

SPECTRAL-GAUGE CUTS FOR SEMIDEFINITE PROGRAMMING

ANTONIO SASAKI*, SOPHIE DEMASSEY*, AND VALENTINA SESSA*

Abstract. We use symmetric gauge theory to develop a general class of cutting-plane algorithms for semidefinite programming. We formulate a separation problem based on spectral normalizations induced by gauges and derive a closed-form separation oracle. This oracle yields an implementable cut-generation procedure that, by varying the gauge, recovers standard cut families and generates new ones with tunable spectral structure. We embed the oracle within Kelley’s method and characterize convergence as a function of the chosen gauge and initial conic relaxation. Numerical experiments on small and large instances of box-constrained quadratic programming and sparse principal component analysis illustrate the versatility and performance of the proposed framework.

Key words. semidefinite programming, cutting-plane algorithm, symmetric gauges, sparse principal component analysis, box-constrained quadratic programming

MSC codes. 90C22, 90C25, 90C57, 15A18

1. Introduction. We consider semidefinite programs (SDP) in the form [36]

$$(1.1) \quad \begin{aligned} \min_{X \in \mathbb{S}^n} \quad & \langle C, X \rangle \\ \text{s.t.} \quad & X \in \mathcal{P}_{\text{lin}}, \\ & X \succeq 0, \end{aligned}$$

where $C \in \mathbb{S}^n$ and $\mathcal{P}_{\text{lin}} \subseteq \mathbb{S}^n$ is a polyhedron describing the linear side constraints. The space \mathbb{S}^n is equipped with the Frobenius inner product $\langle C, X \rangle := \text{tr}(C^\top X) = \sum_{i,j} C_{ij} X_{ij}$. The notation $X \succeq 0$ means that X is positive semidefinite (PSD), i.e., $v^\top X v \geq 0$ for all $v \in \mathbb{R}^n$. Since \mathcal{P}_{lin} is the intersection of finitely many affine hyperplanes and closed halfspaces in \mathbb{S}^n , the feasible set $\mathcal{F} = \mathcal{P}_{\text{lin}} \cap \mathbb{S}_+^n$ is a spectrahedron [29, 8], where $\mathbb{S}_+^n := \{X \in \mathbb{S}^n : X \succeq 0\}$ is the PSD cone.

Interior-point methods are among the most reliable generic algorithms for solving SDPs to high accuracy [24, 5]. Their scalability, however, is limited by the dense linear systems that arise at each iteration. In contrast with linear programming, exploiting sparsity in SDPs is often delicate and may require introducing many auxiliary variables [2]. First-order alternatives, including augmented Lagrangian methods [26, 23, 12] and the alternating direction method of multipliers [35, 4], avoid expensive Hessian operations and scale to larger instances. Their main limitation is reduced numerical precision, which can be problematic when tight dual bounds are needed in combinatorial or nonconvex optimization [14, 6, 22, 40].

A widely used alternative is Kelley’s cutting-plane approach [19], which replaces the PSD constraint $X \succeq 0$ with a tractable relaxation that is iteratively refined by separating valid inequalities, called *cuts* [21, 38, 7, 32]. This approach is particularly useful for convex, nonconvex, and discrete quadratic optimization, where SDP models appear as relaxations of the quadratic relation $X = x x^\top$ [33]. In this setting, a linear or conic cutting-plane method that generates a sequence of dual bounds can be kept computationally light by limiting the number of iterations and can be embedded effectively within global-optimization frameworks such as branch-and-bound [15]. A standard cutting-plane framework relies on the semi-infinite linear description

$$\mathbb{S}_+^n = \mathcal{P}_1 := \{X \in \mathbb{S}^n : v^\top X v \geq 0, \forall v \in \mathbb{R}^n, \|v\|_2 = 1\}.$$

*CMA, Mines Paris – PSL, France.

At a given iterate $\bar{X} \in \mathbb{S}^n$, separation amounts to finding a negative eigenvalue $\lambda \in \mathbb{R}$, typically the smallest one, together with an associated normalized eigenvector $v \in \mathbb{R}^n$ such that $v^\top \bar{X} v = \lambda < 0$. The resulting valid inequality $\langle vv^\top, X \rangle \geq 0$ is an *eigenvalue cut* (or *eigencut* for short). These cuts, introduced by Ramana [28], were also considered in [34] and in the reformulation-linearization technique [32], and have since been used regularly in global optimization [31, 15]. Adding many eigenvalue cuts to an initial linear relaxation of (1.1) can, however, lead to numerical instability in the master LP. To mitigate this issue, sparse eigenvalue cuts restrict the separation problem to eigenvectors with at most m nonzero entries [3, 27, 15].

Bertsimas and Cory-Wright [7] follow a complementary approach by seeking tighter outer approximations in SDP relaxations for machine-learning applications. They relate the empirical performance of cutting-plane methods to the strength of the initial relaxation and use second-order cone relaxations for sparse principal component analysis. They also introduce a deeper class of cuts for nuclear-norm minimization by exploiting the self-duality of the PSD cone

$$\mathbb{S}_+^n = \mathcal{P}_\infty := \{X \in \mathbb{S}^n : \langle S, X \rangle \geq 0, \forall S \in \mathbb{S}_+^n\}.$$

This identity yields a semi-infinite linear reformulation of (1.1) that extends eigenvalue cuts to inequalities of the form

$$(1.2) \quad \mathcal{C}_S := \langle S, X \rangle \geq 0,$$

with S of arbitrary rank. The eigenvalue inequalities defining \mathcal{P}_1 are the rank-one members of this family. The nuclear cuts of [7] can be seen as the most violated inequalities of this type when the separator is normalized in the spectral norm.

Our contributions. In line with the growing field of convex algebraic geometry [8], we use symmetric gauges to give a unified treatment of these cut families and of the continuum between them. Since the inequalities \mathcal{C}_S are invariant under positive scaling, the separator S can be restricted to a compact representative slice of \mathbb{S}_+^n . We study the case in which this slice is the unit ball of a unitarily invariant matrix norm. Such norms are induced by symmetric gauges ϕ , i.e., permutation- and sign-invariant vector norms, through the spectral mapping λ . Although the eigenvalue map is generally nonsmooth, unitarily invariant norms inherit useful variational and dual properties from their inducing gauges [13].

Our first contribution is to characterize the canonical separation problem over the unit ball of a unitarily invariant norm $\phi \circ \lambda$. We express the maximum violation in closed form as the image of the negative eigenvalues of the current iterate through the dual gauge ϕ° , and build the cut from the associated eigenvectors. This yields a practical cut-generation procedure for what we refer to as *spectral-gauge cuts*.

Our second contribution is a convergence analysis for Kelley’s method with these spectral-gauge cuts. The analysis uses two geometric quantities: the size of the initial relaxation and the depth of the cuts induced by the chosen gauge. The proof extends the arguments of [7] to arbitrary symmetric gauges and initial relaxations.

The ℓ_p norms provide a concrete hierarchy of spectral-gauge cut templates, from eigenvalue cuts for $\phi = \ell_1$, to nuclear cuts for $\phi = \ell_\infty$, and a continuum for $1 < p < \infty$, where the separation oracle essentially requires the computation of negative eigenvalues and an associated eigenbasis.

We then add two restrictions on the separators: by enforcing entrywise sparsity, we connect our framework to both sparse eigencuts [27, 3, 15] and the factor-width matrix cone hierarchy [9, 25]; by enforcing spectral sparsity with a rank limit k , we derive a lighter oracle only involving the k most negative eigenvalues.

Numerical experiments on small and large instances of box-constrained quadratic programming and sparse principal component analysis illustrate the practical impact of the framework under various configurations. Overall, LP relaxation combined with ℓ_1 -eigenvalue cuts or ℓ_2 -Frobenius cuts, with or without rank limit, strikes a good balance between bound quality and computational cost. However, no spectral-gauge cut template is strictly dominated by the others, and each can be selected for its respective strengths, depending on the class and size of a problem, or the iteration number allowed for computing the exact SDP bound or a cheaper dual bound.

Organization of the paper. Section 2 introduces notation and preliminary material on symmetric gauges and unitarily invariant norms. Section 3 presents the spectral-gauge separation problem and proves the closed-form separation theorem. Section 4 specializes the theorem to the ℓ_p hierarchy, sparse-support cuts, and rank-limit separators. Section 5 gives the generic cutting-plane algorithm and its convergence analysis. Section 6 reports numerical experiments.

2. Notation and preliminaries. Scalars and vectors are denoted by lowercase letters, and matrices by uppercase letters. Let $[n] := \{1, \dots, n\}$ for short. We use $t_+ := \max\{t, 0\}$ and $|t|$ for the positive part and absolute value of a scalar $t \in \mathbb{R}$. For a vector $u \in \mathbb{R}^n$, we define $u_+ \in \mathbb{R}_+^n$ and $|u| \in \mathbb{R}_+^n$ componentwise by $(u_+)_i = (u_i)_+$ and $|u|_i = |u_i|$. The support of u , denoted $\text{supp}(u)$, is the set of indices corresponding to nonzero components. By extension, we define the support of a symmetric matrix $X \in \mathbb{S}^n$ as the set of row indices with nonzero components. We denote by $e_i \in \mathbb{R}^n$ the i -th canonical vector. For $u \in \mathbb{R}^n$, let u^\uparrow and u^\downarrow denote the nondecreasing and nonincreasing rearrangements of u , respectively.

The *spectral mapping* λ maps a symmetric matrix to the vector of its eigenvalues, arranged in nondecreasing order and repeated according to multiplicity. For $X \in \mathbb{S}^n$, the positive and negative parts of its eigenvalue vector $\lambda(X) = (\lambda_i(X))_i$ by

$$\lambda_+(X) := (\lambda_i(X)_+)_i \in \mathbb{R}_+^n \text{ and } \lambda_-(X) := (-\lambda_i(X)_+)_i \in \mathbb{R}_+^n,$$

and the associated diagonal matrices in \mathbb{S}^n are $\Lambda(X) := \text{diag}(\lambda(X))$, $\Lambda_+(X) := \text{diag}(\lambda_+(X))$, and $\Lambda_-(X) := \text{diag}(\lambda_-(X))$. Throughout the paper, we denote the minimum eigenvalue of the matrix X with both $\lambda_1(X)$ and $\lambda_{\min}(X)$.

Since X is symmetric, there exists a real orthogonal matrix Q , i.e., $Q^\top Q = QQ^\top = I_n$, such that $X = Q\Lambda(X)Q^\top$. We define the *spectral positive part* $X_+ \in \mathbb{S}^n$ and *spectral negative part* $X_- \in \mathbb{S}^n$ as

$$(2.1) \quad X_+ := Q\Lambda_+(X)Q^\top, \quad X_- := Q\Lambda_-(X)Q^\top.$$

Then $X_+ \in \mathbb{S}_+^n$, $X_- \in \mathbb{S}_+^n$, $X = X_+ - X_-$, and $X_+X_- = 0$.

Since a symmetric matrix X is normal, its singular values, denoted by $\sigma(X)$ in nondecreasing order and counted with multiplicity, satisfy $\sigma(X) = |\lambda(X)|^\uparrow$ [18, Thm. 2.6.3], and Von Neumann's trace inequality [18, Thm. 7.4.1.1] reads:

$$(2.2) \quad |\langle A, B \rangle| = |\text{tr}(AB)| \leq (|\lambda(A)|^\uparrow)^\top (|\lambda(B)|^\uparrow) \quad \forall A, B \in \mathbb{S}^n.$$

Furthermore, $X \in \mathbb{S}^n$ is positive semidefinite if and only if all its eigenvalues are nonnegative, equivalently $\lambda_-(X) = 0$ and $\lambda(X) = |\lambda(X)|^\uparrow = \sigma(X)$ [18, Thm. 4.1.10].

DEFINITION 2.1 (Dual norm [18, Def. 5.4.12]). *Let ϕ be a norm on \mathbb{R}^n . The dual of ϕ is the function $\phi^\circ : \mathbb{R}^n \rightarrow \mathbb{R}_+$ defined by*

$$(2.3) \quad \phi^\circ(y) := \max_{z \in \mathbb{R}^n} \{ |\langle y, z \rangle| : \phi(z) \leq 1 \}, \quad \forall y \in \mathbb{R}^n.$$

The dual ϕ° defines a norm on \mathbb{R}^n that satisfies the generalized Cauchy-Schwarz inequality [18, Lem. 5.4.13]:

$$(2.4) \quad |y^\top z| \leq \phi^\circ(y) \phi(z) \text{ for all } y, z \in \mathbb{R}^n.$$

DEFINITION 2.2 (Symmetric gauge [18, Def. 7.4.7.1]). *Let ϕ be a norm on \mathbb{R}^n .*

1. ϕ is absolute if $\phi(z) = \phi(|z|) \forall z \in \mathbb{R}^n$.
2. ϕ is a symmetric gauge if it is absolute and $\phi = \phi \circ P$ for every permutation matrix $P \in \{0, 1\}^{n \times n}$, that is, a square real matrix with exactly one entry equal to 1 in each row and column and all other entries equal to 0.

From [18, Thm. 5.4.19], a norm ϕ on \mathbb{R}^n is absolute if and only if it is monotone (i.e., $\phi(x) \leq \phi(z) \forall x, z \in \mathbb{R}^n$ with $|x| \leq |z|$), and its dual ϕ° is also an absolute norm. By monotonicity and sign-invariance, only vectors z with nonnegative entries and support contained in that of y need to be considered in achieving the maximum in (2.3) when y is nonnegative. Hence, the restriction of ϕ° to the nonnegative orthant can be rewritten as

$$(2.5) \quad \phi^\circ(y) = \max\{y^\top z : z \in \mathbb{R}_+^n, \phi(z) \leq 1, \text{supp}(z) \subseteq \text{supp}(y)\}, \quad \forall y \in \mathbb{R}_+^n.$$

The symmetric gauges are also known as *symmetric absolute norms* [18, p. 335]. Specializing Von Neumann's result to real symmetric matrices, the next lemma shows that the symmetric gauges on \mathbb{R}^n are in one-to-one correspondence with the unitarily invariant matrix norms on \mathbb{S}^n , i.e., $\|UAV\| = \|A\|, \forall A \in \mathbb{S}^n, \forall U, V$ orthogonal, as well as their respective duals, through composition with the spectral mapping λ .

LEMMA 2.3 (Unitarily invariant matrix norm on \mathbb{S}^n). *Let ϕ be a symmetric gauge on \mathbb{R}^n . Define*

$$\|S\|_\phi := \phi(\lambda(S)), \quad \forall S \in \mathbb{S}^n.$$

Then $\|\cdot\|_\phi$ is a unitarily invariant matrix norm on \mathbb{S}^n .

Conversely, every unitarily invariant matrix norm on $\mathbb{R}^{n \times n}$ restricted to \mathbb{S}^n can be written as $S \mapsto \phi(\lambda(S))$ for a (unique) symmetric gauge ϕ on \mathbb{R}^n .

The dual matrix norm, defined by (2.3) with respect to the Frobenius inner product on $\mathbb{R}^{n \times n}$, is also unitarily invariant and its restriction to \mathbb{S}^n satisfies:

$$(\phi \circ \lambda)^\circ(X) = \|X\|_\phi^\circ = \max_{S \in \mathbb{S}^n} \{|\langle S, X \rangle| : \|S\|_\phi \leq 1\} = (\phi^\circ \circ \lambda)(X), \quad \forall X \in \mathbb{S}^n,$$

where ϕ° is the dual gauge of ϕ on \mathbb{R}^n .

Proof. Mapping $\phi \circ \sigma$ is an unitarily invariant matrix norm [18, Thm. 7.4.7.2] on $\mathbb{R}^{n \times n}$, then, restricted to \mathbb{S}^n , $\phi(\sigma(S)) = \phi(|\lambda(S)|^\uparrow) = \phi(\lambda(S))$ by symmetry of ϕ .

For the second assertion, we use the result that a norm $\|\cdot\|$ is unitarily invariant if and only if its dual $\|\cdot\|^\circ$ is also unitarily invariant [18, Thm. 5.6.39]. By (2.3), the dual norm is defined on $\mathbb{R}^{n \times n}$ as $\|X\|_\phi^\circ = \max_{A \in \mathbb{R}^{n \times n}} \{|\langle A, X \rangle| : \|A\|_\phi \leq 1\}$ but it is enough to consider symmetric matrices A when computing this maximum, since $S = \frac{A+A^\top}{2} \in \mathbb{S}^n$, and then $\langle S, X \rangle = \langle A, X \rangle$ with $\|S\|_\phi \leq \|A\|_\phi$.

For any feasible $S \in \mathbb{S}^n$ with $\phi(\lambda(S)) \leq 1$, inequalities (2.2) and (2.4) imply

$$|\langle S, X \rangle| \leq (|\lambda(S)|^\uparrow)^\top (|\lambda(X)|^\uparrow) \leq \phi(|\lambda(S)|^\uparrow) \phi^\circ(|\lambda(X)|^\uparrow) \leq \phi^\circ(\lambda(X)).$$

Hence $\|X\|_\phi^\circ \leq \phi^\circ(\lambda(X))$. Conversely, choose a maximizer $z \in \mathbb{R}^n$ of (2.3) defining $\phi^\circ(\lambda(X))$, i.e., with $\phi(z) \leq 1$ and $|z^\top \lambda(X)| = \phi^\circ(\lambda(X))$. If $X = Q \text{diag}(\lambda(X)) Q^\top$, define $S := Q \text{diag}(z) Q^\top \in \mathbb{S}^n$. Then $\|S\|_\phi = \phi(\lambda(S)) = \phi(z) \leq 1$ by symmetry of ϕ and $|\langle S, X \rangle| = |z^\top \lambda(X)| = \phi^\circ(\lambda(X))$. Thus $\|X\|_\phi^\circ = \phi^\circ(\lambda(X))$, for all $X \in \mathbb{S}^n$. \square

Given a symmetric gauge ϕ on \mathbb{R}^n , we denote by $\mathcal{B}_\phi := \{S \in \mathbb{S}^n : \|S\|_\phi = \phi(\lambda(S)) \leq 1\}$ the unit ball of the associated matrix norm $\phi \circ \lambda$. This is a compact set [18, Cor. 5.4.8], therefore, the dual norm $\|\cdot\|_\phi^\circ$ is well defined.

3. Spectral-gauge cuts. This section studies the separation problem for cuts \mathcal{C}_S whose separator S lies in a compact representative slice of \mathbb{S}_+^n . We take this slice to be $\mathbb{S}_+^n \cap \mathcal{B}_\phi$, where \mathcal{B}_ϕ is the unit ball of the unitarily invariant matrix norm induced by a symmetric gauge ϕ on \mathbb{R}^n . We give a constructive formula for the most violated cut at a current iterate $X \in \mathbb{S}^n$ and a closed formula for the maximum violation involving the dual gauge ϕ° and the negative eigenvalue vector $\lambda_-(X)$.

DEFINITION 3.1 (Spectral-gauge separation). *Given a symmetric gauge ϕ on \mathbb{R}^n , the spectral-gauge separation problem at a point $\bar{X} \in \mathbb{S}^n$ is the following optimization:*

$$(3.1) \quad g_\phi(\bar{X}) := \min_{S \in \mathbb{S}_+^n \cap \mathcal{B}_\phi} \langle S, \bar{X} \rangle \quad \text{with } \mathcal{B}_\phi = \{S \in \mathbb{S}^n : \phi(\lambda(S)) \leq 1\}.$$

If S^* solves (3.1) and $g_\phi(\bar{X}) < 0$, then the linear inequality $\mathcal{C}_{S^*} := \langle S^*, X \rangle \geq 0$ is called a spectral-gauge cut and S^* an optimal separator.

Problem (3.1) minimizes a continuous function over a compact set; therefore, its minimum $g_\phi(\bar{X}) \in \mathbb{R}$ is attained at some $S^* \in \mathbb{S}_+^n \cap \mathcal{B}_\phi$. By Lemma 2.3, this value is tightly related to the dual gauge ϕ° evaluated at the spectrum of \bar{X} .

Since every nonzero element $S \in \mathbb{S}_+^n$ has a representative $S' = S/\|S\|_\phi$ in \mathcal{B}_ϕ , and since $\langle S, X \rangle$ and $\langle S', X \rangle$ have the same sign for any $X \in \mathbb{S}^n$, self-duality of the PSD cone can be written as

$$(3.2) \quad \mathbb{S}_+^n = \mathcal{P}_\phi := \{X \in \mathbb{S}^n : \langle S, X \rangle \geq 0, \forall S \in \mathbb{S}_+^n \cap \mathcal{B}_\phi\}.$$

Remark 3.2. Optimizing (3.1) over any subset of \mathbb{S}_+^n leads to a valid inequality \mathcal{C}_{S^*} that separates \bar{X} if the minimum is negative; optimizing over the representative unit ball ensures completeness: if $g_\phi(\bar{X}) \geq 0$ then $\bar{X} \in \mathbb{S}_+^n$.

3.1. Explicit optimal separator. The following theorem establishes that the problem (3.1) is a complete separation oracle for the semi-infinite linear system (3.2), and it provides a closed formula for g_ϕ and an explicit optimal S^* .

THEOREM 3.3 (Optimal spectral-gauge separator). *Let ϕ be a symmetric gauge on \mathbb{R}^n and let ϕ° be its dual norm. For any $X \in \mathbb{S}^n$ with eigendecomposition $X = Q\Lambda(X)Q^\top$ and negative spectrum $\lambda_-(X) \in \mathbb{R}_+^n$, the optimal value of the spectral-gauge separation problem (3.1) is*

$$(3.3) \quad g_\phi(X) = -\phi^\circ(\lambda_-(X)),$$

and there exists an optimal separator $S^* \in \operatorname{argmin}_{S \in \mathbb{S}_+^n \cap \mathcal{B}_\phi} \langle S, X \rangle$ of the form

$$(3.4) \quad S^* = Q \operatorname{diag}(z^*) Q^\top, \quad \text{with } z^* \in \operatorname{argmax}_{z \in \mathbb{R}_+^n} \{\lambda_-(X)^\top z : \phi(z) \leq 1\}$$

such that $\operatorname{supp}(z^*) \subseteq \operatorname{supp}(\lambda_-(X))$ as in (2.5). Then, either $g_\phi(X) = 0$ and $X \in \mathbb{S}_+^n$, or $g_\phi(X) < 0$ and $\mathcal{C}_{S^*} := \langle S^*, X \rangle \geq 0$ separates X from \mathbb{S}_+^n .

Proof. We prove equality (3.3) by considering $X = X_+ - X_-$ as defined in (2.1) by the eigenbasis Q and the positive and negative spectra $\lambda_+(X)$ and $\lambda_-(X)$.

(\geq): Let $S \in \mathbb{S}_+^n \cap \mathcal{B}_\phi$. Since $X_+ \in \mathbb{S}_+^n$, we have $\langle S, X_+ \rangle \geq 0$ by self-duality, and

$$(3.5) \quad \langle S, X \rangle \geq \langle S, -X_- \rangle.$$

Since $X_- \in \mathbb{S}_+^n$, the trace inequality (2.2) and the dual equation (2.5) together with $\phi(\lambda(S)) \leq 1$ yield

$$(3.6) \quad \langle S, X_- \rangle \leq \lambda(X_-)^\top \lambda(S) = (\lambda_-(X)^\dagger)^\top \lambda(S) \leq \phi^\circ(\lambda_-(X)^\dagger) = \phi^\circ(\lambda_-(X)).$$

Minimizing (3.5) and maximizing (3.6) over $S \in \mathbb{S}_+^n \cap \mathcal{B}_\phi$ yield

$$g_\phi(X) \geq g_\phi(-X_-) \geq -\phi^\circ(\lambda_-(X)).$$

(\leq): Conversely, let $z^* \in \mathbb{R}_+^n$ be a solution of the maximization problem defining the dual norm $\phi^\circ(\lambda_-(X))$ in (2.5) with support contained in the support of $\lambda_-(X)$, which means z^* is orthogonal to $\lambda_+(X)$. Define $S^* = Q \operatorname{diag}(z^*) Q^\top$ using the eigenbasis Q of X , then $S^* \in \mathbb{S}_+^n \cap \mathcal{B}_\phi$ and

$$g_\phi(X) \leq \langle S^*, X \rangle = \langle S^*, X_+ - X_- \rangle = \lambda_+(X)^\top z^* - \lambda_-(X)^\top z^* = -\phi^\circ(\lambda_-(X)).$$

Consequently, S^* and z^* satisfy the conditions in (3.4).

The last assertion follows from reformulation (3.2): if $g_\phi(X) \geq 0$ then $g_\phi(X) = 0$ as $0 \in \mathbb{S}_+^n \cap \mathcal{B}_\phi$, otherwise \mathcal{C}_{S^*} is by definition the valid inequality for \mathbb{S}_+^n with the largest violation $|g_\phi(X)|$ at $X \in \mathbb{S}^n \setminus \mathbb{S}_+^n$ with respect to \mathcal{B}_ϕ . \square

Theorem 3.3 provides a constructive method for computing a solution S^* of the gauge-spectral separation problem (3.1), based on the computation of the dual norm of the negative spectrum of X and an eigenbasis Q . Note that, when building S^* in (3.4), we can restrict Q to the eigenvectors corresponding to the support of z^* and thus to the negative eigenvalues of X .

Remark 3.4 (Spectral-gauges from non-self-dual cones). Although Theorem 3.3 is stated for the self-dual cone \mathbb{S}_+^n , the argument behind the separation problem is more general, as it applies to any closed convex cone $\mathcal{K} \subseteq \mathbb{S}^n$, possibly non-self-dual. Indeed, given a symmetric gauge ϕ on \mathbb{R}^n , the unit ball \mathcal{B}_ϕ is compact in \mathbb{S}^n , so is the slice $\mathcal{K} \cap \mathcal{B}_\phi$, and $0 \in \mathcal{K} \cap \mathcal{B}_\phi$. Moreover, every nonzero $S \in \mathcal{K}$ satisfies $S/\phi(\lambda(S)) \in \mathcal{K} \cap \mathcal{B}_\phi$. Hence, the dual cone can be written

$$\mathcal{K}^* = \{X \in \mathbb{S}^n : \langle S, X \rangle \geq 0, \forall S \in \mathcal{K} \cap \mathcal{B}_\phi\},$$

and the problem

$$g_{\phi, \mathcal{K}}(X) := \min_{S \in \mathcal{K} \cap \mathcal{B}_\phi} \langle S, X \rangle$$

is a separation oracle for \mathcal{K}^* : one has $g_{\phi, \mathcal{K}}(X) = 0$ if and only if $X \in \mathcal{K}^*$, while any minimizer $S^* \in \mathcal{K} \cap \mathcal{B}_\phi$ with $g_{\phi, \mathcal{K}}(X) < 0$ yields a valid inequality $\langle S^*, \cdot \rangle \geq 0$ separating X from \mathcal{K}^* .

3.2. Cut quality and normalization across gauges. Cut selection involves several non-equivalent metrics to estimate the strength of an individual cut, or of a family of cuts; see, e.g., [39, 16] and [3] in the PSD-cut setting. Note first that all cuts \mathcal{C}_S with normalized $S \in \mathbb{S}^n$ are *minimal* for \mathbb{S}_+^n in the sense that

$$\{X \in \mathbb{S}^n : \langle S', X \rangle \geq 0, \forall S' \in \mathcal{K}\} \subseteq \{X \in \mathbb{S}^n : \langle S, X \rangle \geq 0\} \implies S \in \mathcal{K},$$

where \mathcal{K} is any finitely generated cone. The separation problem (3.1) selects a cut in $\{\mathcal{C}_S : S \in \mathbb{S}^n\}$ with maximum violation $\frac{\langle S, \bar{X} \rangle}{\|S\|_\phi}$ relative to normalization $\|\cdot\|_\phi$ and current iterate \bar{X} . This metric measures the potential effectiveness of this individual cut within its family defined by gauge ϕ , but not the relative strength of spectral-gauge cuts arising from different symmetric gauges.

The *depth-of-cut* is a scale-invariant criterion used for such comparison [39, 16]. It measures the Euclidean (or Frobenius, in the matrix space) distance between iterate \bar{X} and the separating hyperplane $H_S := \{X \in \mathbb{S}^n : \langle S, X \rangle = 0\}$, that is

$$(3.7) \quad \text{depth}(\bar{X}, S) := \text{dist}_F(\bar{X}, H_S) = \frac{|\langle S, \bar{X} \rangle|}{\|S\|_F}.$$

The above measure can be easily derived considering that the Frobenius projection of \bar{X} onto H_S is $\text{proj}_{H_S}(\bar{X}) = \bar{X} - \frac{\langle S, \bar{X} \rangle}{\|S\|_F^2} S$. The depth has a simple interpretation in our spectral setting. Let $\bar{X} = \bar{X}_+ - \bar{X}_-$ be the decomposition as in (2.1), then using Cauchy–Schwarz, we have for every $S \in \mathbb{S}_+^n$

$$-\frac{\langle S, \bar{X} \rangle}{\|S\|_F} = \frac{\langle S, \bar{X}_- \rangle - \langle S, \bar{X}_+ \rangle}{\|S\|_F} \leq \frac{\langle S, \bar{X}_- \rangle}{\|S\|_F} \leq \|\bar{X}_-\|_F.$$

Equality is attained by any positive multiple of \bar{X}_- , and, in particular, by the optimal spectral-gauge separator for the symmetric gauge ℓ_2 (see Corollary 4.3).

A third criterion is the complexity of the separator. The rank and the sparsity of the separator have a direct impact on the computational cost of the LP reoptimization that follows each oracle call. According to Theorem 3.3, one may choose an optimal separator S^* whose rank is at most the number of negative eigenvalues of the current iterate \bar{X} , denoted p . Hence, after possibly relabeling the negative eigenvectors, S^* admits the spectral decomposition $S^* = \sum_{i \in J} \alpha_i q_i q_i^\top$, $\alpha_i > 0$, for some $J \subseteq [p]$, where the vectors q_i are eigenvectors associated with negative eigenvalues of \bar{X} .

Geometrically, the separator reflects the spectral structure of the iterate: the negative eigenspace identifies directions along which the quadratic form $q_i^\top \bar{X} q_i$ is negative, and each eigenvector q_i associated with a negative eigenvalue corresponds to a principal direction of curvature violating the semidefinite constraint. The rank-one matrices $q_i q_i^\top$ therefore define elementary separating hyperplanes in the space of symmetric matrices, each enforcing the inequality $q_i^\top X q_i \geq 0$. When all p directions are used, the p -rank separator aggregates all such violated spectral directions, producing a cut whose normal vector spans the entire negative eigenspace. Hence, such a full-rank cut reshapes the feasible set more drastically than a rank-one cut and could move the next optimum to a distant region, therefore increasing the cost of reoptimization. Selecting only a subset of eigenvectors yields a separator whose normal lies in a lower-dimensional subspace of this eigenspace.

A distinct approach to reducing computational cost is to sparsify the separator. Note that rank and entrywise sparsity are not related. Indeed, rank-one separators are typically dense because the eigenvectors q_i themselves are dense, so the resulting cut \mathcal{C}_{S^*} is not significantly sparser. This distinction motivates the sparse and low-rank optimal separators developed in Sections 4.2 and 4.3, respectively.

4. Instantiation. This section illustrates the canonical framework. Section 4.1 applies Theorem 3.3 to the ℓ_p norms, which recover standard eigencuts for $p = 1$, nuclear cuts for $p = \infty$, and infinitely many intermediate cuts. Section 4.2 connects our

framework to the factor-width hierarchy [25] and to sparse eigencuts [15]. Section 4.3 characterizes optimal separators under a prescribed rank limit.

4.1. Spectral- ℓ_p cut hierarchy. The ℓ_p norms form a well-known family of symmetric gauges on \mathbb{R}^n that are closed under duality in the following sense: if ϕ is the ℓ_p norm for some $p \in [1, +\infty]$, i.e., $\phi(z) = \|z\|_p = (\sum |z_i|^p)^{1/p}$ or $\phi(z) = \|z\|_\infty = \max |z_i|$, then the dual norm ϕ° is the ℓ_q norm where $1/p + 1/q = 1$ with the convention $1/\infty = 0$ (see, e.g., [18]). The pair (p, q) is called a pair of *dual exponents*. Theorem 3.3 applies to the ℓ_p norms as follows.

PROPOSITION 4.1 (Closed-forms for ℓ_p gauges). *Let $\phi = \|\cdot\|_p$ be the ℓ_p -norm for some dual exponents $p, q \in [1, \infty]$. Then, for any $X \in \mathbb{S}^n$ with eigendecomposition $X = Q\Lambda(X)Q^\top$ and spectrum $\lambda(X) = (\lambda_i)_i$, the optimum value of the spectral-gauge separation problem (3.1) is*

$$g_\phi(X) = -\|\lambda_-(X)\|_q,$$

with optimal solution $S^* = Q \operatorname{diag}(z^*)Q^\top$ for $z^* \in \operatorname{argmax}_{z \in \mathbb{R}_+^n} \{\lambda_-(X)^\top z : \|z\|_p \leq 1\}$.

In particular, if $1 < p < \infty$ and $X \in \mathbb{S}^n \setminus \mathbb{S}_+^n$, then a maximizer is defined componentwise by

$$z_i^* = \left(\frac{\lambda_-(X)_i}{\|\lambda_-(X)\|_q} \right)^{q-1} \quad \forall i \in [n].$$

Proof. The first part is Theorem 3.3 applied to $\phi = \|\cdot\|_p$ and $\phi^\circ = \|\cdot\|_q$. It remains to compute the maximizer z^* when $1 < p < \infty$ and $X \notin \mathbb{S}_+^n$, i.e., $\lambda_-(X) \neq 0$.

Let $y = \lambda_-(X) \in \mathbb{R}_+^n$. The Cauchy-Schwarz inequality (2.4) reads $y^\top z^* \leq \|y\|_q \|z^*\|_p$ with equality if and only if there exists $\alpha > 0$ such that $(z_i^*)^p = \alpha (y_i)^q$ for all $i \in [n]$ [18, p. 559]. Maximum occurs for $\|z^*\|_p = 1$, which yields

$$1 = \|z^*\|_p^p = \sum_{i=1}^n (z_i^*)^p = \alpha \sum_{i=1}^n y_i^q = \alpha \|y\|_q^q.$$

Hence, $z_i^* = \alpha^{1/p} (y_i)^{q/p} = \|y\|_q^{-q/p} (y_i)^{q/p}$, with $q/p = q - 1$. \square

We now give closed forms for the separation problem corresponding to the Manhattan norm ℓ_1 , the Euclidean ℓ_2 norm, and the maximum norm ℓ_∞ , recognizing the standard eigenvalue cuts in the former case.

COROLLARY 4.2 (Eigencuts ($p = 1$)). $\mathbb{S}_+^n \cap \mathcal{B}_{\ell_1} = \{S \in \mathbb{S}_+^n : \operatorname{tr}(S) \leq 1\}$ and

$$g_{\ell_1}(X) = -\|\lambda_-(X)\|_\infty = \min\{\lambda_1(X), 0\}.$$

Moreover, if $\lambda_1(X) < 0$, an optimal separator may be chosen as $S^* = q_1 q_1^\top$, where q_1 is a unit eigenvector associated with the smallest eigenvalue $\lambda_1(X)$.

Proof. For $S \in \mathbb{S}_+^n$, $\|\lambda(S)\|_1 = \operatorname{tr}(S)$. Apply Proposition 4.1 with $p = 1$ and dual exponent $q = \infty$. If $\lambda_1(X) < 0$, then $S^* = q_1 q_1^\top \in \mathbb{S}_+^n$ with $\operatorname{tr}(S^*) = 1$, and $\langle S^*, X \rangle = \lambda_1(X) = g_{\ell_1}(X)$, so S^* is optimal. \square

COROLLARY 4.3 (Frobenius cuts ($p = 2$)). $\mathcal{B}_{\ell_2} = \{S \in \mathbb{S}^n : \|S\|_F \leq 1\}$ and $g_{\ell_2}(X) = -\|X_-\|_F$. If $X_- \neq 0$, an optimal separator is given by

$$S^* = \frac{X_-}{\|X_-\|_F}.$$

Proof. Apply Proposition 4.1 with $p = q = 2$ and $\|X_-\|_F = \|\lambda_-(X)\|_2$. \square

COROLLARY 4.4 (Nuclear cuts ($p = \infty$)). $\mathcal{B}_{\ell_\infty} = \{S \in \mathbb{S}^n : \|S\|_2 \leq 1\}$ and

$$g_{\ell_\infty}(X) = -\|\lambda_-(X)\|_1 = \sum_{\lambda_i(X) < 0} \lambda_i(X).$$

An optimal separator may be chosen as the (unique) orthogonal projector onto the negative eigenspace of X , i.e., $S^* = Q \operatorname{diag}(1_{\{\lambda_i(X) < 0\}}) Q^\top$.

Proof. Apply Proposition 4.1 with $p = \infty$ and $q = 1$. The vector maximization reads $z^* \in \operatorname{argmax}\{\lambda_-(X)^\top z : z \geq 0, \|z\|_\infty \leq 1\}$ and holds for $z_i^* = 1$ if $\lambda_-(X)_i > 0$ and $z_i^* = 0$ otherwise. \square

The latter case $p = \infty$ coincides with the definition of nuclear cut in [7]:

$$\langle W, X \rangle \leq \operatorname{tr}(X), \quad W \in \mathbb{S}_+^n, \quad \|W\|_2 \leq 1.$$

Let $S := I - W$. Since $W \in \mathbb{S}_+^n$ and $\|W\|_2 \leq 1$, its eigenvalues lie in $[0, 1]$. Hence,

$$\max_{\substack{W \in \mathbb{S}_+^n \\ \|W\|_2 \leq 1}} (\langle W, X \rangle - \operatorname{tr}(X)) = \max_{\substack{W \in \mathbb{S}_+^n \\ \|W\|_2 \leq 1}} \langle W - I, X \rangle = - \min_{\substack{S \geq 0 \\ \|S\|_2 \leq 1}} \langle S, X \rangle = -g_{\ell_\infty}(X).$$

Remark 4.5 (Relative strengths). The spectral- ℓ_p cuts can be compared in terms of the amount of violation $-g_{\ell_p}$ at a given iterate X . For $1 \leq p \leq q \leq \infty$, $\|z\|_q \leq \|z\|_p$ holds $\forall z \in \mathbb{R}^n$. Hence, for any $S \in \mathcal{B}_{\ell_p}$, one has $\|\lambda(S)\|_q \leq \|\lambda(S)\|_p \leq 1$, proving $\mathcal{B}_{\ell_p} \subseteq \mathcal{B}_{\ell_q}$ and $g_{\ell_q}(X) \leq g_{\ell_p}(X)$. Alternatively, under the Euclidean depth measure (3.7), Frobenius cuts are the deepest. Hence, spectral- ℓ_∞ cuts may yield larger violation, but spectral- ℓ_2 cuts maximize depth.

4.2. Spectral-gauge cuts with sparse support. The high density of the eigencuts, i.e., the ℓ_1 -spectral gauge cuts, is an issue addressed in [27, 3, 15] by enforcing the separating eigenvector (q_1 in Corollary 4.2) to have small support. We extend this idea to spectral-gauge cuts for arbitrary symmetric gauge ϕ . A direct analogue is to require the separator S to be supported on a block $I \times I$, so that the inequality $\langle S, X \rangle \geq 0$ only involves the principal submatrix X_I . This restriction is closely related to the factor-width matrix cone hierarchy [9, 25] the conic hull of PSD principal blocks of size at most m is the factor-width- m cone, and its dual is the cone of matrices whose principal submatrices of order at most m are all PSD.

We treat this additional sparsity requirement in two steps. First, we fix a support $I \subseteq [n]$ and show that the sparse spectral-gauge oracle reduces to Theorem 3.3 applied to the principal submatrix X_I . Second, we fix a maximum support size m and optimize over all supports I with $|I| \leq m$. This separates the spectral part of the oracle, which remains explicit, from the outer combinatorial search over supports. In the special case $\phi = \ell_1$, this recovers the sparse eigencut separation problem of [15].

Fixed support I . Fix $I \subseteq [n]$ and let $E_I \in \{0, 1\}^{n \times |I|}$ be the principal block selector, i.e., for any $X \in \mathbb{S}^n$, the product $E_I^\top X E_I$ is the principal submatrix of X supported on I , hence an element of $\mathbb{S}^{|I|}$. We denote

$$X_I := E_I^\top X E_I \in \mathbb{S}^{|I|}, \quad \iota_I(T) := E_I T E_I^\top \in \mathbb{S}^n$$

the principal-block projection and the corresponding lifting map, respectively. A matrix $S \in \mathbb{S}^n$ is supported on the principal block $\operatorname{supp}(S) \subseteq I$ if $S = \iota_I(S_I)$ for some $S_I \in \mathbb{S}^{|I|}$. The block PSD cone associated with I and its dual cone in \mathbb{S}^n are

$$\iota_I(\mathbb{S}_+^{|I|}) = \left\{ E_I S_I E_I^\top : S_I \in \mathbb{S}_+^{|I|} \right\} \subseteq \mathbb{S}_+^n, \quad \iota_I(\mathbb{S}_+^{|I|})^* = \{X \in \mathbb{S}^n : X_I \in \mathbb{S}_+^{|I|}\}.$$

since $\langle E_I S_I E_I^\top, X \rangle = \langle S_I, X_I \rangle$.

Given a symmetric gauge ϕ on \mathbb{R}^n , the support I induces a symmetric gauge ϕ_I on $\mathbb{R}^{|I|}$ by zero-padding $\phi_I(u) := \phi((u, 0_{[n] \setminus I}))$ for $u \in \mathbb{R}^{|I|}$, where the positions of the zeros are not relevant due to the permutation invariance of ϕ . Equivalently, its dual gauge ϕ_I° is the restriction of ϕ° to vectors supported on I . Enforcing a fixed-support condition restricts the spectral-gauge separation problem to

$$(4.1) \quad g_{I,\phi}(\bar{X}) := \min \left\{ \langle S, \bar{X} \rangle : S \in \mathbb{S}_+^n \cap \mathcal{B}_\phi, \text{supp}(S) \subseteq I \right\}.$$

Since $\lambda(\iota_I(S_I)) = (\lambda(S_I), 0_{[n] \setminus I})$, the feasible set is the lifting of the lower-dimensional ball:

$$\mathbb{S}_+^n \cap \mathcal{B}_\phi \cap \{S : \text{supp}(S) \subseteq I\} = \left\{ E_I S_I E_I^\top : S_I \in \mathbb{S}_+^{|I|}, \phi_I(\lambda(S_I)) \leq 1 \right\} =: \mathcal{B}_{I,\phi}$$

Consequently, the restricted problem (4.1) projects onto the subspace $\mathbb{S}^{|I|}$, resulting in the spectral-gauge oracle of Theorem 3.3 applied to the principal block $\bar{X}_I \in \mathbb{S}^{|I|}$ of the current iterate \bar{X} . These results are formalized in the following proposition.

PROPOSITION 4.6 (Fixed-support spectral-gauge oracle). *Given $\bar{X} \in \mathbb{S}^n$,*

$$g_{I,\phi}(\bar{X}) = \min_{\substack{S \in \mathbb{S}_+^{|I|} \\ \phi_I(\lambda(S)) \leq 1}} \langle S, \bar{X}_I \rangle = -\phi_I^\circ(\lambda_-(\bar{X}_I)).$$

Thus, $g_{I,\phi}(\bar{X}) = 0 \iff \bar{X} \in \iota_I(\mathbb{S}_+^{|I|})^$. If $g_{I,\phi}(\bar{X}) < 0$ and $\bar{X}_I = Q \text{diag}(\lambda(\bar{X}_I)) Q^\top$,*

$$z^* \in \text{argmax} \left\{ \lambda_-(\bar{X}_I)^\top z : z \in \mathbb{R}_+^{|I|}, \phi_I(z) \leq 1, \text{supp}(z) \subseteq \text{supp}(\lambda_-(\bar{X}_I)) \right\},$$

then an optimal separator for $g_{I,\phi}(\bar{X})$ is

$$S_I^* = E_I Q \text{diag}(z^*) Q^\top E_I^\top \in \mathcal{B}_{I,\phi} \subseteq \iota_I(\mathbb{S}_+^{|I|}).$$

Moreover, $\langle S_I^, X \rangle \geq 0$ for all $X \in \iota_I(\mathbb{S}_+^{|I|})^*$, and in particular for every $X \in \mathbb{S}_+^n$.*

Variable support of bounded size m . We now fix a maximum support size and allow the support to vary. Following the factor-width terminology of Boman et al. [9] and the cone formulation used by Permenter and Parrilo [25], we define for $1 \leq m \leq n$ the factor-width- m cone by

$$\mathcal{FW}_m^n := \left\{ \sum_{\substack{I \subseteq [n] \\ |I| \leq m}} \iota_I(S_I) : S_I \in \mathbb{S}_+^{|I|} \right\} = \text{cone} \left(\bigcup_{\substack{I \subseteq [n] \\ |I| \leq m}} \iota_I(\mathbb{S}_+^{|I|}) \right),$$

where the sum above is the Minkowski sum of cones. This formulation is equivalent to Boman's definition of factor-width, i.e., every term $\iota_I(S_I)$, with $S_I \succeq 0$, is a sum of rank-one matrices vv^\top with $|\text{supp}(v)| \leq m$, and conversely every such rank-one term is supported on a principal block of size at most m [1, Sec. 5.3].

By self-duality of the PSD cone, $\iota_I(\mathbb{S}_+^{|I|})^* = \{X \in \mathbb{S}^n : X_I \succeq 0\}$. Hence

$$(4.2) \quad (\mathcal{FW}_m^n)^* = \bigcap_{\substack{I \subseteq [n] \\ |I| \leq m}} \iota_I(\mathbb{S}_+^{|I|})^* = \{X \in \mathbb{S}^n : X_I \succeq 0 \quad \forall I \subseteq [n], |I| \leq m\} =: \mathbb{S}_+^{n,m}.$$

Define

$$\mathcal{A}_{m,\phi} := \bigcup_{\substack{I \subseteq [n] \\ |I| \leq m}} \mathcal{B}_{I,\phi}, \quad g_{m,\phi}(X) := \min_{S \in \mathcal{A}_{m,\phi}} \langle S, X \rangle.$$

For every $\bar{X} \in \mathbb{S}^n$, the sparse spectral-gauge separation problem is

$$g_{m,\phi}(\bar{X}) = \min_{\substack{I \subseteq [n] \\ |I| \leq m}} g_{I,\phi}(\bar{X}) = - \max_{\substack{I \subseteq [n] \\ |I| \leq m}} \phi_I^\circ(\lambda_-(\bar{X}_I)).$$

Moreover, $g_{m,\phi}(\bar{X}) = 0 \iff \bar{X} \in \mathbb{S}_+^{n,m}$. If the maximum is attained with support I^* , then the optimizer $S_{I^*}^*$ from Proposition 4.6 is optimal for $g_{m,\phi}(\bar{X})$. If $g_{m,\phi}(\bar{X}) < 0$, then $\langle S_{I^*}^*, X \rangle \geq 0$ for all $X \in \mathbb{S}_+^{n,m}$, and in particular for every $X \in \mathbb{S}_+^n$.

For $\phi = \ell_1$, this reduces to the sparse eigencut separation [16]:

$$(4.3) \quad g_{m,\ell_1}(X) = - \max_{\substack{I \subseteq [n] \\ |I|=m}} (-\lambda_{\min}(X_I))_+.$$

This decomposition shows where the complexity of the oracle lies: given a prescribed support I , separation is the spectral-gauge problem restricted to the block \bar{X}_I . The nontrivial part is the outer maximization over I , which ranges over $\binom{n}{m}$ candidate supports and is thus combinatorial.

General sparsity patterns. Following Günlük et al. [17], let $E \subseteq [n] \times [n]$ be a symmetric index set containing the diagonal, and define

$$H_E := \{S \in \mathbb{S}^n : S_{ij} = 0 \ \forall (i,j) \notin E\}, \quad \mathcal{K}_E := \mathbb{S}_+^n \cap H_E.$$

Given a symmetric gauge ϕ on \mathbb{R}^n , consider

$$g_{E,\phi}(X) := \min \left\{ \langle S, X \rangle : S \in \mathcal{K}_E \cap \mathcal{B}_\phi \right\}.$$

The condition $g_{E,\phi}(X) \geq 0$ is equivalent to nonnegativity against all E -supported PSD separators, and thus yields a reformulation of the SDP- E relaxation [17, Eq. (5)]. In particular, when $\phi = \ell_1$, one recovers the problem of [17, Eq. (15)].

For a general E , however, \mathcal{K}_E is not necessarily orthogonally invariant, so the diagonalization argument of Theorem 3.3 no longer implies a closed form. Such a formula is available only when the cone \mathcal{K}_E is compatible with the spectral structure, as in the full PSD cone or in a fixed principal-block restriction.

4.3. Spectral-gauge separators with rank limit. As discussed, sparsifying the separator can significantly complicate the separation problem. Instead, we propose to control the rank of the separator, and show that it preserves tractability while still controlling spectral complexity.

LEMMA 4.7 (Rank and spectral ℓ_0 -function). *Let $S \in \mathbb{S}^n$ with spectral mapping $\lambda(S) \in \mathbb{R}^n$. Then $\text{rank}(S) = \|\lambda(S)\|_0$, where $\|x\|_0 := |\{i \in [n] : x_i \neq 0\}|$.*

Proof. Decompose $S = Q \text{diag}(\lambda(S)) Q^\top$ with Q orthogonal. Since Q is invertible, $\text{rank}(S) = \text{rank}(\text{diag}(\lambda(S)))$, which is the number of nonzero eigenvalues. \square

PROPOSITION 4.8. *Let ϕ be a symmetric gauge on \mathbb{R}^n , let $k \in [n]$, and define*

$$\mathcal{A}_k(\phi) := \left\{ u \in \mathbb{R}^n : \phi(u) \leq 1, \|u\|_0 \leq k \right\}, \quad \mathcal{K}_k(\phi) := \text{conv}(\mathcal{A}_k(\phi)),$$

together with the associated gauge function [11, Eq. (2)]

$$(4.4) \quad \phi^{(k)}(z) := \inf \left\{ t > 0 : z \in t \mathcal{K}_k(\phi) \right\} \quad \forall z \in \mathbb{R}^n.$$

Then $\phi^{(k)}$ is a symmetric gauge on \mathbb{R}^n with unit ball $\mathcal{K}_k(\phi)$.

The proof of Proposition 4.8 is given in Appendix A. For $y \in \mathbb{R}^n$, let $I_k(y) \subseteq [n]$ be the index set of $k > 0$ largest entries of $|y|$, and

$$\text{trun}_k(y) := \sum_{i \in I_k(y)} y_i e_i \in \mathbb{R}^n.$$

PROPOSITION 4.9. *Let ϕ be a symmetric gauge and define $\phi^{(k)}$ by (4.4), then*

$$\begin{aligned} (\phi^{(k)})^\circ(y) &= \max \left\{ y^\top u : \phi(u) \leq 1, \|u\|_0 \leq k, u \in \mathbb{R}^n \right\} \\ &= \max_{\substack{I \subseteq [n] \\ |I| \leq k}} \phi^\circ(y \odot 1_I) = \phi^\circ(\text{trun}_k(y)) \quad \forall y \in \mathbb{R}^n, \end{aligned}$$

where \odot denotes componentwise multiplication and 1_I is the indicator vector of I .

Proof. The linear function $\langle y, \cdot \rangle$ attains its maximum on the unit ball $\mathcal{K}_k(\phi) = \text{conv}(\mathcal{A}_k(\phi))$ in $\mathcal{A}_k(\phi)$, giving the first equality. For fixed I , maximizing over vectors supported on I gives $\phi^\circ(y \odot 1_I)$. By monotonicity and symmetry of ϕ° , the maximum over all $|I| \leq k$ is attained by keeping the k largest magnitudes of y . \square

THEOREM 4.10. *Let ϕ be any symmetric gauge on \mathbb{R}^n and $\phi^{(k)}$ defined in (4.4). Then, for $X \in \mathbb{S}^n$ with spectral decomposition $X = Q \Lambda(X) Q^\top$, the spectral gauge separation of X from \mathbb{S}_+^n reads*

$$g_{\phi^{(k)}}(X) = -(\phi^{(k)})^\circ(\lambda_-(X)) = -\phi^\circ(\text{trun}_k(\lambda_-(X))).$$

Moreover, there exists an optimal separator of the form $S^* = Q \text{diag}(z^*) Q^\top$ with $z^* \in \text{argmax}\{(\text{trun}_k(\lambda_-(X)))^\top z : z \geq 0, \phi(z) \leq 1, \text{supp}(z) \subseteq \text{supp}(\text{trun}_k(\lambda_-(X)))\}$.

In particular, $\text{rank}(S^*) \leq k$. If $X \not\leq 0$, then $\langle S^*, X \rangle < 0$ and the cut \mathcal{C}_{S^*} is valid.

Proof. The formula follows by the application of Theorem 3.3 to $\phi^{(k)}$, and then Proposition 4.9. If z^* solves the vector problem, we may replace it by $z^* \odot 1_I$, where $I \subseteq [n]$ is the support of $\text{trun}_k(\lambda_-(X))$. Then $z^* \in \mathcal{A}_k(\phi)$, so $\phi^{(k)}(z^*) \leq 1$, and thus $S^* = Q \text{diag}(z^*) Q^\top \in \mathbb{S}_+^n \cap \mathcal{B}_{\phi^{(k)}}$ with $\text{rank}(S^*) = \|z^*\|_0 \leq k$ by Lemma 4.7. Finally, $\langle S^*, X \rangle = -(\text{trun}_k(\lambda_-(X)))^\top z^* = -\phi^\circ(\text{trun}_k(\lambda_-(X)))$, so S^* is optimal. \square

Thus, the rank-limit parameterization boils down to evaluating the original dual gauge on the k most negative eigenvalues. In particular, the ℓ_1 eigencuts result from the rank restriction $k = 1$ applied to any ℓ_p -gauge, $1 \leq p \leq \infty$.

Remark 4.11 (Dual k -norm). Specializing Proposition 4.9 and Theorem 4.10 to $\phi = \ell_\infty$ for any $k \in [n]$ gives

$$(\ell_\infty^{(k)})^\circ(y) = \|\text{trun}_k(y)\|_1 = \sum_{i=1}^k |y|_{(i)} =: \|y\|_{(k)} \quad \text{and} \quad g_{\ell_\infty^{(k)}}(X) = -\|\lambda_-(X)\|_{(k)}$$

where $\|\cdot\|_{(k)}$ denotes the vector k -norm, the symmetric gauge defined as the sum of the k largest absolute components. Therefore, these k -rank nuclear cuts coincide with the $\|\cdot\|_{(k)}^\circ$ -spectral gauge cuts.

5. Spectral-gauge cutting-plane algorithm. In Algorithm 1, we integrate the separation oracle (3.1) into a cutting-plane method for solving (1.1). Starting from an outer approximation $\mathcal{R} \subseteq \mathcal{A} \cap \mathcal{P}_{\text{lin}}$ defined on a cone $\mathcal{A} \supseteq \mathbb{S}_+^n$, the proposed algorithm solves a sequence of master problems $(\mathcal{M}_i)_i$

$$(5.1) \quad \begin{aligned} \mathcal{M}_i : \quad & \min_{X \in \mathbb{S}^n} \quad \langle C, X \rangle \\ & \text{s.t.} \quad X \in \mathcal{R} \subseteq \mathcal{A} \cap \mathcal{P}_{\text{lin}}, \\ & \quad \quad \langle S, X \rangle \geq 0 \quad \forall S \in \mathcal{S}_i. \end{aligned}$$

and adds a spectral-gauge cut (for a given symmetric gauge ϕ) whenever the current solution is not PSD (within a tolerance ε). In what follows, we discuss possible initial relaxations \mathcal{A} and boundedness conditions. We prove the convergence of the algorithm in the general context, and illustrate specific cases.

Algorithm 1: Spectral-gauge cutting-plane method
<p>Require: $C \in \mathbb{S}^n$; polyhedron $\mathcal{P}_{\text{lin}} \subseteq \mathbb{S}^n$; outer-approximation \mathcal{A} with $\mathbb{S}_+^n \subseteq \mathcal{A} \subseteq \mathbb{S}^n$; symmetric gauge ϕ; tolerance $\varepsilon \geq 0$.</p> <p>Ensure: An ε-solution X to (1.1), i.e., with $\lambda_{\min}(X) > -\varepsilon$.</p> <p>$\mathcal{S}_0 \leftarrow \emptyset$</p> <p>for $i \leftarrow 0, 1, 2, \dots$ do</p> <div style="margin-left: 20px;"> <p>Solve \mathcal{M}_i: get optimizer $X^{(i)}$ with eigendecomposition</p> $X^{(i)} = Q^{(i)} \text{diag}(\lambda_1^{(i)}, \dots, \lambda_n^{(i)}) (Q^{(i)})^\top, \quad \lambda_1^{(i)} \leq \dots \leq \lambda_n^{(i)}.$ <p>Set $\lambda_-^{(i)} := (-\lambda^{(i)})_+$ and compute</p> $z^{(i)} \in \operatorname{argmax} \left\{ (\lambda_-^{(i)})^\top z : z \geq 0, \phi(z) \leq 1, \operatorname{supp}(z) \subseteq \operatorname{supp}(\lambda_-^{(i)}) \right\}.$ <p>Form $S^{(i)} := Q^{(i)} \text{diag}(z^{(i)}) (Q^{(i)})^\top$.</p> <p>if $\lambda_1^{(i)} > -\varepsilon$ then</p> <div style="margin-left: 20px;"> <p>return $X^{(i)}$</p> </div> <p>else</p> <div style="margin-left: 20px;"> <p>$\mathcal{S}_{i+1} \leftarrow \mathcal{S}_i \cup \{S^{(i)}\}$</p> </div> </div>

5.1. Outer-approximations of the PSD cone. To establish the convergence of the algorithm, we assume that there exists $T > 0$ such that every feasible solution X of (1.1) satisfies $\operatorname{tr}(X) \leq T$. Such a bound is explicit in the computation, if not in the formulation. We then enforce this condition in the master problems \mathcal{M}_i

$$\mathcal{R} := \left\{ X \in \mathbb{S}^n : X \in \mathcal{P}_{\text{lin}}, X \in \mathcal{A}, \operatorname{tr}(X) \leq T \right\},$$

with convex cone $\mathcal{A} \supseteq \mathbb{S}_+^n$. We also assume that every nonzero element X of \mathcal{A} has a strictly positive trace, so that it can be normalized $\|X\|_{\mathcal{X}} \leq \alpha_{\mathcal{X}}(\mathcal{A}) \operatorname{tr}(X)$, where

$$\alpha_{\mathcal{X}}(\mathcal{A}) := \sup \left\{ \|X\|_{\mathcal{X}} : X \in \mathcal{A}, \operatorname{tr}(X) = 1 \right\}$$

measures the size of cone \mathcal{A} in an arbitrary norm $\|\cdot\|_{\mathcal{X}}$ on \mathbb{S}^n .

Let $\mathcal{F} = \mathcal{P}_{\text{lin}} \cap \mathbb{S}_+^n \neq \emptyset$ denote the feasible set of the original SDP in (1.1). If $\alpha_{\mathcal{X}}(\mathcal{A}) < \infty$, then \mathcal{R} is a compact set containing \mathcal{F} and the feasible sets

$$(5.2) \quad \mathcal{R}^{(i)} := \left\{ X \in \mathcal{R} : \langle S^{(t)}, X \rangle \geq 0, t = 1, \dots, i-1 \right\}, \quad \mathcal{R}^{(1)} \supseteq \mathcal{R}^{(2)} \supseteq \dots \supseteq \mathcal{F}.$$

Hence, every master problem \mathcal{M}_i is feasible and attains an optimum.

The dual factor-width- m cone hierarchy in (4.2) provides us with a set of candidates for \mathcal{A} meeting our assumptions: $\mathcal{LP} \supseteq \mathcal{SOC} = \mathbb{S}_+^{n,2} \supseteq \mathbb{S}_+^{n,3} \supseteq \dots \supseteq \mathbb{S}_+^{n,n} = \mathbb{S}_+^n$, where the case $m = 2$ is the second-order cone

$$(5.3) \quad \mathcal{SOC} := \left\{ X \in \mathbb{S}^n : X_{jj} \geq 0 \quad \forall j, \quad X_{jl}^2 \leq X_{jj}X_{ll} \quad \forall j \neq l \right\},$$

and \mathcal{LP} is the linear relaxation of (5.3) obtained from the inequality of arithmetic and geometric means, namely

$$(5.4) \quad \mathcal{LP} := \left\{ X \in \mathbb{S}^n : X_{jj} \geq 0 \quad \forall j, \quad X_{jj} + X_{ll} \pm 2X_{jl} \geq 0 \quad \forall j \neq l \right\}.$$

Note that $\alpha_{\mathcal{X}}(\mathcal{LP}) < \infty$, hence

$$(5.5) \quad \alpha_{\mathcal{X}}(\mathbb{S}_+^n) \leq \alpha_{\mathcal{X}}(\mathbb{S}_+^{n,n-1}) \leq \dots \leq \alpha_{\mathcal{X}}(\mathbb{S}_+^{n,3}) \leq \alpha_{\mathcal{X}}(\mathcal{SOC}) \leq \alpha_{\mathcal{X}}(\mathcal{LP}) < \infty.$$

5.2. Algorithm convergence analysis. Recall the compact set \mathcal{B}_ϕ and g_ϕ from Definition 3.1. The convergence analysis depends on two geometric constants: $\alpha_{\mathcal{X}}(\mathcal{A})$ of the chosen \mathcal{A} , and the uniform bound

$$L_{\mathcal{X}}(\phi) := \max_{S \in \mathbb{S}_+^n \cap \mathcal{B}_\phi} \|S\|_{\mathcal{X}}^\circ.$$

$L_{\mathcal{X}}(\phi)$ is a finite constant and Cauchy-Schwarz inequality (2.4) yields

$$(5.6) \quad |\langle S, X_1 \rangle - \langle S, X_2 \rangle| \leq L_{\mathcal{X}}(\phi) \|X_1 - X_2\|_{\mathcal{X}} \quad \forall X_1, X_2 \in \mathbb{S}^n, \quad \forall S \in \mathbb{S}_+^n \cap \mathcal{B}_\phi.$$

In particular, $L_{\mathcal{X}}(\phi)$ is monotone under enlargements of \mathcal{B}_ϕ , so it is nondecreasing along the ℓ_p gauge and rank- k hierarchies.

LEMMA 5.1 (Finite ε -termination bound). *Assume that $\alpha_{\mathcal{X}}(\mathcal{A}) < \infty$. Let*

$$d := \dim(\mathbb{S}^n) = \frac{n(n+1)}{2}, \quad N_{\mathcal{X}}^\varepsilon(\mathcal{A}, \phi) := 1 + \left\lceil \left(1 + \frac{2\alpha_{\mathcal{X}}(\mathcal{A}) T L_{\mathcal{X}}(\phi) \phi(e_1)}{\varepsilon} \right)^d \right\rceil,$$

where e_1 is any canonical basis vector. Then there exists $i \in [N_{\mathcal{X}}^\varepsilon(\mathcal{A}, \phi)]$ such that one has $\lambda_{\min}(X^{(i)}) > -\varepsilon$. Equivalently, Algorithm 1 terminates, with stopping tolerance $\varepsilon > 0$, after at most $N_{\mathcal{X}}^\varepsilon(\mathcal{A}, \phi)$ iterations.

Proof. Fix $\varepsilon > 0$ and set $\eta := \varepsilon/\phi(e_1)$. Since $e_i/\phi(e_1)$ is feasible in the dual-norm formula (2.5) for every i , we have $\phi^\circ(y) \geq \|y\|_\infty/\phi(e_1)$ for every $y \in \mathbb{R}_+^n$. Hence, by Theorem 3.3, $g_\phi(X) \leq -\|\lambda_-(X)\|_\infty/\phi(e_1)$ for every $X \in \mathbb{S}^n$. If Algorithm 1 has not terminated at iteration i , then $\lambda_{\min}(X^{(i)}) \leq -\varepsilon$, and therefore $g_\phi(X^{(i)}) \leq -\eta$.

Choose $S^{(i)} \in \operatorname{argmin}_{S \in \mathbb{S}_+^n \cap \mathcal{B}_\phi} \langle S, X^{(i)} \rangle$. Then (5.6) gives, for every $Y \in \mathbb{S}^n$,

$$\langle S^{(i)}, Y \rangle \leq g_\phi(X^{(i)}) + L_{\mathcal{X}}(\phi) \|Y - X^{(i)}\|_{\mathcal{X}}.$$

Define $R = \eta/2L_{\mathcal{X}}(\phi)$. Hence, every Y satisfying $\|Y - X^{(i)}\|_{\mathcal{X}} < 2R$ violates the cut added at iteration i , so no later iterate can belong to that open ball. Therefore the family of balls $\mathcal{B}_{\mathcal{X}}(X^{(i)}, R)$ centered at iterates $X^{(i)}$ failing the stopping test, i.e., for all $i \in \mathcal{N}_\varepsilon := \{i \in \mathbb{N} : \lambda_{\min}(X^{(i)}) \leq -\varepsilon\}$, have pairwise disjoint interiors, and they all share the same volume $\operatorname{vol}_d(\mathcal{B}_{\mathcal{X}}(0, R)) = c_{\mathcal{X}}R^d$, in Lebesgue measure, for a constant $c_{\mathcal{X}} > 0$ related to d -space $(\mathbb{S}^n, \mathcal{X})$. By (5.2), all iterates lie in $\mathcal{R} \subseteq \mathcal{B}_{\mathcal{X}}(0, \alpha_{\mathcal{X}}(\mathcal{A}) T)$

so the family of disjoint balls is included in $\mathcal{B}_{\mathcal{X}}(0, \alpha_{\mathcal{X}}(\mathcal{A})T + R)$. Then, comparing volumes gives, for every finite subset $J \subseteq \mathcal{N}_{\varepsilon}$,

$$|J| c_{\mathcal{X}} R^d \leq c_{\mathcal{X}} (\alpha_{\mathcal{X}}(\mathcal{A})T + R)^d.$$

Since $c_{\mathcal{X}} > 0$, dividing by $c_{\mathcal{X}} R^d$ gives

$$|J| \leq \left(1 + \frac{\alpha_{\mathcal{X}}(\mathcal{A})T}{R}\right)^d = \left(1 + \frac{2\alpha_{\mathcal{X}}(\mathcal{A})T L_{\mathcal{X}}(\phi)\phi(e_1)}{\varepsilon}\right)^d.$$

Taking the supremum over all finite subsets $J \subseteq \mathcal{N}_{\varepsilon}$, we obtain

$$|\mathcal{N}_{\varepsilon}| = \sup_{J \subseteq \mathcal{N}_{\varepsilon}} \{|J| : J \text{ finite}\} < 1 + \left[\left(1 + \frac{2\alpha_{\mathcal{X}}(\mathcal{A})T L_{\mathcal{X}}(\phi)\phi(e_1)}{\varepsilon}\right)^d \right] = N_{\mathcal{X}}^{\varepsilon}(\mathcal{A}, \phi).$$

Thus, not all of the first $N_{\mathcal{X}}^{\varepsilon}(\mathcal{A}, \phi)$ iterates can fail the stopping test in Algorithm 1. \square

THEOREM 5.2 (Convergence of cut method). *Let $\mathcal{A} \subseteq \mathbb{S}^n$ be a closed convex cone with $\mathbb{S}_+^n \subseteq \mathcal{A}$, $\alpha_{\mathcal{X}}(\mathcal{A}) < \infty$, and $\text{tr}(X) > 0$ for all $X \in \mathcal{A} \setminus \{0\}$. Then, for any symmetric gauge ϕ , Algorithm 1, run with stopping tolerance $\varepsilon = 0$, either terminates finitely at an optimal solution of the SDP (1.1), or every limit point of the generated infinite sequence $(X^{(i)})_i$ is an optimal solution of the SDP (1.1).*

Proof. If Algorithm 1 terminates finitely, the returned point satisfies $\lambda_{\min}(X^{(i)}) > 0$. Since $X^{(i)}$ solves the relaxed problem \mathcal{M}_i in (5.1) and $\lambda_{\min}(X^{(i)}) > 0$, $X^{(i)}$ is feasible for (1.1). Hence, $X^{(i)}$ solves (1.1). Assume now that the algorithm generates an infinite sequence. Let \bar{X} be any limit point of $\{X^{(i)}\}_{i \geq 1}$, and take a subsequence $X^{(i_j)} \rightarrow \bar{X}$. Define $v_i := \langle C, X^{(i)} \rangle$. Since $\mathcal{R}^{(i+1)} \subseteq \mathcal{R}^{(i)}$, the sequence $(v_i)_i$ is nondecreasing; and since $\mathcal{F} \subseteq \mathcal{R}^{(i)}$ for all i , it is bounded above by

$$v^* := \min\{\langle C, X \rangle : X \in \mathcal{F} \neq \emptyset\}.$$

Hence $v_i \rightarrow \bar{v} \leq v^*$ for some \bar{v} .

Next, the proof of Lemma 5.1 shows that, for each $p \in \mathbb{N}$, only finitely many iterates satisfy $\lambda_{\min}(X^{(i)}) \leq -1/p$. Since the zero-tolerance run does not terminate, we have $\lambda_{\min}(X^{(i)}) \leq 0$ for every i , and therefore, along the subsequence chosen above, $\lambda_{\min}(X^{(i_j)}) \rightarrow 0$. By continuity of λ_{\min} , $\lambda_{\min}(\bar{X}) = 0$, hence $\bar{X} \succeq 0$. Since all iterates lie in the closed set \mathcal{R} and the affine constraints are preserved in the limit, we have $\bar{X} \in \mathcal{F}$. Finally, continuity of the objective gives

$$\langle C, \bar{X} \rangle = \lim_{j \rightarrow \infty} \langle C, X^{(i_j)} \rangle = \lim_{j \rightarrow \infty} v_{i_j} = \bar{v} \leq v^*.$$

But $\bar{X} \in \mathcal{F}$ implies $\langle C, \bar{X} \rangle \geq v^*$. Therefore $\langle C, \bar{X} \rangle = v^*$, and \bar{X} is optimal. \square

5.3. Illustration and comparison. The above analysis follows the proof in [7], extending it to any arbitrary norm \mathcal{X} in place of the Frobenius norm. This allows us, now, to exhibit a strict hierarchy between the considered instantiations of the framework, ℓ_p -gauge cuts for $p \in [1, \infty]$, rank- k cuts for $k \in [n]$, and factor-width- m relaxations for $m \in [n]$, whereas the termination bounds $N_{\mathcal{X}}^{\varepsilon}$ may collapse to equalities in the Frobenius setting. Let us define

$$\|X\|_{\mathcal{X}} := \|X\|_{\mathcal{X}_0} + \sum_{m=3}^n \delta_m \beta_m(X), \text{ where } \|X\|_{\mathcal{X}_0} := \max\{\|Xu\|_2 : u \in \mathbb{R}^n, \|u\|_2 = 1\}$$

is the operator norm of $X \in \mathbb{S}^n$ induced by ℓ_2 on \mathbb{R}^n [18, Def. 5.6.1], and

$$\beta_m(X) := \max_{\substack{I \subseteq [n] \\ |I|=m}} \|X_I\|_{\mathcal{X}_0}^\circ,$$

with $m = 3, \dots, n$, and $0 < \delta_n < \dots < \delta_3 < 1$.

Let $1 \leq p \leq \infty$ and q be the dual exponent. Since $\|S\|_{\mathcal{X}_0}^\circ = \text{tr}(S)$ for $S \in \mathbb{S}_+^n$,

$$L_{\mathcal{X}_0}(\ell_p) = \max \left\{ \mathbf{1}^\top z : z \in \mathbb{R}_+^n, \|z\|_p \leq 1 \right\} = n^{1/q}$$

Moreover, by Theorem 4.10, one has, for any rank $k \in [n]$

$$L_{\mathcal{X}_0}(\ell_p^{(k)}) = \max \left\{ \mathbf{1}^\top z : z \in \mathbb{R}_+^n, \|z\|_p \leq 1, \|z\|_0 \leq k \right\} = k^{1/q}.$$

Therefore, both sequences $(N_{\mathcal{X}}^\varepsilon(\mathcal{A}, \ell_p))_{p \geq 1}$ and $(N_{\mathcal{X}}^\varepsilon(\mathcal{A}, \ell_p^{(k)}))_{k \geq 1}$ for a fixed $p > 1$, are strictly increasing. Note in passing that, for $p = 1$, one has $(\ell_1^{(k)})^\circ(y) = \|\text{trun}_k(y)\|_\infty = \|y\|_\infty$ so the rank limit has no impact on spectral- ℓ_1 separation.

We now build the coefficients of \mathcal{X} so that, for any fixed gauge ϕ , the sequence $(N_{\mathcal{X}}^\varepsilon(\mathcal{A}_m, \phi))_{m \geq 1}$ is strictly decreasing along the factor-width relaxation hierarchy (5.5): $\mathcal{A}_1 = \mathcal{L}\mathcal{P}$, $\mathcal{A}_m = \mathbb{S}_+^{n,m}$ for $m \in \{2, \dots, n\}$. First, consider

$$W^{(2)} := \begin{pmatrix} a & c \\ c & 1-a \end{pmatrix}, \quad 0 < a < 1, \quad \sqrt{a(1-a)} < c \leq \frac{1}{2},$$

then $W^{(2)} \in \mathcal{L}\mathcal{P} \setminus \mathbb{S}_+^{n,2}$, $\text{tr}(W^{(2)}) = 1$, $\|W^{(2)}\|_{\mathcal{X}_0} > 1$, and thus $\alpha_{\mathcal{X}_0}(\mathcal{L}\mathcal{P}) > 1$. On the other hand, if $X \in \mathbb{S}_+^{n,2} = \mathcal{S}\mathcal{O}\mathcal{C}$ and $\text{tr}(X) = 1$, then for every $u \in \mathbb{R}^n$ with $\|u\|_2 = 1$,

$$|u^\top X u| \leq \sum_{j,l} |X_{jl}| |u_j| |u_l| \leq \left(\sum_j \sqrt{X_{jj}} |u_j| \right)^2 \leq \left(\sum_j X_{jj} \right) \left(\sum_j u_j^2 \right) = 1.$$

Hence $\alpha_{\mathcal{X}_0}(\mathbb{S}_+^{n,m}) = 1$ for every $m = 2, \dots, n$, since $\alpha_{\mathcal{X}_0}(\mathbb{S}_+^{n,m}) \leq \alpha_{\mathcal{X}_0}(\mathbb{S}_+^{n,2}) \leq 1$ and equality $u^\top X u = 1$ is attained by any rank-one matrix X with trace 1.

Given $m \in \{2, \dots, n-1\}$, if $X \in \mathbb{S}_+^{n,m+1}$ and $\text{tr}(X) = 1$, then every $(m+1) \times (m+1)$ principal submatrix I is PSD, so $\|X_I\|_{\mathcal{X}_0}^\circ = \text{tr}(X_I) \leq 1$, then $\beta_{m+1}(X) \leq 1$. Conversely, we build a matrix $W \in \mathbb{S}_+^{n,m} \setminus \mathbb{S}_+^{n,m+1}$ such that $\beta_{m+1}(W) > 1$, as follows:

$$\widehat{W} := \frac{1}{m+1} \left((1-\rho_m)I_{m+1} + \rho_m J_{m+1} \right), \quad -\frac{1}{m-1} \leq \rho_m < -\frac{1}{m},$$

where J_{m+1} is the all-ones matrix, and embed \widehat{W} as a principal block of an $n \times n$ matrix W . Then $\text{tr}(W) = 1$ and every $m \times m$ principal submatrix of W is PSD, but $W \notin \mathbb{S}_+^{n,m+1}$ because \widehat{W} has eigenvalue $(1+m\rho_m)/(m+1) < 0$. Moreover,

$$\beta_{m+1}(W) = \|\widehat{W}\|_{\mathcal{X}_0}^\circ = \frac{m-1-2m\rho_m}{m+1} > 1.$$

Therefore the coefficients δ_m can be chosen recursively for $m \geq 3$ so that the sequences $(\alpha_{\mathcal{X}}(\mathcal{A}_m))_{m \geq 1}$ and thus $(N_{\mathcal{X}}^\varepsilon(\mathcal{A}_m, \phi))_{m \geq 1}$ are strictly decreasing.

6. Numerical experiments. In this section, we analyze our proposed cutting-plane framework empirically across three aspects: versatility (the ease of implementing various spectral-gauge templates), relative performance (the individual strengths of different templates, including standard eigencuts), and absolute performance (comparison with an off-the-shelf interior-point SDP solver). We aim to be neither exhaustive, given the huge number of combinations of the SDP problem with outer-approximation cones and gauge templates, nor exclusive, that is, selecting the best combination and refining the method specifically for it.

Hence, our implementation of Algorithm 1 follows the basic Kelley scheme, without any cut-management or acceleration strategy, adding only one cut at each iteration, namely the one with the largest violation at the current iterate. We compare various instantiations of this framework, considering either the LP or SOC relaxations combined with some of the studied spectral-gauge templates, namely eigenvalue, Frobenius, and nuclear cuts, without or with rank restrictions. We evaluate the dual bounds computed by these combinations on two distinct classes of NP-hard problems commonly addressed through SDP relaxations: box-constrained quadratic programs (BoxQP) [10, 15], and sparse principal component analysis (SPCA) [7].

All numerical experiments are conducted on an Apple M3 Pro with a 5-core 4.05 GHz performance CPU, a 6-core 2.75 GHz efficiency CPU, and 36 GB RAM. The code¹ is implemented in Python 3.12, and the eigenvalue decompositions called for separation are computed with `numpy.linalg.eigh`. The master problem \mathcal{M}_i (5.1) is solved using the homogeneous barrier algorithm in Gurobi v13.0, on both LP and SOC formulations. We measure the performance of Algorithm 1 using the *gap closed* [15]:

$$GC^T = 100 \cdot \frac{MP_T - LP}{best - LP},$$

where MP_T is the value of the last master problem solved by Algorithm 1 before the time limit T and LP is the optimal value of the initial LP relaxation (5.4). For small instances (typically, $n \leq 100$), *best* is the optimal value of the SDP relaxation (1.1), computed by running the primal-dual interior-point solver MOSEK v11.1.10 with tolerance 10^{-6} . We indicate with t_{sdp} the corresponding MOSEK computation time, and evaluate the gap closed at $T \in \{t_{sdp}, 60s\}$. Since interior-point methods become expensive in time and memory as n grows, for larger instances, we define *best* as the largest dual bound obtained by any tested cut strategy within 30 minutes. We then evaluate the gap closed relative to this virtual best strategy at $T \in \{60s, 1800s\}$.

Cut strategies. We implement several spectral-gauge cut templates: eigenvalue cuts (Corollary 4.2), Frobenius cuts (Corollary 4.3), nuclear cuts (Corollary 4.4). In the tables below, the results obtained using these strategies are reported in the columns labeled ℓ_1 , ℓ_2 , and ℓ_∞ , respectively. We also generate spectral cuts using a dynamic rank parameterization (Theorem 4.10). This is implemented by selecting the negative eigenvalues belonging to the interval $[\lambda_1(X), 0.9\lambda_1(X)]$, that is, those closest to the smallest one. The columns labeled as $\ell_*^{90\%}$ refer to the results obtained with this strategy. In this same dynamic rank setting, we additionally consider selecting a varying number of the most negative eigenvalues. This number changes across iterations and is selected from a discrete set whose dimension depends on the problem dimension. For instance, for $n = 20$, we randomly select a number in $[1, 5, 10]$ or all eigenvalues, whereas for $n = 250$, we consider $[1, 5, 10, 20]$ or all eigenvalues. We use the notation ℓ_*^{rand} for this last case.

¹The code is available at <https://github.com/sofdem/sdpgauge26>.

TABLE 6.1

BoxQP “basic” instances [37]: average gap closed GC^T (% of the SDP value) for $T = t_{sdp}$.

n	d	#	T	LP				SOC	
				ℓ_1	ℓ_∞	ℓ_2	$\ell_\infty^{90\%}$	ℓ_1	ℓ_2
20	100	3	0.3	99.9	99.8	99.9	99.9	58.9	91.7
30	[60,100]	15	2.3	94.8	94.3	95.6	95.0	46.4	87.2
40	[30,100]	24	9.0	88.6	86.2	90.8	89.5	30.5	83.2
50	[30,50]	9	25.9	71.8	67.0	75.9	74.4	-	-
60	20	3	56.8	33.7	28.0	38.6	42.6	-	-

6.1. Box-constrained quadratic programs.

We consider the problem

$$\min x^\top Qx + c^\top x : x \in [0, 1]^n,$$

where $Q \in \mathbb{S}^n$ is not positive-definite and $c \in \mathbb{R}^n$. A standard way to obtain a tractable bound is to use an SDP relaxation [15, 22]. The lifting approach linearizes the quadratic form $x^\top Qx = \text{tr}(Qxx^\top) = \langle Q, xx^\top \rangle$ by introducing a symmetric matrix variable $X = xx^\top$. Since the set of matrices of the form xx^\top is nonconvex, the equality $X = xx^\top$ is relaxed to $X - xx^\top \succeq 0$. The resulting SDP relaxation is

$$\min \langle Q, X \rangle + c^\top x : X - xx^\top \succeq 0, x \in [0, 1]^n, X \in \mathbb{S}^n.$$

We relax the PSD condition as either (5.3) or (5.4), and add McCormick inequalities:

$$X_{jl} \geq 0, \quad X_{jl} \leq x_j, \quad X_{jl} \leq x_l, \quad X_{jl} \geq x_j + x_l - 1.$$

The experiments rely on the BoxQP library: the “basic” instances with $n \in [20, 60]$ from [37] and newly benchmarked sets from the generator in [10], made of 3 instances for each dimension $n \in \{30, 250\}$ and density $d \in \{25\%, 50\%, 75\%, 100\%$ }.²

Comparison with an SDP interior-point solver. We first evaluate the dual bounds obtained by Algorithm 1 using the different strategies above for the same time $T = t_{sdp}$ required by MOSEK to compute the SDP bound. In Table 6.1, we report the gap closed value on the “basic” instances [37], averaged over the number of instances (column #) per dimension n , and in Table 6.2 (top part), the gap closed on the generated set with dimension $n = 30$, averaged over the 3 instances per density d .

For the smallest instances ($n = 20$), the LP-based cutting-plane methods close a large fraction of the SDP gap within the MOSEK running time. For all other instances, except the largest one, the Frobenius cuts are particularly competitive in terms of the closed gap. In the setting of a very short time limit, the use of the SOC relaxation is discouraged. In fact, only a few iterations of Algorithm 1 are performed, leading to smaller closed gap values. As an example, the number of iterations for the case $n = 20$ is an average of 6 for SOC with eigenvalue cuts, compared with 106 for the LP relaxation. The notation ‘-’ in the table indicates that the allowed computation time was not sufficient to perform even a single iteration for at least one instance.

On the generated set $n = 30$, we can observe the relationship between algorithm performance and instance density. Regardless of the cut strategy and relaxation type, the denser the instance, the larger the gap closed. Indeed, as observed in [15], when the objective function has many zero coefficients, the objective value tends to vary less across iterations of a cutting-plane algorithm. The Frobenius cuts outperform the other strategies in both the LP and SOC relaxations, and the randomized rank-setting with LP relaxation appears to be equally robust.

²The instances and generator are available at https://github.com/sburer/BoxQP_instances/.

TABLE 6.2

BoxQP instances with $n = 30$: average gap closed GC^T (% of the SDP value) for $T = t_{sdp}$ (top) and $T = 60$ seconds (bottom).

d	#	T	LP						SOC		
			ℓ_1	ℓ_∞	ℓ_2	$\ell_\infty^{90\%}$	$\ell_2^{90\%}$	ℓ_∞^{rand}	ℓ_1	ℓ_∞	ℓ_2
25	3	1.4	86.9	85.7	90.9	89.2	87.7	88.4	25.7	53.2	61.0
50	3	1.8	90.4	89.2	91.4	90.7	90.9	91.6	4.1	71.1	79.5
75	3	3.0	94.0	94.1	95.6	94.2	94.3	95.7	35.3	81.4	86.7
100	3	3.0	95.7	95.4	96.9	95.9	96.1	96.8	51.2	84.4	87.2
25	3	60	100	100	100	100	100	100	95.2	94.6	96.1
50	3	60	99.2	96.2	97.2	98.9	98.9	98.2	92.8	92.4	94.3
75	3	60	99.7	98.7	99.0	99.6	99.6	99.3	96.1	95.7	96.6
100	3	60	99.6	98.5	99.0	99.5	99.5	99.3	96.8	96.3	97.4

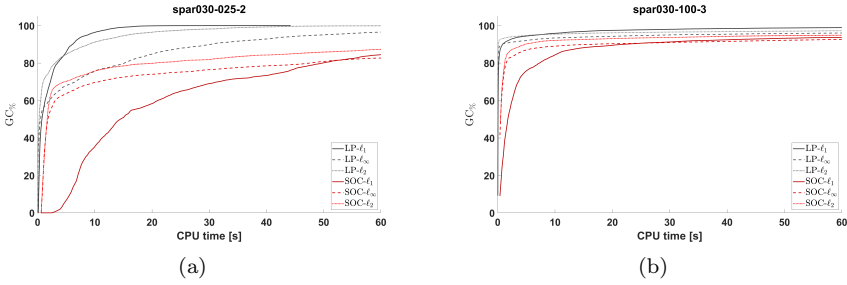


FIG. 6.1. Gap closed evolution for two BoxQP instances with $n = 30$ considering $T = 60$ seconds.

Convergence on the small instances ($n = 30$). We now analyze how the dual bound evolves over the iterations of the cutting-plane algorithm. At the bottom part of Table 6.2, we report the average gap closed when the allowed computation time is set to $T = 60$ seconds on the generated set $n = 30$. In the LP setting, the eigencuts strategy proves to be the most effective across the different density values, and all cut families could close the gap in less than one minute for the sparser instances, except for the full-rank cuts on instance 030-025-2 (ℓ_2 stops after 120 seconds and ℓ_∞ after 263 seconds). This instance is particularly challenging for the cutting-plane algorithm as the gap closed at $T = t_{sdp}$ is around 72% under the best cut strategy, whereas it is around 99.8% for the other two instances with $d = 25\%$.

Figure 6.1 illustrates the evolution of the gap closed over the computation time, for this instance, on the left, and for the full dense case 030-100-3, on the right. We observe that the LP relaxation dominates the SOC relaxation. Regarding cuts, Frobenius cuts are the most effective across the first seconds. After at most 10 seconds, the LP relaxation with eigencuts allows the largest gap closed, and it is the first configuration to terminate, in 44 seconds, on the sparse instance.

We also evaluate the performance of the different cut templates in the LP setting at convergence, by running Algorithm 1 on set $n = 30$ for at most two hours. In Table 6.3, we report the average CPU time to terminate with tolerance $\varepsilon = 10^{-6}$. If the algorithm fails to converge in two hours, we report in parentheses the order of the remaining gap in the worst case. The results confirm that eigenvalue cuts converge fastest to PSD feasibility, although they improve the dual bound slowly during the earliest iterations, as illustrated in Figure 6.1. The denser the instance, the longer it takes to converge. Dynamically varying the rank of the separator, as in the $\ell_\infty^{90\%}$

and $\ell_2^{90\%}$ strategies, also allows Algorithm 1 to converge for two density values and to reach a good approximation of the PSD solution in the other cases.

TABLE 6.3

BoxQP instances ($n = 30$): average CPU time in seconds when convergence occurs ($\varepsilon = 10^{-6}$); otherwise, in parentheses, the order of the largest remaining gap $|\lambda_{\min}(X)|$ after two hours.

d	#	SDP	LP					
		t_{sdp}	ℓ_1	ℓ_∞	ℓ_2	$\ell_\infty^{90\%}$	$\ell_2^{90\%}$	ℓ_∞^{rand}
25	3	1.4	16.4	89.7	40.8	18.0	17.3	31.0
50	3	1.8	1914	(10^{-2})	(10^{-3})	(10^{-5})	(10^{-5})	(10^{-4})
75	3	3.0	1186	(10^{-3})	(10^{-4})	2233	2191	(10^{-5})
100	3	3.0	2491	(10^{-2})	(10^{-3})	(10^{-5})	(10^{-5})	(10^{-3})

Again, we observe substantial variability in the results across instances with the same nominal size and density: although the shape of the gap closed evolution is comparable, the total computation time can differ significantly. For example, for $d = 100$, all tested strategies converge in less than 200 seconds on instance 030-100-2, whereas they require more than 6200 seconds on instance 030-100-3. For $d = 50$, the time for convergence is less than one second on instance 030-050-1 considering all strategies, whereas it exceeds more than 5400 seconds on instance 030-050-3.

Large instances ($n = 250$). For the largest generated BoxQP instances, MOSEK cannot solve the SDP relaxation with the available memory or within a reasonable time, and the SOC relaxation is also too expensive. Hence, Table 6.4 reports the gap closed relative to the best bound obtained by any tested cut strategy in the LP setting within 30 minutes. Frobenius cuts dominate the short-time (i.e., 60 seconds) bound improvement for all density values except the smallest. Over longer runs (i.e., 1800 seconds), randomized rank cuts become competitive with Frobenius cuts. We select two instances, sparse and dense, and show the lower bound improvement over CPU time in Figure 6.2. Frobenius and nuclear cuts quickly improve the bound, then plateau after about 300 seconds for the sparse case and 100 seconds for the dense case, and Frobenius cuts compute better bounds. The eigencut strategy leads to a very slow bound improvement, in particular, for the small instances.

TABLE 6.4

BoxQP instances ($n = 250$): average gap closed GCT relative to the best bound found within 30 minutes, for $T = 60$ seconds (top) and $T = 1800$ seconds (bottom).

d	#	T	LP				
			ℓ_1	ℓ_∞	ℓ_2	$\ell_2^{90\%}$	ℓ_∞^{rand}
25	3	60	0.0	39.4	31.1	0.3	18.2
50	3	60	0.3	59.8	62.1	17.0	40.6
75	3	60	7.9	64.7	69.5	32.8	44.4
100	3	60	17.8	70.5	78.4	44.7	63.8
25	3	1800	5.9	87.3	100	75.1	95.0
50	3	1800	35.6	93.2	100	89.3	97.2
75	3	1800	51.8	95.3	100	91.9	97.6
100	3	1800	58.6	96.3	100	93.5	98.4

Enforcing a maximum rank bound $m \in [n]$ has no impact on the density of the separators, which remain fully dense in these experiments. However, we observe an effect on the LP solution time immediately after the cut is added. As an example, when running ℓ_∞^{rand} for 30 minutes on instance 250-050-2: on average, solving the LP takes 6 seconds after adding an eigencut (14 iterations with rank $k = 1$) and 37 seconds after adding a nuclear cut (14 iterations with average rank $k = 118$).

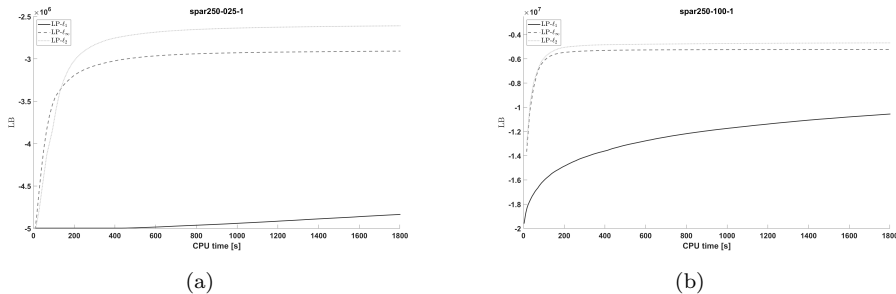


FIG. 6.2. Lower bound evolution for two BoxQP instances with $n = 250$ over $T = 1800$ seconds.

6.2. Sparse principal component analysis. Given a normalized and centered data matrix $A \in \mathbb{R}^{p \times n}$ with p observations in an n -dimensional feature space, let $\Sigma := \frac{1}{p-1} A^\top A \in \mathbb{S}^n$ be the sample covariance matrix. Sparse principal component analysis (SPCA) seeks directions that explain substantial variance while promoting sparsity in the loading vector. The standard formulation is

$$\max x^\top \Sigma x : \|x\|_2 = 1, \|x\|_0 \leq t, x \in \mathbb{R}^n,$$

where $\|x\|_0$ denotes the number of nonzero entries in x . Because of the cardinality constraint $\|x\|_0 \leq t$, this problem is NP-hard, but it admits an SDP relaxation, by setting $X = xx^\top$. The normalization constraint $\|x\|_2 = 1$ implies $\text{tr}(X) = 1$, $X \succeq 0$, and $\text{rank}(X) = 1$, and $\|x\|_0 \leq t$ implies $\|X\|_1 = \sum_{j,l} |X_{jl}| \leq t$. Dropping the rank constraint yields the convex relaxation in [14]:

$$(6.1) \quad \min \langle -\Sigma, X \rangle : \text{tr}(X) = 1, X \succeq 0, \|X\|_1 \leq t, X \in \mathbb{S}^n.$$

The SPCA instances are selected as in [7] from the UCI ML repository [20] and we arbitrarily fix density $t = 10$. For each instance, we first cleaned the data to work only with numerical predictive variables, following the UCI repository documentation. Features with zero variance led to zero standard deviation and NaN values after standardization. We removed these features before constructing Σ (i.e., deleted their rows/columns). This affected the ionosphere, lung, arrhythmia, gait, gastro, and micromass datasets. Reported dimensions reflect this preprocessing.

Small instances ($n < 250$). We first focus on the smallest SPCA instances, for which a lower bound can be computed in seconds by solving the SDP relaxation with MOSEK. As before, we consider two settings for the maximum computation time T . The top of Table 6.5 reports results with $T = t_{sdp}$, while the bottom part reports results obtained with our algorithm running for 60 seconds. Unlike the BoxQP instances, the SOC relaxation is competitive in short-time settings, particularly when using the Frobenius cuts template. When the time limit is extended to 60 seconds, our algorithm shows comparable results across the two relaxations and different templates for the two smallest instances ($n = 13$). For the other instances, eigencuts and dynamic rank strategy yield the best performance in each of the considered relaxations and, overall, in the SOC relaxation. The evolution of the gap-closed is given in Figure 6.3 for the instances `pitprops` ($n = 13$) and `communities` ($n = 101$). For the first case, we can see that once the relaxation is fixed, the eigencuts strategy closes the gap the fastest, followed by the Frobenius cuts and finally the nuclear cuts. We mention that

TABLE 6.5

SPCA small instances with $t = 10$: average gap closed GC^T (% of the SDP value) for $T = t_{sdp}$ (top) and $T = 60$ seconds (bottom).

name	n	T	LP				SOC			
			ℓ_1	ℓ_∞	ℓ_2	$\ell_2^{90\%}$	ℓ_1	ℓ_∞	ℓ_2	$\ell_2^{90\%}$
pitprops	13	0.05	88.8	86.7	89.0	90.3	92.8	83.0	95.2	92.1
wine	13	0.03	90.5	84.6	89.2	90.9	91.3	83.5	95.1	91.3
ionosphere	34	0.34	82.1	58.5	76.9	82.4	48.6	27.2	49.6	48.6
lung	54	1.03	69.1	56.2	65.6	68.6	67.4	49.6	80.2	67.4
geography	68	2.66	74.1	6.7	66.8	75.2	87.2	75.5	85.1	84.9
communities	101	32.3	75.6	34.8	57.5	74.2	79.2	28.7	73.9	79.5
pitprops	13	60	100	100	100	100	100	98.5	99.8	100
wine	13	60	100	100	100	100	100	99.0	99.9	100
ionosphere	34	60	91.2	69.4	85.5	90.8	93.6	62.3	90.9	93.4
lung	54	60	83.1	59.9	77.1	83.1	84.6	52.9	85.0	86.0
geography	68	60	85.3	19.3	73.1	85.0	92.2	76.2	88.2	90.9
communities	101	60	77.2	35.1	59.9	76.1	81.3	29.1	74.8	81.7

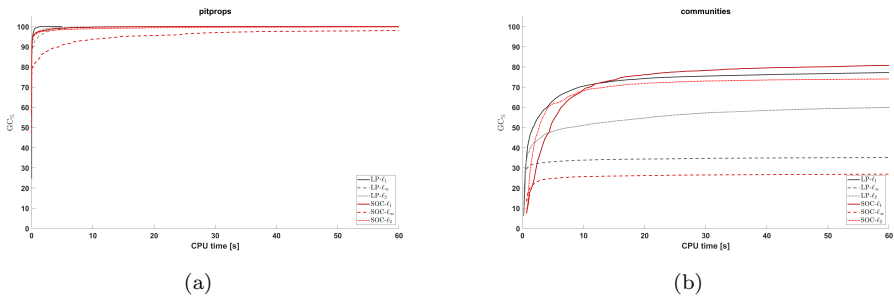


FIG. 6.3. Gap closed evolution for two small SPCA instances with (a) $n = 13$ and (b) $n = 101$.

in the first 0.15 seconds, the SOC relaxation with either the Frobenius cuts or the eigencuts has the highest closed gap. After that, this value is overcome by combining the LP relaxation with the eigencuts, which allows the gap to be closed in less than 10 seconds. This behavior is reversed for the other instance, where the LP relaxation with eigencuts dominates in the first 10 seconds, while the SOC relaxation allows better results in the long run. However, once the relaxation is fixed, the performance of the different cut strategies is the same as in the smallest case.

Large instances ($n \geq 250$). On larger SPCA instances, MOSEK is unable to find a solution due to memory issues, so we compute the gap closed relative to the best value achieved within 30 minutes among all cut strategies. In the upper part of Table 6.6, we report the results of running Algorithm 1 with a time limit set to 60 seconds. The SOC relaxation yields the best early strategy for the smallest instance with Frobenius cuts, but its solution time exceeds 1 minute on the three largest instances. For all instances but two, the LP relaxation with eigencuts provides the largest gap closed. Even with a 1800-second time limit, SOC relaxation is not recommended for the largest instances, where LP relaxation with eigencut or dynamic rank achieves high gap-closed values. Figure 6.4 shows the evolution of the lower bound for two selected instances, underscoring once again that the best combination, that is, relaxation and cut strategy, is instance dependent. Indeed, with $T = 1800$ seconds, SOC relaxation with eigencuts yields the best lower bound for *arrhythmia* but is the worst strategy on the largest instance *micromass*.

TABLE 6.6

SPCA large instances with $t = 10$: average gap closed GC^T (% of the best value) for $T = 60$ seconds (top) and $T = 1800$ seconds (bottom).

name	n	T	LP				SOC			
			ℓ_1	ℓ_∞	ℓ_2	$\ell_2^{90\%}$	ℓ_1	ℓ_∞	ℓ_2	$\ell_2^{90\%}$
arrhythmia	257	60	76.8	53.7	65.5	76.3	65.3	44.0	80.7	65.3
voice	310	60	27.2	3.9	8.6	24.7	3.2	1.3	4.0	3.2
gait	320	60	62.3	22.0	52.3	61.2	32.5	14.9	39.3	33.3
gastro	466	60	12.5	0.5	3.0	12.4	-	-	-	-
parkinson	754	60	15.0	1.8	10.8	15.0	-	-	-	-
micromass	1139	60	19.1	26.1	23.7	19.2	-	-	-	-
arrhythmia	257	1800	95.8	57.5	77.3	96.3	99.6	48.5	95.7	100
voice	310	1800	98.1	4.4	26.3	100	81.1	3.4	48.3	96.2
gait	320	1800	92.1	29.7	63.3	91.4	99.0	24.7	86.7	100
gastro	466	1800	99.9	0.6	21.0	100	38.6	0.5	0.7	63.8
parkinson	754	1800	98.6	2.8	39.4	100	17.9	2.3	18.8	21.2
micromass	1139	1800	100	67.0	79.7	98.1	27.3	60.7	35.3	27.3

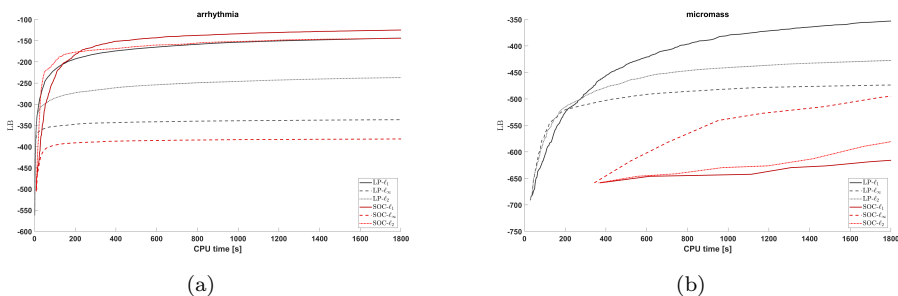


FIG. 6.4. Lower bound evolution for two large SPCA instances with (a) $n = 257$ and (b) $n = 1139$.

7. Conclusions and future work. We introduce spectral-gauge cuts, a canonical family of cutting planes for positive semidefinite (PSD) programming that leverage spectral information through gauge functions. The proposed framework unifies and extends existing outer-approximations of the PSD cone, taking sparsity and rank constraints into account. On the theoretical side, we characterized optimal separators, and we analyzed how the choice of gauge and normalization affects the cut strength.

Numerical experiments on box-constrained quadratic programs and sparse principal component analysis show that, in general, combining an LP relaxation with the classical eigencuts or the proposed Frobenius cuts provides a robust trade-off between bound quality and computational burden. This is even more evident in large instances where interior-point methods encounter memory limitations, our algorithm remains applicable and produces good-quality bounds.

Theory and practice indicate that no spectral-gauge cut template is strictly dominated by the others. Hence, future work could be dedicated to designing adaptive gauge-selection and dynamic-rank strategies.

REFERENCES

- [1] A. A. AHMADI AND A. MAJUMDAR, *DSOS and SDSOS optimization: More tractable alternatives to sum of squares and semidefinite optimization*, SIAM J. Appl. Algebra Geom., 3 (2019), pp. 193–230.

- [2] M. S. ANDERSEN, J. DAHL, Z. LIU, AND L. VANDENBERGHE, *Interior-point methods for large-scale cone programming*, in Optimization for Machine Learning, S. Sra, S. Nowozin, and S. J. Wright, eds., MIT Press, 2011, pp. 55–83.
- [3] R. BALTEAN-LUGOJAN, P. BONAMI, R. MISENER, AND A. TRAMONTANI, *Scoring positive semidefinite cutting planes for quadratic optimization via trained neural networks*. Optimization Online preprint, 2019.
- [4] F. BATTISTA AND M. DE SANTIS, *Dealing with inequality constraints in large-scale semidefinite relaxations for graph coloring and maximum clique problems*, 4OR, 23 (2025), pp. 65–95.
- [5] S. BELLAVIA, J. GONDZIO, AND M. PORCELLI, *A relaxed interior-point method for low-rank semidefinite programming problems with applications to matrix completion*, J. Sci. Comput., 89 (2021).
- [6] L. BERK AND D. BERTSIMAS, *Certifiably optimal sparse principal component analysis*, Math. Program. Comput., 11 (2019), pp. 381–420.
- [7] D. BERTSIMAS AND R. CORY-WRIGHT, *On polyhedral and second-order cone decompositions of semidefinite optimization problems*, Oper. Res. Lett., 48 (2020), pp. 78–85.
- [8] G. BLEKHERMAN, P. A. PARRILO, AND R. R. THOMAS, eds., *Semidefinite optimization and convex algebraic geometry*, vol. 13 of MOS-SIAM Series on Optimization, SIAM, 2012.
- [9] E. G. BOMAN, D. CHEN, O. PAREKH, AND S. TOLEDO, *On factor width and symmetric H-matrices*, Linear Algebra Appl., 405 (2005), pp. 239–248.
- [10] S. BURER AND D. VANDENBUSSCHE, *Globally solving box-constrained nonconvex quadratic programs with semidefinite-based finite branch-and-bound*, Comput. Optim. Appl., 43 (2009), pp. 181–195.
- [11] V. CHANDRASEKARAN, B. RECHT, P. A. PARRILO, AND A. S. WILLSKY, *The convex geometry of linear inverse problems*, Found. Comput. Math., 12 (2012), pp. 805–849.
- [12] S. CHEN, Y.-J. LIU, J. YU, AND W. ZHOU, *A semismooth Newton-based augmented Lagrangian algorithm for Lovász theta SDP problem*, Optimization, (2025), pp. 1–22.
- [13] A. DANILIDIS, D. DRUSVYATSKIY, AND A. S. LEWIS, *Orthogonal invariance and identifiability*, SIAM J. on Matrix Anal. and Appl., 35 (2014), pp. 580–598.
- [14] A. D’ASPROMONT, L. EL GHAOUL, M. I. JORDAN, AND G. R. LANCKRIET, *A direct formulation for sparse pca using semidefinite programming*, SIAM Rev., 49 (2007), pp. 434–448.
- [15] S. S. DEY, A. M. KAZACHKOV, A. LODI, AND G. MUÑOZ, *Cutting plane generation through sparse principal component analysis*, SIAM J. Optim., 32 (2022), pp. 1319–1343.
- [16] S. S. DEY AND M. MOLINARO, *Theoretical challenges towards cutting-plane selection*, Math. Program., 170 (2018), pp. 237–266.
- [17] O. GÜNLÜK, P. JÜNGER, J. LINDEROTH, A. LODI, AND J. LUEDTKE, *Sparse cuts for the positive semidefinite cone*. arXiv:2603.09864 preprint, 2026.
- [18] R. A. HORN AND C. R. JOHNSON, *Matrix analysis*, Cambridge University Press, 2 ed., 2013.
- [19] J. E. KELLEY, JR., *The cutting-plane method for solving convex programs*, J. Soc. Ind. Appl. Math., 8 (1960), pp. 703–712.
- [20] M. KELLY, R. LONGJOHN, AND K. NOTTINGHAM, *The UCI machine learning repository*. <https://archive.ics.uci.edu>.
- [21] K. KRISHNAN AND J. E. MITCHELL, *A unifying framework for several cutting plane methods for semidefinite programming*, Optim. Methods Softw., 21 (2006), pp. 57–74.
- [22] M. LOCATELLI, V. PICCIALI, AND A. M. SUDOSO, *Fix and bound: an efficient approach for solving large-scale quadratic programming problems with box constraints*, Math. Program. Comput., 17 (2025), pp. 231–263.
- [23] J. MALICK, J. POVH, F. RENDL, AND A. WIEGELE, *Regularization methods for semidefinite programming*, SIAM J. Optim., 20 (2009), pp. 336–356.
- [24] Y. NESTEROV AND A. NEMIROVSKII, *Interior-point polynomial algorithms in convex programming*, vol. 13 of SIAM Studies in Applied Mathematics, SIAM, 1994.
- [25] F. PERMENTER AND P. A. PARRILO, *Partial facial reduction: Simplified, equivalent SDPs via approximations of the PSD cone*, Math. Program., 171 (2018), pp. 1–54.
- [26] J. POVH, F. RENDL, AND A. WIEGELE, *A boundary point method to solve semidefinite programs*, Computing, 78 (2006), pp. 277–286.
- [27] A. QUALIZZA, P. BELOTTI, AND F. MARGOT, *Linear programming relaxations of quadratically constrained quadratic programs*, in Mixed Integer Nonlinear Programming, J. Lee and S. Leyffer, eds., vol. 154 of IMA Vol. in Math. and Appl., Springer, 2012, pp. 407–426.
- [28] M. V. RAMANA, *An algorithmic analysis of multiquadratic and semidefinite programming problems*, PhD thesis, Johns Hopkins University, Baltimore, MD, 1993.
- [29] M. V. RAMANA AND A. J. GOLDMAN, *Some geometric results in semidefinite programming*, J. Glob. Optim., 7 (1995), pp. 33–50.
- [30] W. RUDIN, *Functional Analysis*, McGraw–Hill, 2 ed., 1991.

- [31] H. D. SHERALI, E. DALKIRAN, AND J. DESAI, *Enhancing RLT-based relaxations for polynomial programming problems via a new class of v -semidefinite cuts*, *Comput. Optim. Appl.*, 52 (2012), pp. 483–506.
- [32] H. D. SHERALI AND B. M. P. FRATICELLI, *Enhancing RLT relaxations via a new class of semidefinite cuts*, *J. Glob. Optim.*, 22 (2002), pp. 233–261.
- [33] N. Z. SHOR, *Quadratic optimization problems*, *Sov. J. Comput. Syst. Sci.*, 25 (1987), pp. 1–11. Translated from *Tekhnicheskaya Kibernetika*, No. 1 (1987), pp. 128–139.
- [34] N. Z. SHOR, *Nondifferentiable optimization and polynomial problems*, vol. 24 of *Nonconvex Optimization and Its Applications*, Springer, Boston, MA, 1998.
- [35] D. SUN, K.-C. TOH, Y. YUAN, AND X.-Y. ZHAO, *SDPNAL+: A Matlab software for semidefinite programming with bound constraints*, *Optim. Methods Softw.*, 35 (2020), pp. 87–115.
- [36] L. VANDENBERGHE AND S. BOYD, *Semidefinite programming*, *SIAM Rev.*, 38 (1996), pp. 49–95.
- [37] D. VANDENBUSSCHE AND G. NEMHAUSER, *A branch-and-cut algorithm for nonconvex quadratic programs with box constraints*, *Math. Program.*, 102 (2005), pp. 559–575.
- [38] Y. WANG, A. TANAKA, AND A. YOSHISE, *Polyhedral approximations of the semidefinite cone and their application*, *Comput. Optim. Appl.*, 78 (2021), pp. 893–913.
- [39] F. WESSELMANN AND U. H. SUHL, *Implementing cutting plane management and selection techniques*, tech. report, University of Paderborn, 2012.
- [40] E. A. YILDIRIM, *Relaxations of KKT conditions do not strengthen finite RLT and SDP-RLT bounds for nonconvex quadratic programs*, *J. Glob. Optim.*, 94 (2026), pp. 891–918.

Appendix A. Proof of Proposition 4.8. We recall the two notions used below. A set $C \subseteq \mathbb{R}^n$ is *balanced* if $\alpha C \subseteq C$ for every scalar α with $|\alpha| \leq 1$, and it is *absorbing* if, for every $z \in \mathbb{R}^n$, there exists $t > 0$ such that $z \in tC$.

Proof of Proposition 4.8. The set $\mathcal{A}_k(\phi)$ is compact since it is the intersection of the compact ϕ -unit ball with the finite union of coordinate subspaces corresponding to supports of size at most k . Hence $\mathcal{K}_k(\phi) = \text{conv}(\mathcal{A}_k(\phi))$ is compact and convex. Moreover, $\mathcal{A}_k(\phi)$ is balanced. Indeed, if $u \in \mathcal{A}_k(\phi)$ and $|\alpha| \leq 1$, then

$$\phi(\alpha u) = |\alpha| \phi(u) \leq 1, \quad \|\alpha u\|_0 \leq \|u\|_0 \leq k.$$

Therefore, $\mathcal{K}_k(\phi)$ is balanced as well.

Let $c := \phi(e_1) > 0$. Since ϕ is a symmetric gauge, $\phi(e_i) = c$ for every $i \in [n]$, and hence $\pm e_i/c \in \mathcal{A}_k(\phi)$ for every $i \in [n]$. Thus

$$\text{conv} \left\{ \pm \frac{e_i}{c} : i \in [n] \right\} \subseteq \mathcal{K}_k(\phi).$$

The set on the left is a full-dimensional cross-polytope, so $\mathcal{K}_k(\phi)$ is absorbing.

The function $\phi^{(k)}$ in (4.4) is therefore the gauge function of a compact, convex, balanced, absorbing set. It is consequently a seminorm [30, Thm. 1.35].

To see that it is positive definite, note that

$$\mathcal{A}_k(\phi) \subseteq \{u : \phi(u) \leq 1\}.$$

Since the ϕ -unit ball is convex, $\mathcal{K}_k(\phi) \subseteq \{u : \phi(u) \leq 1\}$. Hence, whenever $z \in t\mathcal{K}_k(\phi)$, we have $\phi(z) \leq t$. Taking the infimum over such $t > 0$ yields

$$\phi(z) \leq \phi^{(k)}(z).$$

Thus $\phi^{(k)}(z) > 0$ for every $z \neq 0$, and $\phi^{(k)}$ is a norm.

Finally, $\mathcal{A}_k(\phi)$ is invariant under sign changes and permutations, because both ϕ and $\|\cdot\|_0$ are invariant under these operations. The same is true of $\mathcal{K}_k(\phi)$, and therefore of its gauge. Hence $\phi^{(k)}$ is a symmetric gauge. \square