A Geometric Perspective
on Polynomially Solvable Convex Maximization

Shaoning Han[1], Liangju Li[1], and Yongchun Li[2]

[1] Department of Mathematics, National University of Singapore, Singapore
shaoninghan@nus.edu.sg, e1583542@u.nus.edu
[2] School of Data Science, The Chinese University of Hong Kong, Shenzhen, China
yongchunli@cuhk.edu.cn

Abstract. Convex maximization encompasses a broad class of optimization problems and is generally NP-hard, even for low-rank objectives. While polynomial solvability has been established for several important special cases, a broader structural understanding for mixed-integer settings remains limited. In this paper, we study this question from a geometric perspective and identify a structural property of the feasible region underlying tractability, which we term *comonotonicity*. Under comonotonicity and mild additional conditions, we develop a unified enumerative framework showing that fixed-rank convex maximization is polynomially solvable. This viewpoint recovers known results such as convex matroid maximization and also covers mixed-integer applications that previously required separate analyses. For the more structured class of *standard comonotone* sets, we further refine the analysis in a suitably enlarged parameter space and obtain a square-root improvement in the complexity bound. Applications to several PCA-related problems illustrate the broad applicability and effectiveness of the proposed framework.

Keywords: Comonotone sets, convex maximization, polynomial complexity, hyperplane arrangement, sparse PCA, mixed-integer programming

1 Introduction

In this paper, we consider convex maximization problems of the form:

$$\max_{x \in \mathbb{R}^n} f(Ax)$$
$$\text{s.t. } x \in \mathcal{X} \subseteq \mathbb{R}^n, \tag{1}$$

where $A \in \mathbb{R}^{r \times n}$ has full row rank, the function $f : \mathbb{R}^r \to \mathbb{R}$ is convex, and $\mathcal{X} \subseteq \mathbb{R}^n$ is a generic feasible region. We assume that the set of global optimizers of (1) is nonempty. Following the literature (e.g., [27, 29]), the integer r is often referred to as the *rank* of the objective and serve as a measure of its nonlinearity.

Convex maximization (1) has long been a central topic in global optimization. Many fundamental optimization problems can be formulated directly or reformulated equivalently as convex maximization problems [9], including zero-one integer programming [25, 65], bilinear programming [8], and difference-of-convex programming [35,

Chapter 3]. The broad applicability of (1) in turn makes (1) extremely challenging to solve, both theoretically and computationally. Indeed, even under certain special settings, problem (1) remains NP-hard. For example, Pardalos and Vavasis [64] prove that maximizing a convex quadratic function of rank two ($r = 2$) over a polytope is NP-hard. Moreover, Mittal and Schulz [51] show that (1) generally does not admit a constant-factor approximation algorithm unless $P = NP$.

Despite its general intractability, this paper aims to study *when problem* (1) *becomes polynomially solvable*. For any vector $x \in \mathbb{R}^n$, we denote by $\mathrm{supp}(x) = \{i \in [n] : x_i \neq 0\}$ the support of x. We make the following assumption.

Assumption 1 *For any subset $S \subseteq [n]$, the following restriction of problem* (1) *can be solved in polynomial time* T_1:

$$\max_{x \in \mathcal{X}} \{f(Ax) : \mathrm{supp}(x) = S\}. \tag{2}$$

In other words, Assumption 1 means that problem (1) becomes tractable once the support of decision variables is fixed. Note that the fixed-support set $\{x \in \mathcal{X} : \mathrm{supp}(x) = S\}$ is generally nonclosed. Consequently, problem (2) may not be able to attain its optimum. More precisely, Assumption 1 is understood in the following algorithmic sense: there exists a polynomial-time algorithm that either returns an optimal solution when one exists, or certifies that no optimal solution exists.

To illustrate the modeling scope of Assumption 1, we present three representative settings in which it arises naturally. First, the assumption is automatic in binary settings where $\mathcal{X} \subseteq \{0,1\}^n$. Indeed, for any subset $S \subseteq [n]$, fixing $\mathrm{supp}(x) = S$ uniquely determines the vector x by setting $x_i = 1$ for $i \in S$ and $x_i = 0$ for $i \notin S$. Thus in this case, (2) reduces to evaluating the objective $f(Ax)$. Second, Assumption 1 naturally appears in modern data science problems where support selection itself is part of the decision, such as sparse principal component analysis (SPCA) [37, 38] and variable selection for two-sample tests (2ST) [75]. To improve the effectiveness and interpretability of the underlying statistical model, such problems typically impose a $\|x\|_0$-"norm" constraint to promote sparse solutions, where $\|x\|_0$ denotes the number of nonzero entries of x. Once the support is fixed, (2) amounts to selecting a subset of features or samples and often reduces to a more conventional and tractable optimization problem. Third, Assumption 1 is also closely connected to active set or manifold identification in classical nonlinear programming [31, 59]. Suppose the feasible region $\mathcal{X}$ is described by equality constraints together with nonnegativity constraints $x \geq 0$. This standard form is without loss of generality, since general inequality constraints can be converted into it by introducing slack variables. Then in this context, determining the active constraints is equivalent to determining the support of continuous variables x. Once the active constraints at optimality are identified, problem (1) locally reduces to an equality-constrained problem and is often more tractable in principle.

When $\mathcal{X}$ or $\mathrm{conv}(\mathcal{X})$ is a polytope, where $\mathrm{conv}(\mathcal{X})$ represents the convex hull of $\mathcal{X}$, the convex maximization problem (1) always admits an optimal solution which is an extreme point of $\mathcal{X}$. This fact motivates a naive solution strategy: enumerate all extreme points of $\mathcal{X}$, evaluate the objective value $f(Ax)$ at each such point, and then select the best one. However, the brute-force approach can easily struggle when the number of extreme points of $\mathcal{X}$ becomes prohibitively large. This may happen even for structurally

simple feasible sets; for example, $\mathcal{X} = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i \leq k,\, 0 \leq x \leq 1\}$, where $k \in \mathbb{Z}_+$. A more effective approach exploits the rank of the objective and instead enumerates the extreme points of the lower-dimensional image $A\mathcal{X} \triangleq \{Ax : x \in \mathcal{X}\}$. To illustrate the importance of the rank parameter r for the tractability of (1), consider the case $r = 1$. In this simple case, $\text{conv}(A\mathcal{X})$ is just a bounded interval with two endpoints. Accordingly, solving (1) boils down to evaluating f at these two points. When $r = 2$, one can efficiently enumerate the extreme points of the two-dimensional set $A\mathcal{X}$ using parametric linear programming [33, 5, 78]. The parametric linear programming approach does not extend to higher ranks, where the structure of $\mathcal{X}$ must come into play.

When $\mathcal{X}$ encodes the bases of a matroid and r is fixed, Onn [60] proves that (1) is polynomially solvable. The author constructs a zonotope from the edges (one-dimensional faces) of $\text{conv}(\mathcal{X})$ and reduces problem (1) to solving $\mathcal{O}(n^{2r-2})$ linear programs. With matroids as a prototypical example, Onn and Rothblum [62] subsequently extend the zonotope-reduction technique to more general discrete settings. They show that, for any fixed rank r, (1) can be solved in time polynomial in dimension n and the number of edge directions of $\text{conv}(\mathcal{X})$, provided that these directions can be efficiently constructed and that one has access to an oracle for solving linear programs over $\mathcal{X}$. In particular, for an integer linear set $\mathcal{X} = \{x \in \mathbb{Z}^n : Qx = b,\, \ell \leq x \leq u\}$, where Q, b, ℓ and u are compatible integer matrices and vectors, the *Graver basis* of Q forms a set of all edge directions of $\text{conv}(\mathcal{X})$ [13]. Several refinements and generalizations along this line also exist in the literature [14, 17, 18, 63, 69]. We refer readers to the monograph [61] for a comprehensive treatment of this direction. However, when (1) involves continuous variables and $\text{conv}(\mathcal{X})$ is no longer polyhedral, $\text{conv}(\mathcal{X})$ typically has infinitely many edges and extreme points. It is unclear how to extend the above results to the nonpolyhedral setting.

When $\text{conv}(\mathcal{X})$ is nonpolyhedral, some specialized applications of (1) are known in the literature to be polynomially solvable, most notably SPCA and the least trimmed squares problem (LTS) [67, 26]. In both problems, the main challenge is to optimally select a subset of size k from n candidate features or samples [36, 45]. Once the subset is determined, SPCA and LTS reduce to the standard PCA and ordinary least-square estimation problems, respectively. Given an $n \times n$ covariance matrix of rank r, Asteris et al. [4] show that single-component SPCA can be solved by computing $\mathcal{O}(n^r)$ standard PCA problems. More recently, Del Pia [15] demonstrates that general SPCA can be solved in $\mathcal{O}(n^{\min\{d,r\}(r^2+r)})$ oracle time, where d denotes the number of principal components. Furthermore, [15] also establishes the first polynomial-time complexity for disjoint SPCA on fixed rank matrices. For LTS, Hössjer [36] develops an exact $\mathcal{O}(n^2 \log n)$ algorithm for single-feature instances. Mount et al. [55] later develop a topological plane-sweeping algorithm that solves general LTS problems in $\mathcal{O}(n^{r+2})$ oracle time, where r denotes the number of features in this context. To the best of our knowledge, apart from these specialized results, convex maximization problems with nonpolyhedral $\text{conv}(\mathcal{X})$ have not been systematically investigated in literature.

1.1   Contributions and outline

Following the literature, we also treat the rank r as a constant in this paper. Below we summarize the main contributions of the paper and outline the remaining sections. Our contributions are three-fold:

1. We introduce the notion of comonotonicity—a set property that plays a fundamental role in the polynomial solvability of (1)*, and establish basic characterizations and properties of (standard) comonotone sets.*

Roughly speaking, a set $\mathcal{X}$ is called comonotone if for each linear optimization problem over $\mathcal{X}$, the ordering pattern of its optimal solution depends only on that of the objective. When the two orderings always coincide, the set is further called standard comonotone. It turns out that comonotonicity can capture the symmetric structure of a set and brings several well-known order-related set classes under a common umbrella, including in particular matroids [77, 47, 11, 76] and permutation invariant sets [39, 40, 43, 44]. While this phenomenon of order dependence is prevalent in the literature, comonotonicity has not yet been recognized and studied as a set property in its own right. In this paper, we explicitly identify this structure and investigate its implications.

2. We show that maximizing a fixed-rank convex function over comonotone sets can be solved in polynomial oracle time under Assumption 1.

By leveraging the comonotonicity of $\mathcal{X}$, we develop a general theoretical framework for problem (1) that generates $\mathcal{O}(n^{2r})$ supports of decision variables, one of which coincides with the support of an optimal solution. Consequently, under Assumption 1, (1) can be solved by addressing $\mathcal{O}(n^{2r})$ fixed-support subproblems of the form (2). This framework applies to any comonotone feasible set and, in a unified manner, recovers the known polynomial solvability result of convex matroid maximization [60] and nonpolyhedral applications surveyed above.

3. For standard comonotone sets, we further refine the analysis via a lifting approach and obtain a square-root improvement in oracle complexity.

We work in a suitably lifted space of the parameters under consideration and refine the support-generation framework for standard comonotone sets. This reduces the number of candidate supports from $\mathcal{O}(n^{2r})$ to $\mathcal{O}(n^{r+1})$. As representative illustrations, Table 1 summarizes the resulting complexity bounds when our approach is applied to PCA-related instances of (1), where "–" indicates that no prior polynomial solvability result is known in the literature. Moreover, compared with the existing complexity analyses [4, 15], our approach bypasses intricate problem-specific combinatorial constructions and yields a substantially simpler derivation.

Table 1: Oracle complexity for PCA-related application instances of (1)

	Single SPCA	Nonnegative SPCA	SPCA	Disjoint SPCA	2ST
This paper	$\mathcal{O}(n^r)$	$\mathcal{O}(n^r)$	$\mathcal{O}\left(n^{(r^2+r)/2}\right)$	$\mathcal{O}\left(\left(n(d+1)^2\right)^{\frac{d(r^2+r+2)}{2}-1}\right)$	$\mathcal{O}(n^{r+1})$
Literature	$\mathcal{O}(n^r)$	–	$\mathcal{O}\left(n^{\min\{d,r\}(r^2+r)}\right)$	$\mathcal{O}\left(n^{d^2(r^2+r)/2}\right)$	–

The remainder of the paper is organized as follows. In Section 2, we establish the theoretical foundation of comonotone sets. In Section 3, we develop a unified framework for maximizing convex functions over comonotone sets and obtain polynomial complexity results under mild conditions. In Section 4, we derive improved complexity bounds under standard comonotonicity by means of lifting arguments. We also show there how additional sign conditions can be incorporated to sharpen the bounds further. In Section 5, we extend the framework to settings with quasi-convex objectives or feasible regions that are not exactly comonotone but remain sufficiently close to a comonotone set. Section 6 is devoted to applications. Finally, in Section 7, we conclude the paper.

1.2   Notations

For a positive integer n, we let $\mathbb{R}^n$ and $\mathbb{R}^n_+$ denote the set of all n-dimensional vectors and nonnegative vectors, respectively. We let $\mathcal{Z}_+$ and $\mathcal{Z}_{++}$ be the set of all nonnegative integers and positive integers, respectively. Let $[n] = \{1, 2, \cdots, n\}$. We denote by Π_n the set of all permutations over $[n]$. We let e denote the vector of all ones, and let I denote the identity matrix, with their size being clear in context. For a vector $x \in \mathbb{R}^n$, we let $\|x\|_0$ denote the number of nonzero entries of x, and let $|x| = (|x_1|, \cdots, |x_n|)^\top \in \mathbb{R}^n$ contain the absolute entries of x. For a matrix X and a positive integer d, we let $\|X\|_F$ denote its Frobenius norm, let $\|X\|_0$ denote the number of nonzero rows of X, and let $\|X\|_{(d)}$ denote the sum of its d largest eigenvalues. For a symmetric matrix X, we let $\lambda_{\max}(X)$ denote its largest eigenvalue, and given an index set $S \subseteq [n]$, let $X_{S,S}$ denote a principal submatrix of X indexed by S. For a set $\mathcal{D} \subseteq \mathbb{R}^n$, we denote by $\text{conv}(\mathcal{D})$ its convex hull, by $\text{int}\,\mathcal{D}$ its interior, and by $\dim(\mathcal{D})$ its dimension. Given a hyperplane $H = \{x \in \mathbb{R}^n : a^\top x = b\}$, we denote its two open half-spaces by

$$H^> = \{x \in \mathbb{R}^n : a^\top x > b\}, \text{ and } H^< = \{x \in \mathbb{R}^n : a^\top x < b\}.$$

For a permutation $\pi \in \Pi_n$ and a vector $x \in \mathbb{R}^n$, we say that x is *sorted* by π if $x_{\pi(1)} \geq \cdots \geq x_{\pi(n)}$ holds. Note that the permutation that sorts a given vector x is not unique when there are ties. For example, the vector $x = [1, 0, 1]^\top$ can be sorted by $\pi = (1, 3, 2)$ or $(3, 1, 2)$. We also denote by $x(\pi)$ the vector obtained by permuting the entries of x according to π, that is, $[x(\pi)]_1 \geq [x(\pi)]_2 \geq \cdots \geq [x(\pi)]_n$. Given a permutation π, we define a cone $\mathcal{Z}(\pi) = \{x \in \mathbb{R}^n : x_{\pi(1)} \geq \cdots \geq x_{\pi(n)}\}$ and its nonnegative counterpart $\mathcal{Z}^+(\pi) = \mathcal{Z}(\pi) \cap \mathbb{R}^n_+$. For example, when $n = 2$ and $\pi = (2, 1)$, one has $\mathcal{Z}(\pi) = \{x \in \mathbb{R}^2 : x_2 \geq x_1\}$. Naturally we have $x(\pi) \in \mathcal{Z}(\pi)$ for all $x \in \mathbb{R}^n$ and $\pi \in \Pi_n$.

2   Comonotone sets

In this section, we first formally introduce the notion of comonotonicity. We then establish structural properties and basic characterizations of comonotone sets, with particular emphasis on *standard comonotonicity*. We also present a range of examples to illustrate the definition and to anchor the geometric intuition behind the results. These structural results and examples serve as the geometric cornerstone of our complexity analysis.

We now state the definition of (standard) comonotone sets, which captures the geometric intuition that optimal solutions to a linear program should inherit the sorting pattern of the cost vector. This alignment is formalized through a permutation mapping that specifies the induced ordering by the objective.

Definition 1 (Comonotone sets). *A set $\mathcal{X} \subseteq \mathbb{R}^n$ is called* comonotone *if there exists a permutation mapping $\Psi : \Pi_n \to \Pi_n$ such that for every permutation $\pi \in \Pi_n$ and every cost vector $v \in \mathcal{Z}(\pi)$, whenever the problem*

$$\max_{x \in \mathcal{X}} v^\top x,$$

attains its optimum, it admits an optimal solution x^* sorted by $\Psi(\pi)$, that is,

$$x^* \in \mathcal{Z}\left(\Psi(\pi)\right).$$

If one can choose Ψ to be the identity mapping, namely $\Psi(\pi) = \pi$ for all $\pi \in \Pi_n$, then $\mathcal{X}$ is called standard comonotone.

Below provides a simple example to illustrate the concept.

Example 1. The set $\mathcal{X} = \{x \in \{0,1\}^n : e^\top x = k\}$ with $k \in [n]$ is standard comonotone. Indeed, if $v \in \mathcal{Z}(\pi)$, an optimal solution to $\max\{v^\top x : \ x \in \mathcal{X}\}$ is obtained by setting $x_{\pi(1)} = \cdots = x_{\pi(k)} = 1$ and $x_{\pi(k+1)} = \cdots = x_{\pi(n)} = 0$, which belongs to $\mathcal{Z}(\pi)$. Thus, one can take $\Psi(\pi) = \pi \ \forall \pi \in \Pi_n$. $\qquad\qquad\Box$

While comonotonicity is defined through an admissible permutation mapping Ψ, the mapping is best viewed as a *witness* of an underlying order compatibility built into $\mathcal{X}$, rather than a genuine source of freedom. Theorem 1 below highlights the resulting combinatorial rigidity: when Ψ is surjective, so that no ordering type is missing or collapsed, there is essentially only one substantive realization, that is, $\mathcal{X}$ must be a standard comonotone set. We first recall the rearrangement inequality to be used in the proof of Theorem 1.

Lemma 1 (Rearrangement inequality [30, p. 261]). *For all $x, v \in \mathbb{R}^n$, it holds*

$$[v(\pi)]^\top x(\pi) \geq v^\top x \quad \forall \pi \in \Pi_n.$$

For a permutation mapping $\Psi : \Pi_n \to \Pi_n$ and $k \in \mathbb{Z}_+$, let Ψ^k denote the k-fold composition of Ψ (i.e., $\Psi^k = \underbrace{\Psi \circ \cdots \circ \Psi}_{k \text{ times}}$), where Ψ^0 is understood as the identity mapping by convention.

Theorem 1. *Assume $\mathcal{X} \subseteq \mathbb{R}^n$ is a compact comonotone set under permutation mapping $\Psi : \Pi_n \to \Pi_n$. If Ψ is surjective, then $\mathcal{X}$ is standard comonotone.*

Proof. Note that for a comonotone set, the choice of permutation mapping may not be unique in Definition 1. Although Ψ itself is not necessarily identity, our goal is to show that the identity mapping can be used instead in Definition 1. This is equivalent to prove that for any $\pi \in \Pi_n$ and $v \in \mathcal{Z}(\pi)$, one can find a solution $x^* \in \mathcal{Z}(\pi) \cap \operatorname{argmax}\{v^\top x : \ x \in \mathcal{X}\}$.

Consider an arbitrary but fixed $\pi \in \Pi_n$ and $v \in \mathcal{Z}(\pi)$. We first show that there exists $\widehat{k} \in \mathbb{Z}_+$ such that $\Psi^{\widehat{k}}(\pi) = \pi$. Since Π_n is finite, the iterates

$$\pi, \ \Psi(\pi), \ \Psi^2(\pi), \ \ldots$$

can take only finitely many values, so there exist indices $0 \le a < b$ with $\Psi^a(\pi) = \Psi^b(\pi)$. Because $\Psi : \Pi_n \to \Pi_n$ is surjective and Π_n is finite, Ψ is bijective and hence injective. Applying injectivity repeatedly to the equality $\Psi^a(\pi) = \Psi^b(\pi)$ yields $\pi = \Psi^{b-a}(\pi)$. Setting $\widehat{k} \triangleq b - a$ gives $\Psi^{\widehat{k}}(\pi) = \pi$, as claimed.

Next, we propagate optimality through the iterates of Ψ and carry it back to $\mathcal{Z}(\pi)$. Specifically, we define a sequence of permuted cost vectors by $v^k = v(\Psi^{k-1}(\pi))$ for $k \in \mathbb{Z}_{++}$. It follows that $v^1 = v$ and by construction $v^k \in \mathcal{Z}(\Psi^{k-1}(\pi))$ for all k. Moreover, since $\mathcal{X}$ is comonotone under Ψ, for each $k \in \mathbb{Z}_{++}$, one can choose an optimal solution x^k to $\max_{x \in \mathcal{X}} (v^k)^\top x$ such that x^k is sorted by $\Psi(\Psi^{k-1}(\pi)) = \Psi^k(\pi)$. The optimality of x^k implies that

$$\left\langle v^k, x^k \right\rangle \ge \left\langle v^k, x^{k-1} \right\rangle \quad \forall k. \tag{3}$$

Furthermore, since v^k is a permutation of v^{k-1}, and v^k and x^{k-1} are sorted by the same permutation $\Psi^{k-1}(\pi)$, one can deduce from Lemma 1 that

$$\left\langle v^k, x^{k-1} \right\rangle \ge \left\langle v^{k-1}, x^{k-1} \right\rangle \quad \forall k. \tag{4}$$

Chaining inequalities (3) and (4) and combining $\Psi^{\widehat{k}}(\pi) = \pi$, one obtains

$$\left\langle v^1, x^1 \right\rangle \le \left\langle v^2, x^1 \right\rangle \le \left\langle v^2, x^2 \right\rangle \le \cdots \le \left\langle v^{\widehat{k}+1}, x^{\widehat{k}} \right\rangle = \left\langle v\left(\Psi^{\widehat{k}}(\pi)\right), x^{\widehat{k}} \right\rangle = \left\langle v(\pi), x^{\widehat{k}} \right\rangle.$$

Because $v^1 = v$ and $v_\pi = v$ (due to $v \in \mathcal{Z}(\pi)$), the chain reduces to $\left\langle v, x^1 \right\rangle \le \left\langle v, x^{\widehat{k}} \right\rangle$. Moreover, since x^1 is optimal for the linear coefficients v, this inequality implies that $x^{\widehat{k}}$ is also an optimal solution to $\max\{v^\top x : \ x \in \mathcal{X}\}$. Finally, by construction $x^{\widehat{k}}$ is sorted by $\Psi^{\widehat{k}}(\pi) = \pi$, i.e., $x^{\widehat{k}} \in \mathcal{Z}(\pi)$. Hence we find the required $x^* = x^{\widehat{k}}$, which completes the proof. □

To illustrate Theorem 1, we specialize to the planar case $n = 2$. Since $\Pi_2 = \{(1,2),(2,1)\}$ consists of only two permutations, if Ψ is not surjective, then the compact comonotone set $\mathcal{X}$ must lie entirely in one ordering cone $\mathcal{Z}(1,2)$ or $\mathcal{Z}(2,1)$. In other words, comonotonicity is always standard in $\mathbb{R}^2$ unless the set $\mathcal{X}$ forces a fixed ordering; see Figure 1 for illustration.

2.1   Properties of standard comonotone sets

In this subsection, we study properties of standard comonotone sets. We begin with the planar case to build geometric intuition. While restricted to $n = 2$, it provides a clear lens for understanding the restrictions imposed by standard comonotonicity, or equivalently, the structural information it carries. As shown below, standard comonotonicity in $\mathbb{R}^2$ can be characterized using only two directions, akin to checking extreme rays in polyhedral theory.
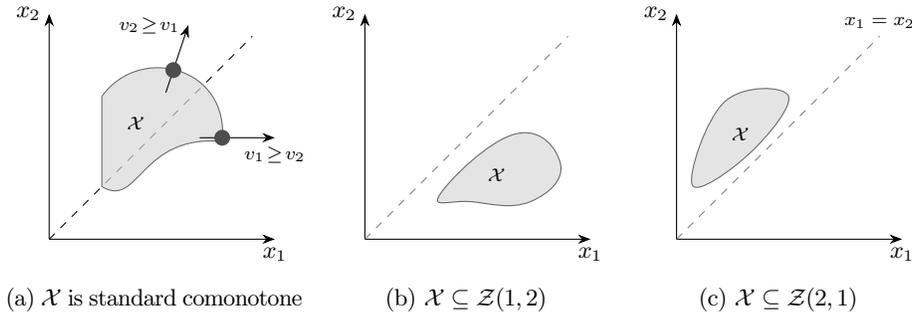
(a) $\mathcal{X}$ is standard comonotone   (b) $\mathcal{X} \subseteq \mathcal{Z}(1, 2)$ (c) $\mathcal{X} \subseteq \mathcal{Z}(2, 1)$

Fig. 1: 2D comonotone sets

Proposition 1. *A compact set $\mathcal{X} \subseteq \mathbb{R}^2$ is standard comonotone if and only if, for all $v \in \{e, -e\}$ and $\pi \in \Pi_2 = \{(1, 2), (2, 1)\}$, the problem $\max\{v^\top x : x \in \mathcal{X}\}$ admits an optimal solution $\bar{x} \in \mathcal{Z}(\pi)$.*

Proof. The necessity follows directly from Definition 1 due to $v \in \mathcal{Z}(1, 2) \cap \mathcal{Z}(2, 1)$. To prove sufficiency, we may assume without loss of generality that $v_1 > v_2$ (the case $v_1 < v_2$ is symmetric after swapping the two coordinates). It remains to prove that there exists an optimal solution $\bar{x} \in \operatorname{argmax}\{v^\top x : x \in \mathcal{X}\}$ such that $\bar{x}_1 \geq \bar{x}_2$.

Take any $\tilde{x} \in \operatorname{argmax}\{v^\top x : x \in \mathcal{X}\}$. Then if $\tilde{x}_1 \geq \tilde{x}_2$, we are done. Otherwise, assume $\tilde{x}_1 < \tilde{x}_2$ and let $\bar{v} = \frac{v_1 + v_2}{2} e$. By our assumption, $\bar{v} \in \mathcal{Z}(1, 2)$ implies there exists $\bar{x} \in \operatorname{argmax}\{\bar{v}^\top x : x \in \mathcal{X}\}$ with $\bar{x}_1 \geq \bar{x}_2$. Next, we show the chain of inequalities

$$v^\top \tilde{x} \leq \bar{v}^\top \tilde{x} \leq \bar{v}^\top \bar{x} \leq v^\top \bar{x}. \tag{5}$$

Indeed, the first inequality follows from

$$v^\top \tilde{x} - \bar{v}^\top \tilde{x} = \frac{1}{2}(v_1 - v_2)(\tilde{x}_1 - \tilde{x}_2) \leq 0,$$

the second from $\bar{x}$ is optimal for the linear objective $\bar{v}$, and the third from

$$\bar{v}^\top \bar{x} - v^\top \bar{x} = -\frac{1}{2}(v_1 - v_2)(\bar{x}_1 - \bar{x}_2) \leq 0.$$

Because $\tilde{x}$ is maximal for the linear objective v, (5) implies that $\bar{x}$ is also an optimal solution for v. The conclusion follows from $\bar{x}_1 \geq \bar{x}_2$.   □

We next give an example illustrating Proposition 1.

Example 2. Consider the sets shown in Figure 2. For the linear objective $x_1 + x_2$, the maximizers are C and D, where $C \in \mathcal{Z}(1, 2)$ and $D \in \mathcal{Z}(2, 1)$; the minimizers are A and B, where $A \in \mathcal{Z}(1, 2)$, and $B \in \mathcal{Z}(2, 1)$. By Proposition 1, all the three sets are standard comonotone. In fact, any set $\mathcal{X}$ satisfying $\{A, B, C, D\} \subseteq \mathcal{X}$ and contained in the strip delimited by the two dotted lines must be standard comonotone.   □
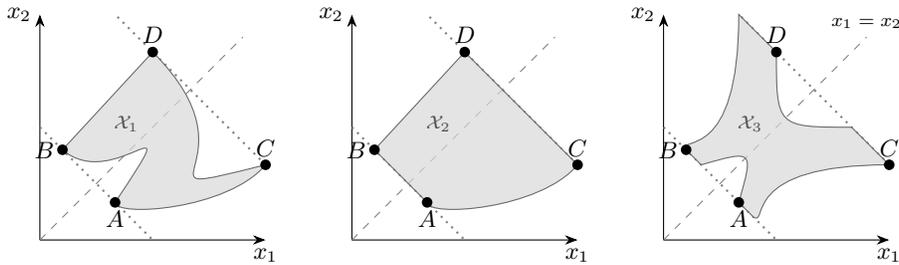
Fig. 2: Characterization of standard 2D comonotone sets

In $\mathbb{R}^2$, ordering information is encoded by only two directions ($x_1 + x_2$ or $-x_1 - x_2$) as indicated by Proposition 1. Unfortunately, the simplification does not directly extend to high dimensions. To see this, consider the distributed lattice $\mathcal{X} = \{x \in \{0,1\}^3 : x_1 \geq x_2 \geq x_3\}$. On the one hand, $0 \in \operatorname{argmax}\{-e^\top x : x \in \mathcal{X}\}$ and $e \in \operatorname{argmax}\{e^\top x : x \in \mathcal{X}\}$, and moreover $\{0, e\} \subseteq \mathcal{Z}(\pi)$ for all $\pi \in \varPi_n$. On the other hand, for $v = (-2, 3, -1)$, the corresponding linear maximization problem has the unique solution $x = (1, 1, 0)$, whose coordinate order is not compatible with that of v.

The geometric intuition from the planar case facilitates the study of standard comonotonicity in higher dimensions. We first state a technical lemma for later use.

Lemma 2. *Let $\mathcal{X} \subseteq \mathbb{R}^n$ be compact. For any $\pi \in \varPi_n$ and $v \in \mathcal{Z}(\pi)$, there exists a sequence $\{v^k\} \subseteq \operatorname{int} \mathcal{Z}(\pi)$ such that $\lim_{k \to \infty} v^k = v$ and for every k, the problem $\max\left\{(v^k)^\top x : x \in \mathcal{X}\right\}$ admits a unique optimal solution.*

Proof. Let $\delta^*(u|\mathcal{X}) \triangleq \max\left\{u^\top x : x \in \mathcal{X}\right\}$ be the *support function* of $\mathcal{X}$, and $\mathcal{D}$ be the set of points where $\delta^*(\cdot|\mathcal{X})$ is differentiable. By [66, Theorem 25.1], $\delta^*(\cdot|\mathcal{X})$ is differentiable at u if and only if the subdifferential $\partial\delta^*(u|\mathcal{X})$ is a singleton. Moreover, [66, Corollary 23.5.3] yields

$$\partial\delta^*(u|\mathcal{X}) = \operatorname{cl}\operatorname{conv}\left(\operatorname{argmax}\left\{u^\top x : x \in \mathcal{X}\right\}\right).$$

Because a set is a singleton if and only if its closed convex hull is a singleton, it follows that that $\mathcal{D}$ is exactly the set of points u where $\{\max u^\top x : x \in \mathcal{X}\}$ admits a unique optimal solution.

Since $\mathcal{X}$ is compact, $\delta^*(\cdot|\mathcal{X})$ is finite-valued and convex on $\mathbb{R}^n$. Hence, by [66, Theorem 25.5], the function $\delta^*(u|\mathcal{X})$ is differentiable almost everywhere, that is, the complement of $\mathcal{D}$ is a set of measure zero. In particular, $\operatorname{int} \mathcal{Z}(\pi) \setminus \mathcal{D}$ has also measure zero. This implies $\operatorname{cl}\left(\operatorname{int} \mathcal{Z}(\pi) \cap \mathcal{D}\right) = \operatorname{cl}\left(\operatorname{int} \mathcal{Z}(\pi)\right) = \mathcal{Z}(\pi)$. Consequently, for any $v \in \mathcal{Z}(\pi)$, there exists a sequence $\{v^k\} \subseteq \operatorname{int} \mathcal{Z}(\pi) \cap \mathcal{D}$ such that $\lim_{k \to \infty} v^k = v$. The conclusion follows. □

Theorem 2 below characterizes standard comonotonicity in $\mathbb{R}^n$ and generalizes Proposition 1.

Theorem 2. *Given a compact set $\mathcal{X} \subseteq \mathbb{R}^n$, the following statements are equivalent:*

(i) The set $\mathcal{X}$ is standard comonotone.

(ii) For all $v \in \mathbb{R}^n$ and all $i, j \in [n]$, there exists a certain $x^* \in \operatorname{argmax}\{v^\top x : x \in \mathcal{X}\}$ such that

$$(v_i - v_j)(x_i^* - x_j^*) \geq 0. \tag{6}$$

(iii) For all $v \in \mathbb{R}^n$, relation (6) holds for all $x^* \in \operatorname{argmax}\{v^\top x : x \in \mathcal{X}\}$ and all $i, j \in [n]$.

(iv) For all $v \in \mathbb{R}^n$ and all $i, j \in [n]$ with $v_i = v_j$, there exist $\bar{x}, \tilde{x} \in \operatorname{argmax}\{v^\top x : x \in \mathcal{X}\}$ such that $\bar{x}_i \geq \bar{x}_j$ and $\tilde{x}_i \leq \tilde{x}_j$.

Proof. We prove the equivalence following the cycle $(i) \Rightarrow (iv) \Rightarrow (iii) \Rightarrow (ii) \Rightarrow (i)$.

$(i) \Rightarrow (iv)$: Since $v_i = v_j$, there exist two permutations $\bar{\pi}$ and $\tilde{\pi}$ such that $v \in \mathcal{Z}(\bar{\pi}) \cap \mathcal{Z}(\tilde{\pi})$, where the index i precedes j in $\bar{\pi}$, and j precedes i in $\tilde{\pi}$. From the definition of standard comonotonicity, there exist optimal solutions $\bar{x} \in \mathcal{Z}(\bar{\pi})$ and $\tilde{x} \in \mathcal{Z}(\tilde{\pi})$ to the problem $\max_{x \in \mathcal{X}} v^\top x$. It follows that $\bar{x}_i \geq \bar{x}_j$ and $\tilde{x}_i \leq \tilde{x}_j$.

$(iv) \Rightarrow (iii)$: For any $v \in \mathbb{R}^n$, let $x^* \in \operatorname{argmax}\{v^\top x : x \in \mathcal{X}\}$ and take arbitrary but fixed $i, j \in [n]$. Because relation (6) holds trivially when $v_i = v_j$, we assume WLOG that $v_i > v_j$. We then construct a new objective vector $\bar{v} \in \mathbb{R}^n$ by

$$\bar{v}_k = \begin{cases} \dfrac{v_i + v_j}{2}, & \text{if } k \in \{i, j\}, \\ v_k, & \text{otherwise.} \end{cases}$$

Since $\bar{v}_i = \bar{v}_j$, condition (iv) implies that there exists a $\bar{x} \in \operatorname{argmax}\{\bar{v}^\top x : x \in \mathcal{X}\}$ such that $\bar{x}_i \geq \bar{x}_j$.

By optimality, we have $v^\top x^* \geq v^\top \bar{x}$ and $\bar{v}^\top \bar{x} \geq \bar{v}^\top x^*$. Summing these two inequalities gives $(v - \bar{v})^\top x^* \geq (v - \bar{v})^\top \bar{x}$, which rearranges to

$$\begin{aligned}
0 \leq &(v - \bar{v})^\top (x^* - \bar{x}) \\
= &\frac{v_i - v_j}{2}(x_i^* - \bar{x}_i) - \frac{v_i - v_j}{2}(x_j^* - \bar{x}_j) \\
= &\frac{1}{2}(v_i - v_j)\left[(x_i^* - x_j^*) - (\bar{x}_i - \bar{x}_j)\right].
\end{aligned}$$

Since $v_i > v_j$, we further obtain $(x_i^* - x_j^*) - (\bar{x}_i - \bar{x}_j) \geq 0$. Together with $\bar{x}_i - \bar{x}_j \geq 0$, we have that $x_i^* - x_j^* \geq 0$. Consequently, $(v_i - v_j)(x_i^* - x_j^*) \geq 0$, which proves (6).

$(iii) \Rightarrow (ii)$: This is immediate.

$(ii) \Rightarrow (i)$: For any fixed $\pi \in \Pi_n$ and $v \in \mathcal{Z}(\pi)$, we take the sequence $\{v^k\}$ in Lemma 2. Then for any v^k, $v^k \in \operatorname{int} \mathcal{Z}(\pi)$ implies $v_{\pi(i)}^k - v_{\pi(j)}^k > 0$ for all $i < j$. Moreover, since the problem $\max\{(v^k)^\top x : x \in \mathcal{X}\}$ has a unique maximizer, which we denote by x^k, one can deduce from (ii) that $x_{\pi(i)}^k \geq x_{\pi(j)}^k$ for all $i < j$. Namely, one has $x^k \in \mathcal{Z}(\pi)$ for all k. Because $\mathcal{X}$ is a compact set, the sequence $\{x^k\}$ has a clustering point. WLOG, we assume $\lim_{k \to \infty} x^k = x^*$; otherwise we apply the arguments to its convergent subsequent. Because $\mathcal{Z}(\pi)$ is closed, it follows that $x^* \in \mathcal{Z}(\pi)$. Moreover, from the optimality of each x^k we have

$$(v^k)^\top x^k \geq (v^k)^\top x, \qquad \forall x \in \mathcal{X}.$$

Taking the limit $k \to \infty$ on both sides leads to

$$v^\top x^* \geq v^\top x, \qquad \forall x \in \mathcal{X},$$

implying $x^* \in \operatorname{argmax}\{v^\top x : x \in \mathcal{X}\}$. This proves (i) and completes the proof. $\qquad\square$

Theorem 2 decomposes global comonotonicity into local pairwise ordering interactions. The complementary roles of parts (ii) and (iv) are worth noting. Since (6) holds trivially when $v_i = v_j$, part (ii) captures only the strict comparison $v_i > v_j$, ensuring that the directional preference in v is faithfully transmitted to the primal solution x. In contrast, part (iv) does *not* impose any strict ordering condition over v. It dictates that if i and j are equivalent in the dual space of cost vectors $(v_i = v_j)$, then they must remain structurally indistinguishable in the primal space of x as well. This requirement embodies a *symmetry via conjugacy*, which becomes even more transparent for convex standard comonotone sets as we discuss next.

Since standard comonotonicity is defined through linear maximization, it is natural to ask whether it is preserved under convexification. Proposition 2 answers this question in the affirmative.

Proposition 2. *Let $\mathcal{X} \subseteq \mathbb{R}^n$ be compact. Then $\mathcal{X}$ is standard comonotone if and only if $\operatorname{conv}(\mathcal{X})$ is standard comonotone.*

Proof. Assume first that $\mathcal{X}$ is standard comonotone. Because

$$\emptyset \neq \operatorname{argmax}\left\{v^\top x : x \in \mathcal{X}\right\} \subseteq \operatorname{argmax}\left\{v^\top x : x \in \operatorname{conv}(\mathcal{X})\right\}, \qquad \forall v \in \mathbb{R}^n, \quad (7)$$

it follows immediately from Definition 1 that $\operatorname{conv}(\mathcal{X})$ is standard comonotone.

Conversely, suppose $\operatorname{conv}(\mathcal{X})$ is standard comonotone. For any given $\pi \in \Pi_n$ and $v \in \mathcal{Z}(\pi)$, let $\{v^k\}$ be the sequence in Lemma 2 converging to v. Because each cost vector v^k admits a unique maximizer x^k over $\operatorname{conv}(\mathcal{X})$, such a solution x^k must also be optimal over $\mathcal{X}$ by (7). Applying the same limiting arguments as in the proof of Theorem 2, one can obtain a solution $x^* \in \mathcal{Z}(\pi) \cap \operatorname{argmax}\{v^\top x : x \in \mathcal{X}\}$. In turn, we deduce from (7) that $x^* \in \mathcal{Z}(\pi) \cap \operatorname{argmax}\{v^\top x : x \in \operatorname{conv}(\mathcal{X})\}$. This proves the standard comonotonicity of $\operatorname{conv}(\mathcal{X})$. $\qquad\square$

Figure 2 illustrates Proposition 2, where $\mathcal{X}_2 = \operatorname{conv}(\mathcal{X}_1)$ and both sets are standard comonotone. Building on the characterization in Theorem 2, we now specialize to the convex setting.

Theorem 3. *Given a compact convex set $\mathcal{X}$, the following statements are equivalent.*

(i) *The set $\mathcal{X}$ is standard comonotone.*
(ii) *For all $v \in \mathbb{R}^n$ and all $i, j \in [n]$ with $v_i = v_j$, there exists a certain $x^* \in \operatorname{argmax}\{v^\top x : x \in \mathcal{X}\}$ such that $x_i^* = x_j^*$.*
(iii) *For all $v \in \mathbb{R}^n$, there exists a certain $x^* \in \operatorname{argmax}\{v^\top x : x \in \mathcal{X}\}$ such that $v_i = v_j$ implies $x_i^* = x_j^*$.*

Proof. The implication $(iii) \Rightarrow (ii)$ is trivial. The direction $(ii) \Rightarrow (i)$ follows directly from Theorem 2. To prove the equivalence, it remains to prove $(i) \Rightarrow (iii)$.

Suppose $\mathcal{X}$ is standard comonotone. For any $v \in \mathbb{R}^n$, define $p(v)$ as the number of distinct pairs (v_i, v_j) of entries of v, namely,

$$p(v) \triangleq \sum_{i=1}^{n} \sum_{j:j>i} \mathbb{I}\{v_i \neq v_j\},$$

where $\mathbb{I}\{v_i \neq v_j\} = 1$ if $v_i \neq v_j$ and 0 otherwise. We prove (iii) by induction on $p(v)$. In the base case where $p(v) = n(n-1)/2$, one has $v_i \neq v_j$ for all $i \neq j \in [n]$ and thus, (iii) holds trivially.

Fix $q \in \mathbb{Z}_{++}$ and assume (iii) holds for all $v \in \mathbb{R}^n$ satisfying $n(n-1)/2 \geq p(v) \geq q$. We aim to show that (iii) also holds for every v with $p(v) = q-1$. If no such v exists, then the claim follows. Otherwise, we consider any fixed v with $p(v) = q-1$, and choose $\alpha \subseteq [n]$ such that $v_i = v_j$ for all distinct $i, j \in \alpha$ while $v_i \neq v_j$ for all $i \in \alpha$ and $j \notin \alpha$.

For each $i \in \alpha$, define a sequence $v^{i,k} = v + \frac{1}{k}e^i$ for $k \in \mathbb{Z}_{++}$. Then it follows that $p(v^{i,k}) \geq q$ for all sufficiently large $k \in \mathbb{Z}_{++}$. By the induction hypothesis, for each $i \in \alpha$ and sufficiently large k, there exists $x^{i,k} \in \operatorname{argmax}\{(v^{i,k})^\top x : x \in \mathcal{X}\}$ such that (iii) holds true for the pair $(v^{i,k}, x^{i,k})$. Particularly, since $v_{i'}^{i,k} = v_{i'}$ for all $i' \in [n]\backslash\{i\}$, we have for sufficiently large k that

$$v_{i'} = v_{j'} \implies x_{i'}^{i,k} = x_{j'}^{i,k} \qquad \forall\, i \in \alpha,\ \forall\, i', j' \in [n] \setminus \{i\}.$$

Moreover, since $\mathcal{X}$ is compact and $\lim_{k\to\infty} v^{i,k} = v$ for all $i \in \alpha$, we may assume WLOG (passing to a convergent subsequence if necessary) that

$$\lim_{k\to\infty} x^{i,k} = \bar{x}^i, \qquad \forall i \in \alpha.$$

Then by construction, the limits $\{\bar{x}^i\}_{i\in\alpha}$ preserve the following ordering properties:

(a) $\bar{x}^i \in \operatorname{argmax}\left\{v^\top x : x \in \mathcal{X}\right\}$ for all $i \in \alpha$.
(b) $v_{i'} = v_{j'}$ implies $\bar{x}_{i'}^i = \bar{x}_{j'}^i$, for all $i \in \alpha$ and $i', j' \in [n]\backslash\{i\}$.
(c) $\bar{x}_i^i \geq \bar{x}_{i'}^i$ for all $i \in \alpha$ and $i' \in \alpha\backslash\{i\}$. Indeed, comonotonicity of $\mathcal{X}$ and $v_i^{i,k} > v_{i'}^{i,k}$ imply $\bar{x}_i^{i,k} \geq \bar{x}_{i'}^{i,k}$ by Theorem 2 for each k. Then it follows by letting $k \to \infty$.

Finally, we construct an optimal solution x^* for v from $\{\bar{x}^i\}_{i\in\alpha}$ and verify that it satisfies (iii). By property (b)-(c), let us first rewrite each $\bar{x}^i = \tilde{x}^i + t_i e^i$, where $t_i = \bar{x}_i^i - \bar{x}_{i'}^i \geq 0$ for a fixed $i' \in \alpha\backslash\{i\}$. Also, by construction, each $\tilde{x}^i$ satisfies the equalities:

$$x_{i'} = x_{j'}, \qquad \forall i', j' \in [n] \text{ s.t. } v_{i'} = v_{j'}. \tag{8}$$

If there exists $i \in \alpha$ such that $t_i = 0$, then $\bar{x}^i = \tilde{x}^i$ satisfies (iii) by property (a) and (8). This completes the inductive step. Hence, we may assume $t_i > 0$ for all $i \in \alpha$ and define

$$\lambda_i = \frac{\prod\limits_{j\in\alpha\backslash\{i\}} t_j}{\sum\limits_{i'\in\alpha} \prod\limits_{j\in\alpha\backslash\{i'\}} t_j}, \qquad \forall i \in \alpha,$$

and set

$$x^* = \sum_{i \in \alpha} \lambda_i \bar{x}^i = \sum_{i \in \alpha} \lambda_i \tilde{x}^i + \frac{\prod\limits_{j \in \alpha} t_j}{\sum\limits_{i' \in \alpha} \prod\limits_{j \in \alpha \setminus \{i'\}} t_j} \left(\sum_{i \in \alpha} e^i \right).$$

Since $\mathcal{X}$ is convex, the set $\mathrm{argmax}\{v^\top x : x \in \mathcal{X}\}$ is also convex. Thus by property (a), the convex combination above yields $x^* \in \mathrm{argmax}\{v^\top x : x \in \mathcal{X}\}$. In addition, each $\tilde{x}^i$ and the vector $\sum_{i \in \alpha} e^i$ satisfy (8), and therefore so does their linear combination x^*. This shows that (iii) holds for v at x^*, completing the induction and the proof.   □

Theorem 3 makes the symmetry viewpoint via conjugacy from Theorem 2 particularly sharp in the convex setting: standard comonotonicity is fully characterized by the implication $v_i = v_j \Rightarrow x_i^* = x_j^* \ \forall i, j \in [n]$.

An extreme subclass of standard comonotone sets comprises permutation-invariant sets. Specifically, following [40], a set $\mathcal{X} \subseteq \mathbb{R}^n$ is called *permutation-invariant* if $x \in \mathcal{X}$ implies $x_\pi \in \mathcal{X}$ for every $\pi \in \Pi_n$. Such sets naturally satisfy the symmetry requirement in Theorem 2(iv). Moreover, permutation invariance yields standard comonotonicity directly from Definition 1 and Lemma 1, since any attained maximizer can be permuted within $\mathcal{X}$ to match the ordering of v without decreasing the objective value. Figure 2 depicts a permutation-invariant set $\mathcal{X}_3$.

Proposition 3. *Every permutation-invariant set $\mathcal{X}$ is standard comonotone.*

We use Example 3 below to illustrate the results established in this section, which shows that for centered ellipsoids, comonotonicity and permutation invariance are equivalent.

Example 3. Let $\mathcal{X} = \{x : x^\top Q x \leq 1\}$, where $Q \in \mathbb{R}^{n \times n}$ is symmetric positive definite. We aim to show that if $\mathcal{X}$ is comonotone, then it is permutation invariant.

Indeed, because 0 lies in the interior of $\mathcal{X}$, one has $\mathrm{int}\, \mathcal{Z}(\pi) \cap \mathcal{X} \neq \emptyset$ for each $\pi \in \Pi_n$. Consequently, if $\mathcal{X}$ is comonotone under a mapping $\Psi : \Pi_n \to \Pi_n$, then Ψ must be surjective. By Theorem 1, it follows that $\mathcal{X}$ is standard comonotone. Now consider any $v \in \mathbb{R}^n$ and $\bar{x} \in \mathrm{argmax}\{v^\top x : x \in \mathcal{X}\}$. Direct calculation yields that $\bar{x} = \frac{1}{\lambda} Q^{-1} v$ or equivalently $v = \lambda Q \bar{x}$, where $\lambda = \sqrt{v^\top Q^{-1} v}$. Combining above and invoking Theorem 3 (ii), $\mathcal{X}$ is comonotone if and only if

$$v_i = v_j \Rightarrow (Q^{-1} v)_i = (Q^{-1} v)_j \quad \forall i, j \in [n], \ \forall v \in \mathbb{R}^n.$$

Expanding this condition implies that for all $v \in \mathbb{R}^n$ and $i, j \in [n]$ with $v_i = v_j = t$,

$$[(Q^{-1})_{ii} + (Q^{-1})_{ij}]t + \sum_{k \in [n] \setminus \{i.j\}} (Q^{-1})_{ik} v_k = [(Q^{-1})_{ij} + (Q^{-1})_{jj}]t + \sum_{k \in [n] \setminus \{i,j\}} (Q^{-1})_{jk} v_k.$$

Since t and $\{v_k\}_{k \neq i,j}$ are arbitrary, we must have $(Q^{-1})_{ii} = (Q^{-1})_{jj}$, $(Q^{-1})_{ik} = Q_{jk}^{-1}$ for all $i, j \in [n]$ and $k \in [n] \setminus \{i, j\}$. Thus all diagonal entries of Q^{-1} are equal and all off-diagonal entries of Q^{-1} are equal. Equivalently, Q^{-1} (and hence Q) is invariant under coordinate permutations, which implies that $\mathcal{X}$ is permutation invariant.   □

2.2   Non-standard comonotone sets and matroids

Moving beyond standard comonotone sets, general comonotonicity can capture combinatorial structures that lack full symmetry. An important source of non-standard comonotone sets is structure closely related to greedy algorithms. Matroids provide a canonical instance of this kind. Let $\mathcal{M} = ([n], \mathcal{B})$ be a matroid over $[n]$ with a collection of bases $\mathcal{B} \subseteq 2^{[n]}$ [77]. For a set $S \subseteq [n]$, its incidence vector $x \in \{0,1\}^n$ is defined as $x_i = 1$ if $i \in S$ and 0 otherwise for all $i \in [n]$. In this paper, we identify $\mathcal{M}$ with the corresponding set of incidence vectors $\mathcal{X} \subseteq \{0,1\}^n$. We also treat checking whether $S \in \mathcal{B}$ or $S \in \mathcal{M}$ as a constant-time operation.

Proposition 4. *If $\mathcal{X} \subseteq \{0,1\}^n$ encodes a matroid or the set of bases of a matroid, then $\mathcal{X}$ is comonotone. Moreover, the associated permutation mapping Ψ can be accessed in $\mathcal{O}(n \log n)$ time.*

Proof. We first assume $\mathcal{X}$ represents the bases of a matroid. Consider any $\pi \in \Pi_n$ and any vector $v \in \mathbb{R}^n$ satisfying $v_{\pi_1} \geq v_{\pi_2} \cdots \geq v_{\pi_n}$. The best-in greedy algorithm scans elements in the order $\pi(1), \ldots, \pi(n)$ and selects an element whenever feasibility is preserved, i.e., it constructs a vector $x^* \in \{0,1\}^n$ by the recursion

$$x^*_{\pi(i)} = 1 \text{ if and only if } [x^*_{\pi(1)}, \cdots, x^*_{\pi(i-1)}, 1, 0, \cdots, 0]^\top \in \mathcal{X}. \qquad (9)$$

By [24], this procedure returns an optimal solution to $\max\{v^\top x : x \in \mathcal{X}\}$. Importantly, the output x^* depends solely on the feasibility check of $\mathcal{X}$ and the permutation π, not on the specific values of v. Therefore, for all v sorted by π, the greedy algorithm produces the same optimal solution x^*. This proves that a matroid is comonotone. To be more specific, given the optimal solution x^* for the natural order $\pi = (1, 2, \ldots, n)$, one can define $\sigma = \Psi(\pi)$ by requiring $\sigma(i) < \sigma(j)$ if either (1) $x^*_{\sigma(i)} = 1$ and $x^*_{\sigma(j)} = 0$, or (2) $x^*_{\sigma(i)} = x^*_{\sigma(j)}$ and $i < j$. Analogously, $\Psi(\pi)$ can be determined for permutation π other than the natural one.

  We next consider the case where $\mathcal{X}$ represents a matroid, i.e., the incidence vectors of independent sets. Let x^* be the vector given by (9). Then the solution obtained by setting $\bar{x}_i = x^*_i$ if $v_i > 0$ and $x_i = 0$ otherwise is optimal for $\max_{x \in \mathcal{X}} v^\top x$ [24]. Because x^* is binary, x^* and $\bar{x}$ share the same order according to above modification. This implies that $\mathcal{X}$ is comonotone under Ψ. In both cases, the time complexity is $\mathcal{O}(n \log n)$, dominated by the sorting step. $\square$

  The connection between greedy solvability and set structure has been extensively studied beyond matroids in literature. Some notable binary generalizations include antimatroids [19], and greedoids with strong exchange axiom and matroid embedding structures [42, 34]. In the continuous setting, Edmonds' classical best-in greedy algorithm extends to polymatroids and base polyhedra of submodular functions [23]. In discrete convex analysis, M-convex sets serve as the integral analogue of these structures [56, Chapter 4]. One can readily verify the comonotonicity of above greedy-related families using the same reasoning as in Proposition 4; we do not repeat the details here.

  While comonotonicity exhibits rich structure and merits further study in its own right, we believe the current level of treatment is sufficient for this paper's purpose of understanding the tractability of (1).

3   Complexity analysis: A unified theoretical framework

In this section, we propose a general analysis framework that identifies a polynomial number of support candidates and guarantees that (1) has an optimal solution supported on one of them. Under Assumption 1, (1) becomes tractable once the support is fixed; hence, enumerating the candidates and solving the corresponding subproblems yields a polynomial-time algorithm. Notably, our analysis applies to an arbitrary comonotone feasible region in (1).

The proposed framework is developed in three steps as summarized in Figure 3 and outlined below:

Step 1. Leveraging the convexity of f, we show that (1) can be reduced to its *linear counterpart* (10), which maximizes a linear function $c^\top Ax$ over the original feasible set $\mathcal{X}$ for a nominal $c \in \mathbb{R}^r$. However, since c relies on the optimal solution to (1) and is unknown a priori, we cannot solve the linear program directly.

Step 2. To circumvent this, we partition the parameter space $\mathbb{R}^r$ of c into a polynomial number of regions based on optimality conditions of (10). These conditions are derived from the comonotonicity structure of $\mathcal{X}$, and they ensure that all cost vectors within the same region share the same optimality certificate for the linear counterpart.

Step 3. Finally, we show that each region created in Step 2 gives rise to a polynomial number of candidate supports. Enumerating them over all regions produces a polynomial-size list that is guaranteed to contain an optimal support for (1).

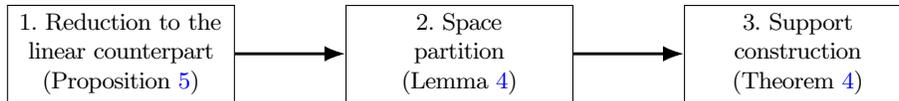| 1. Reduction to the linear counterpart (Proposition 5)	2. Space partition (Lemma 4)	3. Support construction (Theorem 4)

Fig. 3: A theoretical framework of complexity analysis.

Beyond its broad applicability, the proposed framework is also highly adaptable. When additional problem structure is available (e.g., standard comonotonicity), each step can be tailored accordingly to tighten the complexity bounds. To avoid redundancy, we do not repeat the full framework for special families of $\mathcal{X}$ in the subsequent sections; instead, we highlight only the necessary changes.

3.1   Reduction to a linear program

We first establish the reduction from (1) to its linear counterpart.

Proposition 5. *There exists a vector $c \in \mathbb{R}^r$ such that*

$$\max_{x \in \mathcal{X}} c^\top Ax \tag{10}$$

admits an optimal solution, and every optimal solution to the linear counterpart (10) is also optimal for the original problem (1).

Proof. Let $\widehat{x} \in \mathcal{X}$ be an optimal solution to problem (1). Pick any subgradient $\widehat{c} \in \mathbb{R}^r$ of $f(\cdot)$ at $A\widehat{x}$. We first show that $\widehat{x}$ is also optimal for (10) with cost vector $\widehat{c}$. Suppose for contradiction that there exists a solution $x^* \in \mathcal{X}$ such that

$$\widehat{c}^\top A x^* > \widehat{c}^\top A\widehat{x}.$$

By the subgradient inequality of convex f, we have

$$f(Ax^*) \geq f(A\widehat{x}) + \widehat{c}^\top (Ax^* - A\widehat{x}) > f(A\widehat{x}),$$

which contradicts the optimality of $\widehat{x}$ in (1). Hence, $\widehat{x}$ must be optimal for (10) with $\widehat{c}$.

Next, we show that any optimal solution of (10) with the vector $\widehat{c}$ is also optimal for (1). Let $\tilde{x} \in \mathcal{X}$ be any optimal solution of (10) with $\widehat{c}$, and thus $\widehat{c}^\top A\tilde{x} = \widehat{c}^\top A\widehat{x}$ holds. Then, applying the subgradient inequality of f again yields

$$f(A\tilde{x}) \geq f(A\widehat{x}) + \widehat{c}^\top (A\tilde{x} - A\widehat{x}) = f(A\widehat{x}),$$

where the inequality must hold with equality because (1) attains the optimum at $\widehat{x}$. It follows that $\tilde{x}$ also attains the optimal value of (1). We thus complete the proof. $\square$

It is worth noting that Proposition 5 is stronger than the classical results in the literature ([7], [10], [66, Theorem 32.3]). Roughly speaking, the latter basically asserts $\operatorname{argmax}\{c^\top Ax : x \in \mathcal{X}\} \cap \operatorname{argmax}\{f(A^\top x) : x \in \mathcal{X}\} \neq \emptyset$ for a certain $c \in \mathbb{R}^r$. Consequently, even if such a c is known, one still needs to identify, among the maximizers of the linear counterpart (10), a solution that also maximizes $f(A^\top x)$. To avoid such an essentially bilevel program, the literature often explicitly or implicitly ([e.g., 41, 69, 60, 15]) postulates a nondegeneracy assumption to ensure that the linear counterpart has a unique solution. In contrast, because Proposition 5 establishes the stronger inclusion $\operatorname{argmax}\{c^\top Ax : x \in \mathcal{X}\} \subseteq \operatorname{argmax}\{f(A^\top x) : x \in \mathcal{X}\}$, any optimizer of the linear counterpart is automatically optimal for (1). This frees us from above recurring technical nuisance and removes the need for such nondegeneracy assumptions.

3.2   Space partition

Since the required cost vector c in Proposition 5 is typically unknown, we cannot directly work with the linear counterpart (10). Fortunately, under comonotonicity of $\mathcal{X}$, we can instead rely on the following optimality condition for (10)

$$A^\top c \in \mathcal{Z}(\pi) \Rightarrow \operatorname{argmax}\{A^\top c : x \in \mathcal{X}\} \cap \mathcal{Z}(\Psi(\pi)) \neq \emptyset, \tag{11}$$

provided that $\operatorname{argmax}\{A^\top c : x \in \mathcal{X}\} \neq \emptyset$. This condition naturally leads to a partition of the parameter space of $c \in \mathbb{R}^r$, as we explain next.

We first introduce a technique from combinatorial geometry– *hyperplane arrangement*. Hyperplane arrangement is a standard tool for space partition (e.g., [15, 60, 62]). Specifically, let $\mathcal{H} = \{H_i\}_{i \in [p]}$ be a collection of p hyperplanes in $\mathbb{R}^q$, where each H_i denotes a hyperplane. The family $\mathcal{H}$ partitions the space $\mathbb{R}^q$ into a collection of *cells* with dimensions ranging from 0 to q. Each cell R is a possibly nonclosed region of the form $R = \cap_{i \in [p]} \tilde{H}_i$, where $\tilde{H}_i \in \{H_i, H_i^>, H_i^<\}$ represents either the hyperplane

H_i itself or one of its two open half-spaces. We call R a *d-cell* if $\dim(R) \leq d$. The collection of all such cells is called the *arrangement* $\mathcal{A}(\mathcal{H})$ of $\mathcal{H}$. For more details, we refer to [21, 22] or standard monographs in combinatorial geometry (e.g., [20]). Lemma 3 below gives a standard counting bound for hyperplane arrangements.

Lemma 3 ([21]). *Let $\mathcal{H}$ be a finite set of p hyperplanes in $\mathbb{R}^q$. Then the arrangement $\mathcal{A}(\mathcal{H})$ can be constructed in $\mathcal{O}(p^q)$ time and contains $\mathcal{O}(p^d)$ d-cells for each $d \in [q]$. Moreover, if all hyperplanes in $\mathcal{H}$ pass through the origin, then these two bounds tighten to $\mathcal{O}((p-1)^{q-1})$ and $\mathcal{O}((p-1)^{d-1})$, respectively.*

We now apply Lemma 3 to the hyperplane arrangement induced by

$$H_{ij} = \left\{ c \in \mathbb{R}^r : \left(A^\top c \right)_i - \left(A^\top c \right)_j = 0 \right\} \qquad \forall 1 \leq i < j \leq n. \tag{12}$$

Lemma 4. *The hyperplanes (12) partition $\mathbb{R}^r$ into $\mathcal{O}(n^{2r-2})$ cells. Moreover, for each cell R, there exists a permutation $\pi \in \Pi_n$ such that $A^\top c \in \mathcal{Z}(\pi)$ for all $c \in R$.*

Proof. Let $\mathcal{H} = \{H_{ij}\}_{1 \leq i < j \leq n}$ denote the hyperplanes given in (12). Since $|\mathcal{H}| = n(n-1)/2$ and all hyperplanes in $\mathcal{H}$ pass through the origin, Lemma 3 implies $|\mathcal{A}(\mathcal{H})| = \mathcal{O}(|\mathcal{H}|^{r-1}) = \mathcal{O}(n^{2r-2})$.

Each cell R is characterized by $\bigcap_{1 \leq i < j \leq n} \tilde{H}_{ij}$, where $\tilde{H}_{ij} \in \{H_{ij}, H_{ij}^>, H_{ij}^<\}$. By construction, for any fixed cell R and any pair $i < j$, the sign of $(A^\top c)_i - (A^\top c)_j$ is constant over $c \in R$. This implies that the relative ordering between $(A^\top c)_i$ and $(A^\top c)_j$ is invariant within R. Taken over all pairs, these comparisons determine an overall ordering of $A^\top c$ throughout R. Hence, there exists a permutation $\pi \in \Pi_n$ such that $A^\top c \in \mathcal{Z}(\pi)$ for all $c \in R$, completing the proof. □

3.3   Support enumeration

Based on the cell decomposition in Lemma 4, we now turn to support construction and derive complexity results in this subsection. Lemma 5 below shows that a fixed ordering can yield $\mathcal{O}(n^2)$ distinct supports.

Lemma 5. *For any $\sigma \in \Pi_n$, the set $\{\operatorname{supp}(z) : z \in \mathcal{Z}(\sigma)\}$ contains $\mathcal{O}(n^2)$ supports.*

Proof. Since every vector $x \in \mathcal{Z}(\sigma)$ is sorted by σ, its zero entries (if any) appear consecutively and form a contiguous block under this permutation. The support of x is uniquely determined by this zero block. Specifically, if x contains no zero entries, there is exactly one support $\sup(x) = [n]$. Otherwise, there exist indices $1 \leq t_1 \leq t_2 \leq n$ such that

$$x_{\sigma(1)} \geq \cdots \geq x_{\sigma(t_1-1)} > 0, \quad x_{\sigma(t_1)} = \cdots = x_{\sigma(t_2)} = 0, \quad 0 > x_{\sigma(t_2+1)} \geq \cdots \geq x_{\sigma(n)}.$$

For any fixed pair (t_1, t_2), the support of x is then uniquely determined. Since there are $\mathcal{O}(n^2)$ possible choices of (t_1, t_2), the conclusion follows. □

In the rest of this paper, we make the following assumption.

Assumption 2 *The permutation mapping* Ψ *in Definition 1 can be accessed in polynomial time; we denote this cost by* T_2.

We recall that T_1 denotes the time required to solve a fixed-support subproblem in Assumption 1. We are now ready to present the main result in this section.

Theorem 4. *Let* $\mathcal{X}$ *be a comonotone set. Under Assumption 2, the following hold:*

(i) *There exists a collection of* $\mathcal{O}(n^{2r})$ *candidate supports for* (1), *among which at least one support is optimal; and*

(ii) *Under Assumption 1, problem* (1) *admits a polynomial-time algorithm with complexity* $\mathcal{O}(n^{2r-2} \cdot \mathsf{T}_2 + n^{2r} \cdot \mathsf{T}_1)$.

Proof. Our proof is constructive and includes two parts.

Part I. We first prove (i). By Lemma 4, the hyperplanes (12) partition $\mathbb{R}^r$ into $K = \mathcal{O}(n^{2r-2})$ cells $R_1, \cdots, R_K$. In addition, for each $k \in [K]$, there exists a permutation $\pi^k \in \Pi_n$ such that $A^\top c \in \mathcal{Z}(\pi^k)$ for all $c \in R_k$. Let Ψ be the permutation mapping associated with the comonotone set $\mathcal{X}$ and $\sigma^k = \Psi(\pi^k)$ for all $k \in [K]$.

According to Proposition 5, there exists a cost vector $c^* \in \mathbb{R}^r$ such that every optimal solution to (10) with $c = c^*$ is also optimal for (1). Since $\{R_k\}_{k \in [K]}$ forms a partition of $\mathbb{R}^r$, the vector c^* lies in a certain cell $R_{\bar{k}}$ with $\bar{k} \in [K]$, implying $A^\top c^* \in \mathcal{Z}(\pi^{\bar{k}})$ by Lemma 4. Combining the above with the optimality condition (11), we deduce that $\mathcal{Z}(\Psi(\pi^{\bar{k}})) = \mathcal{Z}(\sigma^{\bar{k}})$ contains an optimal solution to (1), and therefore so does $\bigcup_{k \in [K]} \mathcal{Z}(\sigma^k)$. This allows us to equivalently convert (1) into

$$\max_{x \in \mathcal{X}} \left\{ f(A^\top x) : x \in \bigcup_{k \in [K]} \mathcal{Z}(\sigma^k) \right\}. \tag{13}$$

By Lemma 5, each fixed ordering $\mathcal{Z}(\sigma^k)$ can yield at most $\mathcal{O}(n^2)$ possible supports. Since (13) involves $K = \mathcal{O}(n^{2r-2})$ orderings, the total number of candidate supports is $\mathcal{O}(n^{2r})$. This proves (i).

Part II. We next prove (ii). Let $S_1, \ldots, S_L$ be the support candidates constructed above, where $L = \mathcal{O}(n^{2r})$. Then (13) admits the equivalent reformulation

$$\max_{\ell \in [L]} \max_{x \in \mathcal{X}} \left\{ f(A^\top x) : \text{supp}(x) = S_\ell \right\}. \tag{14}$$

The running time consists of two parts: constructing $\{S_\ell\}_{\ell \in [L]}$ and solving all fixed-support subproblems, i.e., the inner maximizations in (14). Because accessing $\sigma^k = \Psi(\pi^k)$ costs T_2 per cell and there are $\mathcal{O}(n^{2r-2})$ cells, the total time for support construction is $\mathcal{O}(n^{2r-2} \cdot \mathsf{T}_2 + n^{2r})$. Moreover, under Assumption 1, each fixed-support subproblem in (14) can be solved in T_1 time. Thus, addressing all subproblems requires $L \cdot \mathsf{T}_1 = \mathcal{O}(n^{2r}) \cdot \mathsf{T}_1$ time. Consequently, the overall complexity is given by $\mathcal{O}\left(n^{2r-2} \cdot \mathsf{T}_2 + n^{2r} \cdot \mathsf{T}_1\right)$. This completes the proof. $\qquad\square$

When $\mathcal{X}$ is a matroid or the bases of a matroid, the complexity bounds from Theorem 4 can be further improved as shown in Corollary 1 below, which recovers the result of [60, Theorem 1.4] from a different perspective.

Corollary 1. *If $\mathcal{X} \in \{0,1\}^n$ encodes a matroid or the bases of a matroid, then problem* (1) *admits a polynomial-time algorithm with complexity $\mathcal{O}(n^{2r-1}\log n)$.*

Proof. Assume $\mathcal{X}$ represents the bases of a matroid. Then as long as the cost vector $A^\top c$ of (10) is sorted by the same permutation, the best-in greedy algorithm (9) returns the same optimal solution x^* to (10). It is thus enough to evaluate one support per permutation cell. Since the arrangement in Theorem 4 has $\mathcal{O}(n^{2r-2})$ cells, this yields $\mathcal{O}(n^{2r-2})$ candidate supports. Furthermore, Assumption 1 is satisfied due to $\mathcal{X} \subseteq \{0,1\}^n$. Hence, the claimed complexity follows from $\mathsf{T}_2 = \mathcal{O}(n \log n)$.

We next turn to the case where $\mathcal{X}$ represents a matroid. In this setting, we refine the arrangement in Theorem 4 by additionally including the coordinate hyperplanes

$$\widehat{H}_i = \{c \in \mathbb{R}^r : (A^\top c)_i = 0\}, \qquad \forall i \in [n].$$

On each induced cell R, both the relative order of $A^\top c$ (from (12)) and the sign pattern of $A^\top c$ (from $\widehat{H}_i$) remain fixed. Thereby, the solution $\bar{x}$ constructed in Proposition 4 is optimal for all $c \in R$, implying that each cell contributes a single support candidate. Since the total number of hyperplanes is $n(n-1)/2 + n = \mathcal{O}(n^2)$, Lemma 3 yields $\mathcal{O}(n^{2r-2})$ cells and results in an overall complexity of $\mathcal{O}(n^{2r-2}) \cdot \mathsf{T}_2 = \mathcal{O}(n^{2r-1}\log n)$. □

4 Improved complexity via lifting under standard comonotonicity

Lifting is a widely used technique in mixed-integer programming to simplify algebraic representation by working in an extended space where the object in question is easier to describe and then interpreting the original object as its projection. In this section, we present a lifting technique of this kind to refine the complexity analysis for standard comonotone sets. Rather than constructing the hyperplane arrangement directly in the space of cost vectors c, we augment it with a few additional parameters and build an arrangement in a lifted space. Projecting the cells of the lifted arrangement back onto the original space yields the same partition as the pairwise arrangement (12). Crucially, the lifted arrangement is defined by only $\mathcal{O}(n)$ hyperplanes, as opposed to $\mathcal{O}(n^2)$ in (12). This reduction in turn leads to improved complexity bounds under standard comonotonicity.

To illustrate the lifting idea, consider a specific standard comonotone set $\mathcal{X} = \{x \in [0,1]^n : e^\top x = k\}$, where $0 \le k \le n$. Since (10) is a linear program, strong duality guarantees the existence of an optimal multiplier λ such that (10) is equivalent to

$$\max_{x \in [0,1]^n} c^\top Ax - \lambda e^\top x + \lambda k = \max_{x \in [0,1]^n} \left(A^\top c - \lambda e\right)^\top x + \lambda k.$$

Clearly, any optimal solution x^* to (10) must satisfy the thresholding rule

  (i) If $\left(A^\top c\right)_i > \lambda$, then $x_i^* = 1$; and
  (ii) If $\left(A^\top c\right)_i < \lambda$, then $x_i^* = 0$.

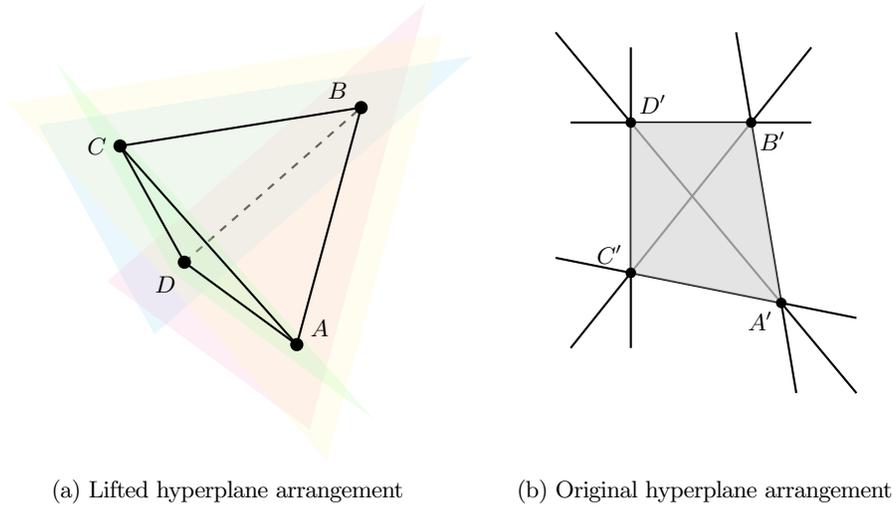(a) Lifted hyperplane arrangement   (b) Original hyperplane arrangement

Fig. 4: Comparison of hyperplane arrangements ($r = 3, n = 4$, slice $c_3 = 1$)

The thresholding rule motivates us to introduce a parameter λ and consider the lifted arrangement in $\mathbb{R}^r \times \mathbb{R}$ induced by the $\mathcal{O}(n)$ hyperplanes $\{(c, \lambda) \in \mathbb{R}^r \times \mathbb{R} : (A^\top c)_i - \lambda = 0\}$, $i \in [n]$.

Figure 4 illustrates an instance with $r = 3$ and $n = 4$. For visualization, we only depict the cross-section obtained by fixing $c_3 = 1$. In the original space of cost vectors, the pairwise arrangement in Fig. 4b is induced by $\binom{n}{2} = 6$ straight lines. In contrast, in the lifted space shown in Fig. 4a, it suffices to consider $n = 4$ planes, which correspond to the four facets of a tetrahedron. The shaded region in Fig. 4b depicts the projection of this tetrahedron onto the original c-space. We formalize the above lifting idea in the remainder of the section.

4.1   New optimality conditions

We first derive refined optimality conditions for (10) over a general standard comonotone feasible set $\mathcal{X}$. The conditions show that comparing each component of $A^\top c$ against two scalar thresholds determines the sign pattern of an optimal solution.

Definition 2. *For any vector $c \in \mathbb{R}^r$ and parameters $\underline{\lambda} \leq \overline{\lambda}$, define the set*

$$\mathcal{Q}\left(c, \overline{\lambda}, \underline{\lambda}\right) \triangleq \{x \in \mathbb{R}^n : \text{(a) sign constraints, (b) ordering constraints}\},$$

where

(a) Sign constraints: For all $i \in [n]$,

$$x_i > 0 \text{ if } \left(A^\top c\right)_i > \overline{\lambda}, \quad x_i = 0 \text{ if } \underline{\lambda} < \left(A^\top c\right)_i < \overline{\lambda}, \quad x_i < 0 \text{ if } \left(A^\top c\right)_i < \underline{\lambda},$$
$$x_i \geq 0 \text{ if } \left(A^\top c\right)_i = \overline{\lambda} \text{ and } \overline{\lambda} > \underline{\lambda}, \quad x_i \leq 0 \text{ if } \left(A^\top c\right)_i = \underline{\lambda} \text{ and } \overline{\lambda} > \underline{\lambda}.$$

(b) *Ordering constraints:* For all $i, j \in [n]$ with $i < j$,

$$x_i \geq x_j \ \text{ if } \ \left(A^\top c\right)_i = \left(A^\top c\right)_j \in \left\{\overline{\lambda}, \underline{\lambda}\right\}.$$

For convenience, we treat $\mathcal{Q}(c, \overline{\lambda}, \underline{\lambda}) = \emptyset$ if $\underline{\lambda} > \overline{\lambda}$.

Lemma 6. *Suppose that $\mathcal{X}$ is standard comonotone. Then, for any $c \in \mathbb{R}^r$, there exist parameters $\underline{\lambda} \leq \overline{\lambda}$ such that $\mathcal{Q}\left(c, \overline{\lambda}, \underline{\lambda}\right)$ contains an optimal solution x^* of (10), i.e.,*

$$x^* \in \mathcal{Q}\left(c, \overline{\lambda}, \underline{\lambda}\right),$$

whenever (10) attains the optimum.

Proof. The vector $A^\top c$ may admit multiple sorting permutations $\pi \in \Pi_n$ due to ties. We consider one such π using the lexicographical tie-breaking rule: if $(Ac)_i = (Ac)_j$ and $i < j$, then $\pi_i < \pi_j$. By the standard comonotonicity of $\mathcal{X}$, (10) admits an optimal solution x^* such that

$$x_i^* \geq x_j^* \quad \forall 1 \leq i < j \leq n \text{ s.t. } \left(A^\top c\right)_i = \left(A^\top c\right)_j. \tag{15}$$

Next, we partition the indices according to the sign of x^*:

$$\mathcal{I}^+ = \{i : x_i^* > 0\}, \ \ \mathcal{I}^0 = \{i : x_i^* = 0\}, \ \ \mathcal{I}^- = \{i : x_i^* < 0\},$$

and introduce

$$\underline{\lambda} = \max_{i \in \mathcal{I}^-} \left(A^\top c\right)_i, \ \ \overline{\lambda} = \min_{i \in \mathcal{I}^+} \left(A^\top c\right)_i,$$

with the conventions $\underline{\lambda} = \min_{i \in [n]}(A^\top c)_i$ if $\mathcal{I}^- = \emptyset$ and $\overline{\lambda} = \max_{i \in [n]}(A^\top c)_i$ if $\mathcal{I}^+ = \emptyset$. It is evident that $x_i^* > x_j^* > x_k^*$ for all $i \in \mathcal{I}^+, j \in \mathcal{I}^0, k \in \mathcal{I}^-$. By Theorem 2 (iii), which holds as a necessary condition for standard comonotonicity even when $\mathcal{X}$ is not compact, one has that

$$(A^\top c)_i \geq \overline{\lambda} \geq (A^\top c)_j \geq \underline{\lambda} \geq (A^\top c)_k \quad \forall i \in \mathcal{I}^+, j \in \mathcal{I}^0, k \in \mathcal{I}^-. \tag{16}$$

Finally, to show $x^* \in \mathcal{Q}(c, \overline{\lambda}, \underline{\lambda})$, we check the conditions in Definition 2:

- If $(A^\top c)_i > \overline{\lambda}$, then $i \notin \mathcal{I}^- \cup \mathcal{I}^0$ in view of (16), implying $x_i^* > 0$. Similarly, one can verify that $(A^\top c)_i < \underline{\lambda}$ implies $x_i^* < 0$, and $\underline{\lambda} < (A^\top c)_i < \overline{\lambda}$ implies $x_i^* = 0$.
- If $\overline{\lambda} > \underline{\lambda}$ and $\left(A^\top c\right)_i = \overline{\lambda}$, then one also has $\left(A^\top c\right)_i > \underline{\lambda}$. Thus, (16) implies $i \notin \mathcal{I}^-$, i.e., $x_i^* \geq 0$. Similarly, $\overline{\lambda} > \underline{\lambda}$ and $\left(A^\top c\right)_i = \underline{\lambda}$ imply $x_i^* \leq 0$.
- The ordering constraints in Definition 2 hold by (15).

This completes the proof.   □

The optimality condition in Lemma 6 is more straightforward than the one in Lemma 4. The former explicitly determines the sign pattern of an optimal solution to (10), whereas the latter only provides the ordering information.

4.2   Extended space partition and improved complexity

In this subsection, building on the optimality condition in Lemma 6, we propose a new partition of the extended parameter space $\mathbb{R}^r \times \mathbb{R} \times \mathbb{R}$ that corresponds to c, $\underline{\lambda}$, and $\overline{\lambda}$, respectively. We then show how this partition leads to an improved complexity bound under standard comonotonicity.

The partition of the extended space is induced by the following $2n+1$ hyperplanes

$$
\begin{aligned}
H_i &= \left\{ \left(c, \underline{\lambda}, \overline{\lambda}\right) \in \mathbb{R}^r \times \mathbb{R} \times \mathbb{R} : \left(A^\top c\right)_i = \overline{\lambda} \right\}, \quad \forall i \in [n], \\
H_{i+n} &= \left\{ \left(c, \underline{\lambda}, \overline{\lambda}\right) \in \mathbb{R}^r \times \mathbb{R} \times \mathbb{R} : \left(A^\top c\right)_i = \underline{\lambda} \right\}, \quad \forall i \in [n], \\
H_{2n+1} &= \left\{ \left(c, \underline{\lambda}, \overline{\lambda}\right) \in \mathbb{R}^r \times \mathbb{R} \times \mathbb{R} : \overline{\lambda} = \underline{\lambda} \right\}.
\end{aligned}
\tag{17}
$$

Recall the notation $\mathcal{Z}^+(\pi) = \mathcal{Z}(\pi) \cap \mathbb{R}^n_+$. We need Lemma 7 below to bound the number of supports per cell.

Lemma 7. *For any $\sigma \in \Pi_n$, the set $\{\operatorname{supp}(x) : x \in \mathcal{Z}^+(\sigma)\}$ contains $\mathcal{O}(n)$ supports.*

Proof. If $x \in \mathcal{Z}^+(\sigma)$ is strictly positive, its support is exactly $[n]$. If $x \in \mathcal{Z}^+(\sigma)$ contains zeros, by nonnegativity, the vector must satisfy

$$
x_{\sigma(1)} \geq \cdots \geq x_{\sigma(t)} > 0, \quad x_{\sigma(t+1)} = \cdots = x_{\sigma(n)} = 0
$$

for some index $t \in [n]$, which yields n distinct supports. $\qquad\square$

Theorem 5. *Suppose that $\mathcal{X}$ is standard comonotone. Then*

  (i) *There exists a collection of $\mathcal{O}(n^{r+1})$ candidate supports for (1), among which at least one support is optimal; and*

  (ii) *Under Assumption 1, problem (1) admits a polynomial-time algorithm with complexity $\mathcal{O}(n^{r+1} \cdot \mathsf{T}_1)$.*

Proof. Denote by $\mathcal{H} = \{H_t\}_{t \in [2n+1]}$ the hyperplanes in (17). Let $\mathcal{A}(\mathcal{H}) = \{R_k\}_{k \in [K]}$ be the resulting arrangement. Within each cell R_k, the signs of

$$
\left(A^\top c\right)_i - \overline{\lambda}, \forall i \in [n], \quad \left(A^\top c\right)_i - \underline{\lambda}, \forall i \in [n], \text{ and } \overline{\lambda} - \underline{\lambda}
$$

are constant. By Definition 2, the set $\mathcal{Q}(c, \overline{\lambda}, \underline{\lambda})$ depends only on these sign patterns and thus is invariant within each cell. More specifically, the following index sets $\mathcal{I}_k^+$, $\mathcal{I}_k^0$, $\mathcal{I}_k^-$, $\mathcal{I}_k^{=\overline{\lambda}}$, and $\mathcal{I}_k^{=\underline{\lambda}}$ remain fixed for any $(c, \overline{\lambda}, \underline{\lambda}) \in R_k$, where

$$
\mathcal{I}_k^+ = \left\{ i : \left(A^\top c\right)_i > \overline{\lambda} \right\}, \quad \mathcal{I}_k^0 = \left\{ i : \underline{\lambda} < \left(A^\top c\right)_i < \overline{\lambda} \right\}, \quad \mathcal{I}_k^- = \left\{ i : \left(A^\top c\right)_i < \underline{\lambda} \right\},
$$
$$
\mathcal{I}_k^{=\overline{\lambda}} = \left\{ i : \left(A^\top c\right)_i = \overline{\lambda} \right\}, \quad \mathcal{I}_k^{=\underline{\lambda}} = \left\{ i : \left(A^\top c\right)_i = \underline{\lambda} \right\}.
$$

Based on these index sets, we define the cell-dependent set:

$$
\tilde{\Upsilon}^k \triangleq \left\{ x \in \mathbb{R}^n : \begin{array}{l} x_i > 0, \forall i \in \mathcal{I}_k^+, \quad x_i = 0, \forall i \in \mathcal{I}_k^0, \quad x_i < 0, \forall i \in \mathcal{I}_k^- \\ x_i \geq x_j, \forall i, j \in \mathcal{I}_k^{=\overline{\lambda}} \text{ with } i < j \text{ or } \forall i, j \in \mathcal{I}_k^{=\underline{\lambda}} \text{ with } i < j \end{array} \right\}. \tag{18}
$$

Since the sign of $\overline{\lambda} - \underline{\lambda}$ is constant on R_k, define Υ^k by setting $\Upsilon^k = \emptyset$ if $\overline{\lambda} < \underline{\lambda}$ on R_k, $\Upsilon^k = \tilde{\Upsilon}^k$ if $\overline{\lambda} = \underline{\lambda}$ on R_k, and

$$\Upsilon^k = \tilde{\Upsilon}^k \cap \left\{ x \in \mathbb{R}^n : x_i \geq 0 \,\forall i \in \mathcal{I}_k^{=\overline{\lambda}} \text{ and } x_i \leq 0 \,\forall i \in \mathcal{I}_k^{=\underline{\lambda}} \right\}$$

if $\overline{\lambda} > \underline{\lambda}$ on R_k. By construction, we have $\Upsilon^k = \mathcal{Q}(c, \overline{\lambda}, \underline{\lambda})$ for every $(c, \overline{\lambda}, \underline{\lambda}) \in R_k$.

Combining Proposition 5 with Lemma 6, there exist c^* and $\underline{\lambda}^* \leq \overline{\lambda}^*$ such that $\mathcal{Q}(c^*, \overline{\lambda}^*, \underline{\lambda}^*)$ contains an optimal solution of (1). Since $(c^*, \overline{\lambda}^*, \underline{\lambda}^*)$ must belong to a certain cell, (1) can be equivalently converted into

$$z^* = \max_{x \in \mathcal{X}} \left\{ f(A^\top x) : x \in \bigcup_{k \in [K]} \Upsilon^k \right\}. \tag{19}$$

Next, we bound the number of supports contributed by each nonempty Υ^k. By (18), the signs of x_i are fixed for $i \in \mathcal{I}_k^+ \cup \mathcal{I}_k^0 \cup \mathcal{I}_k^-$. While the entries indexed by $\mathcal{I}_k^{=\overline{\lambda}} \cup \mathcal{I}_k^{=\underline{\lambda}}$ do not have fixed signs, their relative order is fixed by construction. This leads to three cases:

- *Case 1: $\mathcal{I}_k^{=\overline{\lambda}} \cup \mathcal{I}_k^{=\underline{\lambda}} = \emptyset$.* In this case, each cell R_k can only contribute one single support candidate. The number of such cells is at most $\mathcal{O}(n^{r+1})$ by Lemma 3.
- *Case 2: exactly one of $\mathcal{I}_k^{=\overline{\lambda}}$ and $\mathcal{I}_k^{=\underline{\lambda}}$ is nonempty.* In this case, one must have $\overline{\lambda} > \underline{\lambda}$ and $\dim(R_k) \leq r + 1$. Applying Lemma 7 to the nonempty tie set, each R_k can yield $\mathcal{O}(n)$ supports. The number of such cells is $\mathcal{O}(n^r)$ by Lemma 3.
- *Case 3: both $\mathcal{I}_k^{=\overline{\lambda}}$ and $\mathcal{I}_k^{=\underline{\lambda}}$ are nonempty.* In this case, one has $\dim(R_k) \leq r$. Applying Lemma 5 to $\mathcal{I}_k^{=\overline{\lambda}} \cup \mathcal{I}_k^{=\underline{\lambda}}$, each R_k can yield at most $\mathcal{O}(n^2)$ supports. The number of such cells is $\mathcal{O}(n^{r-1})$ by Lemma 3.

Combining the three cases, the total number of distinct supports for (19) is $\mathcal{O}(n^{r+1}) + \mathcal{O}(n^r) \cdot \mathcal{O}(n) + \mathcal{O}(n^{r-1}) \cdot \mathcal{O}(n^2) = \mathcal{O}(n^{r+1})$. The proof for (ii) is identical to **Part II** of Theorem 4 and thus omitted. $\square$

Compared with Theorem 4, Theorem 5 improves the support complexity by a square-root factor, reducing it from $\mathcal{O}(n^{2r})$ to $\mathcal{O}(n^{r+1})$.

4.3   Complexity under additional sign conditions

In this subsection, we show that additional sign structures can lead to further complexity reductions for (1) over standard comonotone sets.

Nonnegative and standard comonotone sets When $\mathcal{X}$ is nonnegative and standard comonotone, any optimal solution of (10) contains no negative entries. Accordingly, the two thresholds in Definition 2 introduced to separate positive and negative entries are no longer both necessary. Indeed, it suffices to keep a single threshold associated with positive entries, leading to the simplified variant of $\mathcal{Q}$ in Definition 2 below.

Definition 3. *For any vector $c \in \mathbb{R}^r$ and a parameter λ, define*

$$\mathcal{Q}_+ (c, \lambda) \triangleq \{x \in \mathbb{R}^n : (a) \text{ sign constraints, } (b) \text{ ordering constraints}\},$$

where

(a) Sign constraints: For any $i \in [n]$,

$$x_i > 0 \text{ if } \left(A^\top c\right)_i > \lambda, \quad x_i = 0 \text{ if } \left(A^\top c\right)_i < \lambda.$$

(b) Ordering constraints: For any $i, j \in [n]$ with $i < j$,

$$x_i \geq x_j \geq 0 \text{ if } \left(A^\top c\right)_i = \left(A^\top c\right)_j = \lambda.$$

The set $\mathcal{Q}_+$ characterizes the optimality conditions of (10) when $\mathcal{X}$ is nonnegative and standard comonotone. In turn, the conditions on (c, λ) in Definition 3 motivate partitioning the parameter space $\mathbb{R}^r \times \mathbb{R}$ using hyperplanes below

$$H_i = \left\{(c, \lambda) \in \mathbb{R}^r \times \mathbb{R} : \left(A^\top c\right)_i = \lambda\right\}, \quad \forall i \in [n]. \tag{20}$$

Consequently, the simpler partition yields fewer cells than in general settings and leads to an improved complexity bound in Proposition 6.

Proposition 6. *Suppose that $\mathcal{X}$ is nonnegative and standard comonotone. Then*

(i) There exists a collection of $\mathcal{O}(n^r)$ candidate supports for (1), among which at least one support is optimal.

(ii) Under Assumption 1, problem (1) admits a polynomial-time algorithm with complexity $\mathcal{O}(n^r \cdot \mathsf{T}_1)$.

Proof. Since the argument follows that of Theorem 5, we highlight only the necessary changes. Let $\mathcal{A} = \{R_k\}_{k \in [K]}$ be the arrangement induced by the hyperplanes in (20). By adapting the analysis of Lemma 6 and dropping the parameters $\underline{\lambda}$ therein, one can deduce that for any $c \in \mathbb{R}^r$, there exists a parameter λ such that one optimal solution of (10) lies in $\mathcal{Q}_+(c, \lambda)$. Moreover, by construction, $\mathcal{Q}_+(c, \lambda)$ is a constant set (denoted by Υ^k) within each cell R_k. Consequently, as shown in Theorem 5, $\cup_{k \in [K]} \Upsilon^k$ contains an optimal solution to (1). It remains to bound the number of possible supports for $\cup_{k \in [K]} \Upsilon^k$. We first notice that *Case 3* in the proof of Theorem 5 does not exist. Moreover, since the lifted parameter space has one fewer dimension, the support count estimated in the analysis of Theorem 5 reduces to $\mathcal{O}(n^r)$ by Lemma 3. We thus complete the proof. $\qquad \square$

Sign-invariant and standard comonotone sets A set $\mathcal{X} \subseteq \mathbb{R}^n$ is called *sign-invariant* if each $\bar{x} \in \mathcal{X}$ implies $x \in \mathcal{X}$ for all x satisfying $|x| = |\bar{x}|$. Suppose that $\mathcal{X}$ is sign-invariant and standard comonotone. Then for any $c \in \mathbb{R}^r$ and $x \in \mathcal{X}$, we can construct $\bar{x} \in \mathcal{X}$ such that

$$c^\top A x \leq c^\top A \bar{x} = \left(|A^\top c|\right)^T |x|.$$

Specifically, for each $i \in [n]$, we let $\bar{x}_i = x_i$ if $(A^\top c)_i x_i \geq 0$, and $\bar{x}_i = -x_i$ otherwise. Consequently, if $x^* \in \mathcal{X}$ is optimal to (10), then $|x^*|$ must be optimal to

$$\max_{x \in \mathcal{X}} \left(\left|A^\top c\right|\right)^T x. \tag{21}$$

Because $|x^*| \in \mathcal{X}$ is nonnegative and shares the same support as x^*, most of the analysis developed earlier for nonnegative standard comonotone sets also applies, with (21) in place of (10). The main difference is that the objective in (21) involves absolute values of the cost vector. Accordingly, we modify $\mathcal{Q}_+(c, \lambda)$ by replacing $A^\top c$ in Definition 3 with $|A^\top c|$, and denote the resulting set by $\mathcal{Q}^{\mathrm{sign}}(c, \lambda)$. We then refine the space partition to guarantee that the set $\mathcal{Q}^{\mathrm{sign}}(c, \lambda)$ remains constant within each cell. Specifically, to achieve component-wise comparisons of absolute values against λ, we partition $\mathbb{R}^r \times \mathbb{R}$ using the following $2n$ hyperplanes

$$H_i = \left\{(c, \lambda) \in \mathbb{R}^r \times \mathbb{R} : \left(A^\top c\right)_i = \lambda\right\}, \quad \forall i \in [n],$$
$$H_{i+n} = \left\{(c, \lambda) \in \mathbb{R}^r \times \mathbb{R} : \left(A^\top c\right)_i = -\lambda\right\}, \quad \forall i \in [n].$$

The analysis of Proposition 6 extends directly and leads to Proposition 7.

Proposition 7. *Suppose that $\mathcal{X}$ is sign-invariant and standard comonotone. Then*

(i) *There exists a collection of $\mathcal{O}(n^r)$ candidate supports for (1), among which at least one support is optimal; and*

(ii) *Under Assumption 1, problem (1) admits a polynomial-time algorithm with complexity $\mathcal{O}(n^r \cdot \mathsf{T}_1)$.*

We illustrate further uses of the lifting idea in the subsequent sections.

5   Beyond Comonotonicity and Convexity

In previous sections, we assume a finite convex objective function and a comonotone feasible region. In this section, we extend the established complexity results to two settings where these assumptions fail.

5.1   Quasi-convex objective

In some applications, such as minimal cost-reliability ratio spanning tree problem [12], the objective in (1) is a *quasi-convex* rather than convex function. To be more precise, a function $g : \mathcal{D} \to \mathbb{R}$ is called *quasi-convex*, where $\mathcal{D} \subset \mathbb{R}^n$ is a convex open set, if the level set $\{x \in \mathcal{D} : g(x) \leq t\}$ is convex for all $t \in \mathbb{R}$. An important source of quasi-convex functions arises in fractional programming [27, 50, 70, 32]. Specifically, the function $g(x) = \frac{g_1(x)}{g_2(x)}$ is quasi-convex when $g_1 : \mathcal{D} \to \mathbb{R}_+$ is convex and $g_2 : \mathcal{D} \to \mathbb{R}_{++}$ is concave. Moreover, composing a quasi-convex function with a univariate monotone nondecreasing function preserves quasi-convexity. Similar to Proposition 5, under additional regularity conditions, a quasi-convex maximization problem can be reduced to a nominal linear optimization problem.

Proposition 8. *Assume $f : \mathcal{D} \to \mathbb{R}$ is a upper semicontinuous and quasi-convex function, where $\mathcal{D} \subseteq \mathbb{R}^r$ is open and convex, and the set $A\mathcal{X} \triangleq \{Ax : x \in \mathcal{X}\} \subseteq \mathcal{D}$ is compact. Then there exists a vector $c \in \mathbb{R}^r$ such that every optimal solution to the linear counterpart* (10) *is also optimal for the original problem* (1).

Proof. Since f is upper semicontinuous and $A\mathcal{X}$ is compact, the problem $\max\limits_{y \in A\mathcal{X}} f(y)$ admits an optimal solution which is denoted by $\bar{y}$. Let $f_{\max} = f(\bar{y})$. Because f is quasi-concave and upper semicontinuous, the strict sublevel set $\mathcal{S} \triangleq \{y : f(y) < f_{\max}\}$ is an open convex set. Moreover, since $\bar{y} \notin \mathcal{S}$, one can deduce from the renowned hyperplane separating theorem that there exists a $c \in \mathbb{R}^r$ such that

$$\mathcal{S} \subseteq \{y : c^\top (y - \bar{y}) \leq 0\}.$$

Because $\mathcal{S}$ is open, the inclusion can be strengthened to

$$\mathcal{S} \subseteq \{y : c^\top (y - \bar{y}) < 0\}. \tag{22}$$

Now let y^* be any optimal solution to $\max\limits_{y \in A\mathcal{X}} c^\top y$ which exists because $A\mathcal{X}$ is compact. Then $c^\top y^* \geq c^\top \bar{y}$, implying $y^* \notin \mathcal{S}$ by (22). Hence, by the definition of $\mathcal{S}$, one has $y^* \in \operatorname*{argmax}\limits_{y \in A\mathcal{X}} f(y)$. Finally, any optimal solution x^* for (10) leads to an optimal solution $Ax^* \in \operatorname*{argmax}\limits_{y \in A\mathcal{X}} f(y)$. $\qquad\square$

Because the concavity is only used in Proposition 5 to derive the results in Section 3 and 4, Proposition 8 immediately implies the following remark.

Remark 1. The theoretical results established in Section 3 and 4 continue to hold for a quasi-convex objective f, provided that the conditions of Proposition 8 are satisfied.

We emphasize that that the upper semicontinuity and compactness assumptions imposed in Proposition 8 are necessary, which we illustrate in Example 4 and 5 below. Therefore, Proposition 8 does not subsume Proposition 5.

Example 4. Let $A = I$, $n = 1$, $\mathcal{X} = \mathbb{R}$, and $f(x) = \min\{|x|, 1\}$. One can verify that the maximizers of f are $(-\infty, -1] \cup [1, \infty)$ and are thus unbounded. In contrast, for any nonzero $c \in \mathbb{R}$, the linear problem $\max\limits_{x \in \mathcal{X}} c^\top x$ is unbounded and hence has no optimal solution. If $c = 0$, then every $x \in \mathcal{X}$ is optimal for the linear problem. Thus, without compactness of $A\mathcal{X}$, the conclusion of Proposition 8 may fail. $\qquad\square$

Example 5. Let $n = 2$ and $A = I$. Define

$$\mathcal{X} = \left\{(x_1, x_2) : 0 \leq x_1 \leq \sqrt{1 - x_2^2}\right\} \cup \left\{(x_1, x_2) : -1 \leq x_1, x_2 \leq 1\right\},$$

and

$$f(x_1, x_2) = \begin{cases} 1 & \text{if } x_2 > 1 \\ 0 & \text{if } x = (0, 1) \\ -1 & \text{otherwise.} \end{cases}$$

One can verify that f is quasi-convex and admits a unique maximizer $\bar{x} = (0,1)$ over $\mathcal{X}$. However, $\bar{x}$ is not an exposed point of $\mathcal{X}$ (see Figure 5). As a result, no linear objective can single out $\bar{x}$. $\square$
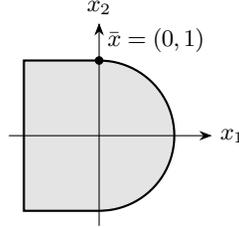

Fig. 5: The optimal solution cannot be exposed in $\mathcal{X}$.

5.2   Affine restriction of binary comonotone sets

In this subsection, we no longer assume that $\mathcal{X}$ is itself a comonotone set. Instead, we study the case where $\mathcal{X} = \mathcal{S} \cap \mathcal{P}$ satisfying Assumption 3 below.

Assumption 3 $\operatorname{conv}(\mathcal{S} \cap \mathcal{P}) = \operatorname{conv}(\mathcal{S}) \cap \mathcal{P}$, where $\mathcal{S} \subseteq \{0,1\}^n$ is comonotone and $\mathcal{P} = \{x \in \mathbb{R}^n : Mx \leq b\}$ for some $M \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.

Assumption 3 holds for example in either of the following cases: (i) $\mathcal{X}$ is a matroid and $\mathcal{P}$ is a matroid polytope [23], or (ii) the affine constraints defining $\mathcal{P}$ are *facial* for $\mathcal{X}$; see [6, Section 3.1] for a formal definition and further discussion.

The additional constraints imposed by $\mathcal{P}$ can destroy the comonotone structure of $\mathcal{X}$. Proposition 9 shows that their effect can be absorbed into a modified linear objective over $\mathcal{S}$.

Proposition 9. *Suppose $\mathcal{X} = \mathcal{S} \cap \mathcal{P}$ satisfies Assumption 3. Then for every $v \in \mathbb{R}^n$, there exists $\gamma \in \mathbb{R}_+^m$ such that $\bar{x} \in \underset{x \in \mathcal{X}}{\operatorname{argmax}} \, v^\top x$ if and only if*

$$\bar{x} \in \mathcal{X} \cap \mathcal{P}_=(\gamma) \cap \left[\underset{x \in \mathcal{S}}{\operatorname{argmax}} \left(v - M^\top \gamma \right)^\top x \right], \tag{23}$$

where $\mathcal{P}_=(\gamma) \triangleq \{x \in \mathbb{R}^n : (Mx)_i = b_i \ \forall i \in [m] \ s.t. \ \gamma_i \neq 0\}$.

Proof. First, suppose that $\bar{x} \in \underset{x \in \mathcal{X}}{\operatorname{argmax}} \, v^\top x$, and denote the optimal value by $\bar{u}$. Then we also have that $\bar{u} = \underset{x \in \operatorname{conv}(\mathcal{X})}{\max} \, v^\top x$. By Assumption 3, $\operatorname{conv}(\mathcal{X}) = \operatorname{conv}(\mathcal{S}) \cap \mathcal{P}$, which implies that

$$\bar{u} = \underset{x \in \mathbb{R}^n}{\max} \, v^\top x$$
$$\text{s.t. } Mx \leq b \tag{24}$$
$$x \in \operatorname{conv}(\mathcal{S}).$$

Let $\gamma \in \mathbb{R}_+^m$ be the optimal Lagrangian multipliers associated with the constraints $Mx \leq b$ in (24). Define the Lagrangian $\ell(x) = (v - M^\top \gamma)^\top x + b^\top \gamma$. It follows that

$$\bar{u} = \max_{x \in \text{conv}(\mathcal{S})} \ell(x) = \max_{x \in \mathcal{S}} \ell(x),$$

implying $x \in \underset{x \in \mathcal{S}}{\arg\max}\, \ell(x)$. Moreover, complementary slackness implies that $\bar{x} \in \mathcal{P}_=(\gamma)$. Together with $\bar{x} \in \mathcal{X}$, this yields (23).

We now prove the converse. Suppose that $\bar{x}$ satisfies (23). Since $\bar{x} \in \mathcal{P}_=(\gamma)$ implies $\gamma^\top (Mx - b) = 0$, we have $\ell(\bar{x}) = v^\top \bar{x}$. Taking any $x \in \mathcal{X} \subseteq \mathcal{S}$, one can deduce from (23) that

$$\ell(\bar{x}) \geq \ell(x) = v^\top x - \gamma^\top (Mx - b) \geq v^\top x,$$

where the last inequality is due to $Mx \leq b$ and $\gamma \geq 0$. Because $\bar{x} \in \mathcal{X}$ is feasible, combining these relations proves that $\bar{x}$ is optimal for $\max_{x \in \mathcal{X}} v^\top x$. $\square$

Next, we consider a setting where in addition to Assumption 3, we assume $\mathcal{S}$ is standard comonotone. With this additional structure, Proposition 9 enables us to reduce the analysis to linear optimization over $\mathcal{S}$ and to derive the corresponding complexity results. We present another application of Proposition 9 in Section 6.4.

As in the previous sections, we begin by introducing a core set that captures the optimality conditions for cost vectors in a hyperplane arrangement cell to be defined later.

Definition 4. *For any $c \in \mathbb{R}^r$, $\lambda \in \mathbb{R}$ and $\gamma \in \mathbb{R}^m$, define the set*

$$\mathcal{Q}^{Aff}(c, \lambda, \gamma) \triangleq \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{l} x_i = 0 \text{ if } \left(A^\top c - M^\top \gamma\right)_i > \lambda \\ x_i = 1 \text{ if } \left(A^\top c - M^\top \gamma\right)_i < \lambda \end{array} \right\}.$$

We denote by LP the worst-case time for both minimizing and maximizing $e^\top x$ over $\mathcal{Q}^{\text{Aff}}(c, \lambda, \gamma) \cap \mathcal{X} \cap \mathcal{P}_=(\gamma)$.

Lemma 8. *Suppose $\mathcal{X} = \mathcal{S} \cap \mathcal{P}$ satisfies Assumption 3 and $\mathcal{S}$ is standard comonotone. Then for all $c \in \mathbb{R}^r$, there exists $\gamma \in \mathbb{R}_+^m$ and $\lambda \in \mathbb{R}_+$ such that the following holds:*

(1) If $\bar{x} \in \underset{x \in \mathcal{X}}{\arg\max}\, c^\top A x$, then $\bar{x}$ is either a minimizer or a maximizer of $e^\top x$ over $\mathcal{Q}^{Aff}(c, \lambda, \gamma) \cap \mathcal{X} \cap \mathcal{P}_=(\gamma)$.

(2) Conversely, if $x^{\max}$ and $x^{\min}$ are any maximizer and minimizer of $e^\top x$ over $\mathcal{Q}^{Aff}(c, \lambda, \gamma) \cap \mathcal{X} \cap \mathcal{P}_=(\gamma)$ respectively, then $\{x^{\max}, x^{\min}\}$ contains an optimal solution to $\max_{x \in \mathcal{X}} c^\top A x$.

Proof. Fix $c \in \mathbb{R}^r$ and let $\bar{x} \in \underset{x \in \mathcal{X}}{\arg\max}\, c^\top A x$. Applying Proposition 9 to $v = A^\top c$, one can deduce that there exists $\gamma \in \mathbb{R}_+^m$ such that $\bar{x}$ is optimal for

$$\max_{x \in \mathcal{S}} \left(A^\top c - M^\top \gamma\right)^\top x. \tag{25}$$

Moreover, since $\mathcal{S}$ is a binary standard comonotone set, Proposition 6 yields a threshold $\lambda \in \mathbb{R}$ such that $\bar{x} \in \mathcal{Q}^{\mathrm{Aff}}(c, \lambda, \gamma)$. Let $\mathcal{I}^= = \{i \in [n] : (A^\top c - M^\top \gamma)_i = \lambda\}$. Because for any $x \in \mathcal{S} \cap \mathcal{Q}^{\mathrm{Aff}}(c, \lambda, \gamma)$, the coordinates x_i are fixed for all $i \notin \mathcal{I}^=$, the problem (25) is equivalent to

$$\max_{x \in \mathcal{S} \cap \mathcal{Q}^{\mathrm{Aff}}} \lambda \sum_{i \in \mathcal{I}^=} x_i,$$

which in turn further amounts to

$$\max_{x \in \mathcal{S} \cap \mathcal{Q}^{\mathrm{Aff}}} \lambda e^\top x. \tag{26}$$

Consequently, $\bar{x}$ is an optimal solution to (26). Depending on the sign of λ and thanks to $\bar{x} \in \mathcal{X} \cap \mathcal{P}_=(\gamma)$, the first conclusion holds true.

To prove the second part, we assume WLOG that $\lambda \geq 0$ since the other case can be proved similarly. Let $x^{\max}$ be a maximizer of $e^\top x$ over $\mathcal{Q}^{\mathrm{Aff}} \cap \mathcal{X} \cap \mathcal{P}_=(\gamma)$. Then $x^{\max}$ is also optimal for (25) by the equivalence above. Consequently, we deduce from Proposition 9 and $x^{\max} \in \mathcal{X} \cap \mathcal{P}_=$ that $x^{\max} \in \operatorname*{argmax}_{x \in \mathcal{X}} c^\top A x$. □

We now state the resulting complexity bound.

Proposition 10. *Suppose $\mathcal{X} = \mathcal{S} \cap \mathcal{P}$ satisfies Assumption 3 and $\mathcal{S}$ is standard comonotone. Then problem (1) can be solved in time $\mathcal{O}((n+m)^{r+m} \cdot (\mathsf{T}_1 + \mathsf{LP}))$.*

Proof. Consider the hyperplane arrangement $\mathcal{H}$ defined by

$$H_i = \left\{(c, \lambda, \gamma) \in \mathbb{R}^{r+m+1} : (A^\top c - M^\top \gamma)_i - \lambda = 0\right\} \qquad \forall i \in [n]$$
$$H'_j = \left\{(c, \lambda, \gamma) \in \mathbb{R}^{r+m+1} : \gamma_j = 0\right\} \qquad \forall j \in [m].$$

Then by construction, for all tuples (c, λ, γ) lying in the same hyperplane arrangement cell, the induced linear optimization problems

$$\min / \max_{x \in \mathbb{R}^n} e^\top x$$
$$\text{s.t. } x \in \mathcal{Q}^{\mathrm{Aff}}(c, \lambda, \gamma) \cap \mathcal{X} \cap \mathcal{P}_=(\gamma) \tag{27}$$

share a common minimizer and a common maximizer.

On the other hand, Proposition 8 yields a cost vector $\bar{c}$ such that any optimal solution to

$$\max (A^\top \bar{c})^\top x$$
$$\text{s.t. } x \in \mathcal{X} = \mathcal{S} \cap \mathcal{P}$$

is optimal for the original problem (1). Combining this with Proposition 9 and Lemma 8, there exists $\bar{\gamma} \in \mathbb{R}^m$ and $\bar{\lambda} \in \mathbb{R}$ such that either the minimizer or the maximizer of (27) associated with $(\bar{c}, \bar{\lambda}, \bar{\gamma})$ is optimal for (1). Because $\mathcal{H}$ gives rise to a partition of the parameter space (c, λ, γ), collecting the minimizers and maximizers over all regions must include one optimal solution to (1). Hence, the total number of candidate solutions is at most $2|\mathcal{H}|$. The conclusion follows from Lemma 3 that $|\mathcal{H}| = \mathcal{O}((n+m)^{r+m})$. □

Proposition 10 shows that for almost binary standard comonotone feasible regions (i.e., with small m), problem (1) remains tractable.

6   Applications

The theoretical results developed in the previous sections are fairly general and are applicable to a broad range of problems, including the applications highlighted in Section 1. Rather than presenting a long catalogue of examples, many of which follow directly once cast as instances of (1) (for example, matroid-constrained instances), we focus in this section on SPCA-related applications. This family offers a clean and representative setting for illustrating the implications of our general theory. Beyond serving as an illustration, the analysis of these SPCA-related problems is of interest in its own right.

Throughout this section, we assume the sample covariance matrix in question admits a rank-r factorization $AA^\top \in \mathbb{R}^{n \times n}$ with $A \in \mathbb{R}^{n \times r}$. For each application, we either (i) match or improve the best-known computational complexity, or (ii) provide the first-known polynomial-time guarantee in fixed-rank settings. Table 1 summarizes our results and compares them with existing bounds.

6.1   SPCA with a single component and its nonnegative variant

Since the work of [37], principle component analysis (PCA) has been a widely-used tool for dimensionality reduction in statistics and machine learning, but its principal components typically involve all features, which can hinder interpretability and lead to unstable estimates. SPCA addresses these issues by restricting the number of features used in each component [38]. In contrast to PCA, which can be solved directly using eigenvalue decomposition, SPCA is NP-hard and even inapproximable in general [48]. However, in the fixed-rank setting, SPCA admits polynomial-time algorithms [15].

This subsection studies SPCA with a single component and its nonnegative variant. The general SPCA problem is treated separately in Section 6.3 because it requires a different analysis. Formally, SPCA with a single component is defined as:

$$\max_{x \in \mathbb{R}^n} \left\{ \left\| A^\top x \right\|_2^2 : \|x\|_2 = 1,\ \|x\|_0 \le s \right\}, \qquad \text{(Single SPCA)}$$

It is evident that Single SPCA fits within the framework of (1), with the objective $f(A^\top x) = \|A^\top x\|_2^2$. In this case, the rank of the function f coincides with the rank r of the covariance matrix $AA^\top$. Moreover, the feasible set of Single SPCA is permutation- and sign-invariant with respect to x, which makes Proposition 7 applicable.

As shown in [52] (see also the proof of Lemma 9 below), once the support of features is fixed, Single SPCA reduces to calculating the largest eigenvalue of an $s \times s$ principal submatrix of $AA^\top$, which can be computed in $\mathsf{T}_1 = \mathcal{O}(s^2)$ time [71]. Combining this result with Proposition 7, we obtain the following complexity bound for Single SPCA.

Proposition 11. *The following hold for Single SPCA:*

   (i) There exists a collection of $\mathcal{O}(n^r)$ candidate supports for Single SPCA, among which at least one support is optimal; and

  (ii) Single SPCA admits a polynomial-time algorithm with complexity $\mathcal{O}(n^r \cdot s^2)$.

Proposition 11 matches the best-known bound in [4] and improves on the $\mathcal{O}(n^{r^2+r})$ bound of [15].

Nonnegative SPCA   In many applications, it is natural to impose nonnegativity constraints on the loadings, leading to *nonnegative PCA* and its sparse variant. Two common modeling motivations are (i) to reflect physical or domain constraints where the latent direction is inherently nonnegative (e.g., intensities, concentrations, gene expression, metabolite abundances) [1, 72], and (ii) to avoid components that rely on positive–negative cancellations (i.e., contrast directions) [79]. By incorporating this additional structure, nonnegativity can often improve interpretability [2] and reduce the estimation error of underlying statistical models; see [53] and literature therein for more details.

Even without sparsity constraints, a nonnegative PCA problem remains NP-hard, as it includes the matrix copositivity testing problem as a special case, which is known to be NP-complete [57]. We show below that when the covariance matrix $AA^\top$ has a fixed rank, the nonnegative SPCA problem is polynomially solvable. Formally, Nonnegative SPCA (NN-SPCA) is defined as

$$\max_{x \in \mathbb{R}^n} \left\{ \left\| A^\top x \right\|_2^2 : x \geq 0, \|x\|_2^2 = 1, \|x\|_0 \leq s \right\}. \qquad \text{(NN-SPCA)}$$

We consider the fixed-support subproblem for NN-SPCA. Given a support set $\text{supp}(x) = S$, NN-SPCA reduces to

$$\begin{aligned}
\max_{x \in \mathbb{R}^n} \ & \|A^\top x\|_2^2 \\
\text{s.t. } & \|x\|_2^2 = 1 \\
& x_i > 0 \qquad \forall i \in S \\
& x_i = 0 \qquad \forall i \notin S.
\end{aligned} \qquad (28)$$

Because the feasible region is not closed, the optimal solution to (28) need not exist.

Lemma 9. *There exists a polynomial algorithm that produces a feasible solution $\bar{x}$ to NN-SPCA. Moreover, whenever (28) attains the optimum, the returned $\bar{x}$ is optimal for (28).*

Proof. Without loss of generality, we assume $S = [n]$; otherwise we can substitute out $x_i = 0 \ \forall i \notin S$ and apply the same argument to the corresponding principle submatrix of $AA^\top$. For any optimal solution $\bar{x} > 0$ to (28), it must satisfy the KKT conditions for smooth optimization problems involving open sets (see [49, Section 11.5]), implying that there exists $\lambda \in \mathbb{R}$ such that

$$\nabla_x L(\bar{x}, \lambda) = 0,$$

where the Lagrangian $L(x, \lambda) = \|A^\top x\|_2^2 - \lambda \|x\|_2^2$. Thus, one has

$$AA^\top \bar{x} = \lambda \bar{x} \text{ and } \|A^\top \bar{x}\|_2^2 = \lambda. \qquad (29)$$

For each eigenvalue λ of $AA^\top$, let $V(\lambda) \triangleq \{x \in \mathbb{R}^n : AA^\top x = \lambda x\}$ denote the associated eigenspace. Then by (29), solving (28) boils down to finding the largest eigenvalue λ such that $V(\lambda) \cap \mathbb{R}_{++}^n \neq \emptyset$.

To test whether $V(\lambda)$ contains a strictly positive vector, consider the linear program

$$
\begin{aligned}
v(\lambda) \triangleq \max_{t,x} \ & t \\
\text{s.t. } & x_i \geq t \ \forall i \in [n] \\
& \sum_{i=1}^{n} x_i = 1 \\
& x \in V(\lambda),
\end{aligned}
\tag{30}
$$

where the second constraint ensures that $v(\lambda)$ is finite. Because at optimality it holds $t = \min_{i \in [n]} x_i$, we have $v(\lambda) > 0$ if and only if $V(\lambda) \cap \mathbb{R}^n_{++} \neq \emptyset$. Let $x(\lambda)$ be the optimal solution to (30). If there exists an eigenvalue with $v(\lambda) > 0$, let $\bar{\lambda}$ be the largest such eigenvalue and return $\bar{x} = x(\bar{\lambda})/\|x(\bar{\lambda})\|_2$. Then $\bar{x}$ is feasible for NN-SPCA, and it is optimal for (28) whenever (28) attains an optimum. If $v(\lambda) \leq 0$ for all eigenvalues λ, then the optimal solution to (28) does not exist. In this case, we set $\bar{x}$ as the first coordinate vector, which is trivially feasible for NN-SPCA.

Since $AA^\top$ has at most n distinct eigenvalues and each instance of (30) is a linear program, the procedure runs in polynomial time. $\square$

In parallel to Proposition 11, we obtain the following complexity bound for NN-SPCA.

Proposition 12. *The following hold for NN-SPCA:*

(i) *There exists a collection of $\mathcal{O}(n^r)$ candidate supports for NN-SPCA, among which at least one support is optimal.*
(ii) *NN-SPCA admits a polynomial-time algorithm with complexity $\mathcal{O}\left(n^r(\mathsf{FS} + s^2)\right)$, where $\mathcal{O}(n^r) \cdot \mathsf{T}_1$, where T_1 is the running time of the fixed-support routine in Lemma 9.*

Proof. Because the feasible region of NN-SPCA is nonnegative and standard comonotone, one can use Lemma 9 in place of Assumption 1 and follow the same argument as in Proposition 11. $\square$

As a corollary of Proposition 12, we obtain the following result for nonnegative PCA (i.e., without the sparsity constraint).

Corollary 2. *The nonnegative PCA problem*

$$
\max_{x \in \mathbb{R}^n} \left\{ \left\|A^\top x\right\|_2^2 : x \geq 0, \ \|x\|_2^2 = 1 \right\}.
\tag{NN-PCA}
$$

can be solved in polynomial time when the rank of A is fixed.

The proof of Corollary 2 follows by setting $s = n$ in NN-SPCA.

6.2   Variable selection for two-sample tests

Two-Sample Tests (2ST) aim to determine whether two collections of samples are drawn from the same distribution, and they have found broad applications in bioinformatics, finance, healthcare, and machine learning [28, 73]. To enhance both statistical efficiency and interpretability, recently Wang et al. [75] proposed selecting a subset of informative variables to conduct 2ST based on the maximum mean discrepancy statistic. This leads to the optimization problem:

$$z^{2st} \triangleq \max_{x \in \mathbb{R}^n} \left\{ \|A^\top x\|_2^2 + a^\top x : \|x\|_2 = 1, \|x\|_0 \le s \right\}, \tag{2ST}$$

where $A \in \mathbb{R}^{n \times r}$.

It is evident that 2ST and Single SPCA share the same feasible set. In fact, Single SPCA is a special case of 2ST: the only difference is the additional linear term $a^\top x$ in the objective, which increases the the objective rank to $r + 1$. More importantly, this term breaks the eigenvalue-based structure of Single SPCA. As a result, existing complexity results (see, e.g., [4, 15]), which rely on eigenvalue properties, do not extend to 2ST. Our general analysis framework can easily accommodate this setting. Observe that for any fixed support S, the corresponding subproblem for 2ST is a trust-region-type subproblem, which can be solved by first computing an eigendecomposition of $(AA^\top)_{S,S}$ and then solving a one-dimensional secular equation; see [54] or Section 4.3 of [58] for details. Since the eigen-decomposition dominates the cost, the fixed-support subproblem can be solved in $\mathcal{O}(s^3)$ time.

As a direct application of Proposition 7, we obtain the first polynomial-time complexity bound for 2ST when r is fixed.

Proposition 13. *The following hold for 2ST:*

(i) *There exists a collection of $\mathcal{O}(n^{r+1})$ candidate supports for 2ST, among which at least one support is optimal.*

(ii) *2ST admits a polynomial-time algorithm with complexity $\mathcal{O}(n^{r+1}s^3)$.*

6.3   General SPCA

In this subsection, we study the general SPCA problem (see [16, 74]):

$$z^{spca} \triangleq \max_{U \in \mathbb{R}^{n \times d}} \left\{ \|A^\top U\|_F^2 : U^\top U = I_d, \|U\|_0 \le s \right\}, \tag{SPCA}$$

where the *row-sparsity* constraint $\|U\|_0 \le s$ enforces the matrix U to contain at most s nonzero rows. SPCA is also more specifically referred to as *Row-Sparse PCA* in the literature [16, 45].

Since the general SPCA problem does not admit a standard comonotone feasible region, the results in Section 4 cannot be invoked verbatim as in Single SPCA. Nevertheless, as shown in Li and Xie [45], SPCA can be reformulated as the following convex maximization problem over a binary permutation-invariant set:

$$z^{spca} = \max_{x \in \{0,1\}^n} \left\{ \left\| \sum_{i \in [n]} x_i a_i a_i^\top \right\|_{(d)} : \sum_{i \in [n]} x_i \le s \right\}, \tag{31}$$

where for each $i \in [n]$, $a_i \in \mathbb{R}^r$ represents the ith row vector of A, and $\|\cdot\|_{(d)}$ is a convex function given by the sum of the d largest eigenvalues of its matrix argument. Shishkin et al. [71] shows that $\|\cdot\|_{(d)}$ for an $s \times s$ symmetric matrix can be computed in time $\mathsf{T}_1 = \mathcal{O}(ds^2)$. Then since the feasible region is a binary set, (31) naturally satisfies Assumption 1. Moreover, since $\sum_{i \in [n]} x_i a_i a_i^\top \in \mathbb{R}^{r \times r}$ is symmetric and depends linearly on x, the objective in (31) is convex and has rank $(r^2 + r)/2$. Consequently, problem (31) falls into the nonnegative and standard comonotone setting with rank $(r^2 + r)/2$ and a fixed support size, and its complexity result immediately follows Proposition 6.

Proposition 14. *The following hold for SPCA:*

(i) *There exists a collection of $\mathcal{O}(n^{(r^2+r)/2})$ candidate supports for SPCA, among which at least one support is optimal; and*

(ii) *SPCA admits a polynomial-time algorithm with complexity $\mathcal{O}(n^{(r^2+r)/2} \cdot ds^2)$.*

Compared with the support complexity $\mathcal{O}(n^{\min\{d,r\}(r^2+r)})$ derived in [15], our new bound $\mathcal{O}(n^{(r^2+r)/2})$ in Proposition 14 is independent of the number of principal components d.

6.4   Disjoint SPCA

In this subsection, we study another variant of SPCA where the principal components are sparse and have disjoint supports. As originally proposed by [3], we consider

$$\max_{\substack{Z \in \{0,1\}^{n \times d}: \\ Ze \leq e,\, Z^\top e \leq s}} \ \max_{U \in \mathbb{R}^{n \times d}} \left\{ \left\|A^\top U\right\|_F^2 : U^\top U = I_d,\, U_{ij}(1 - Z_{ij}) = 0, \forall i \in [n], j \in [d] \right\},$$

(Disjoint SPCA)

where each column of the binary matrix variable $Z \in \{0,1\}^{n \times d}$ encodes the support of each principal component, and $s \in \mathbb{Z}_+^d$ specifies the sparsity budget for each component. The constraints $Ze \leq e$ model that the supports of components are disjoint. We note that Single SPCA is also a special case of Disjoint SPCA at $d = 1$. We first reformulate Disjoint SPCA as a binary optimization problem.

Lemma 10. *Disjoint SPCA is equivalent to*

$$\max_Z \ \sum_{j \in [d]} \lambda_{\max} \left(\sum_{i \in [n]} Z_{ij} a_i a_i^\top \right)$$

$$s.t. \ \sum_{j=1}^{d+1} Z_{ij} = 1 \qquad\qquad \forall i \in [n]$$

$$\sum_{i=1}^{n} Z_{ij} \leq s_j \qquad\qquad \forall j \in [d] \qquad (32)$$

$$Z \in \{0,1\}^{n \times (d+1)}.$$

Proof. In Disjoint SPCA, the matrix U has disjoint column supports, which simplifies the orthonormal constraint $U^\top U = I_d$ to $U_j^\top U_j = 1$ for all $j \in [d]$, where U_j is denotes the jth column of U. Consequently, the inner maximization problem over U decomposes into d independent subproblems. According to [46], the optimal value of the jth subproblem is

$$\max_{U_j \in \mathbb{R}^n} \left\{ \left\| A^\top U_j \right\|_F^2 : U_j^\top U_j = 1, U_{ij}(1 - Z_{ij}) = 0, \forall i \in [n] \right\} = \lambda_{\max} \left(\sum_{i \in [n]} Z_{ij} a_i a_i^\top \right).$$

Plugging these values into Disjoint SPCA yields the objective of (32). Finally, introducing a slack column $Z_{i,d+1} \in \{0,1\}$ for each $i \in [n]$ to indicate that index i is not assigned to any component converts the constraints $Ze \le e$ into $\sum_{j=1}^{d+1} Z_{ij} = 1$. This completes the proof. □

Because the feasible set of (32), denoted by $\mathcal{X}$, is generally not comonotone, the complexity results from Section 3 and 4 do not apply directly. Fortunately, we can invoke Proposition 9 to derive the optimality condition for (32).

Definition 5. *For any $V \in \mathbb{R}^{n \times (d+1)}$ and $\gamma \in \mathbb{R}^{d+1}$, define*

$$\mathcal{Q}^{DS}(V, \gamma) \triangleq \left\{ Z \in \mathcal{X} \middle| \begin{array}{l} Z_{ij} = 0 \quad \forall i \in [n], \; j \notin \mathrm{argmax}\{V_{ij'} - \gamma_{j'} : \; j' \in [d+1]\} \\ \sum_{i=1}^n Z_{ij} = s_j \; \text{if } \gamma_j > 0 \; \forall j \in [d] \end{array} \right\}.$$

Lemma 11. *For any $V \in \mathbb{R}^{n \times (d+1)}$, there exists a $\gamma \in \mathbb{R}_+^{d+1}$ with $\gamma_{d+1} = 0$ such that $\bar{Z}$ is an optimal solution to $\max_{Z \in \mathcal{X}} \langle V, Z \rangle$ if and only if $\bar{Z} \in \mathcal{Q}^{DS}(V, \gamma)$.*

Proof. We write $\mathcal{X} = \mathcal{S} \cap \mathcal{P}$, where

$$\mathcal{S} = \left\{ Z \in \{0,1\}^{n \times (d+1)} : \sum_{j=1}^{d+1} Z_{ij} = 1 \; \forall i \in [n] \right\}$$

is the set of bases of a transversal matroid, and

$$\mathcal{P} = \left\{ Z \in \mathbb{R}^{n \times (d+1)} : \sum_{i=1}^n Z_{ij} \le s_j \; \forall j \in [d] \right\}$$

is the corresponding transversal matroid polytope. The decomposition implies that $\mathrm{conv}(\mathcal{S} \cap \mathcal{P}) = \mathrm{conv}(\mathcal{S}) \cap \mathcal{P}$, and hence Assumption 3 holds for (32). By Proposition 9, there exists $\gamma \in \mathbb{R}_+^d$ such that $\bar{Z} \in \mathrm{argmax}_{Z \in \mathcal{X}} \langle V, Z \rangle$ if and only if (i) $\bar{Z} \in \mathcal{X}$, (ii) $\sum_{i=1}^n \bar{Z}_{ij} = s_j$ for all $j \in [d]$ with $\gamma_j > 0$, and (iii)

$$\bar{Z} \in \mathrm{argmax}_{Z \in \mathcal{S}} \langle V, Z \rangle - \gamma^\top (Z^\top e - s) = \mathrm{argmax}_{Z \in \mathcal{S}} \sum_{i=1}^n \left[\sum_{j=1}^{d+1} (V_{ij} - \gamma_j) Z_{ij} \right].$$

Since $\mathcal{S}$ imposes a $(d+1)$-choose-one constraint for each row i, condition (iii) amounts to requiring that $\bar{Z}_{ij}$ can be nonzero only for indices j attaining $\text{argmax}_{j' \in [d+1]}\{V_{ij'} - \gamma_{j'}\}$. Together with (i) and (ii), we obtain $\bar{Z} \in \mathcal{Q}^{DS}(V, \gamma)$. $\qquad \square$

We now state the complexity result for Disjoint SPCA.

Proposition 15. *Disjoint SPCA can be solved in $\mathcal{O}\left((n(d+1)^2)^{\frac{d(r^2+r+2)}{2}-1} dn^2\right)$ time.*

Proof. Note that the objective of (32) is a sum of d terms, where the jth term returns the largest eigenvalue of $\sum_{i \in [n]} Z_{ij} a_i a_i^\top$. Analogous to the objective function of (31), each term is convex in x and has rank $(r^2+r)/2$. Consequently, the objective function in (32) has rank $\tilde{r} \triangleq d(r^2+r)/2$ and can be written in the form $f(\mathcal{A}Z)$, where f is convex, and $\mathcal{A} : \mathbb{R}^{n \times d} \to \mathbb{R}^{\tilde{r}}$ is a linear operator. Let $\mathcal{A}^\top : \mathbb{R}^{\tilde{r}} \to \mathbb{R}^{n \times d}$ denote its adjoint.

Let $\mathcal{H}$ be the hyperplane arrangement induced by

$$H_{ij\ell} = \left\{(c, \gamma) \in \mathbb{R}^{\tilde{r}} \times \mathbb{R}^d : (\mathcal{A}^\top c)_{ij} + \gamma_j = (\mathcal{A}^\top c)_{i\ell} + \gamma_\ell\right\}, \quad \forall i \in [n], \ j \neq \ell \in [d+1]$$

$$H_j = \left\{(c, \gamma) \in \mathbb{R}^{\tilde{r}} \times \mathbb{R}^d : \gamma_j = 0\right\}, \qquad\qquad\qquad \forall j \in [d],$$

where we treat $\gamma_{d+1} = 0$ and $(\mathcal{A}^\top c)_{i,d+1} = 0 \ \forall i \in [n]$ as fixed constants for notational convenience. Then it follows from Lemma 3 that

$$|\mathcal{H}| = \mathcal{O}\left(\left(\frac{d(d+1)}{2}n + d\right)^{\tilde{r}+d-1}\right) = \mathcal{O}\left(\left(n(d+1)^2\right)^{\frac{d(r^2+r+2)}{2}-1}\right). \tag{33}$$

Combining Proposition 5 and Lemma 11, we obtain a certain $c \in \mathbb{R}^{\tilde{r}}$ and $\gamma \in \mathbb{R}^{\tilde{r}}$ such that $\mathcal{Q}^{DS}(\mathcal{A}^\top c, \gamma) \neq \emptyset$ and any member of $\mathcal{Q}^{DS}(\mathcal{A}^\top c, \gamma)$ is optimal for (32). Furthermore, by the same reasoning as in the proof of Lemma 4, the set $\mathcal{Q}^{DS}(\mathcal{A}^\top c, \gamma)$ is constant within each region of $\mathcal{H}$. Therefore, selecting one representative point from $\mathcal{Q}^{DS}(\mathcal{A}^\top c, \gamma)$ per region and collecting all such points yields a set of candidates that contains an optimal solution to (32).

To prove the complexity, it remains to bound the cost per region. We first note that the feasibility problem over $\mathcal{Q}^{DS}(\mathcal{A}^\top c, \gamma)$ is essentially a transshipment problem, which can be transformed into a maximum flow problem using standard techniques; see [68, Section 11.6]. Therefore, for each region, we can find a representative $Z \in \mathcal{Q}^{DS}(\mathcal{A}^\top c, \gamma)$ in $O(dn^2)$ time using the Ford-Fulkerson Algorithm. Second, for each solution candidate Z, evaluating the objective of (32) amounts to computing the largest eigenvalues of d symmetric matrices in $\mathbb{R}^{r \times r}$, which costs $\mathsf{T}_1 = \mathcal{O}(dr^2)$ in total. Together with (33), the overall running time is

$$|\mathcal{H}| \cdot (O(dn^2) + \mathcal{O}(dr^2)) = \mathcal{O}\left(\left(n(d+1)^2\right)^{\frac{d(r^2+r+2)}{2}-1} dn^2\right).$$

This completes the proof. $\qquad \square$

Compared with [15], Proposition 15 provides a simpler analysis and improves the complexity bound $\mathcal{O}\big((dn)^{d^2(r^2+r)/2}(dnr^2 + d^3n^5 \log n)\big)$ reported therein.

7   Conclusion

In this paper, we take a geometric view of the fixed-rank convex maximization problem (1). We demonstrate that when the feasible region satisfies comonotonicity, a generalized symmetry introduced via conjugacy, the problem is polynomially solvable. This theoretical framework unifies and strengthens existing complexity results in the literature. In the standard comonotone case, we further sharpen the complexity bounds by partitioning an appropriate lifted parameter space.

Beyond these theoretical contributions, our analysis also yields a considerably simpler hyperplane-arrangement-based algorithm with weaker dependence on the rank parameter than prior constructions in the literature. In particular, for standard comonotone sets, the dominant support search scales as $\mathcal{O}(n^r)$ rather than $\mathcal{O}\big(n^{r(r+1)/2}\big)$. The reduced rank sensitivity and the simpler construction pave the way for practical implementations for solving reasonable low-rank approximations or convex relaxations of (1). Building such implementations and further using low-rank surrogates to tackle the generic problem are left for future work.

References

[1] Allen, G.I., Maletić-Savatić, M.: Sparse non-negative generalized pca with applications to metabolomics. Bioinformatics 27(21), 3029–3035 (09 2011), https://doi.org/10.1093/bioinformatics/btr522

[2] Asteris, M., Papailiopoulos, D., Dimakis, A.: Nonnegative sparse pca with provable guarantees. In: International Conference on Machine Learning. pp. 1728–1736. PMLR (2014)

[3] Asteris, M., Papailiopoulos, D., Kyrillidis, A., Dimakis, A.G.: Sparse pca via bipartite matchings. Advances in Neural Information Processing Systems 28 (2015)

[4] Asteris, M., Papailiopoulos, D.S., Karystinos, G.N.: The sparse principal component of a constant-rank matrix. IEEE Transactions on Information Theory 60(4), 2281–2290 (2014)

[5] Atamtürk, A., Narayanan, V.: The submodular knapsack polytope. Discrete Optimization 6(4), 333–344 (2009)

[6] Balas, E.: Disjunctive programming. Springer (2018)

[7] Bector, C.: Some aspects of quasi-convex programming. ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik 50(7-9), 495–497 (1970)

[8] Benson, H.P.: A finite algorithm for concave minimization over a polyhedron. Naval Research Logistics Quarterly 32(1), 165–177 (1985)

[9] Benson, H.P.: Concave minimization: theory, applications and algorithms. In: Handbook of global optimization, pp. 43–148. Springer (1995)

[10] Bereanu, B.: Quasi-convexity, strictly quasi-convexity and pseudo-convexity of composite objective functions. Revue française d'automatique informatique recherche opérationnelle. Mathématique 6(R1), 15–26 (1972)

[11] Calinescu, G., Chekuri, C., Pal, M., Vondrák, J.: Maximizing a monotone submodular function subject to a matroid constraint. SIAM Journal on Computing 40(6), 1740–1766 (2011)

[12] Chandrasekaran, R., Aneja, Y., Nair, K.: Minimal cost-reliability ratio spanning tree. In: North-Holland Mathematics Studies, vol. 59, pp. 53–60. Elsevier (1981)

[13] De Loera, J.A., Hemmecke, R., Onn, S., Rothblum, U., Weismantel, R.: Convex integer maximization via graver bases. Journal of Pure and Applied Algebra 213(8), 1569–1577 (2009)

[14] De Loera, J.A., Hemmecke, R., Onn, S., Weismantel, R.: N-fold integer programming. Discrete Optimization 5(2), 231–241 (2008)

[15] Del Pia, A.: Sparse pca on fixed-rank matrices. Mathematical Programming 198(1), 139–157 (2023)

[16] Dey, S.S., Molinaro, M., Wang, G.: Solving sparse principal component analysis with global support. Mathematical Programming 199(1), 421–459 (2023)

[17] Deza, A., Manoussakis, G., Onn, S.: Primitive zonotopes. Discrete & Computational Geometry 60(1), 27–39 (2018)

[18] Deza, A., Pournin, L., Rakotonarivo, R.: The vertices of primitive zonotopes. Contemporary Mathematics 764, 71–81 (2021)

[19] Edelman, P.H., Jamison, R.E.: The theory of convex geometries. Geometriae dedicata 19(3), 247–270 (1985)

[20] Edelsbrunner, H.: Algorithms in combinatorial geometry, vol. 10. Springer Science & Business Media (1987)

[21] Edelsbrunner, H., O'Rourke, J., Seidel, R.: Constructing arrangements of lines and hyperplanes with applications. SIAM Journal on Computing 15(2), 341–363 (1986)

[22] Edelsbrunner, H., Seidel, R., Sharir, M.: On the zone theorem for hyperplane arrangements. SIAM Journal on Computing 22(2), 418–429 (1993)

[23] Edmonds, J.: Submodular functions, matroids, and certain polyhedra. In: Combinatorial Structures and Their Applications, pp. 69–87. Gordon and Breach (1970)

[24] Edmonds, J.: Matroids and the greedy algorithm. Mathematical programming 1(1), 127–136 (1971)

[25] Giannessi, F., Niccolucci, F.: Connections between nonlinear and integer programming problems. In: Symposia Mathematica. vol. 19, pp. 161–176. Academic Press New York (1976)

[26] Gómez, A., Neto, J.: Outlier detection in regression: conic quadratic formulations. INFORMS Journal on Computing (2025)

[27] Goyal, V., Ravi, R.: An fptas for minimizing a class of low-rank quasi-concave functions over a convex set. Operations Research Letters 41(2), 191–196 (2013)

[28] Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A.: A kernel two-sample test. The journal of machine learning research 13(1), 723–773 (2012)

[29] Han, S., Gómez, A.: Compact extended formulations for low-rank functions with indicator variables. Mathematics of Operations Research 50(3), 1992–2016 (2025)

[30] Hardy, G.H., Littlewood, J.E., Pólya, G., Pólya, G., et al.: Inequalities. Cambridge university press (1952)

[31] Hare, W.L., Lewis, A.S.: Identifying active constraints via partial smoothness and prox-regularity. Journal of Convex Analysis 11(2), 251–266 (2004)

[32] Hashizume, S., Fukushima, M., Katoh, N., Ibaraki, T.: Approximation algorithms for combinatorial fractional programming problems. Mathematical programming 37(3), 255–267 (1987)

[33] Hassin, R., Tamir, A.: Maximizing classes of two-parameter objectives over matroids. Mathematics of Operations Research 14(2), 362–375 (1989)

[34] Helman, P., Moret, B.M., Shapiro, H.D.: An exact characterization of greedy structures. SIAM Journal on Discrete Mathematics 6(2), 274–283 (1993)

[35] Horst, R., Tuy, H.: Global optimization: Deterministic approaches. Springer Science & Business Media (2013)

[36] Hössjer, O.: Exact computation of the least trimmed squares estimate in simple linear regression. Computational Statistics & Data Analysis 19(3), 265–282 (1995)

[37] Hotelling, H.: Analysis of a complex of statistical variables into principal components. Journal of educational psychology 24(6), 417 (1933)

[38] Jeffers, J.N.: Two case studies in the application of principal component analysis. Journal of the Royal Statistical Society: Series C (Applied Statistics) 16(3), 225–236 (1967)

[39] Kalra, S., Adnan, M., Taylor, G., Tizhoosh, H.R.: Learning permutation invariant representations using memory networks. In: European conference on computer vision. pp. 677–693. Springer (2020)

[40] Kim, J., Tawarmalani, M., Richard, J.P.P.: Convexification of permutation-invariant sets and an application to sparse principal component analysis. Mathematics of Operations Research 47(4), 2547–2584 (2022)

[41] Klouda, K.: An exact polynomial time algorithm for computing the least trimmed squares estimate. Computational Statistics & Data Analysis 84, 27–40 (2015)

[42] Korte, B., Lovász, L.: Greedoids-a structural framework for the greedy algorithm. In: Progress in combinatorial optimization, pp. 221–243. Elsevier (1984)

[43] Lee, J., Lee, Y., Kim, J., Kosiorek, A., Choi, S., Teh, Y.W.: Set transformer: A framework for attention-based permutation-invariant neural networks. In: International conference on machine learning. pp. 3744–3753. PMLR (2019)

[44] Li, Y., Xie, W.: On the partial convexification for low-rank spectral optimization: Rank bounds and algorithms. arXiv preprint arXiv:2305.07638 (2023)

[45] Li, Y., Xie, W.: Beyond symmetry: Best submatrix selection for the sparse truncated svd. Mathematical Programming 208(1), 1–50 (2024)

[46] Li, Y., Xie, W.: Exact and approximation algorithms for sparse principal component analysis. INFORMS Journal on Computing 37(3), 582–602 (2025)

[47] Lovász, L.: Matroid matching and some applications. Journal of Combinatorial Theory, Series B 28(2), 208–236 (1980)

[48] Magdon-Ismail, M.: Np-hardness and inapproximability of sparse pca. Information Processing Letters 126, 35–38 (2017)

[49] Mangasarian, O.L.: Nonlinear programming. SIAM (1994)

[50] Megiddo, N.: Combinatorial optimization with rational objective functions. In: Proceedings of the tenth annual ACM symposium on Theory of computing. pp. 1–12 (1978)

[51] Mittal, S., Schulz, A.S.: An fptas for optimizing a class of low-rank functions over a polytope. Mathematical Programming 141(1), 103–120 (2013)

[52] Moghaddam, B., Weiss, Y., Avidan, S.: Spectral bounds for sparse pca: Exact and greedy algorithms. Advances in neural information processing systems 18 (2005)

[53] Montanari, A., Richard, E.: Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. IEEE Transactions on Information Theory 62(3), 1458–1484 (2015)

[54] Moré, J.J., Sorensen, D.C.: Computing a trust region step. SIAM Journal on scientific and statistical computing 4(3), 553–572 (1983)

[55] Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: On the least trimmed squares estimator. Algorithmica 69(1), 148–183 (2014)

[56] Murota, K.: Discrete Convex Analysis. SIAM (2003)

[57] Murty, K.G., Kabadi, S.N.: Some np-complete problems in quadratic and nonlinear programming. Mathematical programming 39(2), 117–129 (1987)

[58] Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York, 2nd edn. (2006)

[59] Oberlin, C., Wright, S.J.: Active set identification in nonlinear programming. SIAM Journal on Optimization 17(2), 577–605 (2006)

[60] Onn, S.: Convex matroid optimization. SIAM Journal on Discrete Mathematics 17(2), 249–253 (2003)

[61] Onn, S.: Nonlinear discrete optimization. Zurich Lectures in Advanced Mathematics, European Mathematical Society p. 75 (2010)

[62] Onn, S., Rothblum, U.G.: Convex combinatorial optimization. Discrete & Computational Geometry 32, 549–566 (2004)

[63] Onn, S., Schulman, L.J.: The vector partition problem for convex objective functions. Mathematics of Operations Research 26(3), 583–590 (2001)

[64] Pardalos, P.M., Vavasis, S.A.: Quadratic programming with one negative eigenvalue is np-hard. Journal of Global optimization 1(1), 15–22 (1991)

[65] Raghavachari, M.: On connections between zero-one integer programming and concave programming under linear constraints. Operations Research 17(4), 680–684 (1969)

[66] Rockafellar, R.T.: Convex analysis. Princeton university press (1970)

[67] Rousseeuw, P.J.: Multivariate estimation with high breakdown point. Mathematical statistics and applications 8(283-297), 37 (1985)

[68] Schrijver, A., et al.: Combinatorial optimization: polyhedra and efficiency, vol. 24. Springer (2003)

[69] Scott, J.R., Geunes, J.: A normal fan projection algorithm for low-rank optimization. Mathematical Programming 209(1), 681–702 (2025)

[70] Shigeno, M., Saruwatari, Y., Matsui, T.: An algorithm for fractional assignment problems. Discrete Applied Mathematics 56(2-3), 333–343 (1995)

[71] Shishkin, S.L., Shalaginov, A., Bopardikar, S.D.: Fast approximate truncated svd. Numerical Linear Algebra with Applications 26(4), e2246 (2019)

[72] Sigg, C.D., Buhmann, J.M.: Expectation-maximization for sparse and non-negative pca. In: Proceedings of the 25th international conference on Machine learning. pp. 960–967 (2008)

[73] Stein, C.: A two-sample test for a linear hypothesis whose power is independent of the variance. The Annals of Mathematical Statistics 16(3), 243–258 (1945)

[74] Vu, V., Lei, J.: Minimax rates of estimation for sparse pca in high dimensions. In: Artificial intelligence and statistics. pp. 1278–1286. PMLR (2012)

[75] Wang, J., Dey, S.S., Xie, Y.: Variable selection for kernel two-sample tests. arXiv preprint arXiv:2302.07415 (2023)

[76] White, N.: Matroid applications. No. 40, Cambridge University Press (1992)

[77] Whitney, H.: On the abstract properties of linear dependence. American Journal of Mathematics 57(3), 509–533 (1935)

[78] Yu, Q., Küçükyavuz, S.: Strong valid inequalities for a class of concave submodular minimization problems under cardinality constraints. Mathematical Programming 201(1), 803–861 (2023)

[79] Zass, R., Shashua, A.: Nonnegative sparse pca. In: Schölkopf, B., Platt, J., Hoffman, T. (eds.) Advances in Neural Information Processing Systems. vol. 19. MIT Press (2006), https://proceedings.neurips.cc/paper_files/paper/2006/file/75b9b6dc7fe44437c6e0a69fd863dbab-Paper.pdf