CONVEXIFICATION OF PERMUTATION-INVARIANT SETS

PURDUE UNIVERSITY MANAGEMENT DEPARTMENT WORKING PAPER NO 1315

Jinhak Kim
Mitchell College of Business
University of South Alabama
Mobile, AL 36688
jinhakkim@southalabama.edu

Mohit Tawarmalani
Krannert School of Management
Purdue University
West Lafayette, IN 47906
mtawarma@purdue.edu

Jean-Philippe P. Richard
Industrial and Systems Engineering
University of Minnesota
Minneapolis, MN 55455
jrichar@umn.edu

February 15, 2019

ABSTRACT

In this paper, we characterize the convex hull of a set, which does not change when variables are permuted, as a projection of a set in a higher-dimensional space. In particular, we show that as long as the set can be convexified after imposing an ordering on the constituent variables, the convex hull of the set can be written using a polynomial number of additional variables and constraints. Then, we explore applications of this result in a variety of contexts. We first study permutation-invariant norm balls intersected by a cardinality constraint. The convex hull can be seen as another norm ball and we provide an explicit formula that calculates values of the norm. Furthermore, we present how to separate an arbitrary point from the convex hull. Then, this idea is used to construct an extended formulation for the feasible set of the sparse principal component analysis [11] and present an alternative proof of the formula for K-support norm [1]. This idea can be extended to sets of matrices that are invariant under permutation of singular values. Using the conjugacy result by [19], we show that the convex hull of the set is directly obtained by the convex hull of its correspondence in the vector space. As a special case, we recover the proof of the convex hull result by [15]. We then generalize the result in the context of hyperbolic programming. Furthermore, we show a convex hull result and the semidefinite representability of rank-constrained permutation-invariant sets of matrices. We next use our convexification result to construct convex/concave envelops of permutation-invariant nonlinear functions over a symmetric box. At last, we study sets that are written by certain logical constraints or cardinality constraints. Another application is on sets of rank-one matrices where the generating vectors of the matrices are in a permutation-invariant set. We provide a variety of valid inequalities for the convex hull in a higher dimensional space. As a motivating example, we provide tight SDP relaxation for the sparse principal component analysis and present computational experiments, from which we show that our formulation reduces more than 90% of gaps generated by the baseline formulation by [11].

***K*eywords** Convexification · permutation-invariant sets · majorization

1 Introduction

In this paper, we study the convex hull 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. A more general formal definition is presented in Section 2.

Permutation-invariant sets appear in a variety of optimization problems. Sparse principal component analysis is to find a sparse vector that explains the most variance of the data. The problem to find the first sparse principal component is formulated in [11] as $\max\{x^\intercal \Sigma x \mid \text{card}(x) \leq K, \|x\| \leq 1\}$ where Σ is the covariance matrix of the given data. The feasible set of the formulation is permutation-invariant because the values of the cardinality and the l_2-norm are invariant under permutations. The convex hull of the feasible set is also known as a norm ball associated with the K-support norm [1] in the machine learning community, which is used to construct tighter relaxation for a sparsity set than the elastic net. Important sets of matrices are often represented solely in their singular values (or eigenvalues). Perhaps, one of the simplest of such sets in the context of a graph-invariance, as discussed in [10], is $\{X \in \mathbf{S}^n \mid \lambda(X) = y\}$ where $\mathbf{S}^n$ is the set of $n \times n$ symmetric matrices and $\lambda(X)$ is the vector of eigenvalues sorted in descending order and y is a vector of constants. In [15], the authors studied the set of the form $\{M \in \mathcal{M}_{m,n}(\mathbb{R}) \mid \text{rank}(M) \leq K, \|M\|_{sp} \leq r\}$ where $\mathcal{M}_{m,n}(\mathbb{R})$ is the set of $m \times n$ real matrices and $\|M\|_{sp}$ is the spectral norm of M. The elements of the set are characterized by their spectral values because the rank of a matrix equals the cardinality of the vector of spectral values and the spectral norm of matrix equals the largest spectral value of the matrix. In nonlinear optimization, characterizing or approximating the convex envelope of a multilinear function $\prod_{i=1}^n x_i$ over a symmetric box $[a,b]^n$ is an important tool in global optimization. The convex envelope of the function corresponds to the convex hull of its epigraph over $[a,b]^n$ and the epigraph $\{(x,y) \in \mathbb{R}^n \times \mathbb{R} \mid \prod_{i=1}^n x_i \leq y, x \in [a,b]^n\}$ is permutation-invariant with respect to $x \in \mathbb{R}^n$. In the convex of stochastic ordering, [12] studied the convex hull of the permutation-invariant set $\{x \in \mathbb{R}^n \mid y_{[i]} \geq x_{[i]}, i = 1, \dots, n\}$ for a fixed vector $y \in \mathbb{R}^n$ where $x_{[i]}$ represents the ith largest component of an arbitrary vector x. Another classic example presented in [24] is the set $\{x \in \{0,1\}^n \mid \sum_{i=1}^n x_i \in S\}$ and its convex hull where S is a subset of $\{1, \dots, n\}$.

A permutation-invariant set can be represented as a disjunctive set because it can be seen as a union of $n!$ sets of the form $S \cap \{x \mid x_{\pi(1)} \geq \cdots \geq x_{\pi(n)}\}$ where π is an n-dimensional permutation. It is often observed that each set $S \cap \{x \mid x_{\pi(1)} \geq \cdots \geq x_{\pi(n)}\}$ has polyhedral or polynomial description while S does not. When each set $S \cap \{x \mid x_{\pi(1)} \geq \cdots \geq x_{\pi(n)}\}$ is a polyhedron, an extended formulation for the convex hull of S can be obtained using disjunctive program [2, 3], which obvious is unfavorable because of its excessively high dimensionality. In this paper, we provide an explicit polynomially sized extended formulation for the convex hull of permutation-invariant sets without using disjunctive programming. The outline of the construction is as follows: we first take a permutation-invariant set S and assume that the convex hull conv$(S \cap \Delta)$ can be easily constructed where $\Delta := \{x \in \mathbb{R}^n \mid x_1 \geq \cdots \geq x_n\}$. Then, the convex hull is simply the union of permutahedra where each permutahedron is generated by a point in conv$(S \cap \Delta)$. We model each permutahedron using polynomially many linear equalities and inequalities and obtain an extended formulation for conv(S) using the notion of majorization and convexity and linear representability of sum of j largest entries of a variable. Similar construction can be employed for sign-invariant sets and the convex hull results are summarized in Theorem 2.7. We then show how the results can be applied to recover or improve existing convexification or relaxation results.

The remainder of the paper is organized as follows. The main convexification results for permutation- and/or sign-invariant sets are presented in Section 2. Then, we explore various applications of the results in the following sections. In Section 3, we study cardinality-constrained permutation-invariant norm balls and their convex hulls. The resulting convex hull can be seen as a norm ball and we provide explicit formula that calculates the values of the norm, so that it is easy to determine whether an arbitrary point is in the convex hull or not. Furthermore, we provide separation inequalities for an arbitrary point. Some special cases of this class of sets are also presented. In particular, we study the connection between permutation-invariant sets and sets of matrices characterized only by their singular values. Furthermore, we present semidefinite-representability of rank-constrained sets of matrices. (Outline for the multilinear section needs to be added.) (Outline for the logical constraint section needs to be added.) In Section 7, we study the set of rank-one matrices whose generating vectors lie in a permutation-invariant set. We next construct semidefinite programming relaxations of the convex hull by proposing various valid inequalities derived from the rank-one condition of the matrix and the majorization inequalities. We then report results of computational experiments on sparse principal component analysis to see how tight our relaxation is compared to the classic baseline relaxation proposed by [11].

2 Main Result

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. We next provide definitions for these properties.

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* $(x,z) \in S$ implies that $(Px, z) \in S$ for all permutation matrices $P \in \mathcal{P}_n$. A function $f(x,z) : \mathbb{R}^n \times \mathbb{R}^p \mapsto \mathbb{R}$ is called *permutation-invariant with respect to* x if $f(x,z) = f(Px,z)$ for all permutation matrices $P \in \mathcal{P}_n$. Any permutation-invariant set S can be written as a sublevel set $S = \{(x,z) : f(x,z) \leq 1\}$ of a permutation-invariant function f.

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 $(\bar{x}, z) \in S$ for all $\bar{x}$ that satisfy $|\bar{x}| = |x|$.

Lemma 2.1 gives 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 $T \in \mathbb{R}^{n\times n}$ and let $S \subseteq \{(x,z) \in \mathbb{R}^n \times \mathbb{R}^p\}$ be such that for each $(x,z) \in S$, $(Tx,z) \in S$ as well. Then, if $(x,z) \in \text{conv}(S)$, $(Tx,z) \in \text{conv}(S)$.*

Proof. An arbitrary $(x,z) \in \text{conv}(S)$ can be written as a convex combination $(x,z) = \sum_i \lambda_i(x^i,z^i)$ where $\lambda_i \geq 0$ for all i, $\sum_i \lambda_i = 1$, and $(x^i,z^i) \in S$ for all i. Then, $(Tx,z) = \sum_i \lambda_i(Tx^i,z^i)$. Observe that $(Tx^i,z^i) \in S$ because of the assumed property for S. Therefore, $(Tx,z) \in \text{conv}(S)$. □

It follows easily from Lemma 2.1 that if S is permutation-invariant (*resp.* sign-invariant) that 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 $i = 1,\ldots,n$. Given two vectors $x, y \in \mathbb{R}^n$, we say that x *majorizes* y, a property we denote by $x \geq_m y$, if

1. $\displaystyle\sum_{i=1}^{j} x_{[i]} \geq \sum_{i=1}^{j} y_{[i]}$ for $j = 1,\ldots,n$, and
2. $\displaystyle\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}$.

The result of Lemma 2.2 relates majorization and permutation. Its proof follows directly by combining the result of Hardy, Littlewood, and Pólya's theorem with that of Birkhoff's theorem; see 2.B.2 and 2.A.2 of [22] for descriptions.

Lemma 2.2. *If $x \geq_m y$ then y is a convex combination of x and its permutations.*

We say that x *weakly majorizes y from below* if

$$\sum_{i=1}^{j} x_{[i]} \geq \sum_{i=1}^{j} y_{[i]}, \quad \forall j = 1,\ldots,n.$$

We denote this relation by $x \geq_{wm} y$. Similarly, we say that x *weakly majorizes y from above* if

$$\sum_{i=j}^{n} x_{[i]} \leq \sum_{i=j}^{n} y_{[i]}, \quad \forall j = 1,\ldots,n$$

and denote this relation by $x \geq^{wm} y$.

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

$$Y := \left\{ (x,u,z) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^p \;\middle|\; \begin{array}{l} (u,z) \in K, \\ u \geq_m x, \\ u_1 \geq \cdots \geq u_n \end{array} \right\}$$

is convex.

Proof. First, observe that $\sum_{i=1}^{j} u_{[i]} = \sum_{i=1}^{j} u_i$ since $u_1 \geq \cdots \geq u_n$. 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, $\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} x_i$ and is, therefore, linear. Therefore, Y has the following convex representation:

$$Y = \left\{ (x,u,z) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^p \;\middle|\; \begin{array}{l} (u,z) \in K, \\ \sum_{i=1}^{j} u_i \geq \sum_{i=1}^{j} x_{[i]}, \text{ for } j = 1,\ldots,n-1, \\ \sum_{i=1}^{n} u_i = \sum_{i=1}^{n} x_i, \\ u_1 \geq \cdots \geq u_n \end{array} \right\}.$$

□

Theorem 2.4. *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,*

$$\text{conv}(S) = X := \left\{ (x,z) \;\middle|\; \begin{array}{l} (u,z) \in \text{conv}(S_0), \\ u \geq_m x \end{array} \right\}, \tag{1}$$

where S_0 is any set that satisfies:

$$\text{conv}(S) \cap \{(u,z) \mid u_1 \geq \cdots \geq u_n\} \supseteq S_0 \supseteq S \cap \{(u,z) \mid u_1 \geq \cdots \geq u_n\}.$$

Proof. The convexity of X follows from Lemma 2.3 because $\text{conv}(S_0) \subseteq \text{conv}(S) \cap \{(u,z) : u_1 \geq \cdots \geq u_n\}$ implies that X satisfies $u_1 \geq \cdots \geq u_n$. 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 $u_i = x_{[i]}$ 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 show that $X \subseteq \text{conv}(S)$. Let $(x,z) \in X$. We show that it 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 in x. Now, we show that $(x,z) \in \text{conv}(S)$ by expressing (x,z) as a convex combination of points $(P_i u, z)$ for some permutation matrices P_i. The result follows since $(P_i u, z) \in \text{conv}(S)$ by permutation-invariance of $\text{conv}(S)$ with respect to u. To express (x,z) as a convex combination, observe that $u \geq_m x$ implies that there exists a doubly stochastic matrix Π such that $x = \Pi u$. By Birkhoff's theorem [8], we can write any doubly stochastic matrix as a convex combination of permutation matrices. Hence

$$(x,z) = (\Pi u, z) = \left(\left(\sum_i \lambda_i P_i\right) u, z\right) = \sum_i \lambda_i \left(P_i u, z\right),$$

where P_i are permutation matrices, $\lambda_i \geq 0$ for all i, and $\sum_i \lambda_i = 1$. ☐

Theorem 2.4 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. In order for this explicit description to be useful, we make use of well-known ways to formulate the condition; see Section 3.3.4 of [25] for instance.

A natural way to model the convex function $\sum_{i=1}^{j} x_{[i]}$ is to express it as the value function of an optimization problem. Given $j \in \{1,\ldots,n-1\}$ and real numbers $x_1,\ldots,x_n$, consider the optimization problem

$$\begin{array}{rll} \max & \sum_{i=1}^n x_i s_i & \\ \text{s.t.} & \sum_{i=1}^n s_i = j, & \\ & 0 \leq s_i \leq 1, & i = 1,\ldots,n. \end{array} \tag{2}$$

Formulation (2) is not directly amenable to inclusion in the result of Theorem 2.4. However, we intend to formulate that $\sum_{i=1}^{j} u_i \geq \sum_{i=1}^{n} x_i s_i$ for all s_i in a polytope. The natural way to model this problem is to take the dual of (2) which converts the "for-all" quantifier to "there-exists" quantifier. We include the dual formulation below:

$$\begin{array}{rrll} \text{LS}(j): & \min & jr + \sum_{i=1}^n t_i & \\ & \text{s.t.} & x_i \leq t_i + r, & i = 1,\ldots,n, \\ & & t_i \geq 0, & i = 1,\ldots,n. \end{array} \tag{3}$$

Since (2) is clearly feasible, (3) exhibits no duality gap and also models $\sum_{i=1}^{j} x_{[i]}$. Then, the constraint $\sum_{i=1}^{j} u_i \geq \sum_{i=1}^{j} x_{[i]}$ can be expressed as the requirement that there exists an (r,t) satisfying the feasibility constraints of (3) such that $\sum_{i=1}^{j} u_i \geq jr + \sum_{i=1}^{n} t_i$.

Theorem 2.5. *Suppose $S \subseteq \{(x,z) : \mathbb{R}^n \times \mathbb{R}^p\}$ is permutation-invariant with respect to x. Then,*

$$\text{conv}(S) = \left\{ (x,z) \;\middle|\; \begin{array}{lll} (u,z) \in \text{conv}(S_0), & & \\ u_1 \geq \cdots \geq u_n, & & \\ \sum_{i=1}^n u_i = \sum_{i=1}^n x_i, & & \\ \sum_{i=1}^j u_i \geq jr^j + \sum_{i=1}^n t_i^j, & j = 1,\ldots,n-1, & \\ x_i \leq t_i^j + r^j, & j = 1,\ldots,n-1, & i = 1,\ldots,n, \\ t_i^j \geq 0, & j = 1,\ldots,n-1, & i = 1,\ldots,n, \end{array} \right\}. \tag{4}$$

We next present a similar convexification result for sign-invariant sets.

Theorem 2.6. *Suppose $S \subseteq \{(x,z) \in \mathbb{R}^n \times \mathbb{R}^p\}$ is sign-invariant with respect to x. Then,*

$$\text{conv}(S) = X := \{(x,z) \mid (u,z) \in \text{conv}(S_0), u \geq |x|\} \tag{5}$$

where $S_0 = S \cap \left(\mathbb{R}^n_+ \times \mathbb{R}^p\right)$.

Proof. It is clear that 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 $(u,z) \in S_0 \subseteq \text{conv}(S_0)$ and, by definition, u satisfies $u \geq |x|$.

We next show that $X \subseteq \text{conv}(S)$. Let $(x,z) \in X$. Then, 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 $\{(\bar{x},z) \mid \bar{x}_i \in \{u_i, -u_i\}\} \subseteq \text{conv}(S)$. Therefore, $(x,z) \in \{(\bar{x},z) \mid |\bar{x}_i| \leq u_i\} \subseteq \text{conv}(S)$, where the containment follows from the convexity of $\text{conv}(S)$. ☐

The above convexification results can be easily extended to the sets which are permutation-invariant or/and sign-invariant with respect to multiple sets of variables. For each positive integer n, let $\Delta^n = \{u \in \mathbb{R}^n \mid u_1 \geq \cdots \geq u_n\}$.

Theorem 2.7. *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 all $k = 1, \ldots, m$. Then,*
$$\text{conv}(S) = \left\{(x^1, \ldots, x^m, z) \,\middle|\, \begin{array}{l} (u^1, \ldots, u^m, z) \in \text{conv}(S_0), \\ u^k \geq_m x^k, k = 1, \ldots, m \end{array} \right\} \tag{6}$$
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 all $k = 1, \ldots, m$. Then,*
$$\text{conv}(S) = \left\{(x^1, \ldots, x^m, z) \,\middle|\, \begin{array}{l} (u^1, \ldots, u^m, z) \in \text{conv}(S_0), \\ u^k \geq |x^k|, k = 1, \ldots, m \end{array} \right\} \tag{7}$$
where $S_0 = S \cap \left(\mathbb{R}^{n_1}_+ \times \cdots \times \mathbb{R}^{n_m}_+ \times \mathbb{R}^p\right)$.

3. *Suppose S is permutation-invariant and sign-invariant with respect to x^k for all $k = 1, \ldots, m$. Then,*
$$\text{conv}(S) = \left\{(x^1, \ldots, x^m, z) \,\middle|\, \begin{array}{l} (u^1, \ldots, u^m, z) \in \text{conv}(S_0), \\ u^k \geq_m |x^k|, k = 1, \ldots, m \end{array} \right\} \tag{8}$$
where
$$S_0 = S \cap \left\{(u^1, \ldots, u^m, z) \mid u^k_1 \geq \cdots \geq u^k_{n_k} \geq 0, \;\; k = 1, \ldots, m\right\}.$$

The results above can be generalized significantly. For a set C, we denote the convex hull of $\{T_{i1}u, \ldots, T_{ik}u\}$, for $u \in C$ as $\mathcal{T}_i(u)$, where $T_{ij} \in \mathbb{R}^{n \times n}$. Assume that $C \subseteq \mathcal{T}_1(C) \subseteq \cdots \subseteq \mathcal{T}_r \circ \cdots \circ \mathcal{T}_1(C) = S$. Since each matrix is a constant matrix, the convex hull of matrices at any level can be written as the affine transform of a k-dimensional simplex with each T_{ij} as an extreme point. We denote the affine transform $M_i = \text{conv}(T_{i1}, \ldots, T_{ik})$. Let $X_i = \{T_i u \mid T_i \in M_i, u \in V\}$. Then, it follows by commutativity of convexification with affine transformation that $\text{conv}(X_i) = \mathcal{T}_i(V)$. Being a collection of disjoint bilinear functions, the convex hull is determined by the extreme points of M_i. The convex hull of the set X_i can be obtained using the reformulation-linearization technique. Assume, V is expressed as $Au \leq b$. Then, we write U as the linearization of $u\lambda^T$. We obtain $AU \leq b\lambda^T, Ue = u, x = \sum_{j=1}^k T_{ij}U_j$, $\lambda^T e = 1$, and $\lambda \geq 0$, where U_j is the j^{th} column of U. This can also be obtained using disjunctive programming on the set of points $(u, T_{ij}u)$ for $j \in \{1, \ldots, k\}$. Observe that the number of inequalities is approximately $km+2n+k+1$ and the number of variables is $n \times (k+1) + k$, where $A \in \mathbb{R}^{m \times n}$. It follows that, for a fixed k and r, repeated application of the procedure r times, leads to a polynomial sized formulation. More importantly, if each $T_{ij} = Q_i + R_{ij}$, where R_{ij} has a fixed rank l_{ij} and m is a fixed, then by working with the column space of R_{ij} we can limit the number of variables to $\sum_{j=1}^k l_{ij}$, a fixed quantity. In this case, the step adds a fixed number of variables and constraints. Therefore, r steps, where r is bounded by a polynomial, leads to a polynomial sized formulation. As an example, consider the scheme where we construct the convex hull of (u_1, u_2, u_1, u_2) and then (u_1, u_2, u_2, u_1), where $u_1 \geq u_2$. Then, we obtain

$$\begin{aligned} U_{11} + U_{12} &= u_1 \\ U_{21} + U_{22} &= u_2 \\ U_{11} + U_{21} &= x_1 \\ U_{12} + U_{22} &= x_2 \\ U_{11} &\geq U_{22} \\ U_{12} &\geq U_{21} \end{aligned}$$

We show that the above system projects to $u_1 \geq x_1$, $u_1 \geq x_2$, $u_1 + u_2 = x_1 + x_2$. The validity of these inequalities follows easily from the above constraints. Consider any (u_1, u_2, x_1, x_2) satisfying the latter inequalites, Assume without loss of generality that $x_1 \geq x_2$. Then, we can define $U_{11} = x_1$, $U_{12} = u_1 - x_1 = x_2 - u_2$, $u_{21} = 0$, and $u_{22} = u_2$ and check that these variables satisfy the system defined above. Similarly, if $x_1 \leq x_2$, we let $U_{11} = x_1 - u_2 = u_1 - x_2$, $U_{12} = x_2$, $U_{21} = u_2$, and $U_{22} = 0$. The projected inequalities are precisely the ones used in [13] at each stage. Since permutations are possible using an $O(n \log n)$ sorting network, the author obtains a compact formulation of the permutahedron which can be used to represent the majorization constraints. Similarly, [18] considers the special case where each T_{ij} describes a reflection relationship. In particular, by adding an additional variable, we can write the reflection about any plane as a reflection about a plane passing through the origin. Assume this plane is $\langle a, v\rangle = 0$

and a is a unit vector. Then, it follows that any point v may be written as: $v - \langle a, v\rangle a + \langle v, a\rangle a$ and its reflection as $v - 2\langle a, u\rangle a$. Using this transformation, it suffices to consider in the above example, a single variable u representing $\langle a, v\rangle$ with the constraint $u \geq 0$ and the transformation consists of the matrix $[1]$ and $[-1]$. Then, we have

$$\begin{aligned} U_1 + U_2 &= u \\ U_1 - U_2 &= x \\ U_1 &\geq 0 \\ U_2 &\geq 0 \end{aligned}$$

Indeed, the set can be projected to $-u \leq x \leq u$, and $u \geq 0$. The validity of these constraints follows easily. Moreover, for a given (u, x), we can set $U_1 = \frac{u+x}{2}$ and $U_2 = \frac{u-x}{2}$ to satisfy the above equations. As pointed out by [18], the previous example can also be seen as a reflection by considering reflection about $u_1 - u_2 = 0$. When T_{ij} are symmetric, the constraints for the set V can often be obtained by imposing constraints using eigenvectors of $T_{i1}, \ldots, T_{ik}$ which have negative eigenvalues. In particular, if Λ_{ij} is the matrix with columns being the eigenvectors of T_{ij}, and Λ_{ij} is of full rank, we can require that $[\Lambda_{ij}^{-1}u]_t \geq 0$ for each t such that the eigenvalue corresponding to the t^{th} column of Λ_{ij} is negative. This is because repeated applications of T switches the sign of these entries in $\Lambda_{ij}^{-1}u$. The advantage is that at each stage these constraints can be used to restrict S simplifying the initial convex hull construction.

3 Sparsity theorem

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

$$N^K_{\|\cdot\|_s} = \{x \in \mathbb{R}^n \mid \|x\|_s \leq 1, \text{card}(x) \leq K\}, \tag{9}$$

where $\|\cdot\|_s$ is a sign- and permutation-invariant norm (also known as a symmetric gauge function). When $K = 1$, the convex hull is the well-known l_1-norm ball and hence we assume $1 < K \leq n$. When the associated norm $\|\cdot\|_s$ is the l_2-norm, $N^K_{\|\cdot\|}$ is the feasible set of the sparse principal component analysis problem (sPCA); see [11].

For notational simplicity, we define $\Delta := \Delta^n \cap \mathbb{R}^n_+$ and, for any vector $x \in \mathbb{R}^n$, define x_Δ as $(x_\Delta)_i = |x|_{[i]}$ for $i = 1, \ldots, n$.

By sign- and permutation-invariance of the norm $\|\cdot\|_s$ and the cardinality constraint, $N^K_{\|\cdot\|_s}$ is sign- and permutation-invariant and hence we can apply Theorem 2.7 to obtain its convex hull as a projection of a higher dimensional set as follows:

$$\text{conv}\left(N^K_{\|\cdot\|_s}\right) = \left\{x \in \mathbb{R}^n \,\middle|\, \begin{array}{l} u \in N^K_{\|\cdot\|_s} \cap \Delta, \\ u \geq_m |x| \end{array}\right\} = \left\{x \in \mathbb{R}^n \,\middle|\, \begin{array}{l} \|u\|_s \leq 1, \\ u_1 \geq \cdots \geq u_K \geq 0, \\ u_{K+1} = \cdots = u_n = 0, \\ u \geq_m |x| \end{array}\right\}. \tag{10}$$

The extended formulation (10) can be written in a closed form with $O(nK)$ additional variables and constraints based on the modeling technique that we described in Section 2; see formulations (2) and (3) for modeling details. Other forms of extended formulations are proposed in [20] and [21] when $\|\cdot\|_s$ is an l_p norm. In their papers, two approaches are used to develop the formulations: (i) a dynamic programming concepts and (ii) Goemans' extended formulation for the permutahedron using a sorting network [13].

In this section, we represent 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, we provide a separating hyperplane for an arbitrary point in $\mathbb{R}^n$.

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

Proof. By Lemma 2.2, we can write $y = \sum_i \lambda_i P^i x$ where $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$, and P^i are permutation matrices. Therefore, $f(y) = f\left(\sum_i \lambda_i P^i x\right) \leq \sum_i f(\lambda_i P^i x) = \sum_i \lambda_i f(P^i x) = \sum_i \lambda_i f(x) = f(x)$ where the inequality follows from the triangle inequality for the seminorm f and the second and the third equalities follow from the positive homogeneity and permutation-invariance of f, respectively. □

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 conv$(N^K_{\|\cdot\|_s})$ is a convex body.

Theorem 3.1. conv$(N^K_{\|\cdot\|_s})$ *is a convex body.*

Proof. Convexity of the set is obvious. We first show that conv$(N^K_{\|\cdot\|_s})$ is compact. Let X be the set of $(u, x) \in \mathbb{R}^n \times \mathbb{R}^n$ that satisfy the conditions of the last set in (10). Since each condition in X forms a closed set, X is an intersection of closed sets, showing the closedness of X. Since $(\mathbb{R}^n, \|\cdot\|_s)$ is a finite-dimensional normed vector space, the norm ball $\{u \in \mathbb{R}^n \mid \|u\|_s \leq 1\}$ is compact. Let M be such that $\|u\| \leq M$ for all u in the norm ball. Recall that $|x|$ is a convex combination of some u and its permutations and let u_x be such u. Then, by Lemma 3.1, $\|x\| \leq \|u_x\|$. Therefore, for each $(u, x) \in X$, $\|(u, x)\| \leq \|u\| + \|x\| \leq \|u\| + \|u_x\| \leq 2M$, showing that X is bounded. Finally, conv$(N^K_{\|\cdot\|_s})$ is compact because it is a projection of a compact set.

We next show that conv$(N^K_{\|\cdot\|_s})$ has a non-empty interior. Observe that $\epsilon e_1 \in$ conv$(N^K_{\|\cdot\|_s})$ for a sufficiently small $\epsilon > 0$ because $(\epsilon e_1, \epsilon e_1) \in X$. By sign- and permutation-invariance of conv$(N^K_{\|\cdot\|_s})$, any sign- and permutation-invariants of ϵe_1 are included in the convex hull. This shows that $\{x \mid \|x\|_1 \leq \epsilon\} \subseteq$ conv$(N^K_{\|\cdot\|_s})$. Since 0 is an interior point of the l_1-norm ball, it is also an interior point of conv$(N^K_{\|\cdot\|_s})$. ☐

It is well-known that there exists an explicit one-to-one correspondence between norms in $\mathbb{R}^n$ and compact convex bodies symmetric about 0 and containing 0 in their interior; see [23], for example. In particular, an arbitrary norm can be matched to its unit ball. Conversely, any given compact convex body C symmetric about 0 and containing 0 in its interior can be assigned to the norm:

$$f_C(x) := \min\left\{t > 0 \,\middle|\, \frac{x}{t} \in C\right\}.$$

It is known that the function f_C satisfies the properties of norms and that the convex body C is a sublevel set of f_C, that is, $C = \{x \mid f_C(x) \leq 1\}$.

Since the set conv$(N^K_{\|\cdot\|_s})$ is a compact convex body symmetric about 0 and containing 0 in its interior, we can define a norm associated with conv$(N^K_{\|\cdot\|_s})$. We denote the norm by $\|\cdot\|_c$. Note that $\|\cdot\|_c$ is sign- and permutation-invariant.

Theorem 3.2. conv$(N^K_{\|\cdot\|_s}) = \{x \in \mathbb{R}^n \mid \|x\|_c \leq 1\}$.

Theorem 3.3. *If* card$(x) \leq K$, $\|x\|_c = \|x\|_s$.

Proof. By sign- and permutation-invariance of $\|\cdot\|_c$ and $\|\cdot\|_s$, $\|x\|_c = \|x_\Delta\|_c$ and $\|x\|_s = \|x_\Delta\|_s$. Therefore, it suffices to show that $\|x_\Delta\|_c = \|x_\Delta\|_s$. Define $y = \frac{x_\Delta}{\|x_\Delta\|_c}$ and $z = \frac{x_\Delta}{\|x_\Delta\|_s}$. Since $\|y\|_c = 1$, by Proposition 3.2, $y \in$ conv$(N^K_{\|\cdot\|_s})$. Therefore, there exists $u_y \in \mathbb{R}^n$ such that $\|u_y\|_s \leq 1$, $u \in \Delta$, card$(u) \leq K$, and $u_y \geq_m |y|$. Therefore, by Lemma 3.1, $\|y\|_s \leq \|u_y\|_s \leq 1$. This shows that $\|x_\Delta\|_s \leq \|x_\Delta\|_c$. For the opposite inequality, since $\|z\|_s = 1$ and card$(z) \leq K$, $z \in N^K_{\|\cdot\|_s} \subseteq$ conv$(N^K_{\|\cdot\|_s})$. Thus, by Proposition 3.2, $\|z\|_c \leq 1$ and hence $\|x_\Delta\|_c \leq \|x_\Delta\|_s$. ☐

We next present an explicit formula that evaluates $\|\cdot\|_c$. For an arbitrary $x \in \mathbb{R}^n$, define $s(x) \in \mathbb{R}^K$ as $s(x)_i = \frac{\sum_{j=i}^n |x|_{[j]}}{K-i+1}$ for all $i \in \{1, \ldots, K\}$. Let i_x be the minimum index that minimizes $s(x)_i$ and let $\delta(x) = s(x)_{i_x}$. By convention, we define $s(x)_0 = s(x)_{K+1} = \infty$. Now, define $u(x) \in \mathbb{R}^n$ as

$$u(x)_i = \begin{cases} |x|_{[i]} & i \in \{1, \ldots, i_x - 1\} \\ \delta(x) & i \in \{i_x, \ldots, K\} \\ 0 & \text{Otherwise} \end{cases} \tag{11}$$

Theorem 3.4. *Suppose $s(x), i_x, \delta(x)$, and $u(x)$ are defined as previously. Then,*

1. $s(x)_{i+1} - s(x)_i = \frac{1}{K-i+1}(s(x)_{i+1} - |x|_{[i]}) = \frac{1}{K-i}(s(x)_i - |x|_{[i]})$ *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. $u(x) \geq_m |x|$
4. $u(x)_1 = \max\{|x|_{[1]}, s(x)_1\}$

Proof. For $i = 1, \ldots, K-1$,

$$\begin{aligned} s(x)_{i+1} - s(x)_i &= \tfrac{1}{K-i}\sum_{j=i+1}^n |x|_{[j]} - \tfrac{1}{K-i+1}\sum_{j=i}^n |x|_{[j]} \\ &= \tfrac{1}{(K-i)(K-i+1)}\left\{(K-i+1)\sum_{j=i+1}^n |x|_{[j]} - (K-i)\sum_{j=i}^n |x|_{[j]}\right\} \\ &= \tfrac{1}{(K-i)(K-i+1)}\left\{\sum_{j=i+1}^n |x|_{[j]} - (K-i)|x|_{[j]}\right\} \\ &\left(\text{or } = \tfrac{1}{(K-i)(K-i+1)}\left\{\sum_{j=i}^n |x|_{[j]} - (K-i+1)|x|_{[j]}\right\}\right) \\ &= \tfrac{1}{K-i+1}(s(x)_{i+1} - |x|_{[i]}) \\ &\left(\text{or } = \tfrac{1}{K-i}(s(x)_i - |x|_{[i]})\right) \end{aligned}$$

and hence Part 1 follows.

Part 2 is clear when $K = 2$ and hence we assume that $K \geq 3$. We first show that $s(x)_1 \geq \cdots \geq s(x)_{i_x}$. This is clearly true when $i_x = 1$ or 2. For $i_x \geq 3$, and $i = 1, \ldots, i_x - 2$, by Part 1 of the proposition,

$$s(x)_{i+2} - s(x)_{i+1} = \frac{(s(x)_{i+1} - |x|_{[i+1]})}{K-i-1} \geq \frac{(s(x)_{i+1} - |x|_{[i]})}{K-i-1} = \frac{K-i+1}{K-i-1}(s(x)_{i+1} - s(x)_i). \tag{12}$$

Therefore, the result follows by induction since $s(x)_{i_x} - s(x)_{i_x-1} \leq 0$ by minimality of $s(x)_{i_x}$. The proof of the remainder of the statement is similar and hence we omit it.

We next prove Part 3. We first show that $u(x)$ is nonincreasing. It is clear that it is true when $i_x = 1$ and hence we assume that $i_x \geq 2$. By Part 1, $s(x)_{i_x} - |x|_{[i_x-1]} = (K - i_x)(s(x)_{i_x} - s(x)_{i_x-1}) \leq 0$. Therefore, $|x|_{[i_x-1]} \geq s(x)_{i_x} = \delta(x)$, showing the desired result. Thus, $u(x)_{[i]} = u(x)_i$ 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 $\delta(x)$ and this implies the equality $\sum_{i=1}^n u(x)_i = \sum_{i=1}^n |x|_{[i]}$. We next show that $\sum_{i=1}^j u(x)_i \geq \sum_{i=1}^j |x|_{[i]}$ for all $j = 1, \ldots, n-1$. When $j = 1, \ldots, i_x - 1$, the inequality holds with equality by definition of $u(x)$. We next consider the case $j \geq i_x$. If $i_x = K$, the inequality holds because $\sum_{i=1}^j u(x)_i = \sum_{i=1}^n u(x)_i = \sum_{i=1}^n |x|_{[i]} \geq \sum_{i=1}^j |x|_{[i]}$. Now assume that $i_x < K$. Since $s(x)_{i_x+1} \geq s(x)_{i_x}$ and $s(x)_{i_x+1} - s(x)_{i_x} = \frac{1}{K-i_x}(s(x)_{i_x} - |x|_{[i_x]})$ by Part 1, $s(x)_{i_x} \geq |x|_{[i_x]}$ and hence $\delta(x) = s(x)_{i_x} \geq |x|_{[i]}$ for all $i \geq i_x$. Therefore, $\sum_{i=1}^j u(x)_i - \sum_{i=1}^j |x|_{[i]} = \sum_{i=i_x}^j u(x)_i - \sum_{i=i_x}^j |x|_{[i]} = \sum_{i=i_x}^j (\delta(x) - |x|_{[i]}) \geq 0$.

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, $u(x)_1 = \max\{|x|_{[1]}, s(x)_1\}$. □

By Proposition 3.4, for arbitrary $x \in \mathbb{R}^n$, we can construct a vector $u(x) \in \Delta$ that satisfies the cardinality constraint and majorizes $|x|$. In the following theorem, we show that x and $u(x)$ actually have the same values of c-norm, enabling us to evaluate $\|x\|_c$ if $\|\cdot\|_s$ is calculable.

Theorem 3.2. *For an arbitrary $x \in \mathbb{R}^n$, suppose $s(x), i_x, \delta(x)$, and $u(x)$ are defined as previously. Then, $\|x\|_c = \|u(x)\|_s$.*

Proof. The inequality $\|x\|_c \leq \|u(x)\|_s$ directly follows from Lemma 3.1, Proposition 3.3, and Part 3 of Proposition 3.4. We next show $\|x\|_c \geq \|u(x)\|_s$. Define $w := u(x)/\|u(x)\|_s$ so that w is on the boundary of the norm ball $B_s := \{y \in \mathbb{R}^n \mid \|y\|_s \leq 1\}$. Let β be an optimal solution to

$$\max\{w^\intercal \beta \mid \|\beta\|_{s*} \leq 1\} \tag{13}$$

where $\|\cdot\|_{s*}$ is the dual norm of $\|\cdot\|_s$. Then, $\beta^\intercal y = \beta^\intercal w$ be a supporting hyperplane of B_s that passes through w. It is clear that $\|\cdot\|_{s*}$ is sign- and permutation-invariant. By rearrangement inequality and permutation-invariance of $\|\cdot\|_{s*}$, we can assume, without loss of generality, that β is in descending order. Furthermore, we can also assume that $\beta \geq 0$ because $w \geq 0$ and $\|\cdot\|_{s*}$ is sign-invariant. We next define $\theta \in \mathbb{R}^n$ as

$$\theta_i = \begin{cases} \beta_i & i = 1, \ldots, i_x - 1 \\ \frac{\sum_{j=i_x}^K \beta_j}{K-i_x+1} & i = i_x, \ldots, K \\ 0 & \text{Otherwise} \end{cases}$$

and consider the inequality $\theta^\intercal y \leq \beta^\intercal w$. We claim that the inequality is valid for B_s. First, $\theta \in \Delta$ since $\beta \in \Delta$ and $\beta_{i_x-1} \geq \frac{\sum_{j=i_x}^K \beta_j}{K-i_x+1}$. Therefore, by rearrangement inequality, it suffices to show the validity for $B_s \cap \Delta$. Define

$\bar{\beta} := (\beta_1, \ldots, \beta_K, 0, \ldots, 0)$ and notice that $\bar{\beta} \geq_m \theta$. Then, for each permutation matrix P, $(P\bar{\beta})^\intercal y \leq \beta^\intercal w$ is valid for $B_s \cap \Delta$ because for each $y \in B_s \cap \Delta$, $(P\bar{\beta})^\intercal y \leq \bar{\beta}^\intercal y \leq \beta^\intercal y \leq \beta^\intercal w$ where the first inequality follows from rearrangement inequality. The validity of $\theta^\intercal y \leq \beta^\intercal w$ follows from the fact that θ is a convex combination of $\bar{\beta}$ and its permutations. Next, define $\chi \in \mathbb{R}^n$ by $\chi_i = \beta_i$ for $i = 1, \ldots, i_x - 1$ and $\sum_{i=i_x}^{K} \beta_i/(K - i_x + 1)$ otherwise. Then,

$$\begin{aligned} \chi^\intercal x_\Delta &= \sum_{i=1}^{i_x-1} \beta_i |x|_{[i]} + \frac{\sum_{i=i_x}^{K} \beta_i}{K - i_x + 1} \sum_{i=i_x}^{n} |x|_{[i]} \\ &= \sum_{i=1}^{i_x-1} \beta_i |x|_{[i]} + \sum_{i=i_x}^{K} \beta_i \frac{\sum_{i=i_x}^{n} |x|_{[i]}}{K - i_x + 1} = \beta^\intercal u(x). \end{aligned} \tag{14}$$

We next claim that $\chi^\intercal y \leq \beta^\intercal w$ is valid for $N^K_{\|\cdot\|_s}$. Again, it suffices to show its validity for $N^K_{\|\cdot\|_s} \cap \Delta$. This is clear because $\chi^\intercal y = \theta^\intercal y$ for all $y \in N^K_{\|\cdot\|_s} \cap \Delta$. Therefore, $\chi^\intercal y \leq \beta^\intercal w$ is valid for $\text{conv}(N^K_{\|\cdot\|_s}) = \{y \mid \|y\|_c \leq 1\}$. Finally,

$$\beta^\intercal \frac{u(x)}{\|x\|_c} = \chi^\intercal \frac{x_\Delta}{\|x\|_c} \leq \beta^\intercal \frac{u(x)}{\|u(x)\|_s},$$

concluding that $\|x\|_c \geq \|u(x)\|_s$. □

Theorem 3.5. *For a fixed $x \in \mathbb{R}^n$, suppose $u(x)$ is defined as* (11) *and w, β, θ, and χ are defined as in the proof of Theorem 3.2. Then, $\chi^\intercal y \leq \|\theta\|_{s*}$ is valid for $N^K_{\|\cdot\|_s}$ and it separates x_Δ if $x \notin \text{conv}(N^K_{\|\cdot\|_s})$.*

Proof. In the proof of Theorem 3.2, we showed that $\chi^\intercal y \leq \beta^\intercal w$ is valid for $N^K_{\|\cdot\|_s}$. By definition of β and dual norm, $\beta^\intercal w = \|\theta\|_{s*}$, proving the validity. Next, when $x \notin \text{conv}(N^K_{\|\cdot\|_s})$, $\|u(x)\|_s = \|x\|_c > 1$. Therefore, $\chi^\intercal x_\Delta = \beta^\intercal u(x) > \beta^\intercal \frac{u(x)}{\|u(x)\|_s} = \beta^\intercal w$ where the first equality follows from (14). □

Remark. *In the proof of Proposition 3.5, let T be the transformation (a composition of sign-conversions and permutations) that maps x to x_Δ. Then, the hyperplane that separates x and $N^K_{\|\cdot\|_s}$ is $T^{-1}(\chi)y \leq \beta^\intercal w$.*

Theorem 3.3 (Sparsity Theorem)**.** *For an arbitrary $x \in \mathbb{R}^n$, consider the following optimization problem:*

$$\begin{aligned} \min \quad & \|u\|_s \\ \text{s.t.} \quad & u_1 \geq \cdots \geq u_K \geq 0, \\ & u_{K+1} = \cdots = u_n = 0, \\ & u \geq_m |x| \end{aligned}$$

Then, $u(x)$ is an optimal solution where $u(x)$ is defined as in (11)*.*

Proof. First, $u(x)$ is feasible because of its definition and Part 3 of Proposition 3.4. Then, for any feasible solution u, $\|u\|_s = \|u\|_c \geq \|x\|_c = \|u(x)\|_s$ where the first equality follows from Proposition 3.3 and the first inequality from Lemma 3.1, and the second equality from Theorem 3.2. Therefore, $u(x)$ is an optimal solution. □

Example 3.1. Consider the case where $n = 6$ and $K = 3$. Let $N = \{1, \ldots, 6\}$ and $x := \left(\frac{27}{28}, \frac{5}{28}, \frac{4}{28}, \frac{3}{28}, \frac{2}{28}, \frac{1}{28}\right)$. Notice that $\|x\|_2 = 1$ and $x \in \Delta$. Throughout this example, we want to check whether $x \in \text{conv}(N^3_{\|\cdot\|_2})$ or not and present an explicit separating hyperplane using the procedure described in the proof of Theorem 3.2. First, we construct the vector $s(x) \in \mathbb{R}^3$ as follows:

$$\begin{aligned} s(x)_1 &= \frac{\sum_{j=1}^{6} x_j}{3-1+1} = \frac{1}{3}\left(\frac{27}{28} + \frac{5}{28} + \frac{4}{28} + \frac{3}{28} + \frac{2}{28} + \frac{1}{28}\right) = \frac{13}{28} \\ s(x)_2 &= \frac{\sum_{j=2}^{6} x_j}{3-2+1} = \frac{1}{2}\left(\frac{5}{28} + \frac{4}{28} + \frac{3}{28} + \frac{2}{28} + \frac{1}{28}\right) = \frac{15}{56} \\ s(x)_3 &= \frac{\sum_{j=3}^{6} x_j}{3-3+1} = \frac{1}{1}\left(\frac{4}{28} + \frac{3}{28} + \frac{2}{28} + \frac{1}{28}\right) = \frac{5}{14} \end{aligned}$$

Observe that $s(x)_2 = \min\{s(x)_1, s(x)_2, s(x)_3\}$. Next, we define $u(x) \in \mathbb{R}^6$ is as follows:

$$u(x)_1 = x_1 = \frac{27}{28}, \quad u(x)_2 = u(x)_3 = s(x)_2 = \frac{15}{56}, \quad u(x)_4 = u(x)_5 = u(x)_6 = 0$$

Since $\|u(x)\|_2 = 1.036 \cdots > 1$, we conclude that $x \notin \text{conv}(N^3_{\|\cdot\|_2})$. We will confirm this by calculating a separating hyperplane. We first construct a hyperplane that separates $u(x)$ from $\text{conv}(N^3_{\|\cdot\|_2})$. Notice that $\|\cdot\|_2$ is self-dual and $u(x)/\|u(x)\|_2$ an optimal solution to (13) where $w = u(x)/\|u(x)\|_2$. Therefore, we set $\beta = u(x)/\|u(x)\|_2$. Then, the inequality $\beta^\intercal y \le \beta^\intercal w (= 1)$ (or equivalently $u(x)^\intercal y \le \|u(x)\|_2$) is valid for $N^3_{\|\cdot\|_2}$ because, for any $y \in N^3_{\|\cdot\|_2}$, $u(x)^\intercal y \le \|u(x)\|_2 \|y\|_2 \le \|u(x)\|_2$. Furthermore, it separates $u(x)$ because $u(x)^\intercal u(x) = \|u(x)\|_2^2 > \|u(x)\|_2$. We next construct a hyperplane that separates x from $N^3_{\|\cdot\|_2}$. Define θ and χ as follows:

$$\theta_1 = \beta_1 = \tfrac{1}{\|u(x)\|_2} u(x)_1 = \tfrac{1}{\|u(x)\|_2}\tfrac{27}{28}, \quad \theta_2 = \theta_3 = \tfrac{1}{3-2+1}(\beta_2+\beta_3) = \tfrac{1}{\|u(x)\|_2}\tfrac{15}{56}, \quad \theta_4 = \cdots = \theta_6 = 0$$
$$\chi_1 = \beta_1 = \tfrac{1}{\|u(x)\|_2} u(x)_1 = \tfrac{1}{\|u(x)\|_2}\tfrac{27}{28}, \quad \chi_2 = \cdots = \chi_6 = \tfrac{1}{3-2+1}(\beta_2+\beta_3) = \tfrac{1}{\|u(x)\|_2}\tfrac{15}{56}$$

Observe that $\beta^\intercal w = \|\theta\|_2 = 1$. Now consider the inequality $\chi^\intercal y \le \beta^\intercal w (= 1)$. It is valid for $N^3_{\|\cdot\|_2}$ because for any $y \in N^3_{\|\cdot\|_2}$, $\chi^\intercal y \le \chi^\intercal y_\Delta = \theta^\intercal y_\Delta \le \|\theta\|_2 \|y_\Delta\|_2 \le \|\theta\|_2 = 1$. Moreover, it separates x because $\chi^\intercal x = \frac{1}{\|u(x)\|_2}\left(\frac{27}{28}, \frac{15}{56}, \frac{15}{56}, \frac{15}{56}, \frac{15}{56}, \frac{15}{56}\right)^\intercal \left(\frac{27}{28}, \frac{5}{28}, \frac{4}{28}, \frac{3}{28}, \frac{2}{28}, \frac{1}{28}\right) = 1.036\cdots > 1$.

Next, we consider some special cases of the set (9).

Theorem 3.6. *Let $S = \{x \in \mathbb{R}^q \mid \text{card}(x) \le K, \|x\|_\infty \le r\}$ where $\|x\|_\infty = |x|_{[1]}$. Then,*

$$\text{conv}(S) = \{x \in \mathbb{R}^q \mid \|x\|_1 \le rK, \|x\|_\infty \le r\} \tag{15}$$

where $\|x\|_1 := \sum_{i=1}^q |x_i|$ and $\|x\|_\infty = |x|_{[1]}$.

Proof. Observe that $S/r = \{x \in \mathbb{R}^q \mid \text{card}(x) \le K, \|x\|_\infty \le 1\}$. Then,

$$\text{conv}(S) = r\,\text{conv}(S/r) = \{y \in \mathbb{R}^q \mid \|u(y)\|_\infty \le r\} = \{y \in \mathbb{R}^q \mid \max\{|y|_{[1]}, s(y)_1\} \le r\}$$

where the second equality follows from Proposition 3.2 and Theorem 3.2 and the third equality from Part 4 of Proposition 3.4. Since $s(y)_1 = \frac{1}{K}\sum_{j=1}^q |y|_{[j]}$ by definition of $s(y)$, the result follows. □

When $\|\cdot\|_s$ is the Euclidean norm, the norm $\|\cdot\|_c$ associated with $\text{conv}(N^K_{\|\cdot\|_2})$ is known to be *K-support norm* (or a.k.a. *K-overlap norm*) and the explicit formula for the norm is known in [1]. We provide an alternative proof for the formula using our arguments. For consistency with literature, we denote the K-support norm by $\|\cdot\|^{sp}_K$.

Lemma 3.4. *$r = K - i_x$ is the unique integer in $\{0, \ldots, K-1\}$ that satisfies* (16) *where $|x|_{[0]} = \infty$ by convention.*

$$|x|_{[K-r-1]} > s(x)_{K-r} \ge |x|_{[K-r]}. \tag{16}$$

Proof. We first claim that $r = K - i_x$ satisfies (16). That is, we prove $|x|_{[i_x-1]} > s(x)_{i_x} \ge |x|_{[i_x]}$. By definition of i_x, $s(x)_{i_x} < s(x)_{i_x-1}$ and hence $|x|_{[i_x-1]} > s(x)_{i_x}$ by Part 1 of Proposition 3.4 and the convention $|x|_{[0]} = \infty$. When $i_x \le K-1$ since $s(x)_{i_x+1} \ge s(x)_{i_x}$ implies that $s(x)_{i_x} \ge |x|_{[i_x]}$. When $i_x = K$, $s(x)_K = \sum_{j=K}^n |x|_{[j]} \ge |x|_{[K]}$. Therefore, (16) holds when $r = K - i_x$. We next prove that (16) does not hold for all $r \ne K - i_x$. When $i_x = 1$, $\{s(x)_i\}_{i=1}^K$ is non-decreasing and hence $|x|_{K-r-1} \le s(x)_{K-r}$ for all $r = 0, \ldots, K-2$, violating the first inequality of (16). Now, assume that $i_x \ge 2$. When $r \le K - i_x - 1$, since $s(x)_{K-r} \ge s(x)_{i_x} \ge |x|_{[i_x]} \ge |x|_{[K-r-1]}$, violating the first inequality of (16). We next assume that there exists $r \in \{K - i_x + 1, \ldots, K-1\}$ that satisfies (16). Since $s(x)_{K-r+1} \le s(x)_{K-r}$, $s(x)_{K-r} \le |x|_{[K-r]}$ by Part 1 of Proposition 3.4. Therefore, $s(x)_{K-r} = x_{[K-r]}$, implying that $s(x)_{K-r} = s(x)_{K-r+1}$ by Part 1 of Proposition 3.4. By (12), $s(x)_{K-r} = s(x)_{K-r+1} = \cdots = s(x)_{i_x}$, contradicting the minimality of i_x. □

Theorem 3.7 (Proposition 2.1 of [1])**.**

$$\|x\|^{sp}_K = \left(\sum_{i=1}^{K-r-1} x^2_{[i]} + \frac{1}{r+1}\left(\sum_{i=K-r}^{n} |x|_{[i]}\right)^2\right)^{\frac{1}{2}}. \tag{17}$$

where r is the unique integer in $\{0, \ldots, K-1\}$ satisfying (16)*.*

Proof. By Lemma 3.4, $r = K - i_x$. By Theorem 3.2,

$$\|x\|^{sp}_K = \|u(x)\|_2 = \left(\sum_{i=1}^{i_x-1} |x|^2_{[i]} + (K - i_x + 1)\delta(x)^2\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{i_x-1} |x|^2_{[i]} + \frac{1}{K - i_x + 1}\left(\sum_{i=i_x}^{n} |x|_{[i]}\right)^2\right)^{\frac{1}{2}}.$$

□

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 $\|M\|_{sp} = \sigma_1(M)$ and $\|M\|_* = \sum_{i=1}^q \sigma_i(M)$ 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 $\bar{S} = \{M \in \mathcal{M}_{m,n}(\mathbb{R}) \mid f_i(\sigma(M)) \leq 1, i = 1, \ldots, r\}$ and their convex hulls where each f_i is a sign- and permutation-invariant function. Define $S = \{x \in \mathbb{R}^q \mid f_i(x) \leq 1, i = 1, \ldots, r\}$ where $q = \min\{m, n\}$. It is clear that $M \in \bar{S}$ if and only if $\sigma(M) \in S$.

The following conjugacy result is a key tool to bridge conv$(\bar{S})$ and conv(S).

Lemma 3.5 (Theorem 2.4 of [19])**.** *Suppose $f : \mathbb{R}^q \to \mathbb{R}$ is sign- and permutation-invariant. Then,*

$$(f \circ \sigma)^* = f^* \circ \sigma$$

where the asterisks represent the conjugate operator of the functions.

Theorem 3.6. *Suppose $\bar{S} = \{M \in \mathcal{M}_{m,n}(\mathbb{R}) \mid f_i(\sigma(M), z) \leq 1, i = 1, \ldots, r\}$ and $S = \{x \in \mathbb{R}^q \mid f_i(x, z) \leq 1, i = 1, \ldots, r\}$ where $q = \min\{m, n\}$ and $f_i : \mathbb{R}^q \times \mathbb{R}^p \to \mathbb{R}, i = 1, \ldots, r$ are a sign- and permutation-invariant with respect to $x \in \mathbb{R}^q$. Then,* conv$(\bar{S}) = \{M \in \mathcal{M}_{m,n}(\mathbb{R}) \mid \sigma(M) \in$ conv$(S)\}$.

Proof. Define $H : \mathcal{M}_{m,n}(\mathbb{R}) \to \mathbb{R}$ and $h : \mathbb{R}^q \to \mathbb{R}$ as follows:

$$H(M) = \begin{cases} 0 & M \in \bar{S} \\ \infty & \text{otherwise} \end{cases}, \qquad h(x) = \begin{cases} 0 & x \in S \\ \infty & \text{otherwise} \end{cases}. \tag{18}$$

Then, $\bar{S} = \{M \mid H(M) \leq 1\}$, $S = \{x \mid h(x) \leq 1\}$, and $H = h \circ \sigma$. We next show that conv$(S) = \{x \mid h^{**}(x) \leq 1\}$. Suppose $x \in$ conv(S). Then, $x = \sum_j \lambda_j y_j$ where $\lambda_j \geq 0$ and $y_j \in S$ for all j and $\sum_j \lambda_j = 1$. Since h^{**} is the closed convex envelope of h, $h^{**}(x) \leq \sum_j \lambda_j h^{**}(y_j) \leq \sum_j \lambda_j h(y_j) = 0$. Therefore, $h^{**}(x) \leq 1$. We next consider an arbitrary x such that $h^{**}(x) \leq 1$. Using the fact that epi$(h^{**}) =$ cl conv(epi(h)), we have $(x, h^{**}(x)) = \sum_j \lambda_j (y_j, z_j)$ where $\lambda_j \geq 0$ and $(y_j, z_j) \in$ epi(h) for all j and $\sum_j \lambda_j = 1$. Therefore, $x = \sum_j \lambda_j y_j$ and $h^{**}(x) = \sum_j \lambda_j z_j$. Then, $1 \geq h^{**}(x) = \sum_j \lambda_j z_j \geq \sum_j \lambda_j h(y_j)$. This implies that $y_j \in S$ for all j because otherwise, $\sum_j \lambda_j h(y_j) = \infty$, violating the inequality. This shows that $x \in$ conv(X). We can also prove that conv$(\bar{S}) = \{M \mid H^{**}(M) \leq 1\}$ by using the similar arguments and we omit the proof.

By sign- and permutation-invariance of S, h is sign- and permutation-invariant and hence so is h^*. Then, by Lemma 3.5, $H^* = h^* \circ \sigma$ and $H^{**} = (h^* \circ \sigma)^* = h^{**} \circ \sigma$. Now, for an arbitrary $M \in$ conv$(\bar{S})$, since $H^{**}(M) \leq 1$, $h^{**}(\sigma(M)) \leq 1$ and hence $\sigma(M) \in$ conv(S). Conversely, consider $\sigma(M) \in$ conv(S). Then, $h^{**}(\sigma(M)) \leq 1$ and hence $H^{**}(M) \leq 1$, showing that $M \in$ conv$(\bar{S})$. □

Notice that the rank of a matrix can be represented as the cardinality of the vector of singular values and cardinality is a sign- and permutation-invariant function. Therefore, we have the following result as a special case of Theorem 3.6.

Corollary 3.6.1. *Let $\bar{S} = \{M \in \mathcal{M}_{m,n}(\mathbb{R}) \mid$ rank$(M) \leq K, \|\sigma(M)\|_s \leq r\}$. Then,*

$$\text{conv}(\bar{S}) = \{M \in \mathcal{M}_{m,n}(\mathbb{R}) \mid \|\sigma(M)\|_c \leq r\} \tag{19}$$

Recall that determining if an arbitrary matrix $M \in \mathcal{M}_{m,n}(\mathbb{R})$ is in the convex hull conv$(\bar{S})$ can be easily done when the norm $\|\cdot\|_s$ can be calculable. In particular, when $\|\cdot\|_s$ is the Euclidean norm, a given matrix M is in conv$(\bar{S})$ if $\|\sigma(M)\|_K^{sp} \leq r$. See (17) for an explicit formula for $\|\cdot\|_K^{sp}$. Semidefinite representability of the convex hull will be discussed in Subsection 3.2.

Next, we consider the special case where $\|\cdot\|_s$ is the l_∞ norm. Proposition 3.6 and Theorem 3.6 together give an alternative proof for the following result.

Theorem 3.8 (Theorem 1 of [15])**.** *Let $\bar{S} = \{M \in \mathcal{M}_{m,n}(\mathbb{R}) \mid$ rank$(M) \leq K, \|M\|_{sp} \leq r\}$. Then,* conv$(\bar{S}) = \{M \in \mathcal{M}_{m,n}(\mathbb{R}) \mid \|M\|_* \leq rK, \|M\|_{sp} \leq r\}$.

Theorem 3.6 can be generalized in the context of hyperbolic polynomials. A multivariate polynomial p in $x \in \mathbb{R}^n$ with real coefficients is called *homogeneous* of degree d if it is a linear combination of monomials of degree d. That is, a homogeneous polynomial is of the form

$$p(x) = \sum_{\alpha_1 + \cdots + \alpha_n = d} c_\alpha x_1^{\alpha_1} \ldots x_n^{\alpha_n}$$

where $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{Z}_+^n$. A real homogeneous polynomial $p : \mathbb{R}^n \to \mathbb{R}$ is called *hyperbolic* with respect to $e \in \mathbb{R}^n$ if $p(e) > 0$ and the univariate polynomial $p(te - a)$ has only real zeros for every $a \in \mathbb{R}^n$. Those zeros are called *e-eigenvalues* of a and denoted by $\lambda_k(a), k = 1, \dots, m$ for some $m \in \{1, \dots, n\}$. Without loss of generality, we assume that $\lambda_1(a) \geq \cdots \geq \lambda_m(a)$. Define the *hyperbolicity cone associated with* p as $\Lambda_{++} = \{a \in \mathbb{R}^n \mid \lambda_m(a) > 0\}$ and denote its closure by Λ_+. A hyperbolic program is an optimization of the form $\min\{c^\intercal x \mid Ax \leq b, x \in \Lambda_+\}$ for some hyperbolic polynomial $p \in \mathbb{R}^n$ with respect to $e \in \mathbb{R}^n$. When the associated polynomial is $p(X) = \det(X)$ and the associated direction e is the identity matrix, the closure of the hyperbolicity cone is the positive semidefinite cone. Therefore, hyperbolic programming is a generalization of semidefinite programming. A polynomial $p : \mathbb{R}^n \to \mathbb{R}$ is *isometric*, if for all $y, z \in \mathbb{R}^n$, there exists $x \in \mathbb{R}^n$ such that $\lambda(x) = \lambda(z)$ and $\lambda(x + y) = \lambda(x) + \lambda(y)$. For example, [14] showed that $p = \det$ is an isometric function: hence, the following conjugacy result generalizes the matrix case.

Lemma 3.7 (Fenchel Conjugacy (Theorem 5.4 of [6]))**.** *Suppose f is permutation-invariant and p is isometric. If* $\text{ran}\lambda = \{x \mid x_1 \geq \cdots \geq x_m\}$ *then* $(f \circ \lambda)^* = f^* \circ \lambda$ *where* $\text{ran}\lambda$ *represents the range of the function* λ*.*

Remark.

1. *The authors of [6] defines Fenchel conjugate on convex functions, but the proof can still applied to arbitrary functions.*
2. $p(X) = \det(X)$ *is isometric.*
3. *In semidefinite programming instance, ran*$\lambda = \{x \mid x_1 \geq \cdots \geq x_m\}$.

The following result is a generalization of Theorem 3.6 in the convex of hyperbolic programming. By similarity of the proof with that of Theorem 3.6, we omit the proof.

Theorem 3.8. *Consider an hyperbolic polynomial $p : \mathbb{R}^n \to \mathbb{R}$ with respect to $e \in \mathbb{R}^n$. Let $\lambda(a) = (\lambda_1(a), \dots, \lambda_m(a)) \in \mathbb{R}^m$ be e-eigenvalues of an arbitrary $a \in \mathbb{R}^n$ where $\lambda_1(a) \geq \cdots \geq \lambda_m(a)$. Suppose p is isometric and* $\text{ran}\lambda = \{x \mid x_1 \geq \cdots \geq x_m\}$. *For permutation-invariant functions $f_i : \mathbb{R}^m \to \mathbb{R}, i = 1, \dots, r$, define $\bar{S} = \{a \in \mathbb{R}^n \mid f_i(\lambda(a)) \leq c_i, i = 1, \dots, r\}$. Then,*

$$\text{conv}(\bar{S}) = \{a \in \mathbb{R}^m \mid \lambda(a) \in \text{conv}(S)\}$$

where $S = \{y \in \mathbb{R}^m \mid f_i(y) \leq c_i, i = 1, \dots, r\}$.

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

We presented a convex hull result of a set of matrices $\bar{S}$ that can be represented in their singular values in Corollary 3.6.1. The resulting convex hull is written in a norm $\|\cdot\|_c$ induced by the defining norm $\|\cdot\|_s$ of $\bar{S}$. While we provided an explicit characterization of the membership of the convex hull, it does not guarantee that the convex hull can be modeled in an SDP solver as a feasible set. In this subsection, we discuss the semidefinite representability of a set of the form as follows: $\mathcal{S} = \{M \in \mathcal{M}_{m,n}(\mathbb{R}) \mid \text{rank}(M) \leq K, f(\sigma(M)) \leq r\}$ where $f : \mathbb{R}^q \to \mathbb{R}$ is a permutation-invariant quasiconvex monotone function. A set is called *semidefinite-representable* if it is a projection of a set expressed by a linear matrix inequality. We remark two well-known properties about semidefinite-representability. (see Section 4.2 of [7]).

Lemma 3.9.

1. *The sum of p largest singular values of a rectangular matrix is semidefinite-representable.*
2. *If A and B are semidefinite-representable then so is $A \cap B$ and the representation is obtained by combining two representations.*

Theorem 3.10. *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 $f : \mathbb{R}^q \to \mathbb{R}$ is a permutation-invariant quasiconvex monotone function. Then,* $\text{conv}(\bar{S})$ *is semidefinite-representable if lower level sets of f are semidefinite-representable.*

Proof. Let $S = \{x \in \mathbb{R}^q \mid \text{card}(|x|) \leq K, f(|x|) \leq r\}$. By sign- and permutation-invariance of S and Theorem 2.7,

$$\text{conv}(S) = \left\{ x \in \mathbb{R}^q \;\middle|\; \begin{array}{l} f(u) \leq r, \\ u_1 \geq \cdots \geq u_K \geq 0, \\ u_{K+1} = \cdots = u_n = 0, \\ u \geq_m |x| \end{array} \right\}. \tag{20}$$

By Theorem 3.6,

$$\text{conv}(\bar{S}) = \left\{ M \in \mathcal{M}_{m,n}(\mathbb{R}) \;\middle|\; \begin{array}{l} f(u) \leq r, \\ u_1 \geq \cdots \geq u_K \geq 0, \\ u_{K+1} = \cdots = u_n = 0, \\ u \geq_m |\sigma(M)| \end{array} \right\}. \tag{21}$$

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

$$\begin{cases} f(u) \le r, \\ u_1 \ge \cdots \ge u_K \ge 0, \\ u_{K+1} = \cdots = u_n = 0, \\ \sum_{i=1}^{j} u_i \ge \sum_{i=1}^{j} \sigma_j(M), j = 1, \ldots, K, \\ \sum_{i=1}^{K} u_j = \sum_{i=1}^{q} \sigma_i(M) \end{cases} \tag{22}$$

The semidefinite-representability of (22) follows from Lemma 3.9 and the semidefinite-representability of the level set $\{u \mid f(u) \le r\}$ and linear inequalities. □

Observe that Theorem 3.10 is consistent with Proposition 4.2.2 in [7] except that we considered the rank constraint, but they do not.

Corollary 3.10.1. *Let $\mathcal{S} = \{M \in \mathcal{M}_{m,n}(\mathbb{R}) \mid \text{rank}(M) \le K, \|\sigma(M)\|_s \le r\}$ where $\|\cdot\|_s$ is permutation-invariant monotone norm. Then, $\text{conv}(\mathcal{S})$ is semidefinite-representable. In particular, when $\|\cdot\|_s$ is Ky Fan p-norm for $p = 1, \ldots, \min\{m, n\}$, the convex hull is semidefinite-representable.*

4 Convex envelopes of nonlinear functions

In this section, we explore the use of Theorem 2.7 in the development of relaxations of non-convex functions.

Lemma 4.1. *Let $x' \in P$, where P is permutation-invariant, and π be a permutation of $\{1, \ldots, n\}$ such that for each $i \in \{1, \ldots, n-1\}$, $x'_{\pi(i)} \ge x'_{\pi(i+1)}$. Then, there exists a $u' \ge_m x'$ with $u' \ne x'$ if and only if there does not exist an $a \in \mathbb{R}^n$ such that $a_{\pi(i)} > a_{\pi(i+1)}$ for all $i \in \{1, \ldots, n-1\}$ and $\sum_{i=1}^n a_i(x_i - x'_i) \le 0$ is valid for P.*

Proof. We first show that if such an a and π exist, there cannot be a $u' \in P$, $u' \ne x'$, such that $u' \ge_m x'$. Assume such a u' exists and because P is permutation-invariant, by sorting u' if necessary, we may assume that $u'_{\pi(i)} \ge u'_{\pi(i+1)}$ for all $i \in \{1, \ldots, n-1\}$. Since $u' \ge_m x'$ and $u' \ne x'$, there exists a y', $\theta > 0$, and $k \in \{1, \ldots, n-1\}$ such that $u' \ge_m y' \ge_m x'$, $y'_{\pi(k)} = x'_{\pi(k)} + \theta$, and $y'_{\pi(k+1)} = x'_{\pi(k+1)} - \theta$. Since $u' \in P$, P is convex and permutation-invariant, and y' can be written as a convex combination of u' and its permutations, it follows that $y' \in P$. Therefore, $\sum_{i=1}^n a_i(y'_i - x'_i) \le 0$ or $a_{\pi(k)} - a_{\pi(k+1)} \le 0$. This leads to a contradiction to the assumed ordering of a.

Now, we show that if there does not exist $u' \in P$ such that $u' \ge_m x'$ with $u' \ne x'$ then such an a exists. Choose any π such that $x'_{\pi(i)} \ge x'_{\pi(i+1)}$ for all $i \in \{1, \ldots, n-1\}$. Let $K := x' + \sum_{i=1}^{n-1} \alpha_{\pi(i)}\big(e_{\pi(i)} - e_{\pi(i+1)}\big)$, where $\alpha \ge 0$ and consider $C = P - K$. Since $x' \in P \cap K$, it follows that $0 \in C$. Let $\langle a', x\rangle \le 0$ define the minimal face of C containing 0. We will show that a can be chosen to be a'. Observe that $e_{\pi(i+1)} - e_{\pi(i)} \in C$ for all $i \in \{1, \ldots, n-1\}$. Therefore, $a'_{\pi(i+1)} - a'_{\pi(i)} \le 0$. We now show that the inequality is in fact strict. Assume, on the contrary, that there exists a $k \in \{1, \ldots, n-1\}$ such that $\langle a', e_{\pi(k+1)} - e_{\pi(k)}\rangle = 0$. Then, there exists a small enough $\epsilon > 0$ such that $\epsilon\big(e_{\pi(k)} - e_{\pi(k+1)}\big) \in C$ because 0 is in the relative interior of its face containing $e_{\pi(k+1)} - e_{\pi(k)}$. It follows that there exists $x'' \in P$ and $\alpha' \ge 0$ such that $\epsilon\big(e_{\pi(k)} - e_{\pi(k+1)}\big) = x'' - x' - \sum_{i=1}^{n-1} \alpha'_{\pi(i)}\big(e_{\pi(i)} - e_{\pi(i+1)}\big)$. Therefore, $x'' = x' + \epsilon\big(e_{\pi(k)} - e_{\pi(k+1)}\big) + \sum_{i=1}^{n-1} \alpha'_{\pi(i)}\big(e_{\pi(i)} - e_{\pi(i+1)}\big)$. Since $\epsilon > 0$, $x'' \in P$, $x'' \ge_m x'$ and $x'' \ne x'$, choosing $u' = x''$ contradicts the assumption that such a u' does not belong to P. Therefore, for all $k \in \{1, \ldots, n-1\}$, $a'_{\pi(k+1)} - a'_{\pi(k)} < 0$. □

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

In this section, for any function $\phi : C \mapsto \mathbb{R}$ we denote $\{(x, \phi) \mid \phi(x) \le \phi \le \alpha, x \in P\}$ as $\text{epi}_{\le\alpha}(\phi)$. A common tool in the construction of convex envelopes is to restrict the domain of the function to a smaller subset. We will say that a function $\phi : C \mapsto \mathbb{R}$ can be restricted to X, where $X \subseteq C$ for the purpose of obtaining the $\text{conv}_C(\phi)$ if $\text{conv}_C(\phi|_X) = \text{conv}_C(\phi)$, where $\phi|_X(x)$ is defined as $\phi(x)$ for any $x \in X$ and $+\infty$ otherwise. Theorem 2.7 can play a key role in obtaining the restriction as we illustrate below.

Lemma 4.2. *Let $\phi : P \mapsto \mathbb{R}$ be a Schur-concave function, where P is a permutation-invariant polytope. Let $M := \{x \mid x \in P, \nexists u \in P \text{ s.t. } u \ge_m x, u \ne x\}$. Let $S := \{(x, \phi) \mid \phi(x) \le \phi \le \alpha, x \in P\}$ and $X := \{(x, \phi) \mid \phi(x) \le \phi \le \alpha, x \in M\}$. 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 $i \in \{1, \dots, n\}$. Let $u' := \arg\max\bigl\{\|u - x''\| \mid u \geq_m x', u \in P\bigr\}$. The maximum is achieved because the feasible set is bounded and objective is upper-semicontinuous. Assume there exists a $y' \in P$ such that $y' \geq_m u'$ and $y' \neq u'$. Since u' can be written as a convex combination of 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 $y' \in P$ such that $y' \geq u'$ and $y' \neq u'$. Therefore, $u' \in M$. Since ϕ is Schur-concave, it follows that $\phi(u') \leq \phi(x') \leq \phi' \leq \alpha$. Therefore, $(u', \phi') \in X$. Since $x' \leq_m u'$, it follows that x' can be written as a convex combination of u' and its permutations. Therefore, (x', ϕ') is not an extreme point of the epigraph of S and $S \subseteq \text{conv}(X)$. It follows that $\text{conv}(S) \subseteq \text{conv}(M)$.

□

It is often useful to restrict the set S to a superset of its extreme points before using Theorem 2.4 to construct the convex hull. We discuss such applications. We are interested in sets $S(Z, a, b) := \{(x, z) \mid (x, z) \in Z, x \in [a, b]^n\}$, where Z is compact and permutation-invariant in x. Further, for $\mathcal{F} = \{F_1, \dots, F_r\}$, where F_i are faces of $[a, b]^n$, we define $X(Z, a, b, \mathcal{F}) := \Bigl\{(x, z) \in [a, b]^n \times \mathbb{R}^m \Bigm| (x, z) \in Z, x \in \bigcup_{i=1}^{r} F_i\Bigr\}$. Observe that by choosing $\mathcal{F} = \{[a, b]^n\}$, there is a trivial collection of faces such that $\text{conv}\bigl(S(Z, a, b)\bigr) = \text{conv}\bigl(X(Z, a, b, \mathcal{F})\bigr)$. However, more importantly, as we shall discuss later, there are many situations, where we can identify an exponential collection of faces $\mathcal{F}'$ such that $\text{conv}\bigl(S(Z, a, b)\bigr) = \text{conv}\bigl(X(Z, a, b, \mathcal{F}')\bigr)$ and $\text{conv}\bigl(X(Z, a, b, \{F_i\})\bigr)$ has a polynomial (possibly extended) formulation. For concreteness, consider $Z = \bigl\{(x, z) \mid z = \prod_{i=1}^{n} x_i\bigr\}$. In this case, $\mathcal{F}' = \{a, b\}^n$, the extreme points of $[a, b]^n$ satisfies the preceding hypotheses. Although, in these situations, an extended formulation for $\text{conv}(S)$ can be constructed using disjunctive programming, such results have found limited use, since the size of $\mathcal{F}'$ is often exponential as in our example. Next, we argue that Theorem 2.4 allows the construction of a polynomial-size extended formulation in these instances.

Theorem 4.3. *Let $a, b \in \mathbb{R}$, Z be a compact permutation-invariant set, and $\mathcal{F} = \{F_1, \dots, F_r\}$ be a collection of faces of $[a, b]^n$ such that $\text{conv}\bigl(S(Z, a, b)\bigr) = \text{conv}\bigl(X(Z, a, b, \mathcal{F})\bigr)$. Moreover, assume that $\text{conv}(X(Z, a, b, \{F_i\}))$ has a polynomial-sized compact extended formulation. Then, $\text{conv}\bigl(S(Z, a, b)\bigr)$ has a polynomial-sized extended formulation.*

Proof. For brevity of notation, in this proof, we shall write $S(Z, a, b)$ as S and $X(Z, a, b, \mathcal{F})$ as $X(\mathcal{F})$. We construct $\text{conv}(S)$ using its equivalence to $\text{conv}\bigl(X(\mathcal{F})\bigr)$. We may assume for computing $\text{conv}\bigl(X(\mathcal{F})\bigr)$, by taking the union of all permutations of $X(\mathcal{F})$ w.r.t. x if necessary, that $X(\mathcal{F})$ is permutation-invariant in x. This is because a permutation of $X(\mathcal{F})$ w.r.t. x, say $X_\pi(\mathcal{F}) := \bigl\{(x, z) \mid \pi(x) \in X_\pi(\mathcal{F})\bigr\}$, is contained in $\text{conv}\bigl(X(\mathcal{F})\bigr)$ as is seen from $X_\pi(\mathcal{F}) \subseteq S(Z, a, b) \subseteq \text{conv}\bigl(S(Z, a, b)\bigr) = \text{conv}\bigl(X(\mathcal{F})\bigr)$, 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 2.4 to construct $\text{conv}\bigl(X(\mathcal{F})\bigr)$. We first show that we can limit the faces of $[a, b]^n$ that need to be considered in the construction of S_0. Consider an arbitrary face F_i of $[a, b]^n$, which is determined by setting a set of variables with indices in $B_i \subseteq \{1, \dots, n\}$ to their upper bound b and a disjoint set of variables $A_i \subseteq \{1, \dots, n\}$ to their lower bound a. The only faces, F_i, $i = 1, \dots, r$ that need to be considered are such that B_i and A_i are *hole-free*, *i.e.*, B_i is of the form $\{1, \dots, p\}$ and A_i is of the form $\{q, \dots, n\}$. To see this, let $j(i) = \max\{j \mid j \in B_i\}$ and $X_i = X(\{F_i\}) \cap \{(x, z) \mid x_1 \geq \cdots \geq x_n\}$. Assume i' is the index of a face such that $j(i') > |B_{i'}|$, $X_{i'}$ contains a point which is not in $\bigcup_{i:j(i)=|B_i|} X_i$ and, among all such faces, i' is chosen 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 $A_{i''} = A_{i'}$. Such a face exists in $\mathcal{F}$ since we assumed that for every face $F_i \in \mathcal{F}$, $\mathcal{F}$ contains all faces obtained by permuting the variables, and $F_{i''}$ is obtained from $F_{i'}$ by exchanging the variables x_j and $x_{j(i')}$. Moreover, since $X_{i''} \supseteq X_{i'}$, it contains a point not in $\bigcup_{i:j(i)=|B_i|} X_i$, 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.

Now, it is easy to see that there are at most $\binom{n+2}{2}$ many such faces, one for each choice of (p, q), where $0 \leq p \leq q - 1 \leq n$. Since each $X(\{F_i\})$ has a polynomial-sized compact extended formulation, it follows, by disjunctive programming, that $\text{conv}(S_0)$ has a polynomial-sized compact extended formulation. Then, the result follows directly from Theorem 2.4. □

We record and summarize the extended formulation of $\text{conv}\big(S(Z,a,b)\big)$ for later use in the following result. We say a collection of faces $\mathcal{F}$ is permutation-invariant, if for a face described by an inequality, there is another face in $\mathcal{F}$ that is described by a permutation of coefficients of the inequality.

Corollary 4.3.1. *Let $a, b \in \mathbb{R}$, Z be a compact permutation-invariant set, and $\mathcal{F} = \{F_1, \dots, F_r\}$ be a collection of faces of permutation-invariant faces of $[a,b]^n$. Let $I = \big\{i \mid \exists p, q, p < q,$ s.t. $\forall x \in F_i, x_i = b$ if $i \le p$ and $x_i = a$ if $i \ge q\big\}$. Then,*

$$\text{conv}\big(X(Z,a,b,\mathcal{F})\big) = \Big\{(u,z) \;\Big|\; \text{conv}\Big(\bigcup_{i\in I} X\big(Z,a,b,\{F_i\}\big)\Big), u_1 \ge \dots \ge u_n, u \ge_m x\Big\}.$$

Theorem 4.3 shows that even though the number of faces in $\mathcal{F}$ is exponentially many, we can exploit the permutation-invariance of the set to consider only polynomially many faces in the construction. More explicitly, there are $2^{n-d}\binom{n}{d}$ d-dimensional faces of $[a,b]^n$ and $\binom{n+1}{d+1}$ d-dimensional faces of the simplex $b \ge x_1 \ge \dots \ge x_n \ge a$. But, there are only $n-d+1$ of the faces of the hypercube, namely, F_l for $l \in \{0, \dots, n-d\}$, where $B_l = \{1, \dots, l\}$ and $A_l = \{l-d+1, \dots, n\}$.

Combining Lemmas 4.1 and 4.2, it follows that we may restrict ϕ to the set of points which have no other point majorizing them in the domain. For a point $x' \in M$ as defined in Lemma 4.2, it follows that there must be a vector a such that $\langle a, x - x'\rangle \le 0$ is valid for P, where coefficients of a can be sorted in a monotonic decreasing sequence. Now, construct a graph $G = (V, E)$ where the vertices are labeled $1, \dots, n$. Then, for $\{i, j\}$ connect the vertices with a directed arc labeled with the index of the facet-defining inequality if the coefficient of x_i is larger than that of x_j in the inequality. Then, for a point x' to be in M, it must be tight on inequalities that yield a hamiltonian path through the vertices. For example, if each inequality only yields k arcs, then x' must be tight on $\lceil \frac{n-1}{k} \rceil$ facet-defining inequalities. As such, it will belong to a face of P of dimension at most $n - \lceil \frac{n-1}{k} \rceil$. This is particularly interesting in the case of hypercubes, where $k = 1$ and the result implies that the function can be restricted to one-dimensional faces for the purpose of constructing convex envelope of the function or its level set. In this case, the direct proof is straightforward, and we include it below.

Theorem 4.1. *Consider a Schur-concave function $\phi(x) : [a,b]^n \mapsto \mathbb{R}$ and the set $S^\alpha : \{(x,\phi) \mid \phi(x) \le \phi \le \alpha, x \in [a,b]^n\}$. For any $x \in [a,b]^n$ let $S(x) = \sum_{i=1}^n (x_i - a)$. For any $s \in \mathbb{R}$, define $i^s = \max\big\{i \mid i(b-a) < S(s)\big\}$ and*

$$u_i^x = \begin{cases} b & \text{if } i \le i^x \\ a + s - (b-a)i^x & \text{if } i = i^x + 1 \\ a & \text{otherwise.} \end{cases} \tag{23}$$

Let $\Theta^\alpha := \big\{(x,\phi) \mid \phi\big(u^{S(x)}\big) \le \phi \le \alpha, x \in [a,b]^n\big\}$. Then, $\text{conv}\big(S^\alpha\big) = \text{conv}(\Theta^\alpha)$. Moreover, if $\phi(\cdot)$ is convex when all but one x variable is fixed, Θ^α is convex.

Proof. We first show that $u^{S(x')} \ge_m x'$. This follows because $u^{S(x')}$ simultaneously maximizes the continuous knapsack problems $\max\big\{\sum_{i=1}^j x_i \mid \sum_{i=1}^j x_i = S(x) + na, x \in [a,b]^n\big\}$ 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, $\phi(u^{S(x')}) \le \phi(x') \le \phi \le \alpha$, where the first inequality follows from Schur concavity of ϕ and $u^{S(x')} \ge_m x'$, and the remaining inequalities follow because (x', ϕ') is feasible to S^α. Therefore, $(x', \phi') \in \Theta^\alpha$.

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

To show the last statement, we write Θ^α as $\text{proj}_{(x,\phi)} \Xi^\alpha$, where $\Xi^\alpha = \big\{(x,s,\phi) \mid \varphi(s) \le \phi \le \alpha, x \in [a,b]^n, s = \sum_{i=1}^n (x_i - a)\big\}$ and $\varphi(s) = \phi(u^s)$. The result follows if $\varphi(s)$ is convex over $[0, n(b-a)]$ since Θ^α is expressed as the projection of a convex set, Ξ^α. First, observe that, for $s \in \big(i(b-a), (i+1)(b-a)\big)$, the convexity of $\varphi(s)$ follows from the assumed convexity of $\phi(u^s)$ when u^s varies only along the i^{th} coordinate. Choose $k \in \{0, \dots, n-1\}$ and let $\bar{s} = k(b-a)$. To prove the result, it suffices to check that the left derivative of $\varphi(s)$ at $\bar{s}$ is no more than the corresponding right derivative. For sufficiently small $\epsilon > 0$, observe that $u^{\bar{s}} + \epsilon e_k \ge_m u^{\bar{s}} + \epsilon e_{k+1}$ because

$b = u^{\bar{s}}_k > u^{\bar{s}}_{k+1} = a$. Since $\phi(\cdot)$ is Schur-concave, it follows that $\phi(u^{\bar{s}} + \epsilon e_k) \le \phi(u^{\bar{s}} + \epsilon e_{k+1}) = \varphi(\bar{s} + \epsilon)$. Then, the following chain of inequalities follows

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

where the first equality is by the definition of $\varphi(\cdot)$ and $u^{\bar{s}}$, 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 $\phi(u^{\bar{s}} + \epsilon e_k) \le \varphi(\bar{s} + \epsilon)$ and $\phi(u^{\bar{s}}) = \varphi(\bar{s})$.

□

In essence, Proposition 4.1 shows that we can reduce our attention to the edges of the hypercube in our construction of the convex hull of S^α. A similar result can be shown for upper level sets of quasiconcave functions over general polytopes]. 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] 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 Propostion 4.1 also applies to level sets of the functions while the result in [] only applies to convex envelope construction.

Many of the applications of Theorem 4.3 extend beyond Schur-concave functions. For example, if we consider the convex hull of $\{(x, \alpha) \in [a, b]^n \times \mathbb{R} \mid \prod_{i=1}^n x_i \le \alpha\}$, where a is not necessary positive. Note that the product function is not Schur-concave when some of the variables can be negative, for example consider $x_1x_2x_3$ and observe that although $(1, -1, -3) \ge_m (0, 0, -3)$, the function value is higher at $(1, -1, , -3)$ than at $(0, 0, 3)$.

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

$$\operatorname*{conv}_{[a,b]^n} \phi(x) = \max\bigl\{f(u) \bigm| u \ge_m x, b \le u_1 \ge \cdots \ge u_n \ge a\bigr\}. \tag{24}$$

Proof. Observe that the points in $\{a, b\}^n$ that intersect with $x_1 \ge \cdots \ge x_n$ are precisely the columns $p_{\cdot j}$ described in the statement of the result. Consider the column $p_{\cdot j}$ and observe that $f\bigl(p_{\cdot j}\bigr) = \phi(p_{\cdot j})$. Moreover, $f(x)$ is linear. Let $\Delta = \{x \in \{a, b\}^n \mid b \ge x_1 \ge \cdots \ge x_n \ge a\}$. Then, we show that $f(x) = \text{conv}_{\text{conv}(\Delta)}(\phi|_\Delta)$, where $\phi|_\Delta$ denotes the restriction of ϕ to Δ. Clearly, $f(x) \le \text{conv}_{\text{conv}(\Delta)}(\phi|_\Delta)$ because it matches ϕ_Δ at all the points in the domain and is a convex underestimator. Also, $f(x) \ge \text{conv}_{\text{conv}(\Delta)}(\phi|_\Delta)$ because of Jensen's inequality applied to $\text{conv}_{\text{conv}(\Delta)}(\phi)$, exactness of $f(x)$ at the extreme points of Δ, and affinity of $f(x)$. Now consider Corollary 4.3.1. Let $\mathcal{F}$ be the extreme points of $[a, b]^n$ and $Z = \{(x, z) \mid z \ge \phi(x)\}$. Then, by assumption, $\text{conv}\bigl(S(Z, a, b)\bigr) = \text{conv}\bigl(X(Z, a, b, \mathcal{F})$. Then, in the statement of Corollary 4.3.1, $R = \bigl\{(i, i+1) \bigm| i = 1, \dots, n-1\bigr\}$ and $\text{conv}\Bigl(\bigcup_{(p,q)\in R} X(Z, a, b, \{F_{i_{pq}}\})\Bigr) = \{(x, z) \mid z \ge f(x), b \ge x_1 \ge \cdots \ge x_n \ge a\}$. Then, the result follows from Corollary 4.3.1.

□

Theorem 4.3. *Consider $S = \Bigl\{(x, y) \in H \Bigm| \prod_{i=1}^m y_i^\alpha \ge \prod_{j=1}^n x_j^\beta\Bigr\}$, where $H = [c, d]^m \times [a, b]^n$, with $a \ge 0$, $c \ge 0$, $\alpha > 0$, and $\beta > 0$. Let $k = \min\{m, \lfloor\frac{\beta}{\alpha}\rfloor\}$. Define the convex sets:*

$$\begin{aligned} S_{ij} &= S \cap \Bigl\{(y_r)_1^i = d, y \in \Delta_m, (y_r)_{i+k+1}^m = c;\ (x_s)_1^j = b, (x_s)_{j+2}^n = a\Bigr\} \\ C_j &= S \cap \Bigl\{y \in \Delta_m;\ (x_s)_1^j = b, (x_s)_{j+1}^n = a\Bigr\} \end{aligned}$$

for $i = 0, \dots, m-k$ and $j = 0, \dots, n-1$. Let $T = \bigcup_{i,j} S_{ij} \cup \bigcup_j C_j$. The convex hull S is obtained as:

$$\text{conv}(S) = X := \{(x, y) \mid v \succeq y, u \succeq x, (u, v) \in \text{conv}(T)\}.$$

In particular, if $m\alpha \le \beta$, then the convex hull is given by

$$\text{conv}(S) = X' := \left\{(x, y) \in H \,\middle|\, \prod_{i=1}^m y_i^{\frac{1}{m}} \ge \prod_{j=1}^n u(x)_j^{\frac{\beta}{m\alpha}}\right\}, \tag{25}$$

where $u(x)$ is defined as follows. For any $x \in [a,b]^n$ and $s \in \mathbb{R}$, let $S(x) = \sum_{i=1}^n (x_i - a)$. Then,

$$u(x)_i = \begin{cases} b & \text{if } i \le i^x \\ a + s - (b-a)i^x & \text{if } i = i^x + 1 \\ a & \text{otherwise.} \end{cases} \tag{26}$$

Proof. Let $\phi(x) := \prod_{j=1}^n x_j^\beta$ and consider the set $\Upsilon(\gamma) = \{x \in [a,b]^n \mid \phi(x) \le \gamma\}$. By Theorem 3.A.3 in [22], $\phi(x)$ is Schur-concave over $[a,b]^n$ because it is permutation invariant and $\frac{\partial\phi}{\partial x_1} \le \frac{\partial \phi}{\partial x_2}$ at any point with $x_1 \ge \cdots \ge x_n$. Let $\Upsilon_i(\gamma) = \bigl\{x \bigm| x_i^\beta b^{(i-1)\beta} a^{(n-i)\beta} \le \gamma\bigr\}$. Then, it follows by Proposition 4.1 and Corollary 4.3.1 that $\text{conv}(\Upsilon(\gamma)) = \bigl\{x \bigm| u \ge_m x, u_1 \ge \cdots \ge u_n, u \in \text{conv}\bigl(\bigcup_{i=1}^n \Upsilon_i(\gamma)\bigr)\bigr\}$.

Now, let $\psi(y) := \prod_{i=1}^m y_i$ and consider the set $\Theta = \bigl\{(x,y) \in [a,b] \times [c,d]^m \bigm| \zeta\psi(y) \ge \delta x^{\frac{\beta}{\alpha}}\bigr\}$, where $\delta, \zeta \ge 0$. Consider a point $(x', y') \in \Theta$. Then, by restricting attention to $\bar{y} = \lambda y'$, we obtain the subset Λ of Θ such that $\Lambda = \bigl\{(x,\lambda) \in [a,b] \times [c',d'] \bigm| \lambda\theta \ge \delta' x^{\frac{\beta}{m\alpha}}\bigr\}$, where $\theta = \zeta^{\frac{1}{m}}\psi(y')^{\frac{1}{m}}$, $\delta' = \delta^{\frac{1}{m}}$, $c' = \max\{\lambda \mid \lambda y'_i \le c \text{ for some } i\}$, and $d' = \min\{\lambda \mid \lambda y'_i \ge d \text{ for some } i\}$. If $x' \in \{a,b\}$, the point belongs to the convex subset of Θ obtained by fixing x' at its current value because the defining inequality can be written as $\zeta^{\frac{1}{m}}\psi(y)^{\frac{1}{m}} \ge \delta x'^{\frac{\beta}{\alpha}}$, a convex inequality. Therefore, we may assume that $x' \in (a,b)$. First, consider the case when $y' \in (c,d)^n$. Then, it follows that $c' < 1$ and $d' > 1$ and $(x', 1) \in \Lambda$. Assume $m > \frac{\beta}{\alpha}$ and let $s = \delta' \frac{\beta}{m\alpha}(x')^{\frac{\beta}{m\alpha}-1}$. If $x' \in (a,b)$, then, for $0 < \epsilon \le \frac{\min\{x'-a, b-x'\}}{\max\{\theta, 1\}}$, we show that $(x', 1)$ can be written as a convex combination of $(x' - \epsilon\theta, 1 - s\epsilon)$ and $(x' + \epsilon\theta, 1 + s\epsilon)$. The latter points are feasible in Λ because:

$$\delta'(x' \pm \epsilon\theta)^{\frac{\beta}{m\alpha}} \le \delta' x'^{\frac{\beta}{m\alpha}} + s(x' \pm \epsilon\theta - x') \le \theta(1 \pm s\epsilon),$$

where the first inequality is by concavity of $x^{\frac{\beta}{m\alpha}}$ for $m \ge \frac{\beta}{m\alpha}$ and the second inequality is because $\delta' x'^{\frac{\beta}{m\alpha}} \le \theta$ by the feasibility of $(x', 1)$ in Λ. Since $\psi(y)^{\frac{1}{m}}$ is homogenous, we have expressed (x', y') as a convex combination of $\bigl(x' - \epsilon\theta, (1 - s\epsilon)y'\bigr)$ and $\bigl(x' - \epsilon\theta, (1 + s\epsilon)y'\bigr)$, each of which is feasible to Θ. Since $\epsilon > 0$ and $x' > a \ge 0$ implies $s > 0$, it follows that these points are distinct and that (x', y') is not an extreme point of the feasible region. Therefore, we may assume that there exists an i such that $y'_i \in \{c,d\}$. However, in this case, we can reduce the dimension of the set by fixing y_i at y'_i and effectively reducing m. In other words, we may assume without loss of generality that $m \le \frac{\beta}{\alpha}$. Then, we rewrite the defining inequality of Θ as $\zeta^{\frac{1}{m}}\psi(y)^{\frac{1}{m}} \ge \delta^{\frac{1}{m}} x^{\frac{\beta}{m\alpha}}$ and observe that this is a convex inequality since $\psi(y)^{\frac{1}{m}}$ is a concave function and $x^{\frac{\beta}{m\alpha}}$ is a convex function. Therefore, we need to consider faces where either all x_j are fixed at their bounds or where we fix all y_i except for a set of cardinality $\min\bigl\{m, \bigl\lfloor \frac{\beta}{\alpha} \bigr\rfloor\bigr\}$ and fix all x_j except for one.

Since $S_{ij} \subseteq S$ and $C_j \subseteq S$ 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 by permuting x and/or y variables, it follows that the extreme point belongs to X. Therefore, $X = \text{conv}(S)$.

Now, we consider the case $m\alpha \le \beta$. Clearly, $k = m$. If we fix y at $\bar{y}$, it follows from Proposition 4.1 that the convex hull of this slice is defined by $\prod_{i=1}^m \bar{y}_i^{\frac{1}{m}} \ge \prod_{j=1}^n u(x)_j^{\frac{\beta}{m\alpha}}$. Then, as in the proof of Proposition 4.1, we let $\varphi(s) = \prod_{j=1}^n u(x)_j^{\frac{\beta}{m\alpha}}$, where $s = \sum_{i=1}^n (x_i - a)$, and rewrite the above inequality as $\varphi(s) - \prod_{i=1}^m y_i^{\frac{1}{m}} \le 0$. Since the left-hand-sise is jointly convex in (s,y) and s is a linear function of x, this proves that X' in (25) is convex in (x,y). By Schur-concavity of $\prod_{j=1}^n x_j^{\frac{\beta}{m\alpha}}$, it follows that $\prod_{j=1}^n x_j^{\frac{\beta}{m\alpha}} \ge \prod_{j=1}^n u(x)_j^{\frac{\beta}{m\alpha}}$. This implies that $S \subseteq X' \subseteq \text{conv}(S)$. Since X' is convex, it is the same as $\text{conv}(S)$.

□

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 $\inf_{(x,u)\in C} \|x - \bar{x}\|$, where C is the extended space representation of X. By duality, the optimal value matches $\max_{\|a\|_* \le 1}\bigl\{\langle \bar{x}, a\rangle - h(a)\bigr\}$, 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 $\langle \bar{x}, a^*\rangle - z^* \ge \langle x, a^*\rangle$ for all $x \in \text{proj}_x C$ and this inequality separates $\bar{x}$ from C.

Given the structure of permutation-invariant sets and their extended space representation admits an alternate approach. Assume we are interested in developing the convex envelope of a permutation-invariant function ϕ, such as $x_1 \dots x_n$, over $[a, b]^n$. Let the convex envelope of ϕ at x be obtained by expressing x as a convex combination of u and its permutations, where $u \geq_m x$. Moreover, assume that the convex envelope at u is obtained as a convex combination of the extreme points of the simplex $a \leq x_1 \leq \cdots \leq x_n \leq b$. We can collect the extreme points of this simplex as the columns of a matrix V and write $x = SV\gamma$. Then, it follows that there exists a representation of $x = Su$, where S is a doubly stochastic matrix.

However, such an inequality is typically not facet-defining for conv(X) even when the latter set is polyhedral. We now discuss a separation procedure that can generate facet-defining inequalities. We implement it for convex hulls of multilinear sets over $[a, b]^n$ to evaluate their impact on the quality of the relaxation. For the purpose of illustration, consider the special case of $\prod_{i=1}^n x_i$ over $[a, b]^n$. In this case, (24) reduces to:

$$\begin{aligned} \min \quad & a^n + \sum_{i=1}^n b^{i-1} a^{n-i} (u_i - a) \\ \text{s.t.} \quad & x \preceq_m u \\ & a \leq u_n \leq \cdots \leq u_1 \leq b \end{aligned}$$

Given an $x \in \mathbb{R}^n$ in general position inside $[a, b]^n$, assume that the optimal solution to the above problem is u. Then, we express $x = Su$, where $S \in \mathbb{R}^{n \times n}$ is a doubly-stochastic matrix. Given x and u, this problem can be solved as a linear program. In our implementation, we use this approach, given the simplicity, although S can also be derived as a product of T-transforms (see proof of Lemma 2 in Section 2.19 of *Inequalities, by Hardy, Littlewood, Polya* []). Then, we express S as a convex combination of permutation matrices. Such a representation exists due to Birkhoff Theorem and can be obtained by the following straightforward algorithm. Observe that in such a representation all permutation matrices with non-zero convex multipliers must 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 $\{1, \dots, n\}$ 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 bipartite 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 π to the minimum non-zero value of $S_{ij} P_{ij}$. Observe that $\frac{1}{1-\pi}(S - \pi P)$ 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 and observe that for each such u, the convex envelope is given by the optimal function value of the above linear program. Each permuted u can be expressed as a convex combination of the corner points of the permuted simplex $\{a \leq u_1 \leq \cdots u_n \leq b\}$. The extreme points with non-zero multipliers must all be tight on the inequality. Then, we obtain the inequality by fitting an inequality to be tight at these points.

5 Stochastic Dominance

In this section, we revisit the convexification result in [12]. Consider two random variables X, Y in $(\Omega, 2^\Omega, P)$ where Ω is the sample space and P is a probability measure. Let F_X and F_Y be cumulative distribution function of X and Y respectively. Then, X is said to *dominate in the first order* Y if

$$F_X(t) \leq F_Y(t)$$

for all $t \in \mathbb{R}$. On the other hand, X is said to *dominate in the second order* Y if

$$\int_{-\infty}^t F_X(s)ds \leq \int_{-\infty}^t F_Y(s)ds.$$

Consider random variables X, Y in $(\Omega := \{1, \dots, n\}, 2^\Omega, P)$ where P is defined as $P(k) = 1/n$ for all $k = 1, \dots, n$ and define vectors x and y as $x_i = X(i)$ and $y_i = Y(i)$ for $i = 1, \dots, n$. Then, the notions of the stochastic dominances can be equivalent written in terms of x and y. That is, X dominates in the first order Y if and only if $y_{[i]} \geq x_{[i]}$ and X dominates in the second order Y if and only if $x \geq^{wm} y$.

For a given $y \in \mathbb{R}_n$, define

$$A_1 = \{x : y_{[i]} \geq x_{[i]}, i = 1, \dots, n\}, \quad A_2 = \{x : x \geq^{wm} y\}, \quad A_3 = \{x : y \geq_{wm} x\}.$$

The following lemma demonstrates the existence of an intermediate vector between two vectors where a vector weakly majorizes the other.

Lemma 5.1 ([22], 5.A.9., 5.A.9.a.). *1. Suppose $x \geq_{wm} y$. Then, there exist u and v such that*

$$x \geq_m u \text{ and } u \geq y, \quad x \geq v \text{ and } v \geq_m y.$$

2. Suppose $x \geq^{wm} y$. Then, there exist u and v such that

$$x \geq_m u \text{ and } y \geq u, \quad v \geq x \text{ and } v \geq_m y.$$

We next prove the following convexification result.

Theorem 5.1 (Dentcheva and Ruszczyński, [12]). *For a given $y \in \mathbb{R}^n$,*

$$\text{conv}(A_1) = A_2 = A_3.$$

Proof. Observe that A_1 is permutation-invariant and hence by Theorem 2.7,

$$\text{conv}(A_1) = \left\{ x : \begin{array}{l} y_{[i]} \geq u_i, i = 1, \ldots, n \\ u_1 \geq \cdots \geq u_n, \\ u \geq_m x \end{array} \right\}. \tag{27}$$

In this proof, let K be the right-hand side of (27). For every $x \in K$, let u be satisfying the constraints in (27). Then, for each $j = 1, \ldots, n$, it is clear that

$$\sum_{i=j}^{n} x_{[i]} \leq \sum_{i=j}^{n} u_i \leq \sum_{i=j}^{n} y_{[i]}$$

where the first inequality follows from the majorization inequality. Therefore, $x \in A_2$. On the other hand, let $x \in A_2$ so that $x \geq^{wm} y$. Then, by Lemma 5.1, there exists v such that $y \geq v$ and $v \geq_m x$. By defining u by $u_i = v_{[i]}, i = 1, \ldots, n$, (x, u) satisfies all constraints in (27). This shows that $x \in K$. □

Notice that Theorem 5.1 is a convexification result in a projected space for a given y value. However, Theorem 2.7 enables us to extend the result to the original space so that the convex hull can be obtained in a higher dimensional space.

Theorem 5.2. *Let $B_1 = \{(x, y) : y_{[i]} \geq x_{[i]}, i = 1, \ldots, n\}$. Then,*

$$\text{conv}(B_1) = \left\{ (x, y) : \begin{array}{l} v_i \geq u_i, i = 1, \ldots, n, \\ u_1 \geq \cdots \geq u_n, \\ v_1 \geq \cdots \geq v_n, \\ u \geq_m x, \\ v \geq_m y \end{array} \right\}.$$

As we discussed earlier, the convex hull can be described using linear inequalities after modeling the majorization inequalities.

6 Modeling logical constraints

Next, we discuss the formulation of certain logical requirements in 0-1 variables. In particular, we study a set that generalizes various models described in the literature. We present an extended formulation for the convex hull of this set, which we obtain through the use of the results of Section 2. This characterization provides streamlined convex hull derivations for various sets studied in the literature.

In the ensuing discussion, for a subset T of $P := \{1, \ldots, p\}$, we use the notation e_T to represent the vector in $\mathbb{R}^p$ having entries with index in T equal to 1, and entries with index in $P \backslash T$ equal to 0. We also use e_i as a shorthand notation for $e_{\{i\}}$ and e as a shorthand notation for e_P.

Consider integers $0 \leq k_1 < k_2 < \ldots < k_r \leq m$ and $0 \leq l_1 < l_2 < \ldots < l_s \leq n$. Define $M := \{1, \ldots, m\}$, $N := \{1, \ldots, n\}$, $K := \{k_1, \ldots, k_r\}$, and $L := \{l_1, \ldots, l_s\}$. We do not require that m or n is positive. When $m = 0$ (*resp.* $n = 0$), we consider M (*resp.* N) to be empty.

We are interested in the logical constraint on binary variables $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ that requires that if the number of variables out of $x_1, \ldots, x_m$ that are true belongs to K, then the number of variables out of $y_1, \ldots, y_n$ that are true belongs to L. Formally, we study

$$S_{m,n}(K, L) := \left\{(x, y) \in \{0, 1\}^{m+n} \,\middle|\, e^\intercal x \in K \implies e^\intercal y \in L\right\}.$$

We refer to this set simply as S whenever m, n, K and L are clear from the context. In Proposition 6.1, we obtain a description of conv(S). To the best of our knowledge, the polyhedral structure of this set has not been investigated previously.

In studying conv(S), we pose the following assumptions in the remainder of this section.

ASSUMPTION A1: $k_1 > 0$ and $l_1 > 0$.

ASSUMPTION A2: $k_r < m$ and $l_s < n$.

We next argue that Assumption (A1) is without loss of generality (*wlog*). Suppose that $k_1 = 0$. We may define $K' = \{k_1 + 1, k_2 + 1, \ldots, k_r + 1\}$ and consider the set $S_{m+1,n}(K', L)$ whose variables we denote by (x, x_{m+1}, y). It is simple to verify that $S_{m,n}(K, L) = \text{proj}_{(x,y)}\left(S_{m+1,n}(K', L) \cap \mathcal{H}\right)$ where $\mathcal{H}$ is the hyperplane $\{(x, x_{m+1}, y) \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^n \,|\, x_{m+1} = 1\}$. Since $\mathcal{H}$ defines a face of $S_{m+1,n}(K', L)$, it holds that $\text{conv}(S_{m,n}(K, L)) = \text{proj}_{(x,y)}\Big(\text{conv}\left(S_{m+1,n}(K', L)\right) \cap \mathcal{H}\Big)$. Verifying that Assumption (A2) is wlog is similar. In particular, when $k_r = m$, it suffices to study $S_{m+1,n}(K, L)$ whose variables we denote by (x, x_{m+1}, y) to obtain the convex hull of $S_{m,n}(K, L)$ as $S_{m,n}(K, L) = \text{proj}_{(x,y)}\left(S_{m+1,n}(K, L) \cap \mathcal{H}\right)$ where $\mathcal{H}$ is the hyperplane $\{(x, x_{m+1}, y) \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^n \,|\, x_{m+1} = 0\}$.

Theorem 6.1. *Under Assumptions (A1) and (A2),* $\text{conv}(S_{m,n}(K, L)) = X$ *where*

$$X := \left\{ (x, y) \in \mathbb{R}^{m+n} \,\middle|\, \begin{array}{l} u \in \Delta^m, \quad u_1 = 1, \quad u_m = 0, \quad u \geq_m x \\ v \in \Delta^n, \quad v_1 = 1, \quad v_n = 0, \quad v \geq_m y \\ v_{l_1} - \sum_{i=1}^{s-1}\left(v_{l_i+1} - v_{l_{i+1}}\right) - v_{l_s+1} \geq u_{k_1} - \sum_{i=1}^{r-1}\left(u_{k_i+1} - u_{k_{i+1}}\right) - u_{k_r+1} \end{array} \right\}.$$

Proof. Because S is permutation-invariant with respect to variables x but also with respect to variables y, Theorem 2.4 establishes that

$$\text{conv}(S) = \underset{(x,y)}{\text{proj}} \left\{(x, u, y, v) \in \mathbb{R}^{2m+2n} \,\middle|\, (u, v) \in \text{conv}(S_0), u \geq_m x, v \geq_m y\right\},$$

where $S_0 = \{(u, v) \in \{0, 1\}^{m+n} \,|\, u \in \Delta^m, v \in \Delta^n, e^\intercal u \in K \implies e^\intercal v \in L\}$. We argue next that

$$\text{conv}(S_0) = X_0 := \left\{ (u, v) \in \mathbb{R}^{m+n} \,\middle|\, \begin{array}{ll} u \in \Delta^m, \quad u_1 = 1, \quad u_m = 0 & (28.1) \\ v \in \Delta^n, \quad v_1 = 1, \quad v_n = 0 & (28.2) \\ v_{l_1} + \sum_{i=1}^{s-1}\left(v_{l_{i+1}} - v_{l_i+1}\right) - v_{l_s+1} & \\ \quad \geq u_{k_1} + \sum_{i=1}^{r-1}\left(u_{k_{i+1}} - u_{k_i+1}\right) - u_{k_r+1} & (28.3) \end{array} \right\},$$

which will conclude the proof.

Given $\dot{u} \in \Delta^m_+$, we define $U(\dot{u}) = (\dot{u}_{k_1} - 1) + \sum_{i=1}^{r-1}(\dot{u}_{k_{i+1}} - \dot{u}_{k_i+1}) + (0 - \dot{u}_{k_r+1})$ and $\theta_j(\dot{u}) = \dot{u}_j - \dot{u}_{j+1} \geq 0$ for $j = 1, \ldots, m-1$. It is clear that $\sum_{j=1}^{m-1} \theta_j(\dot{u}) = \dot{u}_1 - \dot{u}_m \leq 1$ and $U(\dot{u}) = -1 + \sum_{i=1}^{r} \theta_{k_i}(\dot{u})$. It follows that $U(\dot{u}) \in [-1, 0]$. In addition, if $\dot{u} \in \{0, 1\}^m$, we have that $U(\dot{u}) \in \{-1, 0\}$ as $\theta_j(\dot{u}) \in \{0, 1\}$ for $j = 1, \ldots, m-1$. Further, in this case, $U(\dot{u}) = 0$ if and only if $\theta_{k_i}(\dot{u}) = 1$ for some $i \in \{1, \ldots, r\}$, *i.e.*, $\sum_{j=1}^{m} \dot{u}_j = k_i \in K$. Similar results hold for $V(\dot{v}) = (\dot{v}_{l_1} - 1) + \sum_{i=1}^{s-1}(\dot{v}_{l_{i+1}} - \dot{v}_{l_i+1}) - \dot{v}_{l_s+1}$ when $\dot{v} \in \Delta^n_+$.

First consider $(\dot{u}, \dot{v}) \in S_0$. Because $\dot{u} \in \Delta^m \cap \{0, 1\}^m$, then $U(\dot{u}) \in \{-1, 0\}$. Further, $U(\dot{u}) = 0$ if and only if $e^\intercal \dot{u} \in K$. Similarly, we have that $V(\dot{v}) \in \{-1, 0\}$ with $V(\dot{v}) = 0$ if and only if $e^\intercal \dot{v} \in L$. We show that $(\dot{u}, \dot{v}) \in X_0$. There are two cases. If $e^\intercal \dot{u} \notin K$, then $U(\dot{u}) = -1$ and (28.3) is trivially satisfied. If $e^\intercal \dot{u} \in K$, then $e^\intercal \dot{v} \in L$. Hence, constraint (28.3) is satisfied as $U(\dot{u}) = V(\dot{v}) = 0$. We conclude that $S_0 \subseteq X_0$, which implies that $\text{conv}(S_0) \subseteq X_0$ as X_0 is convex being polyhedral.

Second consider any extreme point $(\bar{u}, \bar{v})$ of the polytope X_0. We first show that $(\bar{u}, \bar{v}) \in \{0, 1\}^{m+n}$. On the one hand, assume that (28.3) is not tight at $(\bar{u}, \bar{v})$. Then, at least $m + n - 4$ among the $m + n - 2$ constraints of Δ^m and Δ^n must be satisfied at equality. Because no more than $m - 1$ (*resp.* $n - 1$) among the constraints of Δ^m (*resp.* Δ^n) can be satisfied at equality simultaneously in a feasible solution as $u_1 \neq u_m$ (*resp* $v_1 \neq v_n$), we conclude that $\bar{u} = \sum_{i=1}^{m_1} e_i$ for $m_1 \in \{1, \ldots, m-1\}$ and $\bar{v} = \sum_{i=1}^{n_1} e_i$ for $n_1 \in \{1, \ldots, n-1\}$. On the other hand, assume that (28.3) is tight at $(\bar{u}, \bar{v})$. Then, at least $m+n-5$ among the $m+n-2$ constraints of Δ^m and Δ^n must be satisfied at equality. There are two cases. In the first, the number of tight constraints in Δ^m and Δ^n is at least $m-3$ and $n-2$, respectively. Then $\bar{u} = \sum_{j=1}^{m_1} e_j + f \sum_{j=m_1+1}^{m_2} e_j$ where $1 \leq m_1 \leq m_2 \leq m$ and $f \in (0,1)$ and $\bar{v} = \sum_{j=1}^{n_1} e_j$ where $1 \leq n_1 \leq n-1$. We claim that $m_1 = m_2$. Assume not. Because $V(\bar{v}) \in \{-1, 0\}$ and (28.3) is tight, we also have that $U(\bar{u}) \in \{-1, 0\}$. It follows that either $m_1, m_2 \notin K$ or that $m_1, m_2 \in K$. In these cases however, consider the solutions $\check{u} = \sum_{j=1}^{m_2} e_j$ and $\hat{u} = \sum_{j=1}^{m_1} e_j$. We see that $\bar{u} = f\check{u} + (1-f)\hat{u}$. Solutions $\check{u}$ and $\hat{u}$ also make (28.3) tight, showing that $(\bar{u}, \bar{v})$ is not an extreme point of X_0, yielding a contradiction. We conclude that $(\bar{u}, \bar{v}) \in \{0,1\}^{m+n}$. A similar argument shows that, for the second case where the number of tight constraints in Δ^m and Δ^n is at least $m-2$ and $n-3$, respectively, $(\bar{u}, \bar{v}) \in \{0,1\}^{m+n}$. Now observe that $\bar{u} \in \Delta^m \cap \{0,1\}^m$ implies that $U(\bar{u}) \in \{-1, 0\}$. Similarly, $\bar{v} \in \Delta_n \cap \{0,1\}^n$ implies that $V(\bar{v}) \in \{-1, 0\}$. As constraint (28.3) imposes that $V(\bar{v}) \geq U(\bar{u})$, we conclude that $e^\intercal \bar{u} \in K \implies e^\intercal \bar{v} \in L$, showing that $(\bar{u}, \bar{v}) \in S_0$. The above discussion establishes that the extreme points of X_0 are binary vectors that belong to S_0. This proves that $X_0 \subseteq \text{conv}(S_0)$. □

We next obtain a description of the convex hull of $S_{m,n}(K, L)$ in the space of original variables by projecting the formulation we obtained in Proposition 6.1 onto the space of variables x and y. We first observe that the corresponding projection cone $C_{m,n}(K, L)$ is described by the inequalities:

$$\begin{array}{llll}
\sum_{p=i}^{m} \alpha_p - \alpha'_m - \beta_{i-1} + \beta_i - \delta_i^{\pm}(K)\gamma = 0, & & & \forall i = 1, \ldots, m \\
\sum_{q=j}^{n} \bar{\alpha}_q - \bar{\alpha}'_n - \bar{\beta}_{j-1} + \bar{\beta}_j + \bar{\delta}_j^{\pm}(L)\gamma = 0, & & & \forall j = 1, \ldots, n \\
\alpha_i \geq 0, & \beta_i \geq 0, & & \forall i = 1, \ldots, m \\
\bar{\alpha}_j \geq 0, & \bar{\beta}_j \geq 0, & & \forall j = 1, \ldots, n \\
\beta_0 \geq 0, & \bar{\beta}_0 \geq 0, & \gamma \geq 0, &
\end{array}$$

where dual variable α_p is associated with constraint $\sum_{i=1}^{p} u_i \geq \max_{S \subseteq N \,|\, |N|=p} x(S)$, β_p is the dual variable associated with constraint $u_p - u_{p+1} \geq 0$ for $p = 1, \ldots, n-1$, β_0 is the dual variable associated with constraint $1 - u_1 \geq 0$, β_n is the dual variable associated with constraint $u_n \geq 0$, and γ is the dual for the constraint (28.3). We also define $\delta^{\pm}(K) = e_K - e_{K+1}$ and $\bar{\delta}^{\pm} = e_L - e_{L+1}$.

Given an element $(\alpha, \alpha'_m, \beta, \bar{\alpha}, \bar{\alpha}'_n, \bar{\beta}, \gamma)$ of $C_{m,n}(K, L)$, we can construct the valid inequality

$$\sum_{i=1}^{m} \alpha_i \max_{S \subseteq M \,|\, |S|=i} \{x(S)\} + \sum_{j=1}^{n} \bar{\alpha}_j \max_{T \subseteq N \,|\, |T|=j} \{y(T)\} - \alpha'_m x(M) - \bar{\alpha}'_n y(N) \leq \beta_0 + \bar{\beta}_0. \tag{28}$$

Because α and $\bar{\alpha}$ are nonnegative, (28) can be expressed as an exponential collection of linear inequalities using the relationships $\max_{S \subseteq M \,|\, |S|=i} \{x(S)\} \geq \sum_{k \in S'} x_k$, for all $S' \subseteq S$ with $|S'| = i$.

In particular, we note that the only component of vectors β and $\bar{\beta}$ that occurs in the above inequality is β_0 and $\bar{\beta}_0$.

Because γ is the only variable linking variables $(\alpha, \alpha'_m, \beta)$ and $(\bar{\alpha}, \bar{\alpha}'_n, \bar{\beta})$ in $C_{m,n}(K, L)$ and because in any ray of $C_{m,n}(K, L)$ we may assume that either $\gamma = 0$ or $\gamma = 1$, it is possible to obtain the extreme rays of the cone $C_{m,n}(K, L)$ by focusing on the rays of the smaller cones $c_p(v)$

$$\begin{array}{lll}
\sum_{i=r}^{p} \alpha_i - \alpha'_p - \beta_{r-1} + \beta_r = -v_r, & & \forall r = 1, \ldots, p \\
\alpha_i \geq 0, & \beta_i \geq 0, & \forall i = 1, \ldots, p \\
\alpha'_p \geq 0, & &
\end{array}$$

where $v \in \mathbb{R}^n$.

Theorem 6.2. *Vector $(\alpha, \alpha'_m, \beta, \bar{\alpha}, \bar{\alpha}'_n, \bar{\beta}, 1)$ is an extreme ray of $C_{m,n}(K, L)$ if and only if $(\alpha, \alpha'_m, \beta)$ is an extreme point of $c_m(\delta^{\pm}(K))$ and $(\bar{\alpha}, \bar{\alpha}'_n, \bar{\beta})$ is an extreme point of $c_n(-\delta^{\pm}(L))$. Further, Vector $(\alpha, \alpha'_m, \beta, \bar{\alpha}, \bar{\alpha}'_n, \bar{\beta}, 0)$ is an extreme ray of $C_{m,n}(K, L)$ if and only if $(\alpha, \alpha'_m, \beta)$ is an extreme ray of $c_m(0)$ and $(\bar{\alpha}, \bar{\alpha}'_n, \bar{\beta})$ is an extreme ray of $c_n(0)$.*

6.1 Cardinality constraints

For $L \subseteq \{0, \ldots, n\}$, consider the set

$$S_n^{\#}(L) := \{y \in \{0,1\}^n \mid e^\intercal y \in L\}.$$

When $L = \{0, 1, \ldots, l\}$, $S_n^{\#}(L)$ models a *no-more-than-l* cardinality requirement. It is clear in this case that $\text{conv}(S_n^{\#}(L)) = \{y \in [0,1]^n \,|\, e^\intercal y \leq l\}$. When L is the set of even integers between 0 and n, $\text{conv}(S_n^{\#}(L))$ was obtained in [17]. More generally, $\text{conv}(S_n^{\#}(L))$ is described in [9, 24], where it is shown that

$$\text{conv}(S_n^{\#}(L)) = \left\{ y \in [0,1]^n \,\middle|\, \begin{array}{r} l_1 \leq e^\intercal y \leq l_s \\ (l_{p+1} - |S|)e_S^\intercal y - (|S| - l_p)e_{N\setminus S}^\intercal y \leq l_p(l_{p+1} - |S|), \\ \forall p = 1, \ldots, s-1 \\ \forall S \subseteq N \text{ with } l_p < |S| < l_{p+1}, \end{array} \right\}.$$

Observe that $S_n^{\#}(L) = \text{proj}_y\, S_{m,n}(K, L)$ for any positive integer m if $K = \{0, \ldots, m\}$. Proposition 6.1 can therefore be used to obtain the following description of $\text{conv}(S_n^{\#}(L))$. For reasons similar to those justifying Assumptions (A1) and (A2), it is wlog to impose

ASSUMPTION A3: $l_1 > 0$ and $l_s < n$.

Theorem 6.3. *Under Assumptions (A3),* $\text{conv}(S_n^{\#}(L)) = X$ *where*

$$X = \left\{ y \in \mathbb{R}^n \,\middle|\, \begin{array}{l} v \in \Delta^n, \quad v_1 = 1, \quad v_n = 0, \\ v \geq_m y \\ v_{l_1} = 1, \quad v_{l_s} = 0 \\ v_{l_i+1} = v_{l_{i+1}}, \quad \forall i = 1, \ldots, s-1 \end{array} \right\}.$$

Proof. For a positive integer m, construct $S_{m+2,n}(K, L)$ where $K = \{1, \ldots, m+1\}$. We have that $S_n^{\#}(L) = \text{proj}_y \left(S_{m+1,n}(K,L) \cap \mathcal{H} \cap \mathcal{H}'\right)$ where $\mathcal{H}$ is the hyperplane $\{(x, x_{m+1}, x_{m+2}) \in \mathbb{R}^m \times \mathbb{R} \,|\, x_{m+1} = 1\}$ and $\mathcal{H}'$ is the hyperplane $\{(x, x_{m+1}, x_{m+2}) \in \mathbb{R}^m \times \mathbb{R} \,|\, x_{m+2} = 0\}$. It follows that

$$\text{conv}(S_n^{\#}(L)) = \text{conv}\left(\text{proj}_y \left(S_{m+2,n}(K,L) \cap \mathcal{H} \cap \mathcal{H}'\right)\right) = \text{proj}_y \left(\text{conv}(S_{m+2,n}(K,L)) \cap \mathcal{H} \cap \mathcal{H}'\right).$$

A description of $\text{conv}(S_{m+2,n}(K, L))$ is given in Proposition 6.1. Because $K = \{1, \ldots, m+1\}$, $k_{i+1} = k_i + 1$ for $i = 1, \ldots, m$ and $r = m + 1$. Then (28.3) reduces to

$$v_{l_1} + \sum_{i=1}^{s-1}(v_{l_{i+1}} - v_{l_i+1}) - v_{l_s+1} \geq u_{k_1} - u_{k_{r+1}} = 1. \tag{29}$$

The constraints of Δ^n imply that $v_{l_1} \leq 1$, $v_{l_{i+1}} - v_{l_i+1} \leq 0$ for $i = 1, \ldots, s-1$ (as $l_{i+1} \geq l_i + 1$), and $v_{l_s} \geq 0$. As (29) must hold under these conditions, we conclude that (28.3) in fact reduces to $v_{l_1} = 1$, $v_{l_s} = 0$ and $v_{l_{i+1}} - v_{l_i+1} = 0$ for $i = 1, \ldots, s-1$. Because the resulting formulation of $\text{conv}(S_{m+2,n}(K, L))$ does not contain constraints linking variables u and v, $\text{proj}_y\, \text{conv}(S_{m+2,n}(K, L))$ is simply obtained by retaining only the constraints involving v, yielding the result. □

6.2 Logical constraints

We now consider the common logical constraint [28] on binary variables $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$ that requires that if at least k out of $x_1, \ldots, x_m$ are true then at least l out of $y_1, \ldots, y_n$ are true. Formally, we consider

$$S_{m,n}^{\Longrightarrow}(k, l) = \left\{(x, y) \in \{0,1\}^{m+n} \,\middle|\, e^\intercal x \geq k \implies e^\intercal y \geq l\right\}.$$

Textbook formulations of this constraint introduce a binary variable z to indicate whether the number of x variables with true assignment has reached level k. The implication can then be replaced with the following linear constraints $(k-1) + (m-k+1)z \geq e^\intercal x$ and $e^\intercal y \geq lz$. A constructive procedure to obtain the facets of $\text{conv}(S_{m,n}^{\Longrightarrow}(k, l))$ is described in [28]. A closed-form description is obtained in [4] using disjunctive programming arguments. Note that [4] obtained the desired description using the fact that the convex hull of unions of monotone polytopes has a simple description, see also [5], while we our derivation relies on permutation invariance. For reasons similar to those justifying Assumptions (A1) and (A2), it is wlog to impose

ASSUMPTION A4: $k > 0$ and $l > 0$.

Theorem 6.4. *Under Assumption (A4),* $\text{conv}(S_{m,n}^{\Longrightarrow}(k,l)) = X$ *where*

$$X = \left\{ (x,y) \in \mathbb{R}^{m+n} \,\middle|\, \begin{array}{lll} u \in \Delta^m, & u \geq_m x, & u_1 = 1 \\ v \in \Delta^n, & v \geq_m y, & v_1 = 1 \\ v_l \geq u_k & & \end{array} \right\}.$$

Proof. We observe that $S_{m,n}^{\Longrightarrow}(k,l)$ is in fact $\text{proj}_{(x,y)}\left(S_{m+1,n+1}(K,L) \cap \mathcal{H} \cap \mathcal{H}'\right)$ where $K = \{k,\ldots,m\}$, $L = \{l,\ldots,n\}$, $\mathcal{H}$ is the hyperplane $\{(x,x_{m+1},y,y_{n+1}) \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R} \,|\, x_{m+1} = 0\}$ and $\mathcal{H}'$ is the hyperplane $\{(x,x_{m+1},y,y_{n+1}) \in \mathbb{R}^m \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{R} \,|\, y_{n+1} = 0\}$.. Then, $k_i = k+(i-1)$ for $i = 1,\ldots,m-k+1$ and $l_i = l+(i-1)$ for $i = 1,\ldots,n-l+1$. Then, the left-hand side of (28.3) reduces to $v_{l_1} + \sum_{i=1}^{s-1}(v_{l_{i+1}} - v_{l_i+1}) - v_{l_s+1} = v_l - v_{m+1} = v_{l_1}$ while its right-hand side reduces to $u_{k_1} + \sum_{i=1}^{r-1}(u_{k_{i+1}} - u_{k_i+1}) - u_{k_r+1} = u_k - u_{n+1} = u_k$. Therefore this constraint takes the form $v_l \geq u_k$ Projecting out variables x_{m+1}, y_{n+1}, u_{m+1} and v_{n+1} from the formulation is then trivial as it can be verified that for vectors $w, z \in \mathbb{R}^p$, $(w,0) \geq_m (z,0)$ if and only if $w \geq_m z$. □

7 Set of rank-one matrices associated with permutation-invariant sets

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

As a motivating example, sparse PCA is to find a sparse vector that maximizes the variance $x^\intercal \Sigma x$ associated with the component for a given covariance matrix Σ. A semidefinite relaxation of sparse PCA aims to approximate the following set

$$\{(x,X) \in \mathbb{R}^n \times \mathcal{M}_n \mid X = xx^\intercal, \|x\| \leq 1, \text{card}(x) \leq K\} \tag{30}$$

for a positive integer $K \in \{1,\ldots,n-1\}$ which can be represented as M_S for a permutation-invariant set $S = \{x \in \mathbb{R}^n \mid \|x\| \leq 1, \text{card}(x) \leq K\}$. Separation problems associated with the above set are known to be NP-hard and hence their semidefinite relaxations have been considered by relaxing the non-convex constraint $X = xx^\intercal$ with a convex constraint $X \succeq xx^\intercal$. Then, linear valid inequalities in (X,x) are developed by exploring the property that $X = xx^\intercal$. For example, when x is bounded by a box, one can add McCormick constraints to relax the product terms. The authors of [11] proposed a cut $\mathbb{1}^\intercal X \mathbb{1} \leq K$ which can be easily obtained by the valid inequality $\sum_{i=1}^n x_i \leq \sqrt{K}$ and the condition $X = xx^\intercal$.

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

Theorem 7.1. *Suppose* $S \subseteq \mathbb{R}^n$ *is a permutation-invariant set. Let*

$$\mathcal{N} = \bigcup_{P \in \mathcal{P}_n} \left\{ (x,u,X,U) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{M}^n \times \mathcal{M}^n \,\middle|\, \begin{array}{l} X = xx^\intercal, U = uu^\intercal, \\ x = Pu, \\ u \in S \cap \Delta_n \end{array} \right\}.$$

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

Proof. Proof. We first show that $M_S \subseteq \text{proj}_{x,X}\mathcal{N}$. For any arbitrary $(x,X) \in M_S$, there exists $P \in \mathcal{P}_n$ such that $u := P^{-1}x \in \Delta_n$. By permutation-invariance of S, $u \in S$ and hence $u \in S \cap \Delta_n$. This shows that $(x,u,xx^\intercal,uu^\intercal) \in \mathcal{N}$. We next prove the opposite inclusion. Consider a point $(x,u,X,U) \in \mathcal{N}$. Assume $x = Pu$ for some $P \in \mathcal{P}_n$. Since $u \in S \cap \Delta_n \subseteq S$, $x \in S$ by permutation-invariance of S. Moreover, $X = xx^T$. Therefore, $(x,X) \in M_S$. □

By Proposition 7.1, it suffices to assume that $X = xx^\intercal$ for $x \in \mathbb{R}^n$ that is a permutation of an element in $S \cap \Delta_n$ when constructing a valid inequality. For any $(x,u,X,U) \in \mathcal{N}$, observe that

$$X = xx^\intercal = (Pu)(Pu)^\intercal = PUP^\intercal$$

from the relationship $x = Pu$ for a permutation matrix $P \in \mathcal{P}_n$. In other words, we can assume that U is the matrix obtained by permuting columns and the corresponding rows of matrix X. We can derive some valid inequalities based on this idea. Perhaps, the easiest inequalities we can develop are

$$\text{trace}(X) = \text{trace}(U), \tag{31a}$$
$$\mathbb{1}^\intercal X \mathbb{1} = \mathbb{1}^\intercal U \mathbb{1}, \tag{31b}$$

On the other hand, since $u \in \Delta_n$, entries in each row of $uu^\intercal$ is in descending order, deriving the following inequalities.

$$U_{i,j} \geq U_{i,j+1}, \;\; 1 \leq i \leq n, \;\; 1 \leq j < n. \tag{32}$$

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

We next generalize (31a) and (31b) using majorization. Observe that (31a) does not account for the fact that the diagonal entries of U and X are identical up to permutation. Therefore, we consider the following tighter constraint:

$$\text{diag}(U) \geq_m \text{diag}(X). \tag{33}$$

Notice that (31a) is implied by (33) by definition of majorization. Since (33) models the permutahedron with respect to a base vector diag(U), it provides the most compact linear description for the relationship that diag(U) is a permutation of diag(X). We recall from Theorem 2.5 that a majorization inequality of the form $u \geq_m x$ for $u \in \Delta_n$ is modeled using linear inequalities in a higher dimensional space. Similarly, (33) is linearly representable because diag$(U) \in \Delta_n$.

An extension for (31b) is obtainable from the fact that the vector of row sums of X is a permutation of that of U. For any matrix $Y \in \mathcal{M}_{m\times n}$, define R^Y as an m-dimensional vector whose ith component is the sum of ith row of Y. Then, the following majorization inequality is valid.

$$R^U \geq_m R^X. \tag{34}$$

Similarly, (34) is linearly representable because $R^U \in \Delta_n$. By symmetry of U and X, majorization condition for vectors of column sums is redundant.

We next consider a special case where $u \geq 0$. Denote the sum of the k largest elements of x by $s_k(x)$, and the ith row of a matrix Y by Y_i. For a fixed pair $p, q \in \{1, \ldots, n\}$, take the sums of q largest components of each rows of U and X:

$$\begin{bmatrix} s_q(U_1) \\ \vdots \\ s_q(U_n) \end{bmatrix}, \begin{bmatrix} s_q(X_1) \\ \vdots \\ s_q(X_n) \end{bmatrix}.$$

Under the assumption that $X = PUP^\intercal$ for some permutation matrix P, it is clear that one of the above vectors is a permutation of the other. We next consider the sum of p largest components of those two vectors and

$$s_p\left((s_q(U_1), \ldots, s_q(U_n))\right) \geq s_p((s_q(X_1), \ldots, s_q(X_n))). \tag{35}$$

We next argue the linear representability of (35). The left-hand side of (35) is simply written as $\sum_{i=1}^{p}\sum_{j=1}^{q} U_{ij}$. Using the dual arguments that we used in Theorem 2.5, we represent (35) as

$$\begin{array}{ll} \sum_{i=1}^{p}\sum_{j=1}^{q} U_{ij} \geq pr + \sum_{i=1}^{n} t_i & \\ s_q(X_i) \leq t_i + r, & i = 1, \ldots, n \\ t_i \geq 0, & i = 1, \ldots, n \end{array} \tag{36}$$

Observe that the function $s_q(X_i)$ is a convex function and it is placed in the "lower" side of an inequality. Therefore, (36) is a convex representation. By applying the similar procedure to model $s_q(X_i)$ for $i = 1, \ldots, n$, we can obtain a linear representation for (35). It is clear that (35) holds with equality when $p = n$. Unlike linear representation of a majorization inequality, we do not add the constraint

$$s_n\left((s_q(U_1), \ldots, s_q(U_n))\right) = s_n((s_q(X_1), \ldots, s_q(X_n))) \tag{37}$$

because the right-hand side is still written in ordered variables of the form $(X_i)_{[j]}$ and hence it is not a linear equality. Moreover, it is not convex representation because of the equality.

Note that we used two stages of "sum-of-k-largest" modeling techniques to obtain linear representation of (35) in the previous discussion. We next present that a linear representation can be obtained using a transportation problem and its dual in one stage. To this end, we introduce the following lemma:

Lemma 7.1. *For given $w \in \mathbb{R}_+^n$ and $p, q \in \{1, \ldots, n\}$, the optimal value of the following linear program is $(\sum_{i=1}^{p} w_{[i]})(\sum_{j=1}^{q} w_{[j]})$:*

$$\begin{array}{lll} \max & \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j x_{ij} & \\ \text{s.t.} & \sum_{j=1}^{n} x_{ij} \leq q, & i \in \{1, \ldots, n\} \\ & \sum_{i=1}^{n} x_{ij} \leq p, & j \in \{1, \ldots, n\} \\ & \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij} \leq pq & \\ & 0 \leq x_{ij} \leq 1, & i, j \in \{1, \ldots, n\} \end{array} \tag{38}$$

Proof. Proof. Without loss of generality, we assume that $w_i = w_{[i]}$ for all $i = 1, \ldots, n$. Let z^* be the maximum of (38). First, define x' as $x'_{ij} = 1$ if $i \le p$ and $j \le q$ and 0 otherwise. Then, x' is feasible and the objective function value is $(\sum_{i=1}^{p} w_i)(\sum_{j=1}^{q} w_j)$. This shows that $z^* \ge (\sum_{i=1}^{p} w_i)(\sum_{j=1}^{q} w_j)$. Next, we consider the dual formulation of (38) as follows:

$$\begin{aligned}
\min \quad & q\sum_{i=1}^{n}\alpha_i + p\sum_{j=1}^{n}\beta_j + pq\gamma + \sum_{i=1}^{n}\sum_{j=1}^{n}\delta_{ij} \\
\text{s.t.} \quad & \alpha_i + \beta_j + \gamma + \delta_{ij} \ge w_i w_j \qquad i,j \in \{1,\ldots,n\} \qquad (39a)\\
& \alpha_i \ge 0, \beta_j \ge 0, \gamma \ge 0, \delta_{ij} \ge 0 \qquad i,j \in \{1,\ldots,n\} \qquad (39b)
\end{aligned}$$

Since both (38) and (39) are feasible, z^* is the minimum of (39) by strong duality. By convention, we define $w_{n+1} = 0$. Define $(\alpha', \beta', \gamma', \delta') \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \times \mathcal{M}_n$ as follows:

$$\begin{aligned}
\alpha'_i &= \max\{w_i w_{q+1} - w_p w_q, 0\}, \quad i = 1,\ldots,n \\
\beta'_j &= \max\{w_{p+1} w_j - w_p w_q, 0\}, \quad j = 1,\ldots,n \\
\gamma' &= w_p w_q \\
\delta'_{ij} &= \begin{cases} w_i w_j - \gamma' - \alpha'_i - \beta'_j & \text{if } i \le p \text{ and } j \le q \\ 0 & \text{Otherwise} \end{cases}
\end{aligned}$$

We first prove the feasibility of the point. Nonnegativity of α, β, and γ is clear. To prove nonnegativity of δ, consider $i \le p$ and $j \le q$. Assume first that $w_i w_{q+1} - w_p w_q \ge 0$ and $w_{p+1} w_j - w_p w_q \ge 0$. Then,

$$\begin{aligned}
\delta'_{ij} &= w_i w_j - w_p w_q - (w_i w_{q+1} - w_p w_q) - (w_{p+1} w_j - w_p w_q) \\
&= (w_i - w_p)(w_j - w_q) + w_i(w_q - w_{q+1}) + w_j(w_p - w_{p+1}) \ge 0.
\end{aligned}$$

We next assume that either $w_i w_{q+1} - w_p w_q < 0$ or $w_{p+1} w_j - w_p w_q < 0$. Without loss of generality, assume that $w_i w_{q+1} - w_p w_q < 0$. Then, $\delta'_{ij} = w_i w_j - w_p w_q - (w_{p+1} w_j - w_p w_q) = w_j(w_i - w_{p+1}) \ge 0$. We next show that $(\alpha', \beta', \gamma', \delta')$ satisfies (39a). For $i \le p$ and $j \le q$, (39a) holds with equality by definition of δ'_{ij}. We next consider the case where $i > p$ or $j > q$. Without loss of generality, assume that $i > p$ so that $\alpha'_i = 0$. Then,

$$\alpha'_i + \beta'_j + \gamma' + \delta'_{ij} = \beta'_j + \gamma' = \max\{w_{p+1} w_j - w_p w_q, 0\} + w_p w_q \ge w_{p+1} w_j - w_p w_q + w_p w_q \ge w_i w_j.$$

Finally, the objective function value at $(\alpha', \beta', \gamma', \delta')$ is

$$\begin{aligned}
& q\sum_{i=1}^{n}\alpha'_i + p\sum_{j=1}^{n}\beta'_j + pq\gamma' + \sum_{i=1}^{n}\sum_{j=1}^{n}\delta'_{ij} \\
= \;& q\sum_{i=1}^{p}\alpha'_i + p\sum_{j=1}^{q}\beta'_j + pq\gamma' + \sum_{i=1}^{p}\sum_{j=1}^{q}(w_i w_j - \gamma' - \alpha'_i - \beta'_j) \\
= \;& \sum_{i=1}^{p}\sum_{j=1}^{q} w_i w_j = \left(\sum_{i=1}^{p} w_i\right)\left(\sum_{j=1}^{q} w_j\right)
\end{aligned}$$

This shows that $z^* \le (\sum_{i=1}^{p} w_i)\left(\sum_{j=1}^{q} w_j\right)$, concluding that $z^* = (\sum_{i=1}^{p} w_i)\left(\sum_{j=1}^{q} w_j\right)$. □

An alternative linear representation of (35) can be obtained using the dual formulation in the proof of Lemma 7.1. First of all, the left-hand side of (35) is a constant $\sum_{i=1}^{p}\sum_{j=1}^{q} U_{ij}$. Then, we replace its right-hand side with the objective function of the dual (39) and add inequalities constraints (39a) and (39b) in the formulation where the term $w_i w_j$ must be replaced with $|X|_{ij}$. Notice that we cannot replace the right-hand side with the objective function of (38) because (38) is a maximization problem while the dual is applicable because it is a minimization problem.

This idea is extended to model the the largest sum of the p_1-by-p_2-by-…-by-p_k subtensor of a rank-k nonnegative tensor $w \otimes w \otimes \cdots \otimes w$ where $1 \le p, q, r \le n$ for some $w \in \mathbb{R}^n_+$.

In the remainder of the section, we present semidefinite programming relaxations for sparse PCA.

7.1 An SDP relaxation for sparse PCA

Principal Component Analysis is a well-known dimension reduction technique in statistical analysis. A principal component is a linear combination of independent variables and typically it means the coefficient vector of the linear combination. The first principal component is a unit principal component which maximizes the variance and 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 hard to interpret because typically most coefficients are nonzero. Sparse PCA is introduced to resolve this issue by finding linear combinations with only a few explantory variables. That is, it extends this classic PCA by adding a sparsity contraint which allows only a certain number of explanatory variables are used to principal components.

Formally, let Σ be the covariance matrix of the data set and let x be a coefficient vector for a principal component. Then, the following optimization problem find the first sparse principal component that contains at most K nonzero entries:

$$\begin{array}{ll} \max & x^T\Sigma x \\ \text{s.t.} & \|x\| \leq 1, \\ & \text{card}(x) \leq K \end{array} \tag{sparse PCA}$$

where $x \in \mathbb{R}^n$, K is a positive integer with $1 < K < n$, and card(x) is the number of nonzero components of x. Observe that the feasible set of sparse PCA is $N^K_{\|\cdot\|}$ where $\|\cdot\|$ is L^2-norm. sparse PCA is a non-convex optimization problem because the feasible set is nonconvex due to the sparsity constraint. On the other hand, since the objective function is convex, it can be seen as a convex maximization problem over conv$(N^K_{\|\cdot\|})$ provided that an optimal solution to the relaxation can be transformed to an alternate sparse solution. As we have already seen, the convex hull of the feasible set is represented as follows:

$$\text{conv}(N^K_{\|\cdot\|}) = \left\{ x \;\middle|\; \begin{array}{l} \|u\| \leq 1, \\ u_1 \geq \cdots \geq u_K \geq 0, \\ u_{K+1} = \cdots = u_n = 0, \\ u \geq_m |x| \end{array} \right\}. \tag{40}$$

Once a description of conv$(N^K_{\|\cdot\|})$ is obtained, sparse PCA can be reformulated as a convex maximization problem over a compact convex set.

We next present a positive semidefinite relaxation for (sparse PCA). The most commonly used (and, to the best of our knowledge, only) SDP relaxation for sparse PCA was introduce in [11] as follows:

$$\begin{array}{ll} \max & \text{trace}(\Sigma X) \\ \text{s.t.} & \text{trace}(X) \leq 1, \\ & \mathbb{1}^\intercal |X| \mathbb{1} \leq K, \\ & X \succeq 0. \end{array} \tag{41}$$

On the other hand, we develop the following SDP relaxation by adding the majorization constraints that we introduced earlier in this section:

$$\begin{array}{llr}
\max & \text{trace}(\Sigma X) & \\
\text{s.t.} & \text{trace}(U) \leq 1 & \text{(42a)} \\
& \text{trace}(|X|) = \text{trace}(U), & \text{(42b)} \\
& \mathbb{1}^\intercal |X| \mathbb{1} = \mathbb{1}^\intercal U \mathbb{1}, , & \text{(42c)} \\
& U_{i,j} \geq U_{i,j+1} \geq 0, \qquad 1 \leq i \leq n,\ 1 \leq j < n, & \text{(42d)} \\
& \text{diag}(U) \geq_m \text{diag}(|X|) & \text{(42e)} \\
& R^U \geq_m R^{|X|} & \text{(42f)} \\
& \sum_{i=1}^{p} \sum_{j=1}^{q} U_{ij} \geq s_p\left(s_q(|X_1|), \ldots, s_q(|X_n|)\right), \qquad 1 \leq p \leq q \leq n & \text{(42g)} \\
& \|u\| \leq 1 & \text{(42h)} \\
& u_1 \geq \cdots \geq u_K \geq 0 & \text{(42i)} \\
& u_{K+1} = \cdots = u_n = 0 & \text{(42j)} \\
& u \geq_m |x| & \text{(42k)} \\
& U_{ij} = 0 \qquad i > K \text{ or } j > K & \text{(42l)} \\
& X \succeq xx^\intercal & \text{(42m)} \\
& U \succeq uu^\intercal & \text{(42n)} \\
& X = X^\intercal, U = U^\intercal & \text{(42o)}
\end{array}$$

(42a) are from the norm constraints $\|u\| \leq 1$. The constraints that we introduced earlier in this section are in (42b) - (42k). Notice that nonnegativity of entries of U is added in (42d). (42l) is added because the last $n-K$ components of u are zeros and hence any entries outside the K-by-K top-left submatrix are zeros. Lastly, (42m) and (42n) are relaxation of the condition $X = xx^\intercal$ and $U = uu^\intercal$ and these can be easily modeled using Schur complements. In (42e), we use $\text{diag}(X)$ rather than $\text{diag}(|X|)$ because (42m) implies that $\text{diag}(X) \geq 0$. Furthermore, a natural constraint $\text{trace}(X) \leq 1$ is omitted because it is implied by (42a) and (42b). For the sake of exposition, we omit the modeling detail for (42e) - (42g) and (42k) in the formulation.

Theorem 7.2. *All the constraints in* (41) *are implied by constraints in* (42).

Proof. Proof. First, we already showed that $\text{trace}(X) \leq 1$ is implied by (42a) and (42e). Positive semidefiniteness of X is from the fact that $X \succeq xx^\intercal \succeq 0$. We next show that $\mathbb{1}^\intercal|X|\mathbb{1} \leq K$ is implied. By (42l), we only consider the K-by-K upper-left submatrix of U. Let $U_{K,K}$ be the submatrix and let $\mathbb{1}_K$ is the K-dimensional vector of ones. Define $f(x) := x^\intercal U_{K,K} x$. Since $U \succeq 0$ implies that $U_{K,K} \succeq 0$, f is convex. Furthermore, $f(\alpha x) = \alpha^2 f(x)$ for any scalar α. Therefore,

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

where the first inequality follows from the convexity of f. On the other hand, (42f) implies that $\mathbb{1}^\intercal U \mathbb{1} = \mathbb{1}^\intercal|X|\mathbb{1}$. Therefore, $\mathbb{1}^\intercal|X|\mathbb{1} = \mathbb{1}^\intercal U \mathbb{1} \leq K$. ☐

7.2 Computational experiments for sparse PCA

We next report our computational results. We refer to the formulation (42) by the *upper-sum relaxation* and the formulation except the constraints (42g) by the *row-sum relaxation*. We report test results for the row-sum and upper-sum relaxations in Table 1 and 2. z_E^* represents the global optimal value for the sparse PCA and z_D^* represents the optimal value for SDP relaxation (41). We denote the optimal value for the row sum relaxation by z_{rs}^* and that for the upper-sum relaxation by z_{us}^*. We used `SCS` version 2.0.2 [26, 27], a large-scale convex conic solver, to solve SDPs in the experiments. To measure the relative tightness of a relaxation when compared to (41), we calculate "gap closed" as

$$\left(\frac{z_D^* - z_{SDP}^*}{z_D^* - z_E^*}\right) \times 100.$$

where z_{SDP}^* is z_{rs}^* or z_{us}^*.

The output status `Solved/Inaccurate` indicates that `SCS` could not determine the solution within the default numerical tolerance, but returned a solution using a relaxed tolerance.

7.2.1 `pitprops` problem

`pitprops` [16] is one of the most commonly used problems for sparse PCA algorithms. The instance has 13 variables and 180 observations. Table 1 shows the test results for cardinality $K = 3, \ldots, 10$. The numbers in the parantheses represent"`Solved/Inaccurate`" `SCS` output.

K	z_E^*	z_D^*	z_{rs}^*	Gap closed (%)	z_{us}^*	Gap closed (%)
3	2.475	2.522	2.495	57.86	(2.475)	100.00
4	2.937	3.017	2.967	62.83	(2.948)	87.15
5	3.406	3.458	3.407	97.97	(3.406)	100.00
6	3.771	3.814	3.771	100.00	(3.771)	100.00
7	3.996	4.032	3.996	100.00	(3.996)	100.00
8	4.069	4.145	4.073	94.22	(4.072)	95.48
9	4.139	4.206	4.139	100.00	(4.139)	100.00
10	4.173	4.219	(4.177)	91.32	(4.177)	91.41
			Average	88.025	Average	96.76

Table 1: Optimal values and gaps closed for the test problem `pitprops`

Observe that the row-sum (resp. upper-sum) relaxation reduces the gaps of (41) by more than 88% (resp. 96%), returning global optimal solutions for three (resp. five) problems.

7.2.2 Experiments with randomly generated matrices

We next report test results for randomly generated covariance matrices. Random matrices are generated as follows:

1. Choose a random integer $m \in \{1, \dots, n\}$ for the number of nonzero eigenvalues of the matrix by setting $m = \lceil nU \rceil$ where $U \sim \mathcal{U}(0, 1)$.
2. Generate m random vectors $v_i \in \mathbb{R}^n \sim \mathcal{N}(0, I_n), i = 1, \dots, m$ for rank-1 matrices.
3. Generate m positive random eigenvalues $\lambda_i \sim \mathcal{U}(0, 1), i = 1, \dots, m$.
4. Then, construct the desired random covariance matrix as $\Sigma = \sum_{i=1}^{m} \lambda_i v_i v_i^\intercal$.

The tests are performed for problems with size $n \in \{4, \dots, 10\}$ and cardinalities $K \in \{2, \dots, \lfloor n/3 \rfloor\}$. Note that the reported results are based on the test problems with `SCS` outputs status "`Solved`" or "`Solved/Inaccurate`". See Table 2. We observe that our SDP relaxations improve the gaps of the SDP relaxation (41) by more than 90% (on average).

			Average gap closed (%)	
n	K	# Test Problems	z^*_{rs}	z^*_{us}
4	2	100	94.993	95.459
5	2	100	94.184	96.689
6	2	100	91.454	95.163
7	2	50	88.892	93.179
7	3	50	90.285	93.086
8	2	50	88.689	92.481
8	3	20	93.434	95.053
9	2	20	87.928	94.963
9	3	20	78.115	87.835
10	2	20	75.478	85.015
10	3	20	85.036	88.827
10	4	20	77.327	81.311
		Overall Average	90.180	93.559

Table 2: Test results for randomly generated covariance matrices

8 Conclusion

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 and the generating vectors in $\Delta = \{x \in \mathbb{R}^n \mid x_1 \geq \cdots \geq x_n\}$ of the permutahedra lie in a set whose convex hull is obtainable. We then discover a variety of applications for which the results can be used. We present an extended formulation for the convex hull of permutation-invariant norm balls constrained by a cardinality requirement. This result is extended to the sets of matrices that is characterized by their singular values. On the semidefinite programming side, we study 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 problem, we construct tight semidefinite programming relaxation for the sparse principal component analysis and report computational results that show that our relaxation reduces more than 90% of gaps generated by the classical relaxation proposed by [11].

References

[1] A. Argyriou, R. Foygel, and N. Srebro. Sparse prediction with the k-support norm. In *Advances in Neural Information Processing Systems*, pages 1457–1465, 2012.

[2] E. Balas. Disjunctive programming and a hierarchy of relaxations for discrete optimization problems. *SIAM Journal on Algebraic Discrete Methods*, 6(3):466–486, 1985.

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

[4] E. Balas. Logical constraints as cardinality rules: Tight representation. *Journal of Combinatorial Optimization*, 8:115–128, 2004.

[5] E. Balas, A. Bockmayr, N. Pisaruk, and L. Wolsey. On unions and dominants of polytopes. *Mathematical Programming*, 9:223—-239, 2004.

[6] H. H. Bauschke, O. Güler, A. S. Lewis, and H. S. Sendov. Hyperbolic polynomials and convex analysis. *Canadian Journal of Mathematics*, 53(3):470–488, 2001.

[7] A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, volume 2. Siam, 2001.

[8] G. Birkhoff. Tres observaciones sobre el algebra lineal. *Univ. Nac. Tucumán Rev. Ser. A*, 5:147–151, 1946.

[9] P. Camion and J. Maurras. Polytopes à sommets dans l'ensemble $\{0, 1\}^n$. *Cahier du Centre d'Etudes de Recherche Opérationelle*, 24:107–120, 1982.

[10] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky. Convex graph invariants. *SIAM Review*, 54(3):513–541, 2012.

[11] A. d'Aspremont, L. El Ghaoui, M. I. Jordan, and G. R. Lanckriet. A direct formulation for sparse pca using semidefinite programming. *SIAM review*, 49(3):434–448, 2007.

[12] D. Dentcheva and A. Ruszczyński. Convexification of stochastic ordering. *Comptes Rendus de l'Academie Bulgare des Sciences*, 57(4):11, 2004.

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

[14] A. S. G. O. S. H. S. Hauschke H. H., Lewisz. Hyperbolic polynomials and convex analysis. *Research Report CORR 98-29*, 1998.

[15] J.-B. Hiriart-Urruty and H. Y. 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.

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

[17] R. Jeroslow. On defining sets of vertices of the hypercube by linear inequalities. 11:119–124, 1975.

[18] V. Kaibel and K. Pashkovich. Constructing extended formulations from reflection relations. In O. Günlük and G. J. Woeginger, editors, *Integer Programming and Combinatoral 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.

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

[20] C. H. Lim. A note on extended formulations for cardinality-based sparsity. 2017.

[21] C. H. Lim and S. Wright. k-support and ordered weighted sparsity for overlapping groups: Hardness and algorithms. In *Advances in Neural Information Processing Systems*, pages 284–292, 2017.

[22] A. W. Marshall, I. Olkin, and B. Arnold. *Inequalities: theory of majorization and its applications*. Springer Science & Business Media, 2010.

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

[24] J. Maurras. An example of dual polytopes in the unit hypercube. *Annals of Discrete Mathematics*, 1:391–392, 1977.

[25] A. Nemirovski. Introduction to linear optimization isye 6661. *Lecture Notes in Georgia Institute of Technology*, 2012.

[26] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd. Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications*, 169(3):1042–1068, June 2016.

[27] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd. SCS: Splitting conic solver, version 2.0.2. `https://github.com/cvxgrp/scs`, Nov. 2017.

[28] H. Yan and J. N. Hooker. Tight representation of logical constraints as cardinality rules. *Mathematical Programming*, 85:363–377, 1999.