $a = 2$ and $b = 5$. A similar result can be shown for upper level sets of quasiconcave functions over general polytopes [32]. Symmetric quasiconcave functions are a subclass of Schur-concave functions. In other words, both the results show that for symmetric quasiconcave functions over permutation-invariant polytopes it suffices to consider the edges of the polytope to construct the convex hull. However, the result in [32] applies to general quasiconcave functions over arbitrary polytopes while Proposition 4.1
applies to Schur-concave functions over a hypercube. Perhaps more importantly, the result in Proposition 4.1 also applies to level sets of the functions while the result in [32] only applies to convex envelope construction.

Permutation-invariance also helps with constructing compact extended formulations of certain nonlinear sets. To motivate this statement and introduce the following result, consider the set \( Z' = \{(x,z) \in [a,b]^n \times \mathbb{R} \mid z = \prod_{i=1}^{n} x_i \} \). It is well-known that, in order to generate \( \text{conv}(Z') \), it suffices to restrict \( x \) to the vertices, \( F' = \{a,b\}^n \), of the hypercube \([a,b]^n\). More precisely, \( \text{conv}(Z') = \{ \bigcup_{x \in F'}(x,\prod_{i=1}^{n} x_i) \} \), where each disjunct in the union is a polytope with compact description (a single point.) A higher-dimensional description of the convex hull of this union can be obtained using classical disjunctive programming results. Because the dimension of this formulation depends linearly on the number of disjuncts, \( |F'| \), such an approach has found limited practical use for the given example, as \( |F'| \) is exponential in \( n \). Surprisingly, taking advantage of the permutation-invariance of \( Z' \) through Theorem 2.1 allows for a much more economical use of disjunctive programming. Intuitively, this is because the number of elements of \( F' \) required to compute \( \text{conv}(S_0) \) in Theorem 2.1, is polynomial in \( n \). As a result, disjunctive programming provides a polynomial formulation for \( \text{conv}(S_0) \) which can then be integrated with the result of Theorem 2.1 to obtain a polynomial-sized higher-dimensional formulation of \( \text{conv}(Z') \). The settings where such polynomially-sized formulations can be obtained extend far beyond the example presented above as we describe next in Theorem 4.1.

Consider now the much more general setup of sets \( S(Z,a,b) := \{(x,z) \in [a,b]^n \times \mathbb{R}^m \mid (x,z) \in Z \} \), where \( Z \) is compact and permutation-invariant in \( x \). Further, for \( F = \{F_1,\ldots,F_r\} \), where \( F_i \) are faces of \([a,b]^n\), define \( X(Z,a,b,F) := \{(x,z) \in [a,b]^n \times \mathbb{R}^m \mid (x,z) \in Z, x \in \bigcup_{i=1}^{r} F_i \} \). Using Theorem 2.1, we show next that a polynomial-size extended formulation can be constructed for any set \( S(Z,a,b) \) for which there exists a collection of faces \( F' \) that completely determines the convex hull, i.e., \( \text{conv}(S(Z,a,b)) = \text{conv}(X(Z,a,b,F')) \) and for which a polynomial (possibly extended) formulation of the set on each of these faces \( F_i \in F' \), i.e., \( \text{conv}(X(Z,a,b,\{F_i\})) \), can be obtained. The strength of this result is that we make no assumption on \( |F'| \) and that this collection may have exponentially many faces.

**Theorem 4.1.** Let \( a,b \in \mathbb{R}, Z \subseteq \{(x,z) \mid x \in \mathbb{R}^n, z \in \mathbb{R}^m \} \) be a compact permutation-invariant set with respect to \( x \), and \( F = \{F_1,\ldots,F_r\} \) be a collection of faces of \([a,b]^n\) such that \( \text{conv}(S(Z,a,b)) = \text{conv}(X(Z,a,b,F)) \). Moreover, assume that \( \text{conv}(X(Z,a,b,\{F_i\})) \) has a polynomial-sized compact extended formulation for each \( F_i \in F \). Then, \( \text{conv}(S(Z,a,b)) \) has a polynomial-sized extended formulation.

**Proof.** Proof. For brevity of notation, in this proof, we shall write \( S(Z,a,b) \) as \( S \) and \( X(Z,a,b,F) \) as \( X(F) \). We construct \( \text{conv}(S) \) using its equivalence to \( \text{conv}(X(F)) \). We may assume for computing \( \text{conv}(X(F)) \), by taking the union of all permutations of \( X(F) \) with respect to \( x \) if necessary, that \( X(F) \) is permutation-invariant in \( x \). This is because a permutation of \( X(F) \) with respect to \( x \), say \( X_\pi(F) := \{(x,z) \mid (\pi(x),z) \in X(F)\} \), is contained in \( \text{conv}(X(F)) \) as is seen from \( X_\pi(F) \subseteq \bigcup_{\pi} S = \text{conv}(X(F)) \), where the first inclusion is by permutation-invariance of \( S \) and the equality is by the assumed hypothesis. Since \( S \) is permutation-invariant with respect to \( x \), by Lemma 2.1, \( \text{conv}(S) \) is also permutation-invariant. We shall use Theorem 2.1 to construct \( \text{conv}(X(F)) \). We first show that we can limit the faces of \([a,b]^n\) that need to be considered in the construction of \( S_0 \); see Theorem 2.1. Consider an arbitrary face \( F_i \) of \([a,b]^n\), which is determined by setting a set of variables with indices in \( B_i \subseteq \{1,\ldots,n\} \) to their upper bound \( b \) and a disjoint set of variables \( A_i \subseteq \{1,\ldots,n\} \) to their lower bound \( a \). We will show that the only faces, \( F_i, i = 1,\ldots,r \) that need to be considered are such that \( B_i \) and \( A_i \) are hole-free, i.e., \( B_i \) is of the form \( \{1,\ldots,p\} \) and \( A_i \) is of the form \( \{q,\ldots,n\} \). To see this, let \( j(i) = \max\{j \mid j \in B_i\} \) and \( X_i = X(F_i) \cap \{(x,z) \mid x_1 \geq \cdots \geq x_n\} \). It follows from Theorem 2.1 and permutation-invariance of \( X(F) \) that it suffices to consider points in \( \bigcup_{i=1}^{r} X_i \) to construct \( \text{conv}(X(F)) \). We will further argue that it suffices to restrict attention to points in \( \bigcup_{j(i) = |B_i|} X_i \). Notice that \( B_i \) is hole-free if and only if \( j(i) = |B_i| \). Assume by contradiction that \( i' \) is the index of a face such that \( j(i') > |B_{i'}| \) and that \( X_{i'} \) contains a point which is not in \( \bigcup_{j(i) = |B_i|} X_i \). Among all such faces, we choose \( i' \) to minimize \( j(i) - |B_i| \). Since \( B_{i'} \) is not hole-free, there exists \( j \notin B_{i'} \) such that \( j < j(i') \). Any point that belongs to \( X_{i'} \) must satisfy \( b \geq x_j \geq x_{j(i')} = b \). Therefore, \( x_j = b \). Since \( X_{i'} \neq \emptyset, j \notin A_i \). Consider now \( i'' \) such that \( B_{i''} = B_i \cup \{j\} \setminus \{j(i')\} \) and \( A_{i''} = A_i \setminus \{j\} \). Such a face exists in \( F \) since we assumed that for every face \( F_i \in F \), \( X_i \) contains all faces obtained by permuting the variables, and \( F_{i''} \) is obtained from \( F_{i''} \) by exchanging the variables \( x_j \) and \( x_{j(i')} \). We next show that \( X_{i''} \supseteq X_{i'} \). For arbitrary \( (x,z) \in X_{i''}, x_i = b \) for all \( i \in B_{i''} \). Then, it follows that \( x_i = b \) for all \( i \leq j(i') \) as \( x_j \geq \cdots \geq x_{j(i')} \). Therefore, \( x_i = b \) for all \( i \in B_{i''} \). Since \( A_{i''} = A_{i'} \) and \( x_j \) is not restricted for other indices \( j, (x,z) \in X_{i''} \). Therefore, \( X_{i''} \) contains a point not in \( \bigcup_{j(i) = |B_i|} X_i \), establishing that \( j(i'') > |B_{i''}| \). However, since \( j(i'') - |B_{i''}| < j(i') - |B_i| \), this contradicts our choice of \( i' \). A similar argument can be used to show that we only need to consider faces \( F_i \) such that \( A_i \) is hole-free.
There are at most \( \binom{n+2}{2} \) such faces, one for each choice of \((p, q)\), where \(0 \leq p \leq q - 1 \leq n\). Since each \(\text{conv} \{F_i\}\) is assumed to have a polynomial-sized compact extended formulation, it follows, by disjunctive programming [Proposition 2.3.5 in 6], that \(\text{conv}(S_0)\) has a polynomial-sized compact extended formulation. The result then follows directly from Theorem 2.1.

We record and summarize the extended formulation of \(\text{conv}(S(Z, a, b))\) for later use in Corollary 4.1. We also observe that our construction applies even when \(Z\) is not compact, as long as the convex hull of \(X(Z, a, b, \{F_i\})\) for each \(F_i\) of interest is available. For a face \(F\) of a polyhedron described by \(a^T x \leq b\) where \(a\) is a permutation of \(a\). We say that a collection of faces \(F\) of a polyhedron is permutation-invariant, if for a face in \(F\) described by an inequality, all its permutations are included in \(F\). For a face \(F\), we define \(l(F) = \{j \in \{1, \ldots, n\} \mid x_j = a, \forall \hat{x} \in F\}\) and \(u(F) = \{j \in \{1, \ldots, n\} \mid x_j = b, \forall \hat{x} \in F\}\).

**Corollary 4.1.** Let \(a, b \in \mathbb{R}, Z \subseteq \{(x, z) \mid x \in \mathbb{R}^n, z \in \mathbb{R}^m\}\) be a permutation-invariant set with respect to \(x\), and \(F = \{F_1, \ldots, F_r\}\) be a permutation-invariant collection of faces of \([a, b]^n\). Let \(I = \{i \in \{1, \ldots, r\} \mid \exists p, q, p < q \text{ s.t. } u(F_i) = \{1, \ldots, p\} \text{ and } l(F_i) = \{q, \ldots, n\}\}\). Then,

\[
\text{conv}(X(Z, a, b, F)) = \left\{(x, z) \mid (u, z) \in \text{conv} \left( \bigcup_{i \in I} X(Z, a, b, \{F_i\}) \right), u_1 \geq \cdots \geq u_n, u \geq_x m x \right\}.
\]

We remark that \(\text{conv}(\bigcup_{i \in I} X(Z, a, b, \{F_i\}))\) can be constructed using disjunctive programming techniques if the recession cone of \(X(Z, a, b, \{F_i\})\) does not depend on \(i\) [31, Corollary 9.8.1]. Theorem 4.1 shows that even though the number of faces in \(\mathcal{F}\) may be exponentially large, we can exploit the permutation-invariance of the set to consider only a polynomial number of faces in the construction. More explicitly, there are \(2^{n-d} \binom{n}{d}\) \(d\)-dimensional faces of \([a, b]^n\) and \(\binom{n+d}{d+1}\) \(d\)-dimensional faces of the simplex \(a \geq x_1 \geq \cdots \geq x_n \geq b\). But, there are only \(n-d+1\) of the \(d\)-dimensional faces of the hypercube, namely, \(F_l\) for \(l \in \{0, \ldots, n-d\}\), that are defined by the hole-free sets \(B_l = \{1, \ldots, l\}\) and \(A_l = \{l+d+1, \ldots, n\}\) with the convention that \(B_0 = A_{n-d} = \emptyset\).

Applications of Theorem 4.1 extend beyond Schur-concave functions. For example, consider the convex hull of \(\{(x, \alpha) \in [a, b]^n \times \mathbb{R} \mid x \sum_{i=1}^{n} x_i \leq \alpha\}\), where \(a\) is not necessarily positive. The product function is not Schur-concave when some of the variables take negative values; for example consider \(x_1 x_2 x_3\) and observe that although \((1, -1, -3) \geq_m (0, 0, -3)\) the function value is higher at \((1, -1, -3)\) than at \((0, 0, -3)\).

There are many functions besides the multilinear monomial that are permutation-invariant and whose envelopes are determined by their values on certain low-dimensional faces – the postulates required for the construction in Theorem 4.1. We discuss some examples next. Observe that all elementary symmetric polynomials satisfy these postulates because, being multilinear, their convex envelope is generated by vertices of \([a, b]^n\). Since Theorem 4.1 allows \(z\) to be multi-dimensional, this result in fact yields the simultaneous convexification of the elementary symmetric monomials over \([a, b]^n\) and thus the convex envelope of positive linear combinations of elementary symmetric polynomials. Next, consider the function \(x_1^t \cdots x_n^t\) over \([a, b]^n\), where \(a, t > 0\). If \(t \leq 1\), the function is concave when all but one variable is fixed and therefore the convex envelope is generated by vertices of \([a, b]^n\). If \(t \geq 1\), then by restricting attention to any face of the hypercube, where two or more variables are not fixed, say \(x_1\) and \(x_2\), we see that the determinant of the Hessian of \(x_1^t x_2^t\) is \(-x_2^{t-2} x_1^{t-2} t^2 (2t-1)\), which is strictly negative. Therefore, there is a direction in which the function is concave and the convex envelope is determined by 1-dimensional faces of the hypercube. Since the function is convex on these faces, the convex envelope can be developed using Corollary 4.1. In fact, when \(t \geq 1\), Proposition 4.1 provides a closed-form expression for the convex envelope. This is because the function \(f(x) = x_1^t \cdots x_n^t\) is Schur-concave because \((x_1 - x_j) \left( \frac{\partial f}{\partial x_j} - \frac{\partial f}{\partial x_i} \right) = f(x_1 - x_j) \frac{2 x_1^{t-2} x_j^{t-2} t^2 (2t-1)}{x_1 x_j} > 0\).

We next expand on the characterizations we gave in Corollary 4.1 to obtain a more streamlined description of the result for the particular cases of permutation-invariant functions whose convex envelopes are completely determined by the extreme points of their domain, which is assumed to be a hypercube.

**Proposition 4.2.** Consider a function \(\phi(x) : [a, b]^n \to \mathbb{R}\), that is permutation-invariant in \(x\) and whose convex envelope remains the same even if its domain is restricted to \([a, b]^n\). For \(i = 1, \ldots, n\) and \(j = 0, \ldots, n\), let \(p_{ij} = a\) if \(i > j\) and \(b\) otherwise and let \(p_{ij}\) denote the \(j^{th}\) column of this matrix. Define \(f(x) := \phi(p_0) + \sum_{i=1}^{n} \frac{x_i-a}{b-a} \left( \phi(p_i) - \phi(p_{i-1}) \right)\). Then, the convex envelope of \(\phi(x)\) over \([a, b]^n\) can be expressed as the value function of the following problem:

\[
\text{conv}_{[a, b]^n}(\phi)(x) = \min \left\{ f(u) \mid u \geq_m x, b \geq u_1 \geq \cdots \geq u_n \geq a \right\}.
\]
Proof. Proof. Observe that the points in \( \{a, b\}^n \) that intersect with \( x_1 \cdots x_n \) are precisely the columns \( p_j \) described in the statement of the result. Consider the column \( p_j \) and observe that \( f(p_j) = \phi(p_j) \). Moreover, \( f \) is affine. Let \( \Gamma = \{x \in \{a, b\}^n \mid x_1 \leq \cdots \leq x_n\} \). Then, we show that \( f = \text{conv}_{\Gamma}(\phi|_{\Gamma}) \), where \( \phi|_{\Gamma} \) denotes the restriction of \( \phi \) to \( \Gamma \). Clearly, \( f(x) \leq \text{conv}_{\Gamma}(\phi|_{\Gamma})(x) \) because it matches \( \phi|_{\Gamma} \) at all the points in the domain and is a convex underestimator. Also, \( f(x) \geq \text{conv}_{\Gamma}(\phi|_{\Gamma})(x) \) because of Jensen's inequality applied to \( \text{conv}_{\Gamma}(\phi|_{\Gamma}) \), observing that \( f(x) \) is exact at the extreme points of \( \Gamma \), and because \( f(x) \) is affine. Now, consider Corollary 4.1. Let \( \mathcal{F} \) be the set of extreme points of \( \{a, b\}^n \) and \( Z = \{(x, z) \mid z \geq \phi(x)\} \). By assumption, \( \text{conv}(S(Z, a, b)) = \text{conv}(X(Z, a, b, F)) \). By Corollary 4.1, it follows that \( \text{conv}(X(Z, a, b, F)) = \{(x, z) \mid \{u, z \in \text{conv}\{U_x \mid X(Z, a, b, \{p_j\}), u \geq m, x\} = \{(x, z) \mid z \geq f(u), b \geq u_1 \geq \cdots \geq u_n \geq a, u \geq m x\}. \]

In the next result, we show how the convex hull of \( \{(x, y) \in [a, b] \times [c, d] \mid \prod_{i=1}^{m} y_i^{\alpha} = \prod_{j=1}^{n} x_j^{\beta}\} \) can be constructed. To do so, it suffices to construct the convex hull of \( S = \{(x, y) \in [a, b] \times [c, d] \mid \prod_{i=1}^{m} y_i^{\alpha} \geq \prod_{j=1}^{n} x_j^{\beta}\} \), since the convex hull for the reverse inequality can be developed by switching the \( x \) and \( y \) variables, and the convex hull for the equality can then be obtained as the intersection of these two convex hulls; see [28]. Such a set occurs in polynomial optimization problems where linearized variables are introduced to relax \( uv = y, u^2 = x_1, \) and \( v^2 = x_2 \) in the form of \( y^2 = x_1 x_2 \). Similar higher dimensional equality constraints also occur and their relaxations can be employed in polynomial optimization problems.

Proposition 4.3. Consider \( S = \{(x, y) \in H \mid \prod_{i=1}^{m} y_i^{\alpha} \geq \prod_{j=1}^{n} x_j^{\beta}\} \), where \( H = [a, b] \times [c, d] \), with \( a \geq 0, c \geq 0, \alpha > 0 \), and \( \beta > 0 \). Let \( k = \min\{m, \frac{\alpha}{\beta}\} \). Define the convex sets:

\[
S_{ij} = S \cap \{ (x, y) \in H \mid (y_j)^{\alpha_j-1} = d, (y_j)^{\alpha_j} = e, \ y \in \Delta^m \}
\]

\[
C_j = S \cap \{ (x, y) \in H \mid (x_j)^{\beta_j} = b, (x_j)^{\beta_j+1} = a, \ y \in \Delta^n \}
\]

where \( S_{ij} \) is defined for \( i = 0, \ldots, m - k \) and \( j = 0, \ldots, n - 1 \) and \( C_j \) is defined for \( j = 0, \ldots, n \). Let \( T = \bigcup_{i,j} S_{ij} \cup \bigcup_{j} C_j \). Then

\[
\text{conv}(S) = X := \{(x, y) \mid v \geq m, u \geq m, (u, v) \in \text{conv}(T)\}
\]

In particular, if \( m \alpha \leq \beta \), then

\[
\text{conv}(S) = X' := \left\{ (x, y) \in H \mid \prod_{i=1}^{m} y_i^{\frac{\alpha}{\beta}} \geq \prod_{j=1}^{n} u(x_j)^{\frac{\beta}{\alpha}} \right\},
\]

where \( u(x) := u^{S(x)} \), defined as in Proposition 4.1.

Before providing the proof, we discuss its architecture. This proof will write the convex hull of \( S \) as a disjunctive hull of convex subsets of \( S \). Since these convex subsets will be obtained by fixing variables at their bounds, even though they are exponentially many, Corollary 4.1 will allow the construction of the convex hull. In the first part of the proof, we consider a slice of \( S \) at a fixed \( y \) and show that it suffices to restrict \( x \) to one-dimensional faces of \( [a, b]^n \) for constructing the convex hull of the slice. Then, we show that there are two cases. If the remaining \( x_j \) variable is also fixed to the bounds, the set is already convex. If not, then we show that any point in such a set cannot be extremal in \( S \) unless at least a certain number of \( y \) variables, specifically \( \max\{0, m - \frac{\alpha}{\beta}\} \) of them, are fixed to their bounds. The sets obtained by fixing these variables are once again convex. Then, it will follow by Theorem 4.1, that we may restrict our attention to the faces in \( T \) and, as a result, \( \text{conv}(S) = X \).

Proof. Proof. Let \( \phi(x) := \prod_{j=1}^{n} x_j^{\beta} \) and consider the set \( \Upsilon(\gamma) = \{x \in [a, b]^n \mid \phi(x) \leq \gamma\} \). By Theorem 3.A.3 in [25], \( \phi(x) \) is Schur-concave over \( [a, b]^n \) because it is permutation-invariant and \( \frac{\partial^2 \phi}{\partial x_i \partial x_j} \leq \cdots \leq \frac{\partial^2 \phi}{\partial x_n \partial x_1} \). At any point with \( x_1 \geq \cdots \geq x_n \),

Let \( \Upsilon(\gamma) = \{x \in [b]^{n-1} \times [a, b] \times [a]^{n-1} \mid x_1^{\beta} \cdots x_j^{\beta(n-j)(\beta(n-j))} \leq \gamma\} \). Then, it follows by Proposition 4.1 and Corollary 4.1 that \( \text{conv}(\Upsilon(\gamma)) = \{x \mid u \geq m, x_1 \geq \cdots \geq x_n, u \in \text{conv} \left( \bigcup_{i=1}^{n} \Upsilon_i(\gamma) \right) \} \).
Since \( n - 1 \) of the \( x_j \) variables can be fixed to bounds in the construction of the convex hull of each slice, this restriction can also be imposed in the construction of \( \text{conv}(S) \). Now, let \( \psi(y) := \prod_{i=1}^{m} y_i \) and consider the slightly more general set \( \Theta \) which will appear when we fix some of the \( y \) variables at their bounds. This set is defined as \( \Theta = \{ (x, y) \in [a, b] \times [c, d]^{m} \mid \zeta \psi(y) \geq \delta \} \), where \( \delta, \zeta \geq 0 \). Consider a point \( (x', y') \in \Theta \). Then, by restricting attention to \( y = \lambda y' \), where \( \lambda \in \mathbb{R} \), we obtain an affine transform of a subset \( \Lambda \) of \( \Theta \) such that \( \Lambda = \{ (x, \lambda) \in [a, b] \times \lambda (c, \alpha) \mid \lambda \theta \geq \delta' x^m \} \), where \( \theta = \zeta^\frac{m}{\alpha} \psi(y')^\frac{1}{\alpha} \), \( \delta' = \delta^\frac{1}{\alpha} \), \( \epsilon' = \max\{\lambda \mid \lambda y'_i \leq c \text{ for some } i\} \), and \( \delta' = \min\{\lambda \mid \lambda y'_i \geq d \text{ for some } i\} \).

Assume \( x' \in (a, b) \) and \( y' \in (c, d)^m \). We will first show that, if \( m > \frac{2}{\alpha} \), such a point is not an extreme point of \( S \). By definition, \( \epsilon' < 1 \), \( \delta' > 1 \), and \( (x', 1) \in \Lambda \). Assume \( m > \frac{2}{\alpha} \) and \( (x', y') \) is an extreme point of \( S \). Define \( s = \frac{\delta'}{\alpha} x^m \). If \( x' \in (a, b) \), then, for sufficiently small \( \epsilon > 0 \), we show that \( (x', 1) \) can be written as a convex combination of \( (x' - \epsilon \theta, 1 - \epsilon \delta') \) and \( (x' + \epsilon \theta, 1 + \epsilon \delta') \). The latter points are feasible in \( \Lambda \) because

\[
\delta' (x' + \epsilon \theta) \frac{n}{\alpha} \leq \delta' (x' + \epsilon \theta) + s(x' + \epsilon \theta - x') \leq \theta (1 + \epsilon s),
\]

where the first inequality is by concavity of \( x^\frac{n}{\alpha} \) for \( m > \frac{2}{\alpha} \) and the second inequality is because \( \delta' x^m \leq \delta \) as \((x', 1)\) belongs to \( \Lambda \). Since \( \Lambda \) is an affine transform of a subset of \( \Theta \), we have expressed \((x', y')\) as a convex combination of \((x' - \epsilon \theta, 1 - \epsilon \delta') y' = \epsilon \) and \((x' + \epsilon \theta, 1 + \epsilon \delta') y' = \epsilon \). Since \( \epsilon > 0 \) and \( \epsilon' > 0 \), it follows that these points are distinct thus contradicting the extremality of \((x', y')\).

Since, \( S \) is compact, in order to construct \( \text{conv}(S) \), we may restrict our attention to the extreme points of \( S \). Therefore, either \( x' \in (a, b) \) or there exists an \( i \) such that \( y'_i \in (c, d) \). If \( x' \in \{a, b\} \), the point belongs to the convex subset of \( \Theta \) obtained by fixing \( x' \) at its current value because the defining inequality can be written as \( \zeta^\frac{m}{\alpha} \psi(y')^\frac{1}{\alpha} \geq \delta \), which is a convex inequality. A permutation of such an inequality separates \( \bar{\lambda} = x^m \) for some \( \lambda \in \mathbb{R} \) by solving the problem.

Now, consider the case where \( s(x_i) \) are fixed at their bounds or where we fix all \( y_i \), except for a subset of cardinality \( \min\{m, \frac{2}{\alpha}\} \) and fix all \( x_j \) except for one; all such sets are, up to permutation, included in one of the \( C_{ij} \)s. Therefore, we may assume without loss of generality that \( m \leq \frac{2}{\alpha} \). Then, we rewrite the defining inequality of \( \Theta \) as \( \zeta^\frac{m}{\alpha} \psi(y')^\frac{1}{\alpha} \geq \delta \), and observe that this is a convex inequality because \( \psi(y')^\frac{1}{\alpha} \) is a concave function and \( x^\frac{n}{\alpha} \) is a convex function. Therefore, we need to consider cases where either all \( x_j \) are fixed at their bounds or we fix all \( y_i \), except for a subset of cardinality \( \min\{m, \frac{2}{\alpha}\} \) and fix all \( x_j \) except for one; all such sets are, up to permutation, included in one of the \( C_{ij} \)s or \( S_{ij} \)s.

Since \( S_{ij} \subseteq S \), \( C_{ij} \subseteq S \), and \( S \) is permutation-invariant, it follows that \( \text{conv}(S) \supseteq X \). Since \( X \) is convex and \( S \) is compact, we only need to show that the extreme points of \( S \) are contained in \( X \). However, we have shown that the extreme points of \( S \) belong to \( T \) or a set obtained from \( T \) by permuting the \( x \) and/or \( y \) variables. Since \( X \) is permutation invariant and contains \( T \), it follows that every extreme point belongs to \( X \). Therefore, \( X = \text{conv}(S) \).

Now, consider the case where \( m \alpha \leq \beta \). Clearly, \( k = m \). If we fix \( y \) at \( \bar{y} \), it follows from the Schur-concavity of \( \prod_{j=1}^{n} x_j^\frac{1}{n} \) over \([a, b]^n\), the convexity of \( x^\frac{n}{\alpha} \) over \([a, b] \), and Proposition 4.1 that the convex hull of this slice is defined by \( \prod_{j=1}^{n} x_j^\frac{1}{n} \geq \prod_{j=1}^{n} u(x_j)^\frac{1}{n} \). This shows that \( X' \subseteq \text{conv}(S) \). By Schur-concavity of \( \prod_{j=1}^{n} x_j^\frac{1}{n} \), it follows that \( \prod_{j=1}^{n} x_j^\frac{1}{n} \geq \prod_{j=1}^{n} u(x_j)^\frac{1}{n} \), or \( S \subseteq X' \). This implies that \( S \subseteq X' \subseteq \text{conv}(S) \). To show that \( X' = \text{conv}(S) \), we only need to show that \( X' \) is convex. As in the proof of Proposition 4.1, let \( \varphi(s) = \prod_{j=1}^{n} u(x_j)^\frac{1}{n} \), where \( s = \sum_{j=1}^{n} (x_i - a) \), and rewrite the above inequality as \( \varphi(s) - \prod_{j=1}^{n} y_j^\frac{1}{n} \leq 0 \). Since the left-hand side is jointly convex in \((s, y)\) and \( s \) is a linear function of \( x \), this proves that \( X' \) is convex in \((x, y)\).

So far in this section, we have given various results where we describe the convex hull of a set in an extended space by introducing variables \( u \). We now discuss how inequalities in the original space can be obtained by solving a separation problem.

Usually, given a set \( X \) and an extended space representation of its convex hull, \( C \), we separate a given point \( \bar{x} \) from \( X \) by solving the problem \( \inf_{x \in C} ||x - \bar{x}|| \). By duality, the optimal value matches \( \max_{||a||, \alpha \leq 1} \{ \langle x, a \rangle - h(a) \} \), where \( h(\cdot) \) is the support-function of \( C \) and \( \| \cdot \| \) is the dual norm. Then, if the optimal value, \( z^* \) is strictly larger than zero and the optimal solution to the dual problem is \( a^* \), we have \( \langle x, a^* \rangle > 0 \) for all \( x \in \text{proj}_x C \) and this inequality separates \( \bar{x} \) from \( \text{proj}_x C \).

However, such an inequality is typically not facet-defining for \( \text{conv}(X) \) even when the latter set is polyhedral. We now discuss a separation procedure that often generates facet-defining inequalities. The structure of permutation-invariant sets and their extended space representation allow for this alternate approach. Assume we are interested in developing
the convex envelope of a permutation-invariant function \( \phi \), such as \( \prod_{i=1}^{n} x_i \), over \([a, b]^{n}\). As in Theorem 2.1, the convex envelope of \( \phi \) at \( x \) is obtained by expressing \( x \) as a convex combination of \( u \) and its permutations, where \( u \geq x \). Moreover, assume that the convex envelope at \( u \) is obtained as a convex combination of the extreme points of the simplex \( a \leq x_1 \leq \cdots \leq x_n \leq b \) with convex multipliers \( \gamma \). Since \( x = Su \) for some doubly stochastic matrix \( S \) and \( u = V \gamma \), where the columns of \( V \) correspond to vertices of the simplex, it follows that \( x = SV \gamma \). Therefore, we can find a representation of \( x \) as a convex combination of vertices in \( V \) and their permutations.

We implement the above procedure for multilinear sets over \([a, b]^{n}\) to evaluate its impact on the quality of the relaxation. For the purpose of illustration, we consider the special case of \( \prod_{i=1}^{n} x_i \) over \([a, b]^{n}\). In this case, (26) reduces to

\[
\min \quad a^n + \sum_{i=1}^{n} b^{i-1} a^{n-i}(u_i - a) \\
\text{s.t.} \quad u \geq x \\
b \geq u_1 \geq \cdots \geq u_n \geq a. \tag{28}
\]

Given \( x \in \mathbb{R}^n \) in general position inside \([a, b]^{n}\), assume that the optimal solution to (28) is \( u \). Then, we express \( x = Su \), where \( S \in \mathcal{M}^{n \times n}(\mathbb{R}) \) is a doubly-stochastic matrix. Given \( x \) and \( u \), this can be done through the solution of a linear program, \( \min \{0 \mid x = Su, S = I, TS = I, S \geq 0\} \). Although \( S \) can also be derived as a product of \( T \)-transforms, see proof of Lemma 2 in Section 2.19 of [13], we use the LP approach in our numerical experiments, given its simplicity of implementation. Then, we express \( S \) as a convex combination of permutation matrices. Such a representation exists due to Birkhoff Theorem. We obtain it using a straightforward algorithm, which we describe next. Observe first that the desired representation is such that all permutation matrices with non-zero convex multipliers have a support that is contained within the support of \( S \). This implies that the bipartite graph, we describe next, has a perfect matching. The bipartite graph is constructed with nodes labeled \( \{1, \ldots, n\} \) in each partition and edges that connect a node \( i \) in the first partition to \( j \) in the second partition if and only if \( S_{ij} > 0 \). Given a perfect matching, we construct a permutation matrix \( P \) so that \( P_{ij} = 1 \) if node \( i \) in the first partition is matched to node \( j \) in the second partition. Then, we associate \( P \) with a convex multiplier \( \pi \) which is chosen to be \( \min \{S_{ij} \mid P_{ij} = 1\} \). If \( \pi = 1 \), we have a representation of \( S \) as a convex combination of permutation matrices. Otherwise, observe that \( \frac{1}{\pi} (S - \pi P) \) is again a doubly-stochastic matrix with one less non-zero entry. Therefore, by recursively using the above approach we obtain \( S \) as a convex combination of at most \( n^2 \) permutation matrices. Then, we permute \( u \) according to these permutation matrices. For each such \( u \), the convex envelope is given by the optimal function value of (28). Each permuted \( u \) can be expressed as a convex combination of the corner points of the permuted simplex \( \{b \geq u_1 \geq \cdots \geq u_n \geq a\} \). We claim that an affine underestimator \( \bar{\phi} \) of \( \phi \) that is such that \( \bar{\phi}(x) = \phi(x) \) must also satisfy \( \bar{\phi}(v^i) = \phi(v^i) \) for every vertex \( v^i \) of \([a, b]^{n}\) that has a non-zero multiplier in the representation of \( x \) computed above. If not, we have convex multipliers \( \lambda_i \), where \( x = \sum_i \lambda_i v^i \), such that

\[
\bar{\phi}(x) = \phi(x) = \sum_i \lambda_i \phi(v^i) = \sum_i \lambda_i \phi(v^i) + \sum_i \lambda_i \bar{\phi}(v^i) = \bar{\phi}(x),
\]

which yields a contradiction. Here, the first equality is by the definition of \( \bar{\phi} \), second equality is because our construction obtains \( \phi(x) \) as a convex combination of \( \sum_i \lambda_i \phi(v^i) \) using multipliers \( \lambda^i \), the third equality is because \( v^i \) are extreme points, the first inequality is because there exists an \( i \) such that \( \phi(v^i) > \bar{\phi}(v^i) \), and the final equality is because \( \bar{\phi} \) is affine. Therefore, it follows that \( \bar{\phi}(v^i) = \phi(v^i) \) for all \( i \). Since \( x \) was assumed to be in general position, these equality constraints uniquely identify \( \bar{\phi} \).

We conclude this section by presenting the results of a numerical experiment that suggests that the bounds obtained when building convex relaxations of \( \psi_n(x) = \prod_{i=1}^{n} x_i \) over \([a, b]^{n}\) using the procedure described above are significantly stronger than those obtained using factorable relaxations. To this end, we consider functions \( \psi_n(x) \) where \( n = 10 \) over two permutation-invariant hyper-rectangles. The first one, \( B_1 = [2, 4]^{10} \), is contained in the positive orthant, while the second, \( B_2 = [-2, 3]^{10} \), contains 0 in its interior. We generate nine sample points uniformly at random inside of \( B_1 \) and \( B_2 \). At each point, we compare the value \( z_r \) of the relaxation obtained using a recursive application of McCormick’s procedure with the value \( z_e \) of the convex envelope, obtained using the results described in this section. We then compute the existing gap (Gap) and relative gap (%Gap), using the formulas \( z_e - z_r \) and \( \frac{z_e - z_r}{z_e} \), respectively. Results are presented in Tables 1 and 2 where it can be observed that the proposed approach leads to substantial improvements in bounds, especially when variables \( x \) take both positive and negative values.
For an arbitrary positive integer $n$, we study the case where the base set $S$ is permutation-invariant. Let $M_n$ denote the set of rank-one matrices associated with permutation-invariant sets $S$. Then, we introduce the valid inequality $x \leq 1$ for some $x \in S$. Similarly, we denote by $BD$ the set $M_n$. The separation problem associated with $M_n$ can be seen as the condition $x \leq 1$. Linear valid inequalities in $M_n$ are then developed by the authors of [33]. Hence semidefinite relaxations have been considered that relax the non-convex constraint $x \leq 1$, which is equivalent to the convex constraint $x \leq 1$. Linear valid inequalities in $x$, $X$, find a sparse vector $x$ that maximizes the variance $x^T \Sigma x$. A semidefinite relaxation of sparse PCA therefore aims to approximate $M := \{(x, X) \in \mathbb{R}^n \times \mathcal{M}^n \mid X = xx^T, x \in S\}$ (29) for a positive integer $K \in \{1, \ldots, n-1\}$. The set $M$ can be seen as $M_S$ by choosing $S$ to be the permutation-invariant set $S = \{x \in \mathbb{R}^n \mid \|x\| \leq 1, \mathrm{card}(x) \leq K\}$. The separation problem associated with $M$ is known to be NP-hard [33].

| Sample | $z_x$ | $z_x$ | Gap | %Gap |
|--------|-------|-------|-----|------|
| 1      | 18943.5 | 7584.8 | 11358.6 | 60.0% |
| 2      | 52904.9 | 21933.9 | 30970.9 | 58.5% |
| 3      | 22754.2 | 8622.1 | 14132.1 | 62.1% |
| 4      | 26299.0 | 8526.7 | 17772.3 | 67.6% |
| 5      | 13817.1 | 5694.1 | 8066.3 | 58.4% |
| 6      | 25028.6 | 8906.2 | 16122.4 | 64.4% |
| 7      | 13852.4 | 5694.1 | 8158.3 | 58.9% |
| 8      | 16059.4 | 8069.1 | 7990.2 | 49.8% |
| 9      | 10122.1 | 4812.2 | 5309.9 | 52.5% |
| Average| 59.1% |

Table 1: Gap at a randomly chosen point for $\prod_{i=1}^n x_i$ on $[2, 4]^{10}$

| Sample | $z_x$ | $z_x$ | Gap | %Gap |
|--------|-------|-------|-----|------|
| 1      | -12314.6 | -25655.4 | 13340.9 | 108.3% |
| 2      | -16221.2 | -29559.4 | 13338.2 | 93.7% |
| 3      | -13247.0 | -29405.9 | 16158.9 | 122.0% |
| 4      | -14069.4 | -28248.4 | 14179.0 | 100.8% |
| 5      | -10660.9 | -23134.2 | 12463.3 | 116.9% |
| 6      | -10979.5 | -21263.1 | 10283.7 | 93.7% |
| 7      | -9367.8 | -21327.4 | 11959.6 | 127.7% |
| 8      | -10245.9 | -24782.6 | 14536.8 | 141.9% |
| 9      | -9182.8 | -21137.0 | 11954.2 | 130.2% |
| Average| 113.7% |

Table 2: Gap at a randomly chosen point for $\prod_{i=1}^n x_i$ on $[-2, 3]^{10}$

## 5 Set of rank-one matrices associated with permutation-invariant sets

For a positive integer $n$ and a given set $S \subseteq \mathbb{R}^n$, define $M_S := \{(x, X) \in \mathbb{R}^n \times \mathcal{M}^n \mid X = xx^T, x \in S\}$ where $\mathcal{M}^n = \mathcal{M}^{n,n}(\mathbb{R})$. For each element $(x, X) \in M_S$, it is clear that $\mathrm{rank}(X) = 1$. Studying $M_S$ is particularly important when constructing valid inequalities for semidefinite relaxations of non-convex optimization problems. In this section, we study the case where the base set $S$ is permutation-invariant.

For an arbitrary positive integer $n$, we denote by $\mathbb{I}_n$ the $n$-dimensional vector of ones. When $n$ is clear in the context, we simply denote it by $\mathbb{I}$. Similarly, we denote by $\mathbb{I}_{n \times n}$ the $n$-by-$n$ matrix of ones.

As a motivating example, consider sparse PCA, which, for a given covariance matrix $\Sigma$, finds a sparse vector $x$ that maximizes the variance $x^T \Sigma x$. A semidefinite relaxation of sparse PCA therefore aims to approximate $M := \{(x, X) \in \mathbb{R}^n \times \mathcal{M}^n \mid X = xx^T, x \in S\}$ (29) for a positive integer $K \in \{1, \ldots, n-1\}$. The set $M$ can be seen as $M_S$ by choosing $S$ to be the permutation-invariant set $S = \{x \in \mathbb{R}^n \mid \|x\| \leq 1, \mathrm{card}(x) \leq K\}$. The separation problem associated with $M$ is known to be NP-hard [33]. Hence semidefinite relaxations have been considered that relax the non-convex constraint $X \leq xx^T$, which is equivalent to the convex constraint $x^T X \mathbb{I} \leq K$. Linear valid inequalities in $(x, X)$ are then developed by exploiting the property that $X = xx^T$. For example, the authors of [8] introduce the valid inequality $\mathbb{I}^T X \mathbb{I} \leq K$ for (29), which is implied by valid inequality $\sum_{i=1}^n x_i \leq \sqrt{K}$ and the condition $X = xx^T$.

We next show that additional valid inequalities can be constructed in a higher dimensional space by using the permutation-invariance of the base set $S$. To this end, we prove the following result.

**Proposition 5.1.** Suppose $S \subseteq \mathbb{R}^n$ is a permutation-invariant set. Let

$$N = \left\{(x, u, X, U) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{M}^n \times \mathcal{M}^n \mid \begin{array}{c} X = xx^T, U = uu^T, \\ x = Pu \text{ for some } P \in \mathcal{P}_n \end{array} \right\}.$$  

Then, $M_S = \text{proj}_{(x, X)} N$.  

22
Proof. To prove $M_S \subseteq \text{proj}_{x,X} \mathcal{N}$, consider $(x, X) \in M_S$. Let $P \in \mathcal{P}_n$ be such that $u := P^{-1}x \in \Delta$ and let $U = uu^\top$. Then, $(x, u, U) \in \mathcal{N}$. To prove $M_S \supseteq \text{proj}_{x,X} \mathcal{N}$, consider $(x, u, X, U) \in \mathcal{N}$ and let $P \in \mathcal{P}_n$ be such that $x = Pu$. By permutation-invariance of $S$, $x \in S$, showing $(x, X) \in M_S$. \hfill \square

Now we develop linear inequalities implied by the conditions in $\mathcal{N}$. For any $(x, u, X, U) \in \mathcal{N}$, observe that $X = xx^\top = (Pu)(Pu)^\top = PU^PU^\top$ for a permutation matrix $P$. Therefore, consider linear inequalities implied by the facts that $x$ is a permutation of $u$, and $X$ is obtained by permuting some columns and rows of $U$ symmetrically. Perhaps, the most straightforward such inequalities are

\begin{align}
\mathbb{1}^\top u & = \mathbb{1}^\top x, \\
\text{trace}(X) & = \text{trace}(U), \\
\mathbb{1}^\top X \mathbb{1} & = \mathbb{1}^\top U \mathbb{1},
\end{align}

More generally, consider any function $\phi : \mathbb{R}^n \times \mathcal{M}^n \to \mathbb{R}$ such that $\phi(x, xx^\top)$ is permutation-invariant with respect to $x$. Then, we can impose the equality $\phi(x, X) = \phi(u, U)$ if $\phi$ is linear in $(x, X)$. In fact, if $\phi$ is linear in $(x, X)$ then we argue that this identity is implied by (30). To see this, observe that $\phi(x, xx^\top)$ is a quadratic function in $x$. Let $\psi(x) = \phi(x, xx^\top) = x^\top C x + d^\top x + e$ for $C \in \mathbb{S}^n$, $d \in \mathbb{R}^n$, and $e \in \mathbb{R}$. By permutation-invariance of $\psi$, the function $\psi(x) - \psi(Px)$ is the zero function in $x$ for every $P \in \mathcal{P}_n$. Observe that $\psi(x) - \psi(Px) = x^\top (C - P^\top CP)x + (d - P^\top d)^\top x \equiv 0$.

Therefore, $d = P^\top d$ and $C = P^\top CP$. For $i \neq j \in \{1, \ldots, n\}$, consider the permutation matrix $P$ such that $(Px)_i = x_j, (Px)_j = x_i$, and $(Px)_k = x_k$ for all $k \neq i, j$. Then, $d_{ij} = d_{ji}, C_{ij} = C_{ji},$ and $C_{ij} = C_{ji}$. Since the choices for $i$ and $j$ are arbitrary, it holds that $d = \rho \mathbb{1}$ for some $\rho \in \mathbb{R}^n$, the diagonal entries of $C$ are identical, and $C$ is symmetric. We next claim that all off-diagonal entries of $C$ are identical. We assume $n \geq 3$ since it is clear otherwise. For any $q \in \{1, \ldots, n\} \setminus \{i, j\}$, $C_{ij} = C_{iq} = C_{qj}$. That is, all entries of $q$th column of $C$ except for $C_{qq}$ are equal. By symmetry of $C$, all entries of $q$th row of $C$ except for $C_{qq}$ are equal. Since $q$ is arbitrary, all off-diagonal entries of $C$ are identical because for any $i < j$ and $p < q$ with $q < j$, $C_{ij} = C_{pj} = C_{pq}$. Therefore,

$$
\psi(x) = x^\top (C_{11} \mathbb{1}) + c_2 \mathbb{1} \mathbb{1} \mathbb{1} + \rho (\mathbb{1}^\top x) = c_1 \text{trace}(xx^\top) + c_1 \text{trace}(Ux) + \rho (\mathbb{1}^\top x).
$$

Now, the desired equality $\phi(x, X) = \phi(u, U)$ is

$$
c_1 \text{trace}(X) + c_2 \mathbb{1}^\top X \mathbb{1} + \rho (\mathbb{1}^\top x) = c_1 \text{trace}(U) + c_2 \mathbb{1}^\top U \mathbb{1} + \rho (\mathbb{1}^\top u),
$$

which is implied by equalities in (30).

Another type of constraints can be obtained when $S \subseteq \mathbb{R}^+_n$. That is, $u \in S$ is chosen to be nonnegative and in descending order. Then, entries in each row of $uu^\top$ are also in descending order, yielding the inequalities.

$$
U_{i,j} \geq U_{i,j+1}, \quad 1 \leq i \leq n, \quad 1 \leq j < n-1.
$$

Similar arguments can be made for column entries. These inequalities, however, are redundant because of the symmetry of $U$.

We next introduce a general framework that constructs tighter linear relaxations by exploring the conceptual relationship that $x$ is a permutation of $u$. This allows us to model or relax identities of the form $\phi(x, xx^\top) = \phi(u, uu^\top)$ where $\phi$ is a certain real-valued nonlinear permutation-invariant function. To this end, for fixed integers $p, q \in \{1, \ldots, n\}$, $r \in \{1, \ldots, \min\{pn, qn\}\}$, and $W \in \mathcal{M}^n$, consider the optimization problem

$$
\max \sum_{i=1}^n \sum_{j=1}^n W_{ij} t_{ij} \\
\text{s.t.} \sum_{j=1}^n t_{ij} \leq q, \quad i \in \{1, \ldots, n\} \\
\sum_{i=1}^n t_{ij} \leq p, \quad j \in \{1, \ldots, n\} \\
0 \leq t_{ij} \leq r \\
0 \leq t_{ij} \leq 1, \quad i, j \in \{1, \ldots, n\}.
$$

Its dual is

$$
\min \sum_{i=1}^n \sum_{j=1}^n \alpha_i + p \sum_{j=1}^n \beta_j + r \gamma + \sum_{i=1}^n \sum_{j=1}^n \delta_{ij} \\
\text{s.t.} \quad \alpha_i + \beta_j + \gamma + \delta_{ij} = W_{ij} + \theta_{ij}, \quad i, j \in \{1, \ldots, n\} \\
\alpha_i \geq 0, \beta_j \geq 0, \gamma \geq 0, \delta_{ij} \geq 0, \theta_{ij} \geq 0, \quad i, j \in \{1, \ldots, n\}.
$$
where dual variables $\alpha$, $\beta$, $\gamma$, and $\delta$ correspond to primal constraints (32a), (32b), (32c), and the upper-bound constraints of (32d), respectively. We denote these optimization problems by $\max \{ f^W(t) \mid t \in \Phi \}$ and $\min \{ g(z) \mid z \in \Omega(W) \}$. Strong duality holds since both (32) and (33) are feasible. We denote $h(W) := \max \{ f^W(t) \mid t \in \Phi \} = \min \{ g(z) \mid z \in \Omega(W) \}$. Observe that $h(ww^T)$, as a function of $w$ is permutation-invariant with respect to $w$. Therefore, if $(x, u, X, U) \in \mathcal{N}$, it holds that $h(U) = h(X)$. Since the linearity of the identity is not guaranteed, we construct linear inequalities in $(x, u, X, U)$ by taking the identity and the conditions in the set description of $\mathcal{N}$ into account. In the following discussion, we assume that $(x, u, X, U) \in \mathcal{N}$.

We first consider the case where a closed-form description of the optimal value $h(ww^T)$ is known, $h(U)$ is linear in $U$, and $h(X)$ is not linear in $X$. Since $h(X)$ and $f^X(t)$ are both nonlinear in $(X, W)$ and $(t, W)$, respectively, we use the dual objective formulation to reformulate the identity $h(U) = h(X)$ because the dual objective function and the constraints are linear in $(z, W)$. We obtain a reformulation by replacing $h(X)$ with $g(z)$ and adding the conditions in the feasible set $\Omega(X)$ into the formulation.

We next consider the case where either a closed-form of $h(ww^T)$ is unknown or $h(U)$ is nonlinear in $U$. Then, we construct the relaxation by replacing both $h(U)$ and $h(X)$ with linear functions $g(z_U)$ and $g(z_X)$ with distinct variables $z_U = (\alpha_U, \beta_U, \gamma_U, \delta_U, \theta_U)$ and $z_X = (\alpha_X, \beta_X, \gamma_X, \delta_X, \theta_X)$ and add the conditions in $\Omega(U)$ and $\Omega(X)$.

Then, we tighten the relaxation by exploring the permutation relationship between $u$ and $x$ and the rank-one conditions. Assume that $x = Pu$ and that $z_U = (\alpha_U, \beta_U, \gamma_U, \delta_U, \theta_U)$ is an optimal solution to the dual with $W = uu^T$. Then, $(P\alpha_U, P\beta_U, P\gamma_U, P\delta_U, P\theta_U)$ is an optimal solution to the dual with $W = xx^T$. Therefore, we can tighten the relaxation by considering conditions $\alpha_X = P\alpha_U, \beta_X = P\beta_U, \gamma_X = P\gamma_U, \delta_X = P\delta_U$ for some $P \in \mathcal{P}_n$. For example, we can add the baseline linear conditions $\sum_i \alpha_{U,i} = \sum_i \alpha_{X,i}, \sum_i \beta_{U,i} = \sum_i \beta_{X,i}$, and $\sum_i \delta_{U,i} = \sum_i \delta_{X,i}$. Then, we can replace the left-hand side with $g(z_U)$ and the equality with the inequality $\geq$. We may add inequalities that capture the permutation relationships. Obviously, we can add $\gamma_U = \gamma_X$. In addition, we can impose the linear reformulations of $\alpha_{U,i} \geq m \alpha_X$ and $\delta_{U,i} \geq m \delta_X$ because $\alpha_U$ and $\delta_U$ are constants. In addition, any linear inequalities implied by the relationship $\delta_{U,i} = P\delta_X$ for some $P \in \mathcal{P}_n$, such as $trace(\delta_{U,i}) \geq trace(\delta_X)$ and $\sum_i \delta_{U,i} = \sum_i \delta_X$, can be considered. Similar relations can also be introduced for $\theta$. More generally, arbitrary linear functions that are permutation-invariant in $\alpha, \beta, \gamma, \delta,$ and $\theta$ can be considered instead of the specific one in the objective of the dual.

We next present some special but important cases where the closed-form of $h(W)$ is known.

**Lemma 5.1.** When $W = ww^T$ for $w \geq 0$ and $p = q = 1$, the optimal value of (32) is $\sum_{i=1}^r w_{[i]}^2$.

**Proof.** Without loss of generality, we assume that $w \in \Delta$ and prove that the optimal value is $\sum_{i=1}^r w_{[i]}^2$. Define $t^r$ as $t_{ij}^r = 1$ if $i = j \leq r$ and $0$ otherwise. Then, $t^r$ is feasible for the primal. Its objective value is $\sum_{i=1}^r w_{[i]}^2$. We next define $z^r := (\alpha', \beta', \gamma', \delta') \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \times \mathcal{M}^n$ as $\alpha'_{i} = \beta'_{i} = \max \left\{ \frac{w_{i}^2 - w_{j}^2}{2}, 0 \right\}$ for $i = 1, \ldots, n$, $\gamma' = \delta' = w_{[r]}^2$, and $\delta'_{ij} = 0$ for $i, j \in \{1, \ldots, n\}$ and prove that $z^r$ is feasible for the dual. The nonnegativity of $z^r$ is clear. We next show that $z^r$ satisfies (33a). First, suppose $i \leq r$ and $j \leq r$. Then, $a_{i}' + b_{i}' + c_{i}' + d_{ij}' = \frac{w_{i}^2 + w_{j}^2}{2} \geq w_{i}w_{j}$. Next, consider the case where $i \leq r$ and $j > r$. Then, $a_{i}' + b_{i}' + c_{i}' + d_{ij}' = \frac{w_{i}^2 + w_{j}^2}{2} \geq w_{i}w_{r} \geq w_{i}w_{j}$, where the last inequality holds because $w \in \Delta$. The case where $i > r$ and $j \leq r$ is symmetrical, and the case where $i > r$ and $j > r$ is clear. Since $t^r$ and $z^r$ satisfy complementarity-slackness conditions, they are optimal solutions to the primal and the dual, respectively. Their common objective value is $\sum_{i=1}^r w_{[i]}^2$.

By Lemma 5.1, when $p = q = 1$, $h(W)$ is the sum of $r$ largest diagonal entries of $W$. While $h(W)$ is nonlinear, $h(U)$ is linear because it is the sum of the first $r$ diagonal entries. On the other hand, the inequalities $h(U) \geq h(X)$ for $r \in \{1, \ldots, n-1\}$ are equivalent to the inequality parts of the majorization $\text{diag}(U) \geq m \text{diag}(X)$. The equality part of the majorization is equivalent to the existing constraint (30b). Therefore, $\text{diag}(U) \geq m \text{diag}(X)$ is a special case of the aforementioned modeling technique.

**Lemma 5.2.** When $W = ww^T$ for $w \geq 0$ and $r = pq$, the optimal value of (32) is $\sum_{i=1}^p \sum_{j=1}^q w_{[i]}w_{[j]}$.

**Proof.** Without loss of generality, we assume that $w \in \Delta_+$. First, define $t^r$ as $t_{ij}^r = 1$ if $i \leq p$ and $j \leq q$ and $0$ otherwise. It is clear that $t^r$ is feasible and that its objective function value is $\left( \sum_{i=1}^p w_{[i]} \right) \left( \sum_{j=1}^q w_{[j]} \right)$. Next, we
We first show that $z'$ is feasible for the dual. The nonnegativity of the variables is clear from their definition. To prove that they satisfy (33a), we first consider the case where $i \leq p$ and $j \leq q$. Then, we compute that $\alpha_i' + \beta_j' + \gamma' + \delta_{ij}' = w_i w_j$, showing the result. Next, consider the case where $i > p$ or $j > q$. Without loss of generality, we assume that $i > p$. When $j > q$, it holds that $\alpha_i' + \beta_j' + \gamma' + \delta_{ij}' = w_{p+1} w_{q+1} \geq w_i w_j$ where the inequality holds because $i > p$, $j > q$, and $w \in \Delta_+$. When $j \leq q$, it holds that $\alpha_i' + \beta_j' + \gamma' + \delta_{ij}' = w_{p+1} w_{q+1} \geq w_i w_j$ where the inequality holds because $i > p$ and $w \in \Delta_+$. Since $t'$ and $z'$ satisfy complementarity-slackness conditions, they are optimal solutions to the primal and dual, respectively. Their common objective value is $\sum_{i=1}^p \sum_{j=1}^q w_i w_j$. □

Lemma 5.2 and (32) can be trivially extended to the case where $W = \bar{w} \bar{w}^T$, where $\bar{w} \in \mathbb{R}^n$, $\bar{w} \in \mathbb{R}^n$, $\bar{w}, \bar{w} \geq 0$, and $r = pq$, in which case the optimal value is $\sum_{i=1}^p \sum_{j=1}^q w[I] w[I]$. However, this is not of direct relevance for our sparse PCA relaxations. While $h(W)$ is nonlinear because the order of $w$ is unknown, $h(U)$ is linear because it is the sum of the entries of the $p$-by-$q$ upper-left submatrix of $U$. In particular, for $q = n$ and $p \in \{1, \ldots, n\}$, the inequalities $h(U) \preceq h(X)$ are equivalent to the inequality parts of $R^U \preceq_m R^X$ where $R^U$ and $R^X$ are the vectors of the row sums of $U$ and $X$, respectively. The equality part of the majorization is equivalent to (30c). Therefore, $R^U \preceq_m R^X$ is a special case of the aforementioned modeling technique.

In the remainder of the section, we introduce various strengthened semidefinite programming relaxations for sparse PCA.

### 5.1 An SDP relaxation for sparse PCA

Principal Component Analysis (PCA) is a well-known dimension reduction technique in statistical analysis. A principal component is a linear combination of independent variables. It also typically stands for the coefficient vector of the linear combination. The first principal component is a unit principal component for which variance is maximized; it is the eigenvector corresponding to the largest eigenvalue of the covariance matrix. Even though the first principal component explains the most variance of the data, it is often hard to interpret because most of its coefficients are nonzero. Sparse PCA is a variant of the approach introduced to resolve this issue by finding linear combinations with fewer explanatory variables.

Formally, let $\Sigma \in \mathbb{S}^n$ be the covariance matrix of the data set. The following optimization problem, which we refer to as sparse PCA and as defined in Section 1, where $x \in \mathbb{R}^n$ is the coefficient vector of the principal component, finds a unit sparse vector with at most $K$ nonzero entries that explains most of the variance of the data:

$$\begin{align*}
\max & \quad x^T \Sigma x \\
\text{s.t.} & \quad \|x\| \leq 1, \\
& \quad \text{card}(x) \leq K,
\end{align*}$$

(Sparse PCA)

where $K$ is a positive integer satisfying $1 < K < n$.

We studied the feasible set of sparse PCA in Section 3 where we denoted it as $N^K_\|\cdot\|$ assuming $\|\cdot\|$ is the $\ell_2$-norm. The feasible region of sparse PCA is non-convex because of the sparsity constraint. We established in Section 3 that

$$\begin{align*}
\text{conv}(N^K_\|\cdot\|) = \left\{ x : \begin{array}{l} \|u\| \leq 1, \\
u_1 \geq \cdots \geq u_K \geq 0, \\
u_{K+1} = \cdots = u_n = 0, \\
u \geq_{w_m} |x| \end{array} \right\}.
\end{align*}$$

(34)

Because sparse PCA maximizes a convex function, we can replace the feasible set with its convex hull, thereby obtaining a new problem formulation. This formulation, however, remains difficult to solve as it is a convex maximization problem. Next, we introduce new positive semidefinite relaxations for sparse PCA. The most commonly used (and, to the best of our knowledge, only) SDP relaxation for sparse PCA was introduced in [8] as follows

$$\begin{align*}
\max & \quad \text{trace}(\Sigma X) \\
\text{s.t.} & \quad \text{trace}(X) \leq 1, \\
& \quad \|X\| \leq K, \\
& \quad X \succeq 0.
\end{align*}$$

(35)
We refer to this as the \textit{D-relaxation}.

We next present a strengthened SDP relaxation based on the convex hull description (34). First, we introduce the matrix variable \( X \) to model the relationship \( X = xx^\top \). Then, we introduce variables \( y \) and \( Y \) to represent \(|x|\) and \(|X|\), respectively. Further, we add the auxiliary variables \( u, v, w, V, Y \) and \( W \) to model the absolute values. The variables and the constraints are

\[
\begin{cases}
  x = v - w, \ y = v + w \\
  v, w, \in \mathbb{R}^n, \ x, y, \in \mathbb{R}^n
\end{cases}
\]  
(36)

and

\[
\begin{cases}
  X = V - W, \ Y = V + W \\
  V, W, \in M^n_{\geq 0}, \ X, Y, \in S^n,
\end{cases}
\]  
(37)

where \( M^n_{\geq 0} \) is the set of \( n \)-by-\( n \) (entry-wise) nonnegative matrices and \( S^n \) is the set of \( n \)-by-\( n \) symmetric matrices. Next, we introduce the vector \( u \) majorizing \( y(= |x|) \) and the matrix \( U \) to model \( uu^\top \). The constraint \( u \in \Delta_+ \) and the constraint (31) in the cardinality setting can be written as

\[
\begin{cases}
  u_1 \geq \cdots \geq u_K \\
  u_1 = 0, \\
  U_{i,1} \geq \cdots \geq U_{i,K}, \quad i = 1, \ldots, K \\
  U_{i,j} = 0, \quad i \geq K+1 \text{ or } j \geq K+1 \\
  u \in \mathbb{R}^n, U \in S^n_{\geq 0},
\end{cases}
\]  
(38)

where \( S^n_{\geq 0} \) is the set of \( n \)-by-\( n \) (entry-wise) nonnegative symmetric matrices. In the following construction, we use the relationship between \( Y \) and \( U \) that \( Y = PU \) for some \( P \in \mathcal{P}_n \). (That is, the entries of \( Y \) and \( U \) equal up to row permutations and the corresponding column permutations.) We impose the constraints (30) as follows:

\[
\begin{cases}
  \text{trace}(U) \leq 1 \\
  1^\top U 1 = 1^\top Y 1.
\end{cases}
\]  
(39a)

(39b)

The nonconvex relationships \( X = xx^\top, Y = yy^\top \), and \( U = uu^\top \) can be relaxed using Schur complements as

\[
\begin{bmatrix}
  X & x \\
  x^\top & 1
\end{bmatrix} \succeq 0, \quad \begin{bmatrix}
  Y & y \\
  y^\top & 1
\end{bmatrix} \succeq 0, \quad \begin{bmatrix}
  U & u \\
  u^\top & 1
\end{bmatrix} \succeq 0.
\]  
(40)

By constraint \( X \succeq xx^\top \), it holds that \( X \succeq 0 \); hence, \( \text{diag}(X) \succeq 0 \). Therefore, the constraint \( \text{trace}(U) = \text{trace}(Y) \) can be replaced with

\[
\text{trace}(U) = \text{trace}(X).
\]  
(41)

We next present modeling details for the majorization constraints using the arguments presented earlier. First, the majorization relationship \( u \geq_m y \) is represented as

\[
\begin{cases}
  \sum_{i=1}^j u_i \geq j r_j + \sum_{j=1}^n t_{ij}, \quad j = 1, \ldots, n - 1 \\
  y_i \leq t_{ij} + r_j, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n - 1
\end{cases}
\]  
(42)

as mentioned in (4). Even though the modeling techniques using (32) and (33) can potentially derive several inequalities, we only present certain representative inequalities in the proposed SDP relaxation for the sake of exposition. First, the row sum majorization \( R^U \succeq_m R^Y \) and the diagonal majorization \( \text{diag}(U) \succeq_m \text{diag}(Y) \) are represented as

\[
\begin{cases}
  R^U_i = \sum_{j=1}^n U_{ij}, \quad R^Y_i = \sum_{j=1}^n Y_{ij} \\
  \sum_{i=1}^j R^U_i \geq j r_j + \sum_{j=1}^n t_{ij}, \quad j = 1, \ldots, n - 1 \\
  R^Y_i \leq t_{ij} + r_j, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n - 1 \\
  r^R \in \mathbb{R}^{n-1}, \quad t^R \in \mathbb{R}^{n \times (n-1)}, \quad R^U, R^Y \in \mathbb{R}^n
\end{cases}
\]  
(43)

and

\[
\begin{cases}
  \sum_{i=1}^j U_{ii} \geq j r^D_j + \sum_{j=1}^n t^D_{ij}, \quad j = 1, \ldots, n - 1 \\
  X_{ii} \leq t^D_{ij} + r^D_j, \quad i = 1, \ldots, n, \quad j = 1, \ldots, n - 1 \\
  r^D \in \mathbb{R}^{n-1}, \quad t^D \in \mathbb{R}^{n \times (n-1)}
\end{cases}
\]  
(44)
and let $/BD_{i,j}^2$ (2-Step relaxation). The proposed SDP relaxation, which we refer to as the submatrix relaxation, this approach provides a weaker bound. First, we introduce variables $v$ ariables $f$, showing that $K$.

Proof. First, trace($X$) = trace($U$) ≤ 1, where the equality and the inequality directly follow from (41) and (39a), respectively. We next show that $\mathbb{1}^T[X] \mathbb{1} \leq K$ is implied. Let $U_K$ be the upper-left $K$-by-$K$ submatrix of $U$ and let $\mathbb{1}_K$ be the $K$-dimensional vector of ones. Define $f : \mathbb{R}^K \rightarrow \mathbb{R}$ as $f(x) = x^T U_{K,K} x$. Since $U \succeq 0$, we have $U_{K,K} \succeq 0$, showing that $f$ is convex. Furthermore, $f(\alpha x) = \alpha^2 f(x)$ for any scalar $\alpha$. Therefore,

$$\mathbb{1}^T U \mathbb{1} = \mathbb{1}^T U_{K,K} \mathbb{1}_K = f(\mathbb{1}_K) = f \left( \sum_{i=1}^K e_i \right) = f \left( K \sum_{i=1}^K \frac{1}{K} e_i \right) \leq K^2 \sum_{i=1}^K f(e_i) = K \text{trace}(U) \leq K,$$

where the inequalities follows from the convexity of $f$ and (39a). Therefore,

$$\mathbb{1}^T |X| \mathbb{1} = \mathbb{1}^T |V - W| \mathbb{1} \leq \mathbb{1}^T (V + W) \mathbb{1} = \mathbb{1}^T Y \mathbb{1} = \mathbb{1}^T U \mathbb{1} \leq K,$$

where the first two equalities and the first inequality follow from (37), the third equality from (39b), and the last inequality from (47). The positive semidefiniteness of $X$ follows from the first Schur complement condition in (40).

Remark 5.1. We present another relaxation of (45) which improves computational efficiency compared to (45), albeit this approach provides a weaker bound. First, we introduce variables $(SR)_{i,q}$ to define the sum of $q$-largest components of $i^{th}$ row of $Y$ as follows:

$$\begin{cases} 
(SR)_{i,q} \geq q r_{i,q} + \sum_{j=1}^n t_{i,j}^q & i = 1, \ldots, n, q = 1, \ldots, K \\
Y_{i,j} \leq t_{i,j} + r_{i,j} & i = 1, \ldots, n, j = 1, \ldots, n, q = 1, \ldots, K \\
t_{i,j} \geq 0 & i = 1, \ldots, n, j = 1, \ldots, n, q = 1, \ldots, K \\
t_{i,j} \in \mathbb{R}^K \times \mathbb{R}^n \times \mathbb{R}^n, r_{i,j} \in \mathbb{R}^K \times \mathbb{R}^n.
\end{cases}$$

Using a specific feasible solution of (32), we relax (45) as follows:

$$\begin{cases} 
\sum_{i=1}^p \sum_{j=1}^q U_{i,j} \geq p t_{i,j}^B + \sum_{i=1}^n t_{i,q}^B & p = 1, \ldots, K, q = 1, \ldots, K \\
(SR)_{i,q} \leq t_{i,j}^B + r_{i,q}^B & p = 1, \ldots, K, q = 1, \ldots, K, i = 1, \ldots, n \\
t_{i,j}^B \geq 0 & i = 1, \ldots, n, j = 1, \ldots, n, q = 1, \ldots, K \\
t_{i,j}^B \in \mathbb{R}^K \times \mathbb{R}^K \times \mathbb{R}^n, r_{i,j}^B \in \mathbb{R}^K \times \mathbb{R}^K.
\end{cases}$$

We refer to this relaxation as the 2-Step relaxation.
5.2 Computational experiments for sparse PCA

We next report our computational results on sparse PCA. First, we summarize the following standard result.

**Proposition 5.2.** The following is a correct formulation for sparse PCA:

\[
\begin{align*}
\text{max} & \quad \text{trace}(\Sigma X) \\
\text{s.t.} & \quad \text{trace}(X) \leq 1 \\
& \quad X \succeq 0 \\
& \quad \text{diag}(X) \leq z \\
& \quad \sum_j T_{ij} = 1 \\
& \quad z \in \{0, 1\}^n.
\end{align*}
\]  

(49a)

(49b)

(49c)

(49d)

(49e)

(49f)

Let \((X^*, z^*)\) be an optimal solution to (49) and \(\lambda_i, x_i, x_i^\top\) be an eigenvalue decomposition of \(X^*\) so that each \(x_i\) is a unit eigenvector. Then, for any \(i'\) such that \(\lambda_{i'} > 0\), \(x_{i'}\) is an optimal solution to sparse PCA.

**Theorem 5.2.** The following constraints are valid for sparse PCA:

\[
\begin{align*}
Y_{ij}^2 & \leq T_{ij} T_{ji} \quad i = 1, \ldots, n, j = i + 1, \ldots, n \\
T_{ii} & = Y_{ii} \quad i = 1, \ldots, n \\
\sum_{j=1}^n T_{ij} & = z_i \quad i = 1, \ldots, n \\
\sum_{i=1}^n T_{ij} & = KY_{jj} \quad j = 1, \ldots, n \\
0 & \leq T_{ij} \leq Y_{jj} \quad i = 1, \ldots, n, j = 1, \ldots, n.
\end{align*}
\]  

(50a)

(50b)

(50c)

(50d)

(50e)

In particular, \(T_{ij}\) may be regarded as a linearization of \(z_i y_{ij}^2\). Using that \(Y_{ij} = Y_{ji}\) for all \((i, j)\), the above constraints imply

1. for all \((i, j)\), \(T_{ij} + T_{ji} \geq 2Y_{ij}\);
2. for all \(i\), \(Y_{ii} \leq z_i\) and, for all \((i, j)\), with \(i \neq j\), \(2Y_{ij} \leq z_i\);
3. for all \(S \subseteq \{1, \ldots, n\}\), \(\sum_{i,j \in S} Y_{ij} \leq \sum_{i \in S} z_i\);
4. for all \(i\), \(\sum_{j=1}^n Y_{ij}^2 \leq z_i Y_{ii}\);
5. for all \(i\), \(\sum_{j=1}^n Y_{ij}^2 \leq T_{ii} \) as long as, for all \(j'\), \(Y_{ij'}^2 \leq Y_{ii} Y_{jj'}\) or, more specifically, \(Y \succeq 0\).
6. \(\|X\| \leq K\), as long as (37) holds.
7. \(\sum_{j=1}^n X_{ij}^2 \leq z_i X_{ii}\) as long as (37), and \(\text{trace}(Y) = \text{trace}(X)\) hold.
8. \(\sum_{j=1}^n X_{ij}^2 \leq T_{ii}\) as long as (37) holds, \(\text{trace}(Y) = \text{trace}(X)\), and \(X \succeq 0\).

**Proof.** Proof. To show that the constraints are valid, we only need to show that \(T_{ij}\) can be chosen to be \(z_i y_{ij}^2\). We may assume that for all \((i, j)\), \(Y_{ij} = y_{ij}\). Constraint (50a) follows since for \(i' \in \{i, j\}\), \(y_{i'} = z_i y_{i'}\) implies that \((y_{y'} y')^2 \leq z_i y_{i'}^2 z_j y_{j'}^2\). Constraint (50b) follows since \(z_i Y_{ii} = Y_{ii}\). Constraint (50c) follows since \(\text{trace}(Y) = 1\) implies \(\sum_{j=1}^n z_i Y_{jj} = z_i\). Constraint (50d) is valid because (49e) implies that \(\sum_{i=1}^n z_i Y_{jj} = KY_{jj}\). Finally, (50e) follows because it relaxes \(y_{i'}^2 = z_i y_{i'}^2 = Y_{ii}\).

We now show the implications claimed. First, we show Statement 1. To see this, observe that we can write (50a) as \((2Y_{ij})^2 + (T_{ij} - T_{ji})^2 \leq (T_{ij} + T_{ji})^2\). Then,

\[2Y_{ij} \leq \sqrt{(2Y_{ij})^2 + (T_{ij} - T_{ji})^2} \leq \sqrt{T_{ij}^2 + T_{ji}^2} = T_{ij} + T_{ji},\]

(51)
where the first inequality is because \((T_{ij} - T_{ji})^2 \geq 0\), the second inequality is because of \((50a)\), and the equality is because, by \((50e)\), \(T_{ij}\) and \(T_{ji}\) are non-negative. Second, we show Statement 2. Clearly, \(Y_{ii} = T_{ii} \leq z_i\), where the first equality is by \((50b)\) and the inequality follows from \((50c)\). If \(i \neq j\), we have:

\[
2Y_{ij} \leq T_{ij} + T_{ji} \leq T_{ij} + Y_{ii} = T_{ij} + T_{ii} \leq z_i ,
\]

where the first inequality is by Statement 1, the second inequality is because of \((50e)\), the equality is because of \((50b)\), and the inequality because \(i \neq j\) and \((50c)\). Third, we show Statement 3. This is because

\[
0 \leq \sum_{i,j \in S} (T_{ij} - Y_{ij}) \leq \sum_{i,j \in S} T_{ij} - \sum_{i,j \in S} Y_{ij} = \sum_{i \in S} z_i - \sum_{i,j \in S} Y_{ij},
\]

where the first inequality is because of Statement 1 and \(Y_{ij} = Y_{ji}\), the second inequality is because of \((50e)\), and the equality is by \((50c)\). Now, we show Statement 4 as follows:

\[
\sum_{j=1}^{n} Y_{ij}^2 \leq \sum_{j=1}^{n} T_{ij} \leq \sum_{j=1}^{n} T_{ij} = z_i Y_{ii},
\]

where the first inequality follows from \((50a)\), the second inequality from \((50e)\), and the equality from \((50c)\). To see Statement 5, observe that

\[
\sum_{j=1}^{n} Y_{ij}^2 \leq \sum_{j=1}^{n} Y_{ii} \sum_{j=1}^{n} Y_{jj} = Y_{ii} = T_{ii},
\]

where the first inequality is because, for all \(j\) we have \(Y_{ij}^2 \leq Y_{ii}Y_{jj}\), the first equality is because \(\text{trace}(Y) = 1\) and the third equality is by \((50b)\). Next, we show Statement 6. We write

\[
\mathbb{1}^T |X| \mathbb{1} = \mathbb{1}^T |V - W| \mathbb{1} \leq \mathbb{1}^T (|V| + |W|) \mathbb{1} = \mathbb{1}^T Y \mathbb{1} \leq \mathbb{1}^T z = K,
\]

where the second equality is because \(Y = V + W, V \geq 0, W \geq 0\), and the first inequality is because of \((53)\) with \(S = \{1, \ldots, n\}\), and the last equality is because of \((49e)\). To show Statement 7, we write

\[
\sum_{j=1}^{n} X_{ij}^2 = \sum_{j=1}^{n} (V_{ij} - W_{ij})^2 \leq \sum_{j=1}^{n} (V_{ij} + W_{ij})^2 = \sum_{j=1}^{n} Y_{ij}^2 \leq z_i Y_{ii} = z_i X_{ii},
\]

where the first two equalities and first inequality are by \((37)\), the second equality is because of Statement 4. The last equality is because \((37)\) implies \(\text{diag}(Y - X) \geq 0\). Together with \(\text{trace}(Y - X) = 0\), this implies that, for all \(i\), \(Y_{ii} = X_{ii}\). Finally, the proof of Statement 8 is similar to that for Statement 5 where \(Y\) is replaced with \(X\), where we also utilize that \(\text{trace}(Y - X) = 0\), \(\text{diag}(Y - X) \geq 0\), and \((50b)\) imply that \(X_{ii} = T_{ii}\) for all \(i\).

We use the following formulation, which we refer to as the \(T\)-formulation, to find the optimal solution for sparse PCA. We chose to develop this formulation around the diagonal relaxation obtained using constraints \((44)\) as it is simple to implement but, as we will show later, it is also strong.

\[
(T) : \max \text{ trace}(\Sigma X)
\]

\[
\text{s.t.} \quad (37), (38), (39), (44), (49e), (49f), (50)
\]

\[
\text{trace}(U) = \text{trace}(Y) = \text{trace}(X) = 1
\]

\[
Y_{ij}^2 \leq Y_{ii}Y_{jj} \quad i = 1, \ldots, n, j = i + 1, \ldots, n
\]

\[
X \succeq 0, U \succeq 0.
\]

To prove that this formulation is correct, we only need to show that \((49d)\) is satisfied. To see this observe that \(X_{ii} \leq Y_{ii} \leq z_i\), where the first inequality follows from \((37)\), and the second inequality from \((52)\). Instead of requiring that \(Y \succeq 0\), we relax it so that \(Y_{ij}^2 \leq Y_{ii}Y_{jj}\) for all \((i, j)\). Notice that the constraints \(Y_{ij}^2 \leq T_{ij}T_{ji}\) and \(Y_{ij}^2 \leq Y_{ii}Y_{jj}\) can be written as SOCP constraints \(||(2Y_{ij}, T_{ij} - T_{ji})||_2 \leq T_{ij} + T_{ji}\) and \(||(2Y_{ij}, Y_{ii} - Y_{jj})||_2 \leq Y_{ii} + Y_{jj}\), respectively. We refer to the relaxation obtained by dropping \((49f)\) as the \(T\)-relaxation.
After the initial draft of this paper [18], the following relaxation for sparse PCA was proposed in [7]:

\[
\begin{align*}
\max & \quad \text{trace}(\Sigma X) \\
\text{s.t.} & \quad (37), (49e), (49f) \\
& \quad \text{trace}(X) = 1 \\
& \quad Y_{ii} \leq z_i \quad i = 1, \ldots, n \quad (57a) \\
& \quad 2Y_{ij} \leq z_i \quad i = 1, \ldots, n, \quad j = i + 1, \ldots, n \quad (57b) \\
& \quad \sum_{j=1}^{n} X_{ij}^2 \leq z_i X_{ii} \quad i = 1, \ldots, n \quad (57c) \\
& \quad \mathbf{1}^T \mathbf{Y} = K \\
& \quad X \succeq 0.
\end{align*}
\]

We refer to the relaxation as the \textit{B-relaxation}. Theorem 5.2 shows that \textit{B-relaxation} is dominated by \textit{T-relaxation} since (57a)-(57c) are implied in the \textit{T-relaxation}.

We remark that the relaxations we develop here do not intrinsically depend on trace\((X) = 1\). In particular, when the constraint (50c) is replaced with \(\sum_{j=1}^{n} T_{ij} \leq z_j \text{trace}(T)\) and the constraint \(\text{trace}(U) = \text{trace}(Y) = \text{trace}(X) = 1\) is replaced with \(\text{trace}(U) = \text{trace}(Y) = \text{trace}(X)\) our relaxations can be used more generally. More specifically, they are valid whenever we seek a rank-one matrix \(X\), that is symmetric and positive-definite, and is such that only \(K\) rows and columns have non-zero values.

We next define the notations used in the computational experiments. We refer to the relaxation obtained by dropping constraints (44) and (45) from (46) as the \textit{rowsum relaxation}. Also, we refer to the relaxation obtained by dropping constraints (43) and (45) from (46) as the \textit{diagonal relaxation}. We denote the optimal value of the \(D-, B-, \text{rowsum, diagonal, 2-step, submatrix, and } T\)-relaxation by \(z_D^*, z_B^*, z_{\text{rowsum}}^*, z_{\text{diag}}^*, z_{\text{2step}}^*, z_{\text{submat}}^*, \) and \(z_T^*\), respectively. We report test results for these relaxations in Tables 3 and 5.

We use CVX [12, 11] version 2.2 or \textsc{YALMIP} [23] version R20210331 to solve SDPs in the experiments. \textsc{Mosek} [2] version 9.2.47 was selected as the SDP solver in both cases. To measure the relative tightness of a relaxation when compared to (35), we calculate \textit{gap closed} as

\[
\left(\frac{z_D^* - z_{\text{SDP}}^*}{z_D^* - z^*}\right) \times 100.
\]

where \(z_{\text{SDP}}^*\) is the target SDP relaxation on which we calculate the gap closed. Here, \(z^*\) denotes the optimal value of sparse PCA and we obtain this value by solving the \(T\)-formulation of sparse PCA, \((T)\), to optimality using the \texttt{bnb} solver of \textsc{Yalmip} with \textsc{Mosek} version 9.1.9.

In addition, we conducted experiments with an alternate formulation. This formulation is based on a compact extended formulation of the permutahedron proposed by Goemans [10], where the construction and the size of the reformulation depend on the choice of a sorting network. While the optimal Ajtai–Komlós–Szemerédi sorting network gives an extended formulation for the permutahedron with \(\Theta(n \log n)\) variables and inequalities, it has limited practical value because the constant hidden in the \(\Theta(\cdot)\) notation is very large. Instead, we used Batcher’s bitonic sorting network that gives an extended formulation with \(\Theta(n \log^2 n)\) variables and inequalities. When it comes to the time comparison between these two reformulations, we use the diagonal relaxation among the aforementioned SDP relaxations because, as we show later, this relaxation produces competitive bounds and is simple to implement. Since the formulation based on a sorting network is equivalent to the duality-based formulation based on (4), the quality of bounds is the same, and we only compare their computation times. The result is summarized in Tables 4 and 7.

The experiments are performed with an Intel Core i5-10400 machine containing a 2.90GHz CPU, 32GB RAM, running Windows 10.

### 5.2.1 pitprops problem

The \textit{pitprops} problem [16] is one of the most commonly used problems for sparse PCA algorithms. The instance has 13 variables and 180 observations. Table 3 shows the test results for cardinality \(K = 3, \ldots, 10\).
Observe that the diagonal (resp. submatrix) relaxation reduces the gaps of (35) by more than 88% (resp. 96%), returning global optimal solutions for three (resp. five) problems. The $T$-relaxation attains the optima except for the instance $K = 10$. On average, it reduces the gaps of (35) by 99.81%.

For all computational times reported in Table 3, we used (44) to model the majorization constraints. However, as we had discussed earlier, these constraints can also be formulated using Batcher’s bitonic sorting network, which only requires $\Theta(n \log^2 n)$ variables. Our primary intention here is to compare the formulation of the permutahedron based on (44) with that obtained via Batcher’s bitonic sorting network. For this comparison, we use the diagonal relaxation and simplify it by dropping the requirement that $Y \succeq 0$. We refer to this simplified relaxation as Diagonal-relaxation. To differentiate the relaxation based on sorting-network, we will refer to it as Diagonal-relaxation-sort. Clearly, both relaxations yield the same bound. Therefore, we only report on the solution times when comparing these relaxations. As $n$ and $K$ are small, no significant difference between the reformulations is observed.

| $K$ | $D$-relaxation | $B$-relaxation | Rowsum relaxation | Diagonal relaxation |
|-----|----------------|----------------|------------------|-------------------|
|     | $z^*$ | $z^*_D$ | $z^*_B$ | Time (sec) | Gap clsd | $z^*_{rowsum}$ | Time (sec) | Gap clsd | $z^*_diag$ | Time (sec) | Gap clsd |
| 3   | 2.4753 | 2.5218 | 2.4947 | 0.20 | 49.60 | 2.5033 | 0.26 | 39.78 | 2.4949 | 0.24 | 57.86 |
| 4   | 2.9375 | 3.0172 | 2.9917 | 0.22 | 31.98 | 2.9766 | 0.27 | 50.89 | 2.9671 | 0.25 | 62.83 |
| 5   | 3.4062 | 3.4581 | 3.4303 | 0.22 | 53.60 | 3.4103 | 0.27 | 91.92 | 3.4072 | 0.26 | 97.96 |
| 6   | 3.7710 | 3.8137 | 3.7886 | 0.22 | 58.80 | 3.7710 | 0.26 | 100.00 | 3.7710 | 0.27 | 100.00 |
| 7   | 3.9962 | 4.0316 | 3.9967 | 0.22 | 98.69 | 3.9962 | 0.24 | 100.00 | 3.9962 | 0.26 | 100.00 |
| 8   | 4.0686 | 4.1448 | 4.0839 | 0.22 | 79.93 | 4.0770 | 0.27 | 88.49 | 4.0721 | 0.29 | 95.48 |
| 9   | 4.1386 | 4.2063 | 4.1388 | 0.20 | 98.28 | 4.1399 | 0.26 | 98.20 | 4.1386 | 0.28 | 100.00 |
| 10  | 4.1726 | 4.2186 | 4.1778 | 0.23 | 88.87 | 4.1807 | 0.26 | 82.42 | 4.1766 | 0.26 | 91.32 |

Table 3: Optimal values and gaps closed for the test problem pitprops

| $K$ | Average | Average | Average | Average |
|-----|---------|---------|---------|---------|
|     | 2.4753  | 2.9477  | 3.4062  | 3.7710  |
|     | 0.68    | 0.82    | 1.23    | 1.52    |
|     | 100.00  | 100.00  | 100.00  | 100.00  |
|     | 2.4753  | 2.9477  | 3.4062  | 3.7710  |
|     | 0.80    | 1.85    | 4.26    | 8.81    |
|     | 100.00  | 100.00  | 100.00  | 100.00  |
|     | 2.4753  | 2.9477  | 3.4062  | 3.7710  |
|     | 0.80    | 1.85    | 4.26    | 8.81    |
|     | 100.00  | 100.00  | 100.00  | 100.00  |
|     | 2.4753  | 2.9477  | 3.4062  | 3.7710  |
|     | 0.80    | 1.85    | 4.26    | 8.81    |
|     | 100.00  | 100.00  | 100.00  | 100.00  |
|     | Average | 2.4753  | 2.9477  | 3.4062  |
|     | 0.68    | 0.82    | 1.23    |
|     | 100.00  | 100.00  | 100.00  |

Table 4: Computation time (in seconds) comparison with the diagonal relaxation for pitprops problems

5.2.2 Experiments with randomly generated matrices

We next report test results for randomly generated covariance matrices. Random matrices are generated using the following procedure. First, we choose a random integer $m \in \{1, \ldots, n\}$ for the number of nonzero eigenvalues of the matrix by setting $m = \lceil nU \rceil$ where $U$ is randomly drawn from the uniform distribution $U(0, 1)$. Second, we generate $m$ random vectors $v_i \in \mathbb{R}^n \sim \mathcal{N}(0, I_n)$, for $i = 1, \ldots, m$ for rank-1 matrices. Third, we generate $m$ positive random eigenvalues $\lambda_i \sim U(0, 1)$, for $i = 1, \ldots, m$. Finally, we construct the desired random covariance matrix as $\Sigma = \sum_{i=1}^m \lambda_i v_i v_i^\top$.

The tests are performed for problems with size $n \in \{5 + 5i \mid i = 1, \ldots, 9\}$, and the cardinality $K = \text{round}(n/6)$ is chosen to reflect the motivation of sparse principal components analysis to produce sparse vectors, where $\text{round}(x)$ represents the integer closest to $x$ (i.e., $\text{round}(x) = \lceil x \rceil$ as long as the fractional part of $x$ is strictly less than 0.5 and $\lfloor x \rfloor$ otherwise). For each $n$ and the associated $K$, 30 instances generated from random covariance matrices are tested. The reported computation times and gaps closed are the averages for the 30 instances. In Table 5, we present average gap closed for the computation times in seconds and gap closed of the relaxations. The computational results show that our SDP relaxations improve the gaps of the SDP relaxation (35) significantly.
Among the relaxations in Table 5, the $T$-relaxation yields the tightest bound on our instances. We remark that the $T$-relaxation bound improves when the constraints for the 2-step relaxation and/or the submatrix relaxation are also imposed. However, we do not report on the performance of these relaxations since they are more complex and more computationally expensive to solve. For larger dimensional instances, we choose the $B$-relaxation and the $T$-relaxation to compare. For $n \in \{10, 15, \ldots, 95, 100\}$, we generate 30 random covariance matrices for each $n$ and summarize the results of the comparison in Table 6 and Figure 2, where the computation times are in minutes.

Table 6: Comparison between the $B$-relaxation and the $T$-relaxation with $n \in \{55, 60, \ldots, 95, 100\}$ and $K = \text{round}(n/6)$

We next show that replacing (44) with a sorting network helps reduce the solution time for the Diagonal relaxation by comparing the Diag′-relaxation and Diag′-relaxation-sort on the synthetic data sets. For this demonstration, we consider dimensions $n \in \{10i \mid i = 1, \ldots, 15\}$ and choose $K = \text{round}(n/6)$. For each choice of $n$ and the associated $K$, 30 randomly generated instances are tested; see Table 7 and Figure 3 for a comparison of the computation times (in minutes). Our results show that the sorting network based formulation reduces the computational time for the relaxation.
In this paper, we present an explicit convex hull description of permutation-invariant sets and applications of the results to various important sets/functions in optimization. The construction of the convex hull is based on the fact that a permutation-invariant set is a union of permutahedra with generating vectors in \( \Delta = \{ x \in \mathbb{R}^n \mid x_1 \geq \cdots \geq x_n \} \) and the convex hull of this union can be described in closed-form. We then applied this result to derive various convexification results. First, we presented an extended formulation for the convex hull of permutation-invariant norm balls constrained by a cardinality requirement. Second, we convexified sets of matrices that are characterized using functions of their singular values. Third, we derived convex/concave envelopes of various nonlinear functions and convex hulls of sets defined using nonlinear functions when bounds for variables are congruent. Fourth, we studied sets of rank-one matrices whose generating vectors lie in a permutation-invariant set. We use majorization inequalities in the space of generating vectors to construct valid inequalities for the convex hull in the matrix space. As a motivating example, we construct tight semidefinite programming relaxations for sparse principal component analysis and report computational results that show that our tightest relaxation reduces more than 99.8\% (resp. 97.9\%) of the gaps for the pitprops data set (resp. synthetic data sets with the dimensions up to \( n = 50 \)). The concept of permutation-invariance can be used to study a variety of other sets, including those arising from logical requirements in 0-1 mixed integer programming; see [18] for additional descriptions.

Acknowledgments.

The second and third authors wish to acknowledge the support from NSF CMMI Award #1727989 and #1917323, respectively.

---

### Table 7: Computation time comparison between the Diagonal'-relaxation and Diagonal'-relaxation-sort for sparse PCA with random covariance matrix, \( n = 10, 20, \ldots, 150 \) and \( K = \text{round}(n/6) \)

|       | \( n \) | 10  | 20  | 30  | 40  | 50  | 60  | 70  | 80  |
|-------|--------|-----|-----|-----|-----|-----|-----|-----|-----|
| Diagonal'-relaxation | # vars | 832 | 3261| 7289| 12917| 20146| 28974| 39402| 51431|
|       | # cons | 384 | 1469| 3254| 5739 | 8924 | 12809| 17394| 22679|
|       | Time (min) | 0.00 | 0.00| 0.01| 0.01 | 0.04 | 0.10 | 0.22 | 0.45 |
| Diagonal'-relaxation-sort | # vars | 848 | 3087| 6125| 11451| 16890| 23528| 34214| 43253|
|       | # cons | 411 | 1452| 2737| 5226 | 7511 | 10296| 15437| 19222|
|       | Time (min) | 0.00 | 0.00| 0.01| 0.01 | 0.03 | 0.07 | 0.17 | 0.36 |

![Figure 3: Computation time comparison between the Diagonal'-relaxation and Diagonal'-relaxation-sort for sparse PCA with random covariance matrix, \( n = 10, 20, \ldots, 150 \) and \( K = \text{round}(n/6) \)](image-url)
References

[1] Miklós Ajtai, János Komlós, and Endre Szemerédi. An $O(n \log n)$ sorting network. In Proceedings of the fifteenth annual ACM symposium on Theory of computing, pages 1–9, 1983.

[2] MOSEK ApS. Mosek optimization suite. 2017.

[3] Andreas Argyriou, Rina Foygel, and Nathan Srebro. Sparse prediction with the $k$-support norm. In Advances in Neural Information Processing Systems, pages 1457–1465, 2012.

[4] Egon Balas. Disjunctive programming. Springer, 2018.

[5] A. I. Barvinok. A course in convexity. American Mathematical Society, 2002.

[6] Ahron Ben-Tal and Arkadi Nemirovski. Lectures on modern convex optimization: analysis, algorithms, and engineering applications, volume 2. Siam, 2001.

[7] Dimitris Bertsimas, Ryan Cory-Wright, and Jean Pauphilet. Solving large-scale sparse PCA to certifiable (near) optimality. arXiv preprint arXiv:2005.05195, 2020.

[8] Alexandre d’Aspremont, Laurent El Ghaoui, Michael I Jordan, and Gert RG Lanckriet. A direct formulation for sparse pca using semidefinite programming. SIAM review, 49(3):434–448, 2007.

[9] Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary, and Ronald de Wolf. Exponential lower bounds for polytopes in combinatorial optimization. Journal of the ACM, 62(2):17:1–17:23, 2015.

[10] Michel X Goemans. Smallest compact formulation for the permutahedron. Mathematical Programming, 153(1):5–11, 2015.

[11] Michael Grant and Stephen Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, Recent Advances in Learning and Control, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. http://stanford.edu/~boyd/graph_dcp.html.

[12] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx, March 2014.

[13] G.H. Hardy, J.E. Littlewood, and G. Pólya. Inequalities. Cambridge Mathematical Library. Cambridge University Press, 1952.

[14] Jean-Baptiste Hiriart-Urruty and Hai Yen Le. Convexifying the set of matrices of bounded rank: applications to the quasiconvexification and convexification of the rank function. Optimization Letters, 6(5):841–849, 2012.

[15] Roger A Horn and Charles R Johnson. Matrix analysis. Cambridge university press, 2nd edition, 2012.

[16] JNR Jeffers. Two case studies in the application of principal component analysis. Applied Statistics, pages 225–236, 1967.

[17] Volker Kaibel and Kanstantsin Pashkovich. Constructing extended formulations from reflection relations. In Oktay Günlük and Gerhard J. Woeginger, editors, Integer Programming and Combinatorial Optimization - 15th International Conference, IPCO 2011, New York, NY, USA, June 15-17, 2011. Proceedings, volume 6655 of Lecture Notes in Computer Science, pages 287–300. Springer, 2011.

[18] Jinhak Kim, Mohit Tawarmalani, and Jean-Philippe P Richard. Convexification of permutation-invariant sets and applications. arXiv preprint arXiv:1910.02573, 2019.

[19] Adrian S Lewis. The convex analysis of unitarily invariant matrix functions. Journal of Convex Analysis, 2(1):173–183, 1995.

[20] Chi-Kwong Li and Paras P Mehta. Permutation invariant norms. Linear algebra and its applications, 219:93–110, 1995.

[21] Cong Han Lim. A note on extended formulations for cardinality-based sparsity. In 10th NIPS Workshop on Optimization for Machine Learning (NIPS 2017), 2017.

[22] Cong Han Lim and Stephen Wright. $k$-support and ordered weighted sparsity for overlapping groups: Hardness and algorithms. In Advances in Neural Information Processing Systems, pages 284–292, 2017.

[23] J. Löfberg. Yalmip : A toolbox for modeling and optimization in matlab. In In Proceedings of the CACSD Conference, Taipei, Taiwan, 2004.

[24] J. Luedtke, M. Namazifar, and J. Linderoth. Some results on the strength of relaxations of multilinear functions. Mathematical Programming, 136(2):325–351, 2012.
[25] Albert W Marshall, Ingram Olkin, and Barry Arnold. *Inequalities: theory of majorization and its applications*. Springer Science & Business Media, 2010.

[26] Jiří Matoušek. *Lectures on discrete geometry*, volume 212. Springer New York, 2002.

[27] Arkadi Nemirovski. Introduction to linear optimization ISYE 6661. *Lecture Notes in Georgia Institute of Technology*, 2012.

[28] Trang T Nguyen, Jean-Philippe P Richard, and Mohit Tawarmalani. Deriving convex hulls through lifting and projection. *Mathematical Programming*, 169(2):377–415, 2018.

[29] Richard Rado. An inequality. *Journal of the London Mathematical Society*, 1(1):1–6, 1952.

[30] Anatoliy D Rikun. A convex envelope formula for multilinear functions. *Journal of Global Optimization*, 10(4):425–437, 1997.

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

[32] M. Tawarmalani and J.-P. P. Richard. Decomposition techniques in convexification of sets. working paper, 2017.

[33] Andreas M Tillmann and Marc E Pfetsch. The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. *IEEE Transactions on Information Theory*, 60(2):1248–1259, 2013.

[34] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2):265–286, 2006.
# Table of Notations

The following table lists some notations used multiple times throughout the paper.

| Notation | Definition | Defined in |
|----------|------------|------------|
| $\text{card}(x)$ | the number of nonzero components of $x$ | Section 1 |
| $\| \cdot \|_1$ | $l_1$-norm of vectors | Section 1 |
| $\mathcal{M}^{m,n}(\mathbb{R})$ | the set of $m$-by-$n$ real matrices | Section 1, Section 3.1 |
| $\| \cdot \|_{sp}$ | spectral norm of matrices | Section 1, Section 3.1 |
| $\Delta_x$ | $\{ x \in \mathbb{R}^n \mid x \geq \cdots \geq x \}$ | Section 1 |
| $\Delta^n(= \Delta)$ | $\{ x \in \mathbb{R}^n \mid x_1 \geq \cdots \geq x_n \}$ | Section 1 (1) |
| $\text{conv}(X)$ | the convex hull of the set $X$ | Section 1 |
| $\mathcal{P}_k$ | $k$-by-$k$ permutation matrices | Section 2 |
| $x_{[i]}$ | $i$th largest element of $x$ | Section 2 Definition 2.1 |
| $x \geq_m y$ | $x$ majorizes $y$ | Section 2 Definition 2.1 |
| $x \geq_{wm} y$ | $x$ weakly majorizes $y$ from below | Section 2 Definition 2.1 |
| $\| \cdot \|_s$ | an arbitrary sign- and permutation-invariant norm of vectors | Section 3 (10) |
| $N^K_x$ | $\{ x \in \mathbb{R}^n \mid \| x \|_s \leq 1, \text{card}(x) \leq K \}$ | Section 3 (10) |
| $B_r(x)$ | $\{ x \in \mathbb{R}^n \mid \| x \|_s \leq r \}$ | Section 3 |
| $\Delta^n = \Delta^n_x$ | $\Delta^n \cap \mathbb{R}^n_+$ | Section 3 (12) |
| $x_{\Delta^n}(= x_{\Delta})$ | $\{ x_{\Delta^n}_i := x_{[i]} \}$ | Section 3 (12) |
| $x_{\Delta^n}(= x_{\Delta})$ | $\{ x_{\Delta^n}_i := | x_{[i]} | \}$ | Section 3 (12) |
| vert($P$) | the set of vertices of the polyhedron $P$ | Section 3 Proposition 3.1 |
| int($X$) | the set of interior points of $X$ | Section 3 Proposition 3.1 |
| $\| \cdot \|_c$ | the norm corresponding to the convex body $N^K_x$ | Section 3 |
| $s(x)$ | $s(x)_i = \sum_{k=i}^{K+i} |x|_k \cdot 1_{x_k \geq 0}$, $i = 1, \ldots, K$, $s(x)_0 = s(x)_{K+1} = +\infty$ | Section 3 |
| $\| x \|_c$ | $\arg \min \{ s(x)_i \mid i = 1, \ldots, K \}$ | Section 3 |
| $\delta(x)$ | $s(x)_i$ | Section 3 |
| $u(x)$ | $u(x)_i := \begin{cases} |x|_i, & i \in \{ 1, \ldots, i_s - 1 \} \\ \delta(x), & i \in \{ i_s, \ldots, K \} \\ 0, & \text{otherwise} \end{cases}$ | Section 3 |
| $\| \cdot \|_{sp}$ | $K$-support norm ($K$-overlap norm) | Section 3 |
| $\sigma(M)$ | the vector of singular value of the matrix $M$ | Section 3.1 |
| $\| \cdot \|_s$ | nuclear norm of matrices | Section 3.1 |
| diag($v$) | the diagonal matrix whose diagonal is $v$ | Section 3.1 |
| $\mathcal{S}^p$ | the set of $p$-by-$p$ symmetric matrices | Section 3.1 |
| $\mathcal{S}_p^+$ | the set of $p$-by-$p$ positive semidefinite matrices | Section 3.1 |
| $\lambda(M)$ | the vector of eigenvalues of the matrix $M$ | Section 3.1 |
| $\text{conv}_{C}(\phi)$ | the convex envelope of $\phi$ over $C$ | Section 4 |
| $\phi|_X$ | $\phi|_X (x) = \phi(x)$ for any $x \in X$ and $+\infty$ otherwise | Section 4 |
| $S(Z, a, b)$ | $\{ (x, z) \in [a, b]^n \times \mathbb{R}^m \mid (x, z) \in Z \}$ | Section 4 |
| $X(Z, a, b, \{ F_i \})_{i=1}^m$ | $\{ (x, z) \in [a, b]^n \times \mathbb{R}^m \mid (x, z) \in Z, x \in \bigcup_{i=1}^m F_i \}$ | Section 4 |
| $\mathcal{M}^n(= \mathcal{M})$ | $\{ (x, X) \in \mathbb{R}^n \times \mathcal{M}^n \mid X = x x^T, x \in S \}$ | Section 5 |
| $\mathcal{M}_n(= \mathcal{M})$ | $n$-dimensional vector of ones | Section 5 |
| $\mathcal{M}_{k,n}$ | $k$-by-$k$ matrix of ones | Section 5 |
| trace($M$) | the trace of the matrix $M$ | Section 5 (30b) |
| diag($M$) | the diagonal vector of the matrix $M$ | Section 5 |
| $\mathcal{R}^m_0$ | the set of row sums of the matrix $M$ | Section 5 |
| $\mathcal{M}^n_0$ | the set of $n$-by-$n$ non-negative matrices | Section 5.1 |
| $\mathcal{S}^n_0$ | the set of $n$-by-$n$ non-negative symmetric matrices | Section 5.1 |

Table 8: Table of notations
This figure "xy1.png" is available in "png" format from:

http://arxiv.org/ps/1910.02573v2
This figure "xy2.png" is available in "png" format from:

http://arxiv.org/ps/1910.02573v2