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 arises in many applications but is generally NP-hard, even for low-rank objectives. This paper introduces a set of broadly applicable conditions that certify when such problems are polynomially solvable. Our main condition is a new property of the feasible set, which we term co-monotonicity. We show that this property holds for two important families: matroids and permutation-invariant sets. Under co-monotonicity and mild additional assumptions, we develop a geometric framework that generates polynomially many candidate solutions, one of which is optimal. This yields a polynomial-time algorithm. We further derive substantially sharper complexity bounds when the feasible set is permutation-invariant. Our framework recovers existing tractable instances and often improves their complexity. It also expands the frontier of tractability by providing the first polynomial-time guarantees for new applications.

1 Introduction

In this paper, we study convex maximization of the form:

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

where the matrix $A \in \mathbb{R}^{r\times n}$ has r linearly independent rows, $r \le n$, and the function $f : \mathbb{R}^r \to \mathbb{R}$ is convex. Without loss of generality, we assume that problem (1) has a global optimal solution. By definition, the integer r is exactly the *rank* of the objective f (see, e.g., [25, 28]).

Convex maximization problem (1) has long been a cornerstone of optimization, known for its interesting mathematical properties and broad applicability. Problem (1) is known to be computationally difficult, as it covers several fundamental $\mathcal{NP}$-hard optimization problems, such as zero-one integer programming [24, 62], bilinear programming [9], and difference-of-convex programming [33]; we refer the reader to the excellent survey by [10] for a complete discussion.

Despite its general intractability, this paper aims to study *when problem* (1) *becomes polynomially solvable*. Our approach is straightforward: if one can construct polynomially many solutions, one of which is optimal for problem (1), then the problem can be solved in polynomial time by enumerating these candidates. To this end, (i) we characterize novel conditions that identify tractable instances of problem (1); and (ii) leveraging these conditions, we propose a general theoretical framework to generate the ideal polynomial-size set of candidate solutions for (1).

1.1 A tractable subclass of problem (1)

In this subsection, we present structural conditions that single out a tractable subclass of problem (1), with particular emphasis on a novel property of the feasible set $\mathcal{X}$. We show in the rest of the paper that any instance of (1) satisfying these conditions admits a polynomial-time algorithm.

Throughout, we assume that the rank r is fixed, which is prevalent in prior studies of tractability for special examples of problem (1) (see, e.g., [15, 58]). To illustrate the significance of this assumption for tractability, consider the simple case $r = 1$. When $\mathcal{X}$ is compact, problem (1) admits an optimal solution x^* such that Ax^* is an extreme point of the convex hull of the set $\{Ax \in \mathbb{R}^r : x \in \mathcal{X}\}$ [11]. For $r = 1$, this convex hull is just a bounded interval with two extreme points. To solve (1), it suffices to evaluate only these two extreme points. This suggests that fixing the rank r may constrain the number of relevant candidate solutions.

Nevertheless, problem (1) remains $\mathcal{NP}$-hard even in the fixed-rank setting. For example, [61] proved that minimizing a concave quadratic function of rank two ($r = 2$) over a polytope is $\mathcal{NP}$-hard. For any hope in tackling problem (1) under a fixed rank, one needs to impose further conditions on the feasible set $\mathcal{X}$. To this end, we now introduce a key combinatorial property of $\mathcal{X}$ – *co-monotonicity*. Intuitively, co-monotonicity requires that whenever a linear objective is sorted according to a permutation, the linear optimization problem over $\mathcal{X}$ admits an optimal solution whose components follow a (perhaps different) permutation that depends only on the ordering pattern of the objective.

We propose *co-monotonicity* as a novel unifying lens for a structural phenomenon that arises across both discrete and continuous optimization, yet has not been explicitly recognized. In particular, we demonstrate that this property is satisfied by two fundamental classes of sets:

(i) *Matroids*. A matroid is a combinatorial structure that generalizes the notion of linear independence in matrices [72, 73]; and
(ii) *Permutation-invariant sets*. A set $\mathcal{X} \subseteq \mathbb{R}^n$ is permutation-invariant if it is symmetric, that is, any reordering of the entries of a vector within $\mathcal{X}$ is still contained in $\mathcal{X}$ [39]. As the demand for handling set-based data has grown, permutation invariance has become a cornerstone of modern machine learning, such as 3D shape recognition and image classification [38, 40, 43]. Furthermore, it plays a key role in classical optimization, from sparsity-constrained optimization to robust optimization (see, e.g., [39, 44, 74]), where the constraints exhibit symmetry.

Consequently, even under the co-monotonicity condition on $\mathcal{X}$, problem (1) remains highly expressive. It encompasses a wide range of problems in optimization, statistics, and machine learning, as exemplified in Section 6.

Besides, 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:*

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

Assumption 1 means that problem (1) is tractable once the support is fixed. Thus, *instead of constructing polynomially many candidate solutions, it suffices to generate polynomially many candidate supports*, each of which can be solved efficiently under Assumption 1. Consequently, our proof strategy centers on constructing supports for (1), where the co-monotonicity of $\mathcal{X}$ plays a key role. Table 1 summarizes the complexity bounds on the number of candidate supports for the co-monotone set and for its two important classes.

Note that the fixed-support set $\{x \in \mathcal{X} : \mathrm{supp}(x) = S\}$ is generally non-closed. Consequently, problem (2) may not be able to attain its optimum. More precisely, by Assumption 1 we mean that there exists a polynomial-time algorithm that either returns an optimal solution when one exists, or certifies that no optimal solution exists.

In the binary setting, given a subset S, the binary variable $x \in \mathcal{X}$ is uniquely determined by setting $x_i = 1$ for $i \in S$ and $x_i = 0$ for $i \notin S$. Thus, we remark

Remark 1. For $\mathcal{X} \subseteq \{0,1\}^n$, Assumption 1 reduces to evaluating the objective f.

Another important class of (1) where Assumption 1 is often valid arises in best subset selection problems, including sparse principal component analysis (SPCA) [35, 37] and variable selection for two-sample tests (2ST) [71]. In these problems, fixing the support is equivalent to specifying a subset of features or samples. This often reduces the original problem to a conventional and tractable optimization form.

1.2 Literature review

The tractability of special cases of (1) has been extensively investigated, particularly for fixed rank r. This line of research broadly splits into two categories based on the structure of $\mathcal{X}$: matroid-constrained optimization and sparsity-constrained optimization.

Matroid-constrained optimization. We begin by explaining a strategy used in the literature on matroids: extreme point enumeration. When $\mathcal{X} \subseteq \{0,1\}^n$ represents a matroid, (1) must admit an optimal solution x^* such that Ax^* is an extreme point of the convex hull of $\{Ax : x \in \mathcal{X}\}$ [11]. If the number of such extreme points is polynomially bounded, then (1) can be solved efficiently by enumeration. Researchers have sought to identify feasible sets that allow for efficient extreme point enumeration, among which matroids stand out as a widely-studied and successful case.

A key property of matroids is that any linear optimization over them is efficiently solvable by the greedy algorithm [23, 42]. This tractability has laid the foundation for developing various combinatorial and geometric techniques to enumerate relevant extreme points. When $\mathcal{X} \subseteq \{0,1\}^n$ represents a matroid and f is a rank-2 polynomial in (1), [31] demonstrated that the number of extreme points is at most $\mathcal{O}(n^2)$ by leveraging parametric linear programming. [6] later applied the same technique to uniform matroids and separable rank-2 functions. However, this parametric linear programming approach does not extend to higher ranks. For any fixed rank r, problem (1) over uniform matroids coincides with the shaped partition problem for two parts [36, 60]. [7] established a lower bound of $\Omega(n^{2\lfloor (r-1)/2 \rfloor})$ for the complexity of this partition problem.

[58, 59] are two papers that are most closely related to our work. The celebrated result of [58] shows that for a matroid $\mathcal{X} \subseteq \{0,1\}^n$, problem (1) admits a polynomial-time algorithm for any fixed rank r. Their proof idea is a reduction of (1) to solving polynomially many linear optimization problems over $\mathcal{X}$ whose optimal solutions cover all extreme points. The reduction leverages several results of zonotopes that represent the dual geometry of a Hyperplane Arrangement (HA) [13, 49]. Importantly, the number of such linear problems is bounded by $\mathcal{O}(n^{2(r-1)})$ [27, 58]. It is noted that each linear problem is greedily solvable based on the matroid property. Subsequently, [59] generalized this tractability beyond matroids to a broader class of sets, which they termed edge-guaranteed. This concept was specifically defined to facilitate the zonotope-based reduction method. By doing so, only a polynomial-time number of linear optimization counterparts over edge-guaranteed sets need to be solved. [59] also provided a solution approach for solving these linear problems.

Sparsity-constrained optimization. A complementary line of work investigates the tractability of continuous optimization in the fixed-rank setting, specifically focusing on SPCA and the linear Least Trimmed Squares (LTS) problem [64]. Both problems impose a sparsity (zero-norm) constraint, where the main challenge is to select a sized-k subset of features or samples from n candidates [15, 34, 45]. Once the subset is determined, SPCA and LTS reduce to the standard PCA and least-square estimation, both of which admit closed-form solutions. Rather than enumerating all $\binom{n}{k}$ possibilities, the literature has focused on characterizing polynomially many subsets, among which lies an optimal one. This reduces the original problem to evaluating a manageable set of subsets.

SPCA is an interpretable dimensionality reduction method that constructs principal components from a small subset of features [12, 16, 37]. Given an $n \times n$ covariance matrix of rank r, by leveraging eigenvalue properties and introducing the auxiliary angle vector, [5] demonstrated that only $\mathcal{O}(n^r)$ candidate subsets suffice for solving single-component SPCA on fixed-rank matrices. In another recent paper, [15] extended this idea by combining eigenvalue properties with the HA technique. He constructed $\mathcal{O}(n^{\min\{d,r\}(r^2+r)})$ subsets for general SPCA, where d denotes the number of principal components. Furthermore, [15] first derived the polynomial-time complexity of disjoint SPCA via a reduction to the maximum-profit integer circulation problem.

LTS is a robust statistical method that fits a linear model to a subset of k samples among n, thus mitigating the influence of $n-k$ potential outliers [63, 64]. For the single feature case, [34] demonstrated that it suffices to evaluate only $\mathcal{O}(n^2)$ subsets. Building on the plane sweep algorithm given by [21], [34] developed an exact $\mathcal{O}(n^2 \log n)$ algorithm to solve LTS to optimality. Later, [54] improved the complexity to $\mathcal{O}(n^2)$ for the single-feature case. In addition, they proposed a more general algorithm for LTS using topological plane sweep, which computes $\mathcal{O}(n^{p+2})$ candidate subsets with p denoting the number of features.

We conclude this literature review by highlighting two key research gaps. First, existing theory primarily focuses on specific special cases of (1) and lacks generalizability to the broader problem class studied in this paper. Second, the proposed exact algorithms in the literature, while polynomial-time in theory, are not directly implementable in practice. Bridging the gap between theoretical tractability and practical solution method remains an open challenge.

1.3 Contributions and outline

Below we list the main contributions and an outline of the remaining paper.

(i) In Section 2, we formalize the definition of the co-monotone set, and demonstrate that matroids and permutation-invariant sets form two canoical classes of co-monotone sets.

For the $n = 2$ case, we establish the necessary and sufficient geometric condition under which a set is co-monotone and the underlying mapping becomes identical mapping.

(ii) In Section 3, by leveraging the convexity of f and the comonotonicity of $\mathcal{X}$, we develop a general theoretical framework for problem (1) that produces $\mathcal{O}(n^{2r})$ supports, one of which matches the support of an optimal solution. For any fixed r, the bound $\mathcal{O}(n^{2r})$ is polynomial in n; together with Assumption 1, it yields a polynomial-time algorithm for (1) with complexity $\mathcal{O}(n^{2r-2} \cdot \mathsf{T}_2 + n^{2r} \cdot \mathsf{T}_1)$ that searches over these supports. This framework applies to any co-monotone feasible set and, in a unified manner, recovers the known tractability of convex matroid maximization and SPCA.

We also explicitly derive the polynomial-time complexity for the matroid case, which coincides with the known bound in [58];

(iii) In Section 4, we refine the support-generation framework for permutation-invariant sets, an important class of co-monotone sets. Specifically, we reduce the number of candidate supports from $\mathcal{O}(n^{2r})$ to $\mathcal{O}(n^{r+3})$, as detailed in Table 1.

Moreover, Section 4.3 shows that equipping permutation-invariant sets with additional sign conditions leads to further reductions in complexity bounds, as summarized in Table 2;

(iv) In Section 5, we broaden our tractability results and show that problem (1) can be still solvable in polynomial time when f is only quasi-convex and $\mathcal{X}$ is not comonotone; and

(v) Finally, Section 6 applies our results to several applications. We recover or improve existing complexity bounds for SPCA and its related from a novel perspective. We also establish, for the first time, the polynomial-time complexity for two additional application examples.

1.4 Notations and definitions

For a positive integer n, we let $\mathbb{R}^n$ and $\mathbb{R}^n_+$ denote the set of all the n-dimensional vectors and nonnegative vectors, respectively, let $[n] = \{1, 2, \cdots, n\}$, let Π_n denote

Table 1. The number of candidate supports for problem (1)

	Co-monotone set	Matroid	Permutation-invariant set
This paper	$\mathcal{O}(n^{2r})$	$\mathcal{O}(n^{2r-2})$	$\mathcal{O}(n^{r+3})$
The literature	–	$\mathcal{O}(n^{2r-2})$	–

the set of all permutations of $[n]$. We let e denote the all-ones vector, 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 $|x| = (|x_1|, \cdots, |x_n|)^\top$ 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 a subset S, 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 $\operatorname{conv}(\mathcal{D})$ its convex hull. 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 $x_{\pi(1)} \leq \cdots \leq x_{\pi(n)}$ 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 $(2, 1, 3)$ or by $(2, 3, 1)$. Given a permutation π, we define a set $\mathcal{Z}(\pi) = \{x \in \mathbb{R}^n : x_{\pi(1)} \leq \cdots \leq x_{\pi(n)}\}$ and its nonnegative counterpart $\mathcal{Z}^+(\pi) = \{x \in \mathbb{R}^n_+ : 0 \leq x_{\pi(1)} \leq \cdots \leq x_{\pi(n)}\}$.

2 Co-monotone sets

In this section, we formally introduce the notion of co-monotone sets and discuss their basic properties.

Definition 1. *A set $\mathcal{X} \subseteq \mathbb{R}^n$ is called* co-monotone *if there is a mapping $\mathcal{M} : \Pi_n \to \Pi_n$ such that for each permutation $\pi \in \Pi_n$ and each 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 $\mathcal{M}(\pi)$, that is,

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

If one can choose $\mathcal{M}$ to be the identity mapping, then $\mathcal{X}$ is called standard *co-monotone.*

Below provides a simple example to illustrate the concept.

Example 1. Let $\mathcal{X} = \{x \in \{0,1\}^n : e^\top x = k\}$ with $k \in [n]$. Then $\mathcal{X}$ is a standard co-monotone set. 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 $\mathcal{M}(\pi) = \pi$.

We note that a cost vector v can be sorted by multiple π in Definition 1 when v has ties (as in the extreme case $v = 0$). However, regardless of the specific π chosen, Definition 1 requires that there always exists an optimal solution of (6) that is sorted by $\mathcal{M}(\pi)$.

We begin by characterizing standard co-monotone sets in the planar case. While elementary, the two-dimensional setting provides a microcosm for understanding the geometric restrictions imposed by co-monotonicity. The following proposition establishes that in $\mathbb{R}^2$, a mild surjectivity condition is sufficient for standard co-monotonicity.

Proposition 1. *Assume $\mathcal{X} \subseteq \mathbb{R}^2$ is a compact co-monotone set under permutation mapping $\mathcal{M}$. If $\mathcal{M}$ is surjective, then $\mathcal{X}$ is standard co-monotone.*

Proof. Let $\pi = (1,2)$ and $\tilde{\pi} = (2,1)$. If $\mathcal{M}$ is the identity, the claim is immediate. Thus, we assume $\mathcal{M}(\pi) = \tilde{\pi}$ and $\mathcal{M}(\tilde{\pi}) = \pi$. It suffices to prove for every $v = (v_1, v_2)$ satisfying $v_1 \geq v_2$, one can find a vector $\widehat{x} \in \mathcal{S} \triangleq \operatorname{argmax}\{v^\top x : x \in \mathcal{X}\}$ such that $\widehat{x}_1 \geq \widehat{x}_2$.

Since $\mathcal{X} \subseteq \mathbb{R}^2$ is co-monotone under $\mathcal{M}$, there exists $\tilde{x} \in \mathcal{S}$ with $\tilde{x}_1 \leq \tilde{x}_2$. Let $\widehat{v} = (v_2, v_1) \in \mathcal{Z}(\pi)$. Then one has

$$\langle v, \tilde{x} \rangle - \langle \widehat{v}, \tilde{x} \rangle = (v_1 - v_2)(\tilde{x}_1 - \tilde{x}_2) \leq 0. \tag{3}$$

Moreover, applying co-monotonicity to $\widehat{v}$ which is sorted by $\tilde{\pi}$, one can deduce a vector $\widehat{x} \in \operatorname{argmax}\{\widehat{v}^\top x : x \in \mathcal{X}\}$ such that $\widehat{x} \in \mathcal{Z}(\pi)$, i.e., $\widehat{x}_1 \geq \widehat{x}_2$. This implies $\langle \widehat{v}, \widehat{x} \rangle \geq \langle \widehat{v}, \tilde{x} \rangle$. Together with (3), one has $\langle v, \tilde{x} \rangle \leq \langle \widehat{v}, \widehat{x} \rangle$. Because $\bar{x}$ is maximal for v, one deduce $\widehat{x} \in \mathcal{S}$ as well. The conclusion follows from $\widehat{x}_1 \geq \widehat{x}_2$. □

Proposition 1 implies that in two-dimensional case, co-monotonicity is always standard unless the set forces a fixed ordering, that is, either $\mathcal{X} \subseteq \mathcal{Z}(1,2)$ or $\mathcal{X} \subseteq \mathcal{Z}(2,1)$. We next provide a condition for verifying standard co-monotonicity using only two directions, akin to checking extreme rays in polyhedral theory.

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

Proof. The necessity condition directly follows from Definition 1 due to $v \in \mathcal{Z}(1,2) \cap \mathcal{Z}(2,1)$. To prove sufficiency, we take an arbitrary $v = (v_1, v_2)$ with $v_1 > v_2$. It remains to prove that there exists $\bar{x}$ with $\bar{x}_1 \geq \bar{x}_2$ which is optimal for the cost vector v. Take any $\tilde{x} \in \mathcal{X}$. Then if $\tilde{x}_1 \geq \tilde{x}_2$ then 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

$$v^\top \tilde{x} \leq \bar{v}^\top \tilde{x} \leq \bar{v}^\top \bar{x} \leq v^\top \bar{x}. \tag{4}$$

Indeed, the first inequality follows

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

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

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

Because $\tilde{x}$ is optimal for the linear objective v, (4) implies that $\bar{x}$ is also an optimal solution for v. The conclusion follows from $\bar{x}_1 \ge \bar{x}_2$. □

In $\mathbb{R}^2$, ordering information is encoded by a small number of "coarse" directions (such as the sum $x_1 + x_2$). Proposition 2 exploits this fact by reducing standard co-monotonicity to checking the existence of ordered optima for only a finite collection of linear objectives.

We next connect co-monotonicity to symmetry. Following [39], a set $\mathcal{X} \subseteq \mathbb{R}^n$ is called *permutation-invariant* if for every vector $x \in \mathcal{X}$ and every permutation $\pi \in \Pi_n$, the permuted vector $[x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)}]^\top$ also belongs to $\mathcal{X}$. Next, we demonstrate that permutation-invariant sets are standard co-monotone.

Lemma 1. *Any permutation-invariant set $\mathcal{X}$ is standard co-monotone.*

Proof. Let $v \in \mathbb{R}^n$ be sorted by $\pi \in \Pi_n$. Suppose that $\widehat{x}$ is an optimal solution to the problem $\max_{x\in\mathcal{X}} v^\top x$. Then, we construct a new vector x^* by re-ordering the entries of $\widehat{x}$ so that

$$x^*_{\pi(1)} \le \cdots \le x^*_{\pi(n)}.$$

The vector x^* still belongs to $\mathcal{X}$ due to permutation-invariance. By the rearrangement inequality [29], we get

$$v^\top \widehat{x} \le v^\top x^*,$$

implying x^* must also be optimal. Thus, we have $\mathcal{M}(\pi) = \pi$ in Definition 1. □

We do not expect that every standard co-monotone set is permutation invariant. However, we can show that if a centered ellipsoid in $\mathbb{R}^2$ is co-monotone, then it must be permutation invariant.

Example 2. Let $n = 2$ and $\mathcal{X} = \{x : \ x^\top Q x \le 1\}$, where $Q \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix. If $\mathcal{X}$ is co-monotone, then $\mathcal{X}$ is permutation invariant. Indeed, because 0 is an interior of $\mathcal{X}$, if $\mathcal{X}$ is co-monotone under a mapping $\mathcal{M} : \Pi_n \to \Pi_n$, then $\mathcal{M}$ must be surjective, which implies that $\mathcal{M}$ can be chosen to be the identity mapping. Consider any $v \in \mathbb{R}^n$ and $\bar{x} \in \operatorname{argmax}\{v^\top x : \ x \in \mathcal{X}\}$. A standard calculation yields that $\bar{x} = \frac{1}{\lambda} Q^{-1} v$, where $\lambda = \sqrt{v^\top Q^{-1} v}$. Therefore, $\mathcal{X}$ is co-monotone if and only if

$$[(Qx)_i - (Qx)_j]\,(x_i - x_j) \ge 0 \ \forall i \ne j,$$

As $n = 2$, this is equivalent to

$$(Q_{11} - Q_{12})x_1^2 + (Q_{11} + Q_{22} - 2Q_{12})x_1 x_2 + (Q_{22} - Q_{12})x_2^2 \ge 0 \ \forall x,$$

which implies $(Q_{11} + Q_{22} - 2Q_{12})^2 - 4(Q_{11} - Q_{12})(Q_{22} - Q_{12}) = (Q_{11} - Q_{22})^2 \le 0$. Therefore, $Q_{11} = Q_{22}$ and $\mathcal{X}$ is permutation invariant.

While permutation-invariant sets form a substantial subclass of co-monotone sets, the concept also captures combinatorial structures that lack full symmetry. An important source of non-standard co-monotone sets is structure closely related to greedy algorithms. Matroids provide a canonical instance of this kind. Let $M = ([n], \mathcal{B})$ be a matroid over $[n]$ with a collection of bases $\mathcal{B} \subseteq 2^{[n]}$ [73]. 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 M with the corresponding set of incidence vectors $\mathcal{X} \subseteq \{0,1\}^n$.

Lemma 2 ([23]). *If $\mathcal{X} \in \{0,1\}^n$ encodes a matroid or the set of bases of a matroid, then $\mathcal{X}$ is co-monotone. Moreover, the associated permutation mapping $\mathcal{M}$ can be accessed in $\mathcal{O}(n \log n)$ time.*

Proof. Assume $\mathcal{X}$ represents the bases of a matroid. Consider any vector $v \in \mathbb{R}^n$ sorted by $v_{\pi_1} \geq v_{\pi_2} \cdots \geq v_{\pi_n}$ for a $\pi \in \Pi_n$. 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

$$x^*_{\pi(i)} = 1 \text{ if and only if } [x^*_{\pi(1)}, \cdots, x^*_{\pi(i-1)}, 1, 0, \cdots, 0]^\top \in \mathcal{X}. \tag{5}$$

By [23], 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, greedy produces the same optimal solution x^*. This proves that a matroid is co-monotone. To be more specific, we assume WLOG that $\pi = (1, 2, \ldots, n)$ is the natural order. Define $\sigma = \mathcal{M}(\pi)$ 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 $i < j$. For permutation π other than the natural one, $\mathcal{M}(\pi)$ can be defined similarly.

Consider the case where $\mathcal{X}$ represents a matroid, i.e., the incidence vectors of independent sets. Let x^* be the vector given by (5). Then the solution obtained by setting $\bar{x}_i = x^*_i$ if $v_i > 0$ and $x_i = 0$ otherwise is optimal for $\max_{x \in \mathcal{X}} v^\top x$. Because x^* is binary, x^* and $\bar{x}$ share the same order according to above modification. This implies that $\mathcal{X}$ is co-monotone under $\mathcal{M}$. In both cases, the time complexity is $\mathcal{O}(n \log n)$, dominated by the sorting step. ☐

The connection between greedy solvability and set structure has been extensively studied beyond matroids in literature. Other notable binary generalizations include antimatroids [18], and greedoids with strong exchange axiom and matroid embedding structures [41, 32]. In the continuous setting, Edmonds' classical best-in greedy algorithm extends to polymatroids and base polyhedra of submodular functions [22]. In discrete analysis, M-convex sets serve as the integral analogue of these structures [55, Chapter 4]. One can readily verify the co-monotonicity of above greedy-related families using the same reasoning as in Lemma 2; we do not repeat the details here.

3 Polynomial-time solvability

In this section, we propose a general theoretical framework that identifies a polynomial number of candidate supports for (1) and demonstrates the existence of an optimal support among them. Under Assumption 1, (1) is tractable for any fixed support; therefore, evaluating only these candidates yields a polynomial-time algorithm. Notably, our analysis applies to any co-monotone set in (1) and unifies, as special cases, all tractability results surveyed in Section 1.2.

3.1 Complexity analysis: A novel theoretical framework

In this subsection, we reduce the search for an optimal solution to (1) to a polynomial number of candidate supports, which is established through a three-step framework outlined in Figure 1. Under Assumption 1, we show that this framework directly leads to a polynomial-time algorithm for (1).

Step 1. Leveraging the convexity of f, we reduce (1) to its *linear counterpart* (6), which maximizes a linear function $c^\top Ax$ over the original feasible set $\mathcal{X}$ for some $c \in \mathbb{R}^r$. As $\mathcal{X}$ is co-monotone, we derive an optimality condition for (6) that relies only on the vector $A^\top c$.

Step 2. Since c is unknown, we partition its space $\mathbb{R}^r$ into a polynomial number of regions within which the optimality condition for (6) is fixed.

Step 3. Finally, we show that each region in Step 2 admits a polynomial-sized set of possible supports. Thus, at least one of them is optimal to (1).

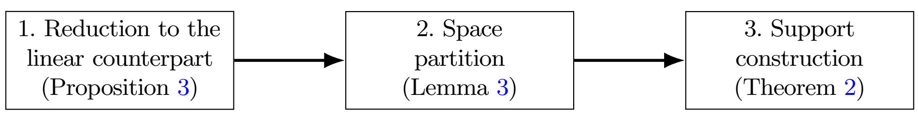

Fig. 1. A theoretical framework of complexity analysis.

We remark that the proposed framework has two key advantages. First, generality: it provides a systematic procedure that establishes polynomial-time solvability of (1) for any co-monotone feasible set $\mathcal{X}$. Second, adaptability: each step can be tailored to specific $\mathcal{X}$ (e.g., permutation-invariant sets) to tighten the complexity bounds. To avoid redundancy, we do not repeat the framework in the subsequent sections for special families of $\mathcal{X}$; instead, we present only the necessary changes.

Proposition 3. *There exists a vector $c \in \mathbb{R}^r$ such that*

$$\max_{x \in \mathcal{X}} c^\top Ax \tag{6}$$

admits an optimal solution, and every optimal solution to the linear counterpart (6) is also optimal for the original problem (1).

Proof. Let $\widehat{x} \in \mathcal{X}$ be an optimal solution to problem (1). Let $\widehat{c} \in \mathbb{R}^r$ be a (sub)gradient of $f(Ax)$ at $A\widehat{x}$. Then, we show that $\widehat{x}$ is also optimal for (6) with the vector $\widehat{c}$. Suppose, by contradiction, that there exists a solution $x^* \in \mathcal{X}$ satisfying

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

Then, we have that

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

where the first inequality results from the convexity of f. The strict inequality contradicts the optimality of $\widehat{x}$ in (1). Hence, $\widehat{x}$ must be optimal for (6) with $\widehat{c}$.

Furthermore, we show that any optimal solution of (6) with the vector $\widehat{c}$ is also optimal for (1). Let $\tilde{x} \in \mathcal{X}$ be an optimal solution of (6) with $\widehat{c}$, and thus $\widehat{c}^\top A\tilde{x} = \widehat{c}^\top A\widehat{x}$ holds. Then, applying the convexity 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. □

It is worth noting that Proposition 3 provides a strong result: each optimal solution of (6) for some c is also optimal for (1). Therefore, our analysis can focus on a single optimal solution of (6) to solve (1). Compared to (1), its linear counterpart (6) can be better controlled, as (6) enjoys an optimality condition derived from the co-monotonicity of $\mathcal{X}$.

Remark 2. Let $\mathcal{X}$ be a co-monotone set. Let $c \in \mathbb{R}^r$, and let $\pi \in \Pi_n$ be a permutation such that

$$\left(A^\top c\right)_{\pi(1)} \leq \cdots \leq \left(A^\top c\right)_{\pi(n)}.$$

That is, $A^\top c$ is sorted by π. Then, by Definition 1, the set $\mathcal{Z}(\mathcal{M}(\pi))$ must contain an optimal solution of (6). ◇

In this context, the permutation π depends on the cost vector $A^\top c$ in (6). We make two remarks about the set $\mathcal{Z}(\mathcal{M}(\pi))$:

(i) By definition, the set $\mathcal{Z}(\mathcal{M}(\pi))$ contains all vectors in $\mathbb{R}^n$ whose entries are sorted by $\mathcal{M}(\pi)$; and

(ii) For a given vector $c \in \mathbb{R}^r$, the permutation π may not be unique. As noted in Section 2, co-monotonicity allows us to fix an arbitrary π. The corresponding set $\mathcal{Z}(\pi)$ is then uniquely characterized and always contains an optimal solution of (6).

To prepare for partitioning the parameter space of $c \in \mathbb{R}^r$, we introduce a technique from discrete geometry–the *hyperplane arrangement Theorem*. 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 set $\mathcal{H}$ partitions the whole space $\mathbb{R}^q$ into *relatively open faces* (or *regions*) of dimensions from 0 to q. In what follows, we will use the terms *face* and *region* interchangeably. Each face can be characterized by $\cap_{i\in[p]}\tilde{H}_i$, where $\tilde{H}_i \in \{H_i, H_i^>, H_i^<\}$ represents either the hyperplane H_i or one of the two open half-spaces corresponding to H_i. The collection of these faces is called the *arrangement* $\mathcal{A}(\mathcal{H})$ of $\mathcal{H}$. For more details, we refer to [19, 20], with [1, Chapter 1] that provides illustrative examples of $\mathcal{A}(\mathcal{H})$ in $\mathbb{R}^2$.

Theorem 1 (Hyperplane arrangement Theorem [19]). *Let $\mathcal{H}$ be a finite set of p hyperplanes in $\mathbb{R}^q$. Then the arrangement $\mathcal{A}(\mathcal{H})$ consists of $\mathcal{O}(p^q)$ relatively open regions (or faces) and can be constructed in $\mathcal{O}(p^q)$ time. Furthermore, when all hyperplanes in $\mathcal{H}$ pass through the origin, this bound tightens to $\mathcal{O}((p-1)^{q-1})$.*

Lemma 3. *The space $\mathbb{R}^r$ can be partitioned into $\mathcal{O}(n^{2r-2})$ regions. For each region R, there exists a fixed permutation π such that, for all $c \in R$, the vector $A^\top c$ is sorted by π.*

Proof. For any pair $i, j \in [n]$ with $i < j$, define the hyperplane

$$H_{ij} = \left\{ c \in \mathbb{R}^r : \left(A^\top c\right)_i - \left(A^\top c\right)_j = 0 \right\}. \tag{7}$$

Let $\mathcal{H} = \{H_{ij}\}_{1 \le i < j \le n}$ denote these hyperplanes, whose number is $n(n-1)/2$. Each hyperplane leads to two open half-spaces: $H_{ij}^{>}$ and $H_{ij}^{<}$. According to Theorem 1, the arrangement $\mathcal{A}(\mathcal{H})$, which forms a partition of $\mathbb{R}^r$, consists of at most $\mathcal{O}(n^{2(r-1)})$ relatively open regions (or faces).

Each region is characterized by $\cap_{1 \le i < j \le n} \tilde{H}_{ij}$, where $\tilde{H}_{ij} \in \{H_{ij}, H_{ij}^{>}, H_{ij}^{<}\}$. Consequently, the relative ordering between any two components $\left(A^\top c\right)_i$ and $\left(A^\top c\right)_j$ is invariant within a region. This establishes the existence of a fixed permutation that sorts $A^\top c$ for all c in the same region. We thus conclude the proof. □

We make two remarks about Lemma 3:

- We characterize the optimality condition of (6) via the set $\mathcal{Z}(\pi|c)$ (see Remark 2); however, the parameter $c \in \mathbb{R}^r$ in (6) is unknown. Importantly, Lemma 3 establishes that, as c ranges over $\mathbb{R}^r$, only $\mathcal{O}(n^{2r-2})$ distinct sets $\mathcal{Z}(\pi|c)$ exist; and
- Hyperplane arrangements and its dual geometry are standard tools for space partition (see, e.g., [15, 58, 59]). Furthermore, in the next section, we introduce a family of hyperplanes in the extended space that yield improved bounds.

In the rest of this paper, we assume that the permutation mapping $\mathcal{M}$ in Definition 1 can be evaluated in polynomial time, which is denoted by T_2. Now, we are ready to show the main results.

Theorem 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 includes two parts.
Part I. By Lemma 3, we can partition $\mathbb{R}^r$ into regions $R_1, \cdots, R_K$ with $K = \mathcal{O}(n^{2r-2})$. In addition, for each $k \in [K]$, there exists a permutation $\pi^k \in \Pi_n$ that sorts $A^\top c$ for all $c \in R_k$. Combing this result with Remark 2, we obtain that

Claim. Let $k \in [K]$. For any $c \in R^k$, the set $\mathcal{Z}(\sigma^k)$ contains an optimal solution of (6), where $\sigma^k := \mathcal{M}(\pi^k)$.

According to Proposition 3, there exists a vector $c^* \in \mathbb{R}^r$ such that every optimal solution of (6) is also optimal for (1). According to Remark 2, (6) admits an optimal solution that belongs to the set $\mathcal{Z}(\mathcal{M}(\pi^*))$, where π^* can be an arbitrary permutation that sorts $A^\top c^*$. Since c^* must belong to some region R_ℓ with $\ell \in [K]$, we have

$$\pi^* = \pi^\ell, \text{ and } \mathcal{Z}(\mathcal{M}(\pi^*)) = \mathcal{Z}(\sigma^\ell).$$

As a result, the union $\bigcup_{k\in[K]} \mathcal{Z}(\sigma^k)$ contains an optimal solution to (1), which allows us to equivalently convert (1) into

$$z^* = \max_{x\in\mathcal{X}} \left\{ f(A^\top x) : x \in \bigcup_{k\in[K]} \mathcal{Z}(\sigma^k) \right\}. \tag{8}$$

Claim 1 *For any permutation $\sigma \in \Pi_n$, the set $\mathcal{Z}(\sigma)$ admits at most $\mathcal{O}(n^2)$ candidate supports.*

Proof. Since every vector $x \in \mathcal{Z}(\sigma)$ is sorted by σ, its zero entries (if any) appear consecutively and form a continuous 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 $[n]$.

If x contains zeros, there exist indices $1 \le t_1 \le t_2 \le n$ such that

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

It is evident that for any fixed pair (t_1, t_2), the support of x is uniquely identified. Since there are $\mathcal{O}(n^2)$ possible choices of (t_1, t_2), the claim follows. ⋄

According to Claim 1, there are at most $\mathcal{O}(n^2)$ possible supports within each set $\mathcal{Z}(\sigma^k)$. Since there are $\mathcal{O}(n^{2r-2})$ such sets in (8), the total number of candidate supports is at most $\mathcal{O}(n^{2r})$, denoted by $S_1, \cdots S_L$ with $|L| = \mathcal{O}(n^{2r})$. Thus, (8) can be reformulated as

$$z^* = \max_{\ell\in[L]} \max_{x\in\mathcal{X}} \left\{ f(A^\top x) : \mathrm{supp}(x) = S_\ell \right\}. \tag{9}$$

Part II. By Definition 1, computing the permutation σ^k for the region R_k takes T_2 for each $k \in [K]$. As shown in Part I, each region yields $\mathcal{O}(n^2)$ supports once the permutation is fixed. Therefore, the overall computational complexity of support generation is $\mathcal{O}(n^{2r-2} \cdot \mathsf{T}_2 + n^{2r})$. Under Assumption 1, each subproblem in (9) is solvable T_1 time. Accordingly, evaluating all candidate supports $\{S_\ell\}_{\ell\in[L]}$ requires

$$\mathcal{O}\left(n^{2r-2} \cdot \mathsf{T}_2 + n^{2r} \cdot \mathsf{T}_1\right)$$

running time. Since both T_1 and T_2 are polynomial, there is a polynomial-time algorithm that operates by evaluating $\{S_\ell\}_{\ell\in[L]}$. This completes the proof. □

We make the following remarks about Theorem 2.

- According to Theorem 2, (1) is polynomially solvable under Assumption 1. This result generalizes existing tractability analyses, which were limited to specific cases of (1) (e.g., [58, 15, 59]). For example, Theorem 2 establishes, for the first time, the polynomial-time solvability of 2ST with the fixed rank in Section 6; and
- The complexity bounds in Theorem 2, while general, are not necessarily tight for all problem classes of (1). We can further improve the complexity bound for specific structure of the feasible set $\mathcal{X}$, as shown in Corollary 1.

For the matroid setting, according to Remark 1 and Lemma 2, the results of Theorem 2 can direcly apply. By further exploiting the matroid structure, we reduce the number of candidate supports required for optimality from $\mathcal{O}(n^{2r})$ to $\mathcal{O}(n^{2r-2})$. Corollary 1 recovers the result of [58, Theorem 1.4] from a different perspective.

Corollary 1. *When $\mathcal{X} \in \{0,1\}^n$ encodes the bases of a matroid, the following hold:*

(i) There exists a collection of $\mathcal{O}(n^{2r-2})$ candidate supports for (1), among which at least one support is optimal; and
(ii) Problem (1) admits a polynomial-time algorithm with complexity $\mathcal{O}(n^{2r-1}\log n)$.

Proof. The best-in greedy algorithm (5) yields an optimal solution x^* to (6) in $\mathsf{T}_2 = \mathcal{O}(n \log n)$ time. In particular, the optimal solution x^* remains the same as long as the cost vector $A^\top c$ of (6) is sorted by the same permutation. Hence, it suffices to evaluate one support per region defined by the permutation of $A^\top c$. Combining this with the $\mathcal{O}(n^{2r-2})$ regions in Theorem 2, we obtain $\mathcal{O}(n^{2r-2})$ candidate supports for (1) that guarantee optimality. Then, the complexity in Part (ii) immediately follows from Remark 1 and Theorem 2. □

4 Improved complexity for permutation-invariant sets

In this section, we focus on permutation-invariant $\mathcal{X}$ for which $\mathsf{T}_2 = \mathcal{O}(1)$.

4.1 Optimality condition

In this subsection, we derive a new optimality condition for (6) tailored to permutation-invariant sets. This condition characterizes the sign pattern of an optimal solution by a thresholding rule, which compares each element of the cost vector $A^\top c$ of (6) against two threshold parameters.

To motivate the thresholding rule, consider the permutation-invariant set $\mathcal{X} = \{x \in [0,1]^n : e^\top x = k\}$. For this case, according to strong duality, there exists an optimal Lagrangian multiplier λ such that (6) 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, the above maximization problem admits an optimal solution x^* satisfying the thresholding rule:

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

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

This example clearly illustrates how a thresholding rule, parameterized by the dual variable λ, yields an optimality condition for (6). Such a rule can be extended to general permutation-invariant sets. It is expected that the resulting optimality condition becomes more complicated than in the illustrative example above. In the general setting, there are two threshold parameters, which arise from the permutation-invariant structure rather than directly using Lagrangian multipliers.

We now formalize the general optimality condition.

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

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

where

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

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

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

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

By convention, if $\underline{\lambda} > \overline{\lambda}$, then $\mathcal{Q}(c, \overline{\lambda}, \underline{\lambda}) = \emptyset$.

Lemma 4. *Suppose that $\mathcal{X}$ is permutation-invariant. Then, for any $c \in \mathbb{R}^r$, there exist parameters $\underline{\lambda} \le \overline{\lambda}$ such that $\mathcal{Q}\left(c, \overline{\lambda}, \underline{\lambda}\right)$ contains an optimal solution x^* of (6), i.e.,*

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

Proof. We begin by deriving two properties of the optimal solution x^* of (6).

Claim 2 *For any $i, j \in [n]$, if $\left(A^\top c\right)_i = \left(A^\top c\right)_j$, then swapping x_i^* and x_j^* remains optimal to (6).*

Proof. Let $\tilde{x}$ be the vector obtained by exchanging the i-th and j-th components of x^*. Since $\mathcal{X}$ is permutation-invariant, we have $\tilde{x} \in \mathcal{X}$. In addition, it is easy to verify that

$$c^\top A\tilde{x} - c^\top Ax^* = \left(\left(A^\top c\right)_j - \left(A^\top c\right)_i\right)(x_i^* - x_j^*) = 0.$$

Hence, $\tilde{x}$ is also optimal. ◇

For any $i < j$ with $\left(A^\top c\right)_i = \left(A^\top c\right)_j$, if $x_i^* < x_j^*$, we swap them such that $x_i^* \ge x_j^*$. Thus, for any $\underline{\lambda} \le \overline{\lambda}$, the updated vector x^* satisfies the ordering constraints in $\mathcal{Q}\left(c, \overline{\lambda}, \underline{\lambda}\right)$ while preserving optimality by Claim 2.

Claim 3 *For any $i, j \in [n]$, if $x_i^* > x_j^*$, then $\left(A^\top c\right)_i \ge \left(A^\top c\right)_j$ must hold.*

Proof. We prove the result by contradiction. Suppose $\left(A^\top c\right)_i < \left(A^\top c\right)_j$. By the permutation-invariance of $\mathcal{X}$, exchanging the components x_i^* and x_j^* of x^* yields another feasible vector $\tilde{x} \in \mathcal{X}$. The resulting change in the objective value is

$$c^\top A\tilde{x} - c^\top Ax^* = \left(\left(A^\top c\right)_j - \left(A^\top c\right)_i\right)(x_i^* - x_j^*) > 0,$$

which contradicts the optimality of x^*. Thus, the condition $\left(A^\top c\right)_i \geq \left(A^\top c\right)_j$ holds. ⋄

We partition the index set based on the signs of the components of x^*:

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

If $\mathcal{I}^0$ is nonempty, define the thresholds:

$$\overline{\lambda} = \max_{i\in\mathcal{I}^0} \left(A^\top c\right)_i, \quad \underline{\lambda} = \min_{i\in\mathcal{I}^0} \left(A^\top c\right)_i.$$

Then, we have $\overline{\lambda} \geq \underline{\lambda}$. The case where $\mathcal{I}^0$ is empty will be discussed separately.

Next, we show that the sign constraints in $\mathcal{Q}\left(c, \overline{\lambda}, \underline{\lambda}\right)$ are also satisfied by x^*.

(i) For any $i \in \mathcal{I}^+$ and $j \in \mathcal{I}^0$, given $x_i^* > 0 = x_j^*$, Claim 3 implies that $v_i \geq v_j$ must hold. Therefore, we have that

$$\left(A^\top c\right)_i \geq \max_{j\in\mathcal{I}^0} \left(A^\top c\right)_j = \overline{\lambda}$$

for all $i \in \mathcal{I}^+$.

(ii) By construction of $\overline{\lambda}$ and $\underline{\lambda}$, it is clear that $\underline{\lambda} \leq \left(A^\top c\right)_i \leq \overline{\lambda}$ for all $i \in \mathcal{I}^0$.

(iii) Analogous to Part (i), we can show that $\left(A^\top c\right)_i \leq \underline{\lambda}$ for all $i \in \mathcal{I}^-$.

Based on Parts (i)-(iii), we conclude that for any $i \in [n]$,

(I) If $\left(A^\top c\right)_i > \overline{\lambda}$, then $i \in \mathcal{I}^+$ must hold, which implies $x_i^* > 0$.

(II) If $\underline{\lambda} < \left(A^\top c\right)_i < \overline{\lambda}$, then $i \in \mathcal{I}^0$ must hold, which implies $x_i^* = 0$.

(III) If $\left(A^\top c\right)_i < \underline{\lambda}$, then $i \in \mathcal{I}^-$ must hold, which implies $x_i^* < 0$.

If $\mathcal{I}^0 = \emptyset$, then set

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

If $\mathcal{I}^- = \emptyset$, then set $\underline{\lambda} = \overline{\lambda}$. If $\mathcal{I}^+ = \emptyset$, then set $\overline{\lambda} = \underline{\lambda}$. Under these settings, the result follows immediately from the above analysis for the case of $\mathcal{I}^0 \neq \emptyset$. Hence, we complete the proof. □

We remark about Lemma 4 that

(i) The optimality condition in Lemma 4 is more straightforward than the one in Remark 2. It explicitly determines the sign pattern of some elements of an optimal solution, whereas Remark 2 only provides the permutation; and

(ii) In addition, the set $\mathcal{Q}(c, \underline{\lambda}, \overline{\lambda})$ relies only on a component-wise comparison between $A^\top c$ and thresholds $\underline{\lambda}, \overline{\lambda}$, which bypasses the need to determine a specific permutation of $A^\top c$ and compute the mapping $\mathcal{M}$ as in Remark 2.

Leveraging these advantages, we significantly reduce the computational complexity of (1) in the following subsections.

4.2 Extended space partition and improved complexity

In this subsection, we propose a novel partition of the extended parameter space $\mathbb{R}^r \times \mathbb{R} \times \mathbb{R}$ that corresponds to c, $\underline{\lambda}$, $\overline{\lambda}$, respectively.

Lemma 5. *The space $\mathbb{R}^r \times \mathbb{R} \times \mathbb{R}$ can be partitioned into $\mathcal{O}(n^{r+1})$ regions such that for all $(c, \underline{\lambda}, \overline{\lambda}) \in \mathbb{R}^r \times \mathbb{R} \times \mathbb{R}$ in the same region, the set $\mathcal{Q}\left(c, \overline{\lambda}, \underline{\lambda}\right)$ is constant.*

Proof. To begin, we define the following $2n + 1$ hyperplanes:

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

Let $\mathcal{H} = \{H_t\}_{t \in [2n+1]}$. According to Theorem 1, the arrangement $\mathcal{A}(\mathcal{H})$ divides the space $\mathbb{R}^r \times \mathbb{R} \times \mathbb{R}$ into at most $\mathcal{O}((2n)^{r+1}) = \mathcal{O}(n^{r+1})$ relatively open regions. Each region is defined as an intersection of the form

$$\bigcap_{t \in [2n+1]} \tilde{H}_t, \text{ where } \tilde{H}_t \in \left\{ H_t, H_t^{>}, H_t^{<} \right\}.$$

Within each region, where the choices of $\tilde{H}_t$ are fixed for all $t \in [2n + 1]$, 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 fixed. By Definition 2, the set $\mathcal{Q}(c, \overline{\lambda}, \underline{\lambda})$ depends only on these sign patterns. Consequently, it is invariant within each region. ☐

Compared to Lemma 3, Lemma 5 reduces the number of regions from $\mathcal{O}(n^{2r-2})$ to $\mathcal{O}(n^{r+1})$, which yields a complexity linear in r. This reduction directly leads to an improved complexity for (1), as we demonstrate below.

Theorem 3. *Suppose that $\mathcal{X}$ is permutation-invariant. Then*

(i) There exists a collection of $\mathcal{O}(n^{r+3})$ 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 $\mathcal{O}(n^{r+3} \cdot \mathsf{T}_1)$.

Proof. The proof includes two parts.

Part I. According to Lemma 5, the space $\mathbb{R}^r \times \mathbb{R} \times \mathbb{R}$ can be partitioned into finitely many regions $\{R_k\}_{k \in [K]}$ with $K = \mathcal{O}(n^{r+1})$. Moreover, Lemma 5 shows that for any $(c, \overline{\lambda}, \underline{\lambda}) \in R_k$, the set $\mathcal{Q}(c, \overline{\lambda}, \underline{\lambda})$ is identical. That is, the sign and ordering constraints that characterize $\mathcal{Q}(c, \overline{\lambda}, \underline{\lambda})$ in Definition 2 are fixed. 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 unchanged for any $(c, \overline{\lambda}, \underline{\lambda}) \in R_k$, where

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

These index sets are disjoint and satisfy

$$\mathcal{I}_k^+ \cup \mathcal{I}_k^0 \cup \mathcal{I}_k^- \cup \mathcal{I}_k^{=\overline{\lambda}} \cup \mathcal{I}_k^{=\underline{\lambda}} = [n].$$

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

$$\Upsilon^k := \left\{ x \in \mathbb{R}^n : \begin{array}{l} x_i > 0, \forall i \in \mathcal{I}_k^+, \quad x_i = 0, \forall i \in \mathcal{I}_k^0, \quad x_i < 0, \forall i \in \mathcal{I}_k^- \\ x_i \geq x_j, \forall i, j \in \mathcal{I}_k^{=\overline{\lambda}} \text{ with } i < j \text{ or } \forall i, j \in \mathcal{I}_k^{=\underline{\lambda}} \text{ with } i < j \end{array} \right\}. \quad (10)$$

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 3 with Lemma 4, we can find 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 some region, (1) can be equivalently converted into

$$z^* = \max_{x \in \mathcal{X}} \left\{ f(A^\top x) : x \in \bigcup_{k \in [K]} \Upsilon^k \right\}. \quad (11)$$

Next, we analyze the possible supports of all vectors in Υ^k. By (10), the sign of each x_i is fixed for any $i \in \mathcal{I}_k^+ \cup \mathcal{I}_k^0 \cup \mathcal{I}_k^-$. In contrast, the entries indexed by $\mathcal{I}_k^{=\overline{\lambda}} \cup \mathcal{I}_k^{=\underline{\lambda}}$ do not have specified signs, which results in various supports. Fortunately, their relative order is fixed. Since the set $\mathcal{I}_k^{=\overline{\lambda}} \cup \mathcal{I}_k^{=\underline{\lambda}}$ has cardinality at most n, it generates at most $\mathcal{O}(n^2)$ supports, as established in Claim 1. Combining this with $K = \mathcal{O}(n^{r+1})$, the total number of distinct supports for (11) is $\mathcal{O}(n^{r+3})$.

Part I. The proof is identical to that of Theorem 2 and thus omitted. □

4.3 Complexity under additional sign conditions

In this subsection, we show how additional sign properties lead to further complexity reductions for (1) over permutation-invariant sets. The corresponding complexity bounds are as summarized in Table 2. While the complexity analysis builds on that for general permutation-invariance, we refine it for each family by leveraging their specific properties.

We next discuss these special families in detail and analyze their perspective refinements.

Table 2. Candidate support complexity of (1) over permutation-invariant sets

	General	Nonnegative / Sign-invariant
General	$\mathcal{O}(n^{r+3})$	$\mathcal{O}(n^{r+1})$
Fixed support size	$\mathcal{O}(n^{r+2})$	$\mathcal{O}(n^{r})$

Nonnegative permutation-invariant sets When $\mathcal{X}$ is nonnegative permutation-invariant, any optimal solution of (6) contains no negative entries. Accordingly, the two thresholding parameters in Definition 2, which distinguish positive and negative entries of an optimal solution, are no longer both necessary. Indeed, it suffices to keep a single thresholding parameter associated with positive entries, which simplifies the set $\mathcal{Q}$ in Definition 2 as follows.

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

$$\mathcal{Q}^{nonneg}(c, \lambda) := \{x \in \mathbb{R}^n : \textit{(a) sign constraints, (b) ordering constraints}\},$$

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 $i < j$,

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

By adapting the analysis of Lemma 4 and dropping the parameter $\underline{\lambda}$ therein, we readily obtain the following result.

Remark 3. Suppose that $\mathcal{X}$ is nonnegative permutation-invariant. Then for any $c \in \mathbb{R}^r$, there exists a parameter λ such that at least one optimal solution of (6) lies in $\mathcal{Q}^{\text{nonneg}}(c, \lambda)$.

Remark 3 means that the simplified set $\mathcal{Q}^{\text{nonneg}}(c, \lambda)$ still characterizes the optimality condition of (6).

As a byproduct of $\mathcal{Q}^{\text{nonneg}}(c, \lambda)$, the extended space to be partitioned reduces from $\mathbb{R}^r \times \mathbb{R} \times \mathbb{R}$ to $\mathbb{R}^r \times \mathbb{R}$, which yields fewer regions than in Lemma 5. Specifically, we construct a collection of n hyperplanes, denoted by $\mathcal{H}^{\text{nonneg}} = \{H_i\}_{i \in [n]}$, where

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

Remark 4. The set $\mathcal{H}^{\text{nonneg}}$ partitions $\mathbb{R}^r \times \mathbb{R}$ into $\mathcal{O}(n^r)$ regions. Within each region, the set $\mathcal{Q}^{\text{nonneg}}(c, \lambda)$ is invariant.

Corollary 2. *Suppose that $\mathcal{X}$ is nonnegative and permutation-invariant. 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. By leveraging Remarks 3 and 4, we split $\mathbb{R}^r \times \mathbb{R}$ into $\mathcal{O}(n^r)$ regions, each associated with a fixed set $\mathcal{Q}^{\text{nonneg}}(c, \lambda)$. As shown in Theorem 3, the union of these sets contains an optimal solution of (1). Furthermore, the nonnegativity of each $\mathcal{Q}^{\text{nonneg}}(c, \lambda)$ leads to a sharper bound on the number of candidate supports. This improvement follows from a refinement of Claim 1, as shown below.

Claim 4 *For any permutation $\sigma \in \Pi_n$, the set $\mathcal{Z}^+(\sigma)$ admits at most $\mathcal{O}(n)$ candidate 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

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

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

Combining $\mathcal{O}(n)$ bound per region with the $\mathcal{O}(n^{2r-2})$ regions gives $\mathcal{O}(n^{2r-1})$ total candidate supports. We thus complete the proof. □

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

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

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, (6) can be written as

$$\max_{|x| \in \mathcal{X}} \left(\left|A^\top c\right|\right)^T |x| = \max_{x \in \mathcal{X} \cap \mathbb{R}^n_+} \left(\left|A^\top c\right|\right)^T x \tag{12}$$

Since the set $\mathcal{X} \cap \mathbb{R}^n_+$ in (12) is nonnegative and permutation-invariant, the analysis from the previous subsection can be partially applied. The main difference is that the objective of (12) involves absolute values of the cost vector. To characterize its optimality condition, it is natural to modify $\mathcal{Q}^{\text{nonneg}}(c, \lambda)$ by replacing $A^\top c$ and λ in Definition 3 with their absolute values. We denote the resulting set by $\mathcal{Q}^{\text{sign}}(c, \lambda)$ and omit its formal definition for brevity.

Accordingly, we revise the space partition approach to guarantee that the set $\mathcal{Q}^{\text{sign}}(c, \lambda)$ remains fixed within each regions. Specifically, we now construct $2n$ hyperplanes in $\mathbb{R}^r \times \mathbb{R}$ to achieve component-wise comparisons of absolute values. Define $\mathcal{H}^{\text{sign}} = \{H_i\}_{i \in [2n]}$, where

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

The set $\mathcal{H}^{\text{sign}}$ results in $\mathcal{O}((2n)^r) = \mathcal{O}(n^r)$ distinct sets of $\mathcal{Q}^{\text{sign}}(c, \lambda)$, which yields the same order of complexity as in Remark 4, but for this sign- and permutation-invariant case. Hence, Corollary 2 extends directly to give the following result.

Corollary 3. *Suppose that $\mathcal{X}$ is sign- and permutation-invariant. 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 $\mathcal{O}(n^{r+1} \cdot \mathsf{T}_1)$.

Fixed support size Inspired by practical applications (e.g., see Section 6), we are also interested in the setting where the support size of $\mathcal{X}$ is fixed to $s \in [n]$, which means that any solution in $\mathcal{X}$ must have exactly s nonzero entries. Note that fixed support size does not affect the optimality condition for (6); hence, hyperplane arrangements remain the same. This additional property refines the complexity analysis by reducing the number of candidate supports per region. For a given region, this number is of the same order as the number of candidate supports in either $\mathcal{Z}(\sigma)$ or $\mathcal{Z}^+(\sigma)$, depending on the structure of $\mathcal{X}$ (see Claim 1 and Claim 4). When the support size is fixed, we show that this order is reduced by one for both $\mathcal{Z}(\sigma)$ and $\mathcal{Z}^+(\sigma)$.

Claim 5 *For any permutation $\sigma \in \Pi_n$, the set $\mathcal{Z}(\sigma)$ with fixed support size contains at most at most $\mathcal{O}(n)$ supports, and the set $\mathcal{Z}^+(\sigma)$ with fixed support size contains at most at most $\mathcal{O}(1)$ supports.*

Proof. As shown in Claim 1, the set $\mathcal{Z}(\sigma)$ can generate up to $\mathcal{O}(n^2)$ candidate supports, since each support is determined by a pair of indices $1 \leq t_1 \leq t_2 \leq n$ therein. With a fixed support size s, we can replace t_2 with $t_1 + n - s - 1$, which thus reduce the number of possible supports to $\mathcal{O}(n)$.

For the nonnegative case, the set $\mathcal{Z}^+(\sigma)$ admits a unique support when the support size is fixed. □

Claim 5 enables an additional reduction in complexity, as summarized in the following principle.

Remark 5. For any specific family of $\mathcal{X}$ with fixed support size, the order of the number of candidate supports decreases by one.

5 Beyond Permutation-mapped Property and Convexity

In previous sections, we assume a finite convex objective function and a co-monotone 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 [14], 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 [25, 50, 66, 30]. 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 3, under additional regularity conditions, a quasi-convex maximization problem can be reduced to a nominal linear optimization problem.

Proposition 4. *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 (6) is also optimal for the original problem (1).*

Proof. Since f is upper semicontinuous and $A\mathcal{X}$ is compact, the problem $\max_{y \in A\mathcal{X}} f(y)$ 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

$$\mathcal{S} \subseteq \{y : c^\top (y - \bar{y}) < 0\}. \tag{13}$$

Now let y^* be any optimal solution to $\max_{y \in A\mathcal{X}} c^\top y$ which exists because $A\mathcal{X}$ is compact. Then $c^\top y^* \geq c^\top \bar{y}$, implying $y^* \notin \mathcal{S}$ by (13). Hence, by the definition of $\mathcal{S}$, one has $y^* \in \operatorname{argmax}_{y \in A\mathcal{X}} f(y)$. Finally, any optimal solution x^* for (6) leads to an optimal solution $Ax^* \in \operatorname{argmax}_{y \in A\mathcal{X}} f(y)$. □

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

Remark 6. The theoretical results established in Section 3 and 4 continue to hold for a quasi-convex objective f, provided that the conditions of Proposition 4 are satisfied.

We emphasize that that the upper semicontinuity and compactness assumptions imposed in Proposition 4 are necessary, which we illustrate in Example 3 and 4 below. Therefore, Proposition 4 does not subsume Proposition 3.

Example 3. 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 $\max_{x \in \mathcal{X}} c^\top x$ 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 4 may fail.

Example 4. 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 4). As a result, no linear objective can single out $\bar{x}$:

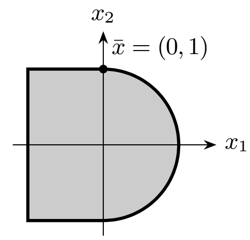

Fig. 2. The optimal solution cannot be exposed in $\mathcal{X}$.

5.2 Affine restriction of binary permutation invariant sets

In this subsection, we no longer assume that $\mathcal{X}$ is itself a co-monotone set. Instead, we study the case where $\mathcal{X} = \mathcal{S} \cap \mathcal{P}$ satisfying Assumption 2 below.

Assumption 2 $\mathrm{conv}(\mathcal{S} \cap \mathcal{P}) = \mathrm{conv}(\mathcal{S}) \cap \mathcal{P}$, *where* $\mathcal{S} \subseteq \{0,1\}^n$ *is co-monotone 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 2 holds for example in either of the following cases: (i) $\mathcal{X}$ is a matroid and $\mathcal{P}$ is a matroid polytope [22], or (ii) the affine constraints defining $\mathcal{P}$ are *facial* for $\mathcal{X}$; see [8, Section 3.1] for a formal definition and further discussion.

The additional constraints imposed by $\mathcal{P}$ can destroy the co-monotone structure of $\mathcal{X}$. Proposition 5 shows that their effect can be absorbed into a modified linear objective over $\mathcal{S}$.

Proposition 5. *Suppose* $\mathcal{X} = \mathcal{S} \cap \mathcal{P}$ *satisfies Assumption 2. Then for every* $v \in \mathbb{R}^n$, *there exists* $\gamma \in \mathbb{R}^m_+$ *such that* $\bar{x} \in \operatorname*{argmax}_{x \in \mathcal{X}} v^\top x$ *if and only if*

$$\bar{x} \in \mathcal{X} \cap \mathcal{P}_=(\gamma) \cap \left[\operatorname*{argmax}_{x \in \mathcal{S}} \left(v - M^\top \gamma \right)^\top x \right], \tag{14}$$

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 $\bar{x} \in \operatorname*{argmax}_{x \in \mathcal{X}} v^\top x$, and denote the optimal value by $\bar{u}$. Then we also have that $\bar{u} = \max_{x \in \mathrm{conv}(\mathcal{X})} v^\top x$. By Assumption 2, $\mathrm{conv}(\mathcal{X}) = \mathrm{conv}(\mathcal{S}) \cap \mathcal{P}$, which implies that

$$\begin{aligned} \bar{u} = \max_{x \in \mathbb{R}^n} \ & v^\top x \\ \text{s.t. } & Mx \leq b \\ & x \in \mathrm{conv}(\mathcal{S}). \end{aligned} \tag{15}$$

Let $\gamma \in \mathbb{R}^m_+$ be the optimal Lagrangian multipliers associated with the constraints $Mx \le b$ in (15). Define the Lagrangian $\ell(x) = (v - M^\top\gamma)^\top x + b^\top\gamma$. It follows that

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

implying $x \in \operatorname*{argmax}_{x \in \mathcal{S}} \ell(x)$. Moreover, complementary slackness implies that $\bar{x} \in \mathcal{P}_=(\gamma)$. Together with $\bar{x} \in \mathcal{X}$, this yields (14).

We now prove the converse. Suppose that $\bar{x}$ satisfies (14). 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 (14) that

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

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

Next, we consider a setting where in addition to Assumption 2, we assume $\mathcal{S}$ is permutation invariant. With this additional structure, Proposition 5 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 5 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 region 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 $\mathcal{Q}^{\mathrm{Aff}}(c, \lambda, \gamma) \cap \mathcal{X} \cap \mathcal{P}_=(\gamma)$.

Lemma 6. *Suppose $\mathcal{X} = \mathcal{S} \cap \mathcal{P}$ satisfies Assumption 2 and $\mathcal{S}$ is permutation invariant. 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 $\bar{x} \in \operatorname*{argmax}_{x \in \mathcal{X}} c^\top Ax$, 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 $\max_{x \in \mathcal{X}} c^\top Ax$.

Proof. Fix $c \in \mathbb{R}^r$ and let $\bar{x} \in \operatorname*{argmax}_{x \in \mathcal{X}} c^\top Ax$. Applying Proposition 5 to $v = A^\top c$, one can deduce that there exists $\gamma \in \mathbb{R}^m_+$ such that $\bar{x}$ is optimal for

$$\max_{x \in \mathcal{S}} \left(A^\top c - M^\top\gamma\right)^\top x. \tag{16}$$

Moreover, since $\mathcal{S}$ is a binary permutation invariant set, Remark 3 yields a threshold $\lambda \in \mathbb{R}$ such that $\bar{x} \in \mathcal{Q}^{\text{Aff}}(c, \lambda, \gamma)$. Let $\mathcal{I}^{=} = \{i \in [n] : (A^\top c - M^\top \gamma)_i = \lambda\}$. Because for any $x \in \mathcal{S} \cap \mathcal{Q}^{\text{Aff}}(c, \lambda, \gamma)$, the coordinates x_i are fixed for all $i \notin \mathcal{I}^{=}$, the problem (16) is equivalent to

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

which in turn further amounts to

$$\max_{x \in \mathcal{S} \cap \mathcal{Q}^{\text{Aff}}} \lambda e^\top x. \tag{17}$$

Consequently, $\bar{x}$ is an optimal solution to (17). 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 $\mathcal{Q}^{\text{Aff}} \cap \mathcal{X} \cap \mathcal{P}_{=}(\gamma)$. Then $x^{\max}$ is also optimal for (16) by the equivalence above. Consequently, we deduce from Proposition 5 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 6. *Suppose $\mathcal{X} = \mathcal{S} \cap \mathcal{P}$ satisfies Assumption 2 and $\mathcal{S}$ is permutation invariant. Then problem (1) can be solved in time $\mathcal{O}((n+m)^{r+m} \cdot (\texttt{T}_1 + \texttt{LP}))$.*

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

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

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

$$\begin{aligned} \min_{x \in \mathbb{R}^n} / \max \; & e^\top x \\ \text{s.t. } & x \in \mathcal{Q}^{\text{Aff}}(c, \lambda, \gamma) \cap \mathcal{X} \cap \mathcal{P}_{=}(\gamma) \end{aligned} \tag{18}$$

share a common minimizer and a common maximizer.

On the other hand, Proposition 4 yields a cost vector $\bar{c}$ such that any optimal solution to

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

is optimal for the original problem (1). Combining this with Proposition 5 and Lemma 6, there exists $\bar{\gamma} \in \mathbb{R}^m$ and $\bar{\lambda} \in \mathbb{R}$ such that either the minimizer or the maximizer of (18) 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 Theorem 1 that $|\mathcal{H}| = \mathcal{O}((n+m)^{r+m}$. □

6 Applications

In this section, we present several applications of (1), including SPCA, its variants, and variable selection for two-sample tests. Throughout this section, we assume the sample covariance matrix admits a rank-r factorization $AA^\top \in \mathbb{R}^{n\times n}$ with $A \in \mathbb{R}^{n\times r}$. We show that when r is fixed, these problems can be solved in polynomial time. For each application, we either (i) match or improve the best-known computational complexity, or (ii) provide the first-known polynomial-time guarantee. Table 3 summarizes our results and compares them with existing bounds.

Table 3. Number of candidate supports for application examples

	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	$\mathcal{O}(n^r)$	–	$\mathcal{O}\left(n^{\min\{d,r\}(r^2+r)}\right)$	$\mathcal{O}\left(n^{d^2(r^2+r)/2}\right)$	–

6.1 SPCA with a single component and its nonnegative variant

Since the work of [35], 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 [37]. In contrast to PCA, which can be solved directly using eigenvalue decomposition, SPCA is NP-hard and even inapproximable in general [47]. 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:

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

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 Corollary 3 applicable.

As shown in [51], once the support of features is fixed, Single SPCA reduces to a standard PCA problem on the corresponding principle submatrix of $AA^\top$. Leveraging this result, we next illustrate that candidate supports can be efficiently evaluated. Let $\{S_\ell\}_{\ell\in[L]}$ be the collection of feasible supports satisfying $|S_\ell| \le s$. Then, Single SPCA can be reformulated as

$$\begin{aligned} z^{spca} &= \max_{\ell\in[L]} \max_{x\in\mathbb{R}^n} \left\{ \left\|A^\top x\right\|_2^2 : \|x\|_2^2 = 1, \operatorname{supp}(x) = S_\ell \right\} \\ &= \max_{\ell\in[L]} \lambda_{\max}\left(\left(AA^\top\right)_{S_\ell,S_\ell} \right) = \max_{\ell\in[L]} \left\{ \lambda_{\max}\left(\left(AA^\top\right)_{S_\ell,S_\ell} \right) : |S_\ell| = s \right\}, \end{aligned} \qquad (19)$$

where the second equality follows from the closed-form solution of PCA [51], and the last equality follows from the monotonicity of the objective function in the subset S_ℓ. The reformulation (19) serves two purposes:

(i) It verifies Assumption 1 for Single SPCA: once the support is fixed, the subproblem reduces to computing the largest eigenvalue of an $s \times s$ matrix, which can be done in $\mathsf{T}_1 = \mathcal{O}(s^2)$ time [67].
(ii) It restricts the search to supports of size exactly s, which simplifies the support-construction procedure and reduces the overall complexity, as noted in Remark 5.

Combining Corollary 3 with the results above, we obtain the following complexity bound for Single SPCA.

Theorem 4. *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 $\mathcal{O}(n^r \cdot s^2)$.

Note that Theorem 4 matches the best-known bound in [5] and improves on the $\mathcal{O}(n^{r^2+r})$ bound of [15].

Nonnegative SPCA with a single component In many applications, it is natural to impose an entrywise nonnegativity constraint on the principal component, 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) [2, 68], and (ii) to avoid components that rely on positive–negative cancellations (i.e., contrast directions) [75]. By incorporating this additional structure, nonnegativity can often improve interpretability [3] and reduce the estimation error of underlying statistical models; see [52] 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 [56]. We show below that when the design matrix A has a fixed rank, the nonnegative SPCA problem is polynomially solvable. Formally, Nonnegative SPCA (NN-SPCA) is defined as

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

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

$$\begin{aligned} \max_{x \in \mathbb{R}^n}\ & \|A^\top x\|_2^2 \\ \text{s.t. }\ & \|x\|_2^2 = 1 \\ & x_i > 0 \qquad \forall i \in S \\ & x_i = 0 \qquad \forall i \notin S. \end{aligned} \tag{20}$$

Because the feasible region is not closed, the optimal solution to (20) need not exist.

Lemma 7. *There exists a polynomial algorithm that produces a feasible solution $\bar{x}$ to NN-SPCA. Moreover, whenever (20) attains an optimum, the returned $\bar{x}$ is optimal for (20).*

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 (20), it must satisfy the KKT conditions for smooth optimization problems involving open sets (see [48, 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

$$AA^\top \bar{x} = \lambda \bar{x} \text{ and } \|A^\top \bar{x}\|_2^2 = \lambda. \tag{21}$$

For each eigenvalue λ of $AA^\top$, let $V(\lambda) \triangleq \{x : AA^\top x = \lambda x\}$ denote the associated eigenspace. Then by (21), solving (20) 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

$$\begin{aligned} v(\lambda) \triangleq \max_{t,x}\ & t \\ \text{s.t. } & x_i \geq t\ \forall i \in [n] \\ & \sum_{i=1}^n x_i = 1 \\ & x \in V(\lambda), \end{aligned} \tag{22}$$

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 (22). 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 (20) whenever (20) attains an optimum. If $v(\lambda) \leq 0$ for all eigenvalues λ, then the optimal solution to (20) 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 (22) is a linear program, the procedure runs in polynomial time. □

In parallel to Theorem 4, we obtain the following complexity guarantee for NN-SPCA.

Theorem 5. *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 $\mathcal{O}\left(n^r(\mathsf{LP} + s^2)\right)$, where $\mathcal{O}\big(n^r(\mathsf{LP} + s^2)\big)$, where LP is the running time of the fixed-support routine in Lemma 7.

Proof. Because the feasible region of NN-SPCA is permutation invariant, one can use Lemma 7 in place of Assumption 1 and follow the same argument as in Theorem 4. □

As a corollary of Theorem 5, nonnegative PCA (i.e., without the sparsity constraint) is solvable in polynomial time when A has fixed rank, by setting $s = n$.

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 [26, 69]. To enhance both statistical efficiency and interpretability, recently Wang et al. [71] proposed selecting a subset of informative variables to conduct 2ST based on the maximum mean discrepancy statistic. This leads to the optimization problem:

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

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., [5, 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 eigen-decomposition of $(AA^\top)_{S,S}$ and then solving a one-dimensional secular equation; see [53] or Section 4.3 of [57] 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 Corollary 3, we obtain the first polynomial-time complexity bound for 2ST when r is fixed.

Theorem 6. *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 [17, 70]):

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

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 [17, 45].

In contrast to Single SPCA, the general SPCA problem does not admit a permutation- and sign-invariant feasible region; therefore, the results in Section 4 do not apply directly. By leveraging eigenvalue properties, [45] reformulated SPCA as the following convex maximization problem over a binary permutation-invariant set:

$$z^{spca} = \max_{x \in \{0,1\}^n} \left\{ \Big\| \sum_{i\in[n]} x_i a_i a_i^\top \Big\|_{(d)} : \sum_{i\in[n]} x_i = s \right\}, \tag{23}$$

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. [67] shows that $\|\cdot\|_{(d)}$ for an $s \times s$ symmetric matrix can be computed in time $\mathsf{T}_1 = \mathcal{O}(ds^2)$. Then, by Remark 1, (23) 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 (23) is convex and has rank $(r^2+r)/2$. Consequently, problem (23) falls into the sign- and permutation-invariant setting with rank $(r^2+r)/2$ and a fixed support size, and its complexity result immediately follows Corollary 2 and Remark 5.

Theorem 7. *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 $\mathcal{O}(n^{(r^2+r)/2} \cdot ds^2)$.

Importantly, Theorem 7 reduces the number of candidate supports from the bound $\mathcal{O}(n^{\min\{d,r\}(r^2+r)})$ in [15] to the significantly smaller $\mathcal{O}(n^{(r^2+r)/2})$. The theoretical reduction in complexity in turn increases the feasibility of developing practical, exact algorithms for SPCA.

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 [4], we consider

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

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 $Ze \le e$ models that the supports of components are disjoint. We note that Single SPCA is also a special case Disjoint SPCA at $d = 1$. We first reformulate Disjoint SPCA as a purely binary optimization problem.

Lemma 8. *Disjoint SPCA is equivalent to*

$$\begin{aligned}
\max_{Z}\ & \sum_{j\in[d]} \lambda_{\max}\left(\sum_{i\in[n]} Z_{ij} a_i a_i^\top\right) \\
s.t.\ & \sum_{j=1}^{d+1} Z_{ij} = 1 && \forall i \in [n] \\
& \sum_{i=1}^{n} Z_{ij} \le s_j && \forall j \in [d] \\
& Z \in \{0,1\}^{n\times(d+1)}.
\end{aligned} \tag{24}$$

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

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

Plugging these values into Disjoint SPCA yields the objective of (24). 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 (24), denoted by $\mathcal{X}$, is generally not co-monotone, the complexity results from previous sections do not apply directly. Instead, we invoke Proposition 5 to derive the optimality condition for (24).

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

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

Lemma 9. *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

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

is the corresponding transversal matroid polytope. The decomposition implies that $\text{conv}(\mathcal{S} \cap \mathcal{P}) = \text{conv}(\mathcal{S}) \cap \mathcal{P}$, and hence Assumption 2 holds for (24). By Proposition 5, there exists $\gamma \in \mathbb{R}^d_+$ such that $\bar{Z} \in \underset{Z\in\mathcal{X}}{\text{argmax}}\langle V, Z\rangle$ 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)

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

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), this is precisely $\bar{Z} \in \mathcal{Q}^{DS}(V, \gamma)$. □

We now state the complexity result for Disjoint SPCA.

Proposition 7. *Disjoint SPCA can be solved in* $\mathcal{O}\Big((n(d+1)^2)^{\frac{d(r^2+r+2)}{2}-1} dn^2 \Big)$ *time.*

Proof. Note that the objective of (24) 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 (23), each term is convex in x and has rank $(r^2+r)/2$. Consequently, the objective function in (24) 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

$$\begin{aligned} H_{ij\ell} &= \left\{ (c, \gamma) \in \mathbb{R}^{\tilde{r}} \times \mathbb{R}^d : \ (\mathcal{A}^\top c)_{ij} + \gamma_j = (\mathcal{A}^\top c)_{i\ell} + \gamma_\ell \right\} \quad i \in [n], \ j \ne \ell \in [d+1] \\ H_j &= \left\{ (c, \gamma) \in \mathbb{R}^{\tilde{r}} \times \mathbb{R}^d : \ \gamma_j = 0 \right\} \qquad j \in [d], \end{aligned}$$

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 Theorem 1 that

$$|\mathcal{H}| = \mathcal{O}\left(\left(\frac{d(d+1)}{2} n + d \right)^{\tilde{r}+d-1} \right) = \mathcal{O}\left(\left(n(d+1)^2 \right)^{\frac{d(r^2+r+2)}{2}-1} \right). \qquad (25)$$

Combining Proposition 3 and Lemma 9, we obtain a certain $c \in \mathbb{R}^{\tilde{r}}$ and $\gamma \in \mathbb{R}^{\tilde{r}}$ such that $\mathcal{Q}^{DS}(\mathcal{A}^\top c, \gamma) \ne \emptyset$ and any member of $\mathcal{Q}^{DS}(\mathcal{A}^\top c, \gamma)$ is optimal for (24). Furthermore, by the same reasoning as in the proof of Lemma 3, 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 (24).

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 [65, 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 (24) 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 (25), the overall running time is

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

This completes the proof. □

References

[1] Aguiar, M., Mahajan, S.: Topics in hyperplane arrangements, vol. 226. American Mathematical Soc. (2017)

[2] 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

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

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

[5] 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)

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

[7] Aviran, S., Onn, S.: Momentopes, the complexity of vector partitioning, and davenport—schinzel sequences. Discrete & Computational Geometry 27(3), 409–417 (2002)

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

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

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

[11] Bertsekas, D., Nedic, A., Ozdaglar, A.: Convex analysis and optimization, vol. 1. Athena Scientific (2003)

[12] Bertsimas, D., Cory-Wright, R., Pauphilet, J.: Solving large-scale sparse pca to certifiable (near) optimality. Journal of Machine Learning Research 23(13), 1–35 (2022)

[13] Brown, K.S., Diaconis, P.: Random walks and hyperplane arrangements. Annals of Probability pp. 1813–1854 (1998)

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

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

[16] Dey, S.S., Mazumder, R., Wang, G.: Using ℓ_1-relaxation and integer programming to obtain dual bounds for sparse pca. Operations Research 70(3), 1914–1932 (2022)

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

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

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

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

[21] Edelsbrunner, H., Souvaine, D.L.: Computing least median of squares regression lines and guided topological sweep. Journal of the American Statistical Association 85(409), 115–119 (1990)

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

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

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

[25] 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)

[26] 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)

[27] Gritzmann, P., Sturmfels, B.: Minkowski addition of polytopes: computational complexity and applications to gröbner bases. SIAM Journal on Discrete Mathematics 6(2), 246–269 (1993)

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

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

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

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

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

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

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

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

[36] Hwang, F.K., Onn, S., Rothblum, U.G.: A polynomial time algorithm for shaped partition problems. SIAM Journal on Optimization 10(1), 70–81 (1999)

[37] 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)

[38] 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)

[39] 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)
[40] Kimura, M., Shimizu, R., Hirakawa, Y., Goto, R., Saito, Y.: On permutation-invariant neural networks. arXiv preprint arXiv:2403.17410 (2024)
[41] Korte, B., Lovász, L.: Greedoids-a structural framework for the greedy algorithm. In: Progress in combinatorial optimization, pp. 221–243. Elsevier (1984)
[42] Lee, J.: A first course in combinatorial optimization, vol. 36. Cambridge University Press (2004)
[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] Magdon-Ismail, M.: Np-hardness and inapproximability of sparse pca. Information Processing Letters 126, 35–38 (2017)
[48] Mangasarian, O.L.: Nonlinear programming. SIAM (1994)
[49] McMullen, P.: On zonotopes. Transactions of the American Mathematical Society 159, 91–109 (1971)
[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] Moghaddam, B., Weiss, Y., Avidan, S.: Spectral bounds for sparse pca: Exact and greedy algorithms. Advances in neural information processing systems 18 (2005)
[52] 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)
[53] Moré, J.J., Sorensen, D.C.: Computing a trust region step. SIAM Journal on scientific and statistical computing 4(3), 553–572 (1983)
[54] 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)
[55] Murota, K.: Discrete Convex Analysis. SIAM (2003)
[56] Murty, K.G., Kabadi, S.N.: Some np-complete problems in quadratic and nonlinear programming. Mathematical programming 39(2), 117–129 (1987)
[57] Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York, 2nd edn. (2006)
[58] Onn, S.: Convex matroid optimization. SIAM Journal on Discrete Mathematics 17(2), 249–253 (2003)
[59] Onn, S., Rothblum, U.G.: Convex combinatorial optimization. Discrete & Computational Geometry 32, 549–566 (2004)
[60] Onn, S., Schulman, L.J.: The vector partition problem for convex objective functions. Mathematics of Operations Research 26(3), 583–590 (2001)

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

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

[63] Rousseeuw, P.J.: Least median of squares regression. Journal of the American statistical association 79(388), 871–880 (1984)

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

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

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

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

[68] 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)

[69] 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)

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

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

[72] White, N.: Theory of matroids. No. 26, Cambridge University Press (1986)

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

[74] Wu, C.F.: On the robustness and efficiency of some randomized designs. The Annals of Statistics pp. 1168–1177 (1981)

[75] 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