

A predictor-corrector algorithm for semidefinite programming that uses the factor width cone

Authors	Kirschner,Felix; de Klerk,Etienne
Published in	Vietnam Journal of Mathematics
DOI	10.1007/s10013-023-00666-8
Publication Date	2025-07
Link	https://research.tilburguniversity.edu/en/publications/957e76ec-7f75-4e6e-adc4-c19d773cca44
Citation	Kirschner, F & de Klerk, E 2025, 'A predictor-corrector algorithm for semidefinite programming that uses the factor width cone', Vietnam Journal of Mathematics, vol. 53, no. 3, pp. 495-515. https://doi.org/10.1007/s10013-023-00666-8
Download Date	2026-04-13 08:24:37
Rights	<p>General rights</p> <p>Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.</p> <ul style="list-style-type: none"> - Users may download and print one copy of any publication from the public portal for the purpose of private study or research. - You may not further distribute the material or use it for any profit-making activity or commercial gain - You may freely distribute the URL identifying the publication in the public portal" <p>Take down policy</p> <p>If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.</p>

A predictor-corrector algorithm for semidefinite programming that uses the factor width cone

Felix Kirschner^{1*}† and Etienne de Klerk^{1†}

^{1*}Department for Econometrics and OR, Tilburg University,
Warandelaan 2, Tilburg, 5037 AB, Netherlands.

*Corresponding author(s). E-mail(s):

f.c.kirschner@tilburguniversity.edu;

Contributing authors: e.deklerk@tilburguniversity.edu;

†These authors contributed equally to this work.

Abstract

We propose an interior point method (IPM) for solving semidefinite programming problems (SDPs). The standard interior point algorithms used to solve SDPs work in the space of positive semidefinite matrices. Contrary to that the proposed algorithm works in the cone of matrices of constant *factor width*. We prove global convergence and provide a complexity analysis. Our work is inspired by a series of papers by Ahmadi, Dash, Majumdar and Hall, and builds upon a recent preprint by Roig-Solvas and Sznaier [arXiv:2202.12374, 2022].

Keywords: Conic optimization, factor width cone, semidefinite programming, interior point methods

1 Introduction

Semidefinite programming problems (SDPs) are a generalization of linear programming problems (LPs). While capturing a much larger set of problems, SDPs are solvable up to fixed precision in polynomial time in terms of the input data, and linear in terms of the precision [17]; see [10] for the complexity in the Turing model of computation.

2 An IPM for SDP using the factor width cone

Practical computation is, however, more complicated. While we are able to solve linear programs with millions of variables and constraints routinely, SDPs become intractable already for a few tens of thousands of constraints and for $n \times n$ matrix variables of the order $n \approx 1,000$. The reason is that each iteration of a typical interior point algorithm for SDP requires $\mathcal{O}(n^3m + n^2m^2 + m^3)$ operations, where n is the size of the matrix variable and m is the number of equality constraints; see e.g. [15] or [3]. However, solving large instances of SDPs is of growing interest, due to applications in power flow problems on large power grids, SDP-based hierarchies for polynomial and combinatorial problems, etc (see [13, 23, 24]). In the following we will revisit a relaxation of a given SDP, where the cone of positive semidefinite matrices is replaced by a more tractable cone, namely the cone of matrices of constant factor width [7]. The simplest examples of matrices of constant factor width are non-negative diagonal matrices (corresponding to linear programs), and scaled diagonally dominant matrices (corresponding to second order cone programming) [4]. We then review how iteratively rescaling the cone and solving the given optimization problem over this new set leads to a non-increasing sequence of optimal values lower bounded by the optimum of the sought SDP. This iterative procedure, due to [1], does not lead to a convergent algorithm. However, its essence can be used to construct a convergent predictor-corrector interior point method, as was done in [19]. Our paper is inspired by ideas from [1, 2, 4, 5, 19]. In particular, we will extend the results in [19], and give a more concise complexity analysis in our extended setting.

1.1 Iterative approximation scheme

Let the set of symmetric $n \times n$ matrices be given by \mathbb{S}^n , where $n \in \mathbb{N}$ is a positive integer. We write $[m]$ for the set $\{1, 2, \dots, m\}$, where $m \in \mathbb{N}$. Consider a set $\{A_i \in \mathbb{S}^n : i \in [m]\}$ of symmetric data matrices and define the linear operator

$$\mathcal{A}(X) = (\langle A_1, X \rangle, \dots, \langle A_m, X \rangle) \in \mathbb{R}^m,$$

where $\langle X, Y \rangle := \text{tr}(XY)$ for $X, Y \in \mathbb{S}^n$. Further, define for $b \in \mathbb{R}^m$ the affine subspace

$$L = \{X \in \mathbb{S}^n : \mathcal{A}(X) = b\}. \quad (1)$$

Consider the following semidefinite program

$$v_{\text{SDP}}^* = \inf \{ \langle A_0, X \rangle : \mathcal{A}(X) = b, X \in \mathbb{S}_+^n \}, \quad (2)$$

which we assume to be strictly feasible. Replacing the cone of positive semidefinite (psd) matrices in (2) by a cone $\mathcal{K} \subseteq \mathbb{S}_+^n$, which is more tractable, leads to the following program

$$v_{\mathcal{K}} = \inf \{ \langle A_0, X \rangle : \mathcal{A}(X) = b, X \in \mathcal{K} \}, \text{ where } \mathcal{K} \subseteq \mathbb{S}_+^n. \quad (3)$$

Clearly, $v_{\mathcal{K}} \geq v_{\text{SDP}}^*$. The quality of the approximation depends on the chosen cone \mathcal{K} . In [4], while focusing on sums-of-squares optimization the authors

consider the cones of *diagonally dominant* and *scaled diagonally dominant* matrices. Ahmadi and Hall developed the idea of replacing the psd cone by a simpler cone further in [1], leveraging an optimal solution of the relaxation. Essentially, the idea is as follows. Define the feasible set for (2) as

$$\mathcal{F}_{\text{SDP}} = \{X \succeq 0 : \mathcal{A}(X) = b\}.$$

We will consider a sequence of strictly feasible points for (3), denoted by X_ℓ for $\ell = 0, 1, \dots$. Since $X_\ell \succeq 0$, the matrix $X_\ell^{1/2}$ is well-defined. One can *update* the data matrices in the following way

$$A_i^{(\ell)} = X_\ell^{1/2} A_i X_\ell^{1/2} \quad (i \in \{0, 1, \dots, m\}, \ell = 0, 1, \dots),$$

giving rise to a new linear operator

$$\mathcal{A}^{(\ell)}(X) = (\langle A_1^{(\ell)}, X \rangle, \dots, \langle A_m^{(\ell)}, X \rangle) \in \mathbb{R}^m.$$

We may also refer to this operation as *rescaling* with respect to X_ℓ . Via this rescaling one obtains the following sequence of reformulations of (2):

$$v_{\text{SDP}}^* = \min \left\{ \langle A_0^{(\ell)}, X \rangle : \mathcal{A}^{(\ell)}(X) = b, X \in \mathbb{S}_+^n \right\}, \quad (4)$$

whose feasible set we define as

$$\mathcal{F}_{\text{SDP}_\ell} = \left\{ X \succeq 0 : \mathcal{A}^{(\ell)}(X) = b \right\}.$$

For each ℓ the identity matrix is feasible, i.e., we have $X = I \in \mathcal{F}_{\text{SDP}_\ell}$. To see this, note that for all $i \in [m]$ we have

$$\langle A_i^{(\ell)}, I \rangle = \langle (X_\ell)^{\frac{1}{2}} A_i (X_\ell)^{\frac{1}{2}}, I \rangle = \langle A_i, X_\ell \rangle = b_i.$$

Similarly, the identity leads to the same objective value in (4) as X_ℓ in (3). Let X_0 be an optimal solution to (3). Rescaling with respect to X_0 we find by the same reasoning that $v_{\mathcal{K}}^{(0)} \leq v_{\mathcal{K}}$, where

$$v_{\mathcal{K}}^{(\ell)} = \min \left\{ \langle A_0^{(\ell)}, X \rangle : \mathcal{A}^{(\ell)}(X) = b, X \in \mathcal{K} \right\}. \quad (5)$$

Reiterating this procedure leads to a non-increasing sequence of values $\left\{ v_{\mathcal{K}}^{(\ell)} \right\}_{\ell \in \mathbb{N}}$ lower bounded by v_{SDP}^* . Unfortunately, this procedure does not always converge to the true optimum of (2) if \mathcal{K} is a cone of matrices of constant factor width, as mentioned in [19]. Indeed, it can happen that $\liminf_{\ell \rightarrow \infty} v_{\mathcal{K}}^{(\ell)} > v_{\text{SDP}}^*$. The rest of this paper is devoted to the development and analysis of an interior point algorithm, which converges to the optimal value v_{SDP}^* . We

thereby refine and extend results from [19], where a different interior point method (based on the factor width cone) was introduced. Our contribution is to give a concise polynomial-time convergence analysis, since the iteration complexity bounds given in [19] involve constants that depend on the data, but the dependence is not made explicit; see e.g. [19, Theorem 4.12]. Moreover, the authors of [19] only consider factor width at most 2 (i.e. the scaled, diagonally dominant matrices), while we analyse the general case.

Outline of the paper

This paper is conceptually divided into two parts. The first part contains sections 1 and 2 and is devoted to introducing the setting as well as the algorithm. Our aim with the first part is to convey the concept in a comprehensible way. The second part consists of the remaining sections 3-6. It is more technical and contains the derivation of objects used in the algorithm as well as the formal complexity analysis.

1.2 The factor width cone

Fix $n \in \mathbb{N}$. The cone of $n \times n$ matrices of *factor width* k , denoted by $\text{FW}_n(k)$, is defined as

$$\text{FW}_n(k) = \left\{ Y \in \mathbb{S}^n : Y = \sum_{i \in \mathbb{N}} x_i x_i^T \text{ for } x_i \in \mathbb{R}^n, \text{supp}(x_i) \leq k, \forall i \right\}.$$

The notion of factor width was first used in [7] where the authors proved that $\text{FW}_n(2)$ is the cone of scaled diagonally dominant matrices. Trivially, $\text{FW}_n(1)$ is the cone of non-negative $n \times n$ diagonal matrices. Clearly, we have that

$$\text{FW}_n(k) \subseteq \text{FW}_n(k+1) \subseteq \mathbb{S}_+^n \quad \forall k \in [n-1].$$

Moreover, $\text{FW}_n(n) = \mathbb{S}_+^n$. It is easy to see these cones are proper. As they define an inner approximation of the cone \mathbb{S}_+^n we may use them in the aforementioned iterative scheme. Define $\mathcal{J} := \{J \subset [n] : |J| = k\}$ for fixed $n, k \in \mathbb{N}$ with $k|n$.

$$\mathbb{S}^{(n,k)} := (\mathbb{S}^k)^{\mathcal{J}} \quad \text{and} \quad \mathbb{S}_+^{(n,k)} := (\mathbb{S}_+^k)^{\mathcal{J}}$$

An optimization problem over the cone $\text{FW}_n(k)$ may be formulated as an optimization problem over the cone product $\mathbb{S}_+^{(n,k)}$. To see this we need to consider principal submatrices. For a matrix $S \in \mathbb{R}^{n \times n}$ we define the principal submatrix $S_{J,J}$ for $J \subseteq [n]$ to be the restriction of S to rows and columns whose indices appear in J . Further, for a set $J = \{i_1, \dots, i_{|J|}\} \subseteq [n]$ and a matrix $S \in \mathbb{R}^{J \times J}$ we define the $n \times n$ matrix $S_J^{\rightarrow n}$ as follows for $i, j \in [n]$

$$(S_J^{\rightarrow n})_{i,j} = \begin{cases} S_{k,l} & \text{if } i = i_k, j = i_l \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

In other words, $S_J^{\rightarrow n}$ has S_J as principal sub-matrix indexed by J , and zeros elsewhere. Now, to write a program over $\text{FW}_n(k)$ as an SDP note the following lemma. It is easy to see that, for any $X \in \text{FW}_n(k)$, we have

$$X = \sum_{J \in \mathcal{J}} Y_J^{\rightarrow n}$$

for suitable $Y_J \in \mathbb{S}_+^k$ indexed by \mathcal{J} . Thus, we can write

$$\inf \{ \langle C, X \rangle : \mathcal{A}(X) = b, X \in \text{FW}_n(k) \} \quad (7)$$

as

$$\inf \left\{ \sum_{|J|=k} \langle C_{J,J}, Y_J \rangle : \sum_{|J|=k} \langle (A_i)_{J,J}, Y_J \rangle = b_i, Y_J \in \mathbb{S}_+^k, \forall |J| = k \right\}. \quad (8)$$

It is straightforward to show that the dual cone is given by

$$\text{FW}_n(k)^* = \{ S \in \mathbb{S}^n : S_{J,J} \succeq 0 \text{ for } J \subseteq [n], |J| = k \}.$$

The dual cone has been studied in the context of semidefinite optimization in [8], where it was shown that the distance of $\text{FW}_n(k)^*$ and \mathbb{S}_+^n in the Frobenius norm can be upper bounded by $\frac{n-k}{n+k-2}$ for matrices of trace 1. For $k \geq 3n/4$ and $n \geq 97$ this bound can be improved to $O(n^{-3/2})$ (see [8]).

2 Interior point methods and the central path

Interior point methods (IPMs) are among the most commonly used algorithms to solve conic optimization problems in practice. Notable software for IPMs include Mosek [16], CSDP [9], SDPA [12, 22], SeDuMi [20] and SDPT3 [21]. In the remainder of this section, we will closely follow the notation used in [18], since we will make use of several results from this book. Consider the following conic optimization problem for a proper convex cone $\mathcal{K} \subset \mathbb{R}^n$:

$$\min \{ \langle c, x \rangle : \langle a_i, x \rangle = b_i, i \in [m], x \in \mathcal{K} \}.$$

In IPMs the cone membership constraint is replaced by adding a convex penalty function f to the objective. This function f is a so-called self-concordant barrier function. Loosely speaking, the function f returns larger values the closer the input is to the boundary of the cone and tends to infinity as the boundary is approached. In order to formally define self-concordant barrier functionals, let $f : \mathbb{R}^n \supset D_f \rightarrow \mathbb{R}$ be such that its Hessian $H(x)$ is positive definite (pd) for all $x \in D_f$. With respect to this function, we can define a *local inner product* as follows

$$\langle u, v \rangle_x := \langle u, H(x)v \rangle,$$

6 An IPM for SDP using the factor width cone

where $u, v \in \mathbb{R}^n$ and $\langle \cdot, \cdot \rangle$ is some reference inner product. Let $B_x(y, r)$ be the open ball centered at y with radius $r > 0$ whose radius is measured by $\|\cdot\|_x$, i.e., the norm arising from the local inner product at x .

Definition 1 (see [18, § 2.2.1]) A functional f is called (*strongly non-degenerate*) *self-concordant* if for all $x \in D_f$ we have that $B_x(x, 1) \subset D_f$ and whenever $y \in B_x(x, 1)$ we have

$$1 - \|y - x\|_x \leq \frac{\|v\|_y}{\|v\|_x} \leq \frac{1}{1 - \|y - x\|_x} \text{ for all } v \neq 0.$$

A functional f is called a *self-concordant barrier functional* if f is self-concordant and additionally satisfies

$$\vartheta_f := \sup_{x \in D_f} \|H(x)^{-1}g(x)\|_x^2 < \infty,$$

where $g(x)$ is the gradient of f .

We refer to ϑ_f as the complexity value of f (see [18, p. 35]), which will become crucial in our complexity analysis. Henceforth, let f be a self-concordant barrier functional for \mathcal{K} and consider the following family of problems for positive $\eta \in \mathbb{R}_+$

$$\begin{aligned} z_\eta &= \operatorname{argmin} \eta \langle c, x \rangle + f(x) \\ &\text{s.t. } \langle a_i, x \rangle = b_i \quad i \in [m]. \end{aligned} \tag{9}$$

The minimizers z_η of (9) define a curve, parametrized by η in the interior of \mathcal{K} . This curve is called the *central path*. For $\eta \rightarrow \infty$ one can show that $z_\eta \rightarrow x^*$, where x^* denotes an optimal solution. Interior point methods work by subsequently approximating a sequence of points $\{z_{\eta_i} : i = 1, \dots, N\}$ on the central path, where $\eta_1 < \eta_2 < \dots$ such that z_{η_N} is within the desired distance to the optimal solution. The type of interior point method we consider is an adaptation of the (primal) *predictor-corrector method* (see [18, § 2.4.4]). This method uses the ordinary *affine scaling direction* to produce a new point inside the cone with decreased objective value. Afterwards, a series of corrector steps is performed to obtain feasible solutions with the same objective value that lie increasingly close to the central path. Interior point methods typically rely on Newton's method in each step, where the convergence rate depends on the so-called Newton decrement.

Definition 2 If $f : \mathbb{R}^n \rightarrow \mathbb{R}$ has a gradient $g(x)$ and positive definite Hessian $H(x) \succ 0$ at a point x in its domain, then the Newton decrement of f at x is defined as

$$\Delta(f, x) = \sqrt{\langle g(x), H^{-1}(x)g(x) \rangle} = \|H^{-1}(x)g(x)\|_x.$$

For self-concordant functions f , a sufficiently small value of $\Delta(f, x)$, e.g., $\Delta(f, x) < 1/9$, implies that x is close to the minimizer of f (cf. [18, Theorem 2.2.5]).

Suppose we are given a starting point x_0 , which is close to z_{η_0} for some $\eta_0 \in \mathbb{R}$. The affine-scaling direction is given by $-H(x_0)^{-1}c$ and points approximately tangential to the central path in the direction of decreasing the objective value $\langle c, x \rangle$ ($-H^{-1}(z_{\eta_0})c$ is exactly tangential to the central path). The predictor step moves from x_0 a fixed fraction $\sigma \in (0, 1)$ of the distance towards the boundary of the feasible set in the affine-scaling direction, thereby producing a new point x_1 satisfying $\langle c, x_1 \rangle < \langle c, x_0 \rangle$. The new point x_1 is not necessarily close to the central path. The algorithm then proceeds to produce a sequence of feasible points x_2, x_3, \dots satisfying $\langle c, x_i \rangle = \langle c, x_1 \rangle$ for $i = 2, 3, \dots$ while each x_i for $i = 2, 3, \dots$ is closer to the central path than its predecessor x_{i-1} . In other words, the algorithm targets the point z_{η_1} on the central path with the same objective value as x_1 and produces a sequence of points converging to z_{η_1} . Once an x_j is found such that $\Delta(f, x_j) < 1/9$, the next predictor step is taken. This procedure is repeated until an ε -optimal solution is found. The corrector phase works by minimizing the self-concordant barrier restricted to the feasible affine space intersected with the set of all $x \in \mathbb{R}^n$ such that $\langle c, x \rangle = \langle c, x_i \rangle$, where x_i is the point produced by the most recent predictor step. This minimization problem is solved iteratively by performing line searches along the direction given by the Newton step for the restricted functional. We provide a visualization of the predictor-corrector method in Figure 1.

Newton decrements for functions restricted to subspaces

If a self-concordant function f is restricted to a (translated) linear subspace L , and denoted by $f|_L$, then the Newton decrement at x becomes

$$\Delta(f|_L, x) = \|P_{L,x}H^{-1}(x)g(x)\|_x,$$

where $\|\cdot\|_x$ is the norm induced by the inner product $\langle u, v \rangle_x = \langle u, H(x)v \rangle$, and $P_{L,x}$ is the orthogonal projection onto L for the $\|\cdot\|_x$ norm; see [18, § 1.6].

Note that we have

$$\begin{aligned} \Delta(f, x) &= \langle g(x), H^{-1}(x)g(x) \rangle^{1/2} = \langle g(x), -n(x) \rangle^{1/2} \\ &= \langle n(x), n(x) \rangle_x^{1/2} = \|n(x)\|_x = \sup_{\|d\|_x=1} \langle d, n(x) \rangle_x, \end{aligned}$$

where $n(x)$ is the Newton step at x , i.e., $n(x) = -H(x)^{-1}g(x)$. Hence, restricting the function f to a subspace L we find

$$\begin{aligned} \Delta(f|_L, x) &= \sup_{\|d\|_x=1} \langle d, P_{L,x}n(x) \rangle_x = \sup_{\substack{\|d\|_x=1 \\ d \in L}} \langle d, n(x) \rangle_x \\ &= \sup_{0 \neq d \in L} \frac{\langle d, n(x) \rangle_x}{\|d\|_x} \geq \frac{\langle d, n(x) \rangle_x}{\|d\|_x} \text{ for all } d \in L \setminus \{0\}. \end{aligned} \tag{10}$$

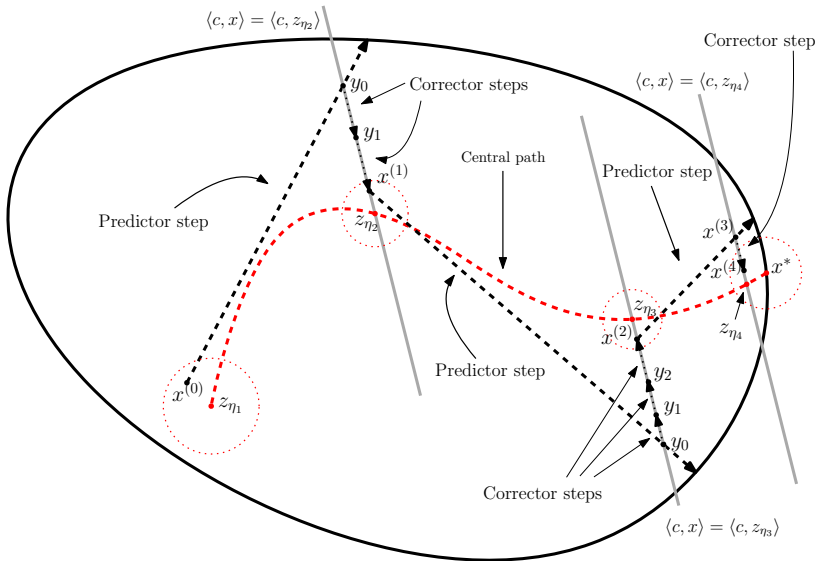


Fig. 1 Visualization of predictor-corrector method. Initial feasible solution close to central path (red) is given by x_1 . Algorithm performs predictor step returning x_2 . Corrector steps are taken until point close enough to central path (x_4) is found. Next predictor step returns x_5 . Corrector steps are taken until x_8 is found, which is close enough to central path to perform next predictor step returning x_9 . After one corrector step the final point x_{10} is ε -close to x^* .

2.1 A predictor-corrector method using $FW(k)$

In this subsection we propose our algorithm which makes use of the rescaling introduced in section 1.1; see Algorithm 1 below. Our aim is to provide a comprehensible exposition, while the details are postponed to the second part of the paper, beginning with section 3.

Algorithm 1 is an adaption of the predictor-corrector method as described in [18, § 2.2.4]. Before describing the algorithm in detail we fix some notation. Let $\mathcal{Y} \in \mathbb{S}^{(n,k)}$ be a collection of $\binom{n}{k}$ matrices of size $k \times k$. We define the operator Ψ as

$$\Psi(\mathcal{Y}) = \sum_{J \in \mathcal{J}} Y_J \rightarrow^n,$$

where we made use of the notation defined in (6). Hence, if \mathcal{Y} is a collection of positive semidefinite $k \times k$ matrices, then $\Psi(\mathcal{Y}) \in FW_n(k)$. Furthermore, let

$$Y_0 = \left\{ Y_J = \binom{n-1}{k-1}^{-1} I_{k \times k} : J \subset [n], |J| = k \right\}, \quad (11)$$

so that $\Psi(Y_0) = I$. Now let X_ℓ be a strictly feasible solution to a problem of form (2) and rescale the data matrices with respect to X_ℓ . Recall that the

feasible set of the resulting SDP is contained in the following affine space

$$L_\ell = \{X \in \mathbb{S}^n : \mathcal{A}_0^{(\ell)}(X) = b\}. \quad (12)$$

Likewise, the feasible set of the factor width relaxation written over $\mathbb{S}_+^{(n,k)}$ (cf. (8)) is constrained to lie in the affine space

$$L_\ell^\Psi = \{\mathcal{Y} \in \mathbb{S}^{(n,k)} : (\mathcal{A}^{(\ell)} \circ \Psi)(\mathcal{Y}) = b\}. \quad (13)$$

Note that $I \in L_\ell$ and $\mathcal{Y}_0 \in L_\ell^\Psi$. We emphasize that, by definition, for any element $\mathcal{Y} \in L_\ell^\Psi$ we have $\Psi(\mathcal{Y}) \in L_\ell$.

Main method

The algorithm requires a feasible starting point X_0 close to the central path, which is used in the first rescaling step. We also require an $\varepsilon > 0$, i.e., our desired accuracy as well as a $\sigma \in (0, 1)$ used in the predictor step. In the following let $f^{\text{FW}(k)}$ be a self-concordant barrier function for $\mathbb{S}_+^{(n,k)}$ (we postpone its derivation to section 3, for now we assume it exists and is efficiently computable). In the algorithm we denote the restriction of $f^{\text{FW}(k)}$ to the subspace $\text{null}(\mathcal{A}^{(\ell)} \circ \Psi)$ by $f_{|\text{null}(L_\ell^\Psi)}^{\text{FW}(k)}$. The algorithm initializes $\ell = 0$. The outer while loop repeats until an ε optimal solution is found. If after rescaling with respect to X_ℓ the Newton decrement at \mathcal{Y}_0 satisfies

$$\Delta \left(f_{|\text{null}(L_\ell^\Psi)}^{\text{FW}(k)}, \mathcal{Y}_0 \right) \leq 1/14$$

the predictor subroutine is called. Here, the affine-scaling direction is projected onto the null space of L_ℓ^Ψ , call it \mathcal{Z} . Clearly, $\mathcal{Y}_0 + s\mathcal{Z} \in L_\ell^\Psi$ for all $s \in \mathbb{R}$. Then the subroutine computes

$$s^* = \sup \left\{ s : \mathcal{Y}_0 - s\mathcal{Z} \in \mathbb{S}_+^{(n,k)} \right\}, \quad (14)$$

which provides the necessary notion of distance to the boundary in terms of \mathcal{Y}_0 and \mathcal{Z} . The returned point $\mathcal{Y}_\ell := \mathcal{Y}_0 + \sigma s^* \mathcal{Z}$ is feasible and decreases the objective value, as shown in section 5.

If the Newton decrement is not small enough, the corrector subroutine is called. Let $v_\ell = \langle A_0, X_\ell \rangle$, i.e., the objective value of the previous iteration, and define

$$L_\ell^\Psi(v_\ell) = \{\mathcal{Y} \in \mathbb{S}_+^{(n,k)} : \langle A_0, \Psi(\mathcal{Y}) \rangle = v_\ell, \mathcal{A}^{(\ell)}(\Psi(\mathcal{Y})) = b\}.$$

Let $x_0 := \mathcal{Y}_0$. Denote by $n_{|L_\ell^\Psi(v_\ell)}(x_i)$ the Newton step of $f_{|L_\ell^\Psi(v_\ell)}^{\text{FW}(k)}$ at a point x_i . The corrector step now computes

$$x_{i+1} = \operatorname{argmin}_t f^{\text{FW}(k)} \left(x_i + t n_{|L_\ell^\Psi(v_\ell)}(x_i) \right)$$

until x_{i+1} is close enough to the central path of the rescaled problem over $\mathbb{S}_+^{(n,k)}$ and returns $\mathcal{Y}_\ell := x_{i+1}$. We will prove in section 4 how this leads to a decrease in distance to the central path of the original SDP. Note that multiple calls of the corrector step may be necessary as after rescaling the Newton decrement might not be small enough anymore. However, as we prove later on, the maximum number of corrector steps can be bounded in terms of the problem data. Let \mathcal{Y}_ℓ be the point returned by one of the subroutines. We set

$$X_{\ell+1} = X_\ell^{1/2} \Psi(\mathcal{Y}_\ell) X_\ell^{1/2}.$$

Then

$$\langle A_i^{(\ell+1)}, I \rangle = \langle A_i^{(\ell)}, \Psi(\mathcal{Y}_\ell) \rangle = \langle A_i, X_{\ell+1} \rangle$$

for all $i = 0, 1, \dots, m$.

Termination criterion

In the predictor as well as in the corrector subroutine we solve a linear system for $y \in \mathbb{R}^m$. The solution of this linear system may be interpreted as a dual feasible solution provided the current iterate is sufficiently close to the central path. Hence, we can approximate the duality gap of our problem by calculating the difference

$$\langle A_0, X_\ell \rangle - y^T b \geq 0,$$

where y is calculated in every subroutine call. We may use this as a termination criterion. Once this quantity falls below some $\varepsilon > 0$ chosen beforehand, we terminate with an ε optimal solution.

Algorithm 1 Predictor-Corrector SDP algorithm using $\text{FW}_n(k)$

Require: $\varepsilon > 0$, $\sigma \in (0, 1)$, X_0 close to CP

$\ell \leftarrow 0$

while Duality gap $> \varepsilon$ **do**

$A_i^{(\ell)} \leftarrow (X_\ell)^{1/2} A_i (X_\ell)^{1/2}$, for $i = 0, 1, \dots, m$

if $\Delta \left(f_{|\text{null}(L_\Psi)}^{\text{FW}(k)}, \mathcal{Y}_0 \right) \leq \frac{1}{14}$ **then**

$\mathcal{Y}_\ell \leftarrow \text{Predictor_Step}(\mathcal{A}^{(\ell)}, A_0^{(\ell)}, \sigma)$

else

$\mathcal{Y}_\ell \leftarrow \text{Corrector_Step}(\mathcal{A}_0^{(\ell)} \circ \Psi, f^{\text{FW}(k)}, \mathcal{Y}_0)$

end if

$X_{\ell+1} \leftarrow (X_\ell)^{1/2} \Psi(\mathcal{Y}_\ell) (X_\ell)^{1/2}$

$\ell \leftarrow \ell + 1$

end while

return X_ℓ

Algorithm 2 Subroutine Predictor_Step

Require: $\mathcal{A}, A_0, \sigma \in (0, 1)$
 Solve for y : $\mathcal{A}A_0 = \mathcal{A}\mathcal{A}^*y$
 $\mathcal{Z} = \Psi^\dagger(\mathcal{A}^*y - A_0)$
 $s^* \leftarrow \sup\{s : \mathcal{Y}_0 - s\mathcal{Z} \in \text{FW}_n(k)\}$
 $\mathcal{Y} \leftarrow \mathcal{Y}_0 - \sigma s^* \mathcal{Z}$
return \mathcal{Y}

Algorithm 3 Subroutine Corrector_Step

Require: $\mathcal{A}, f, x^{(0)} : \Delta(f|_L, x^{(0)}) > \frac{1}{14}, (L = \text{null}(\mathcal{A}))$
 $j \leftarrow 0$
while $(f|_L, x^{(j)}) > \frac{1}{14}$ **do**
 Solve for y : $\mathcal{A}H(x^{(j)})^{-1}\mathcal{A}^*y = \mathcal{A}H(x^{(j)})^{-1}g(x^{(j)})$
 $n|_L(x^{(j)}) \leftarrow H(x^{(j)})^{-1}(\mathcal{A}^*y - g(x^{(j)}))$
 $x^{(j+1)} \leftarrow \text{argmin}_t f(x^{(j)} + tn|_L(x^{(j)}))$
 $j \leftarrow j + 1$
end while
return $x^{(j-1)}$

3 Barrier functionals for \mathbb{S}_+^n and $\text{FW}_n(k)$

In this section we derive the self-concordant barrier functional for the cone $\mathbb{S}_+^{(n,k)}$ which is used in the algorithm. Note that the ordinary self-concordant barrier for \mathbb{S}_+^n is given by $f^{\text{SDP}}(X) = -\log(\det(X))$. We will emphasize parallels to the work of Roig-Solvas and Sznaier [19].

In order to construct a self-concordant barrier function for our underlying set, we introduce the notions of hyper-graphs and edge colorings as well as a well-known result about these objects.

Definition 3 A hyper-graph $\mathcal{H} = (V, E)$ consists of a set $V = \{1, \dots, n\}$ of vertices and a set of hyper-edges $E \subseteq \{J \subseteq V : |J| \geq 2\}$, which are subsets of the vertex set V . If all elements in E contain exactly k vertices, we call the corresponding hyper-graph *k-uniform*.

Definition 4 Let $\mathcal{H} = (V, E)$ be a hyper-graph. A proper hyper-edge coloring with m colors is a partition of the hyper-edge set E into m disjoint sets (color classes), say $E = \cup_{i \in [m]} S_i$ such that $S_i \cap S_j = \emptyset$ if $i \neq j$, and two hyper-edges that share a vertex are not in the same color class. In other words, a proper hyper-edge coloring assigns a color to every hyper-edge such that, if a given vertex appears in two different hyper-edges, they have different colors.

Theorem 1 (Baranyai's theorem [6]) *Let $k, n \in \mathbb{N}$ be such that $k \geq 2$ and $k|n$, and let K_k^n be the complete k -uniform hyper-graph on n vertices. Then K_k^n has a proper hyper-edge coloring using $\binom{n-1}{k-1}$ colors.*

In (8) we wrote a program over $\text{FW}_n(k)$ as an equivalent program over the cone product $\mathbb{S}_+^{(n,k)}$. The algorithm uses a self-concordant barrier function over said cone product. The mapping Ψ from $\mathbb{S}_+^{(n,k)}$ to $\text{FW}_n(k)$ is surjective, but not bijective, since multiple elements in the former may give rise to the same element in the latter set.

Assumption 1 Throughout we will assume $k|n$ for some given $n \in \mathbb{N}$ and $2 \leq k \in \mathbb{N}$.

This assumption is not without loss of generality, but one can always border the data matrices of the SDP problem (4) with $(n \bmod k)$ extra rows and columns in a suitable way to ensure the assumption holds.

In the following we will let $\mathcal{Y} \in \mathbb{S}^{(n,k)}$ be a collection of $\binom{n}{k}$ matrices of size $k \times k$. We recall the operator Ψ is defined as

$$\Psi(\mathcal{Y}) = \sum_{J \in \mathcal{J}} Y_J \rightarrow^n.$$

The following generalizes Lemma 4.4 in [19], where a similar result is proved for $k = 2$. It will be crucial in our analysis as it allows us to compare the values taken by the barrier functionals on $\mathbb{S}_+^{(n,k)}$ and \mathbb{S}_+^n at \mathcal{Y} and $\Psi(\mathcal{Y})$, respectively. In particular, it will allow us to bound the reduction in the SDP barrier function in terms of the reduction of the barrier for $\text{FW}(k)$.

Lemma 2 *Let*

$$f^{\text{FW}(k)}(\mathcal{Y}) = - \sum_{J \in \mathcal{J}} \log(\det(Y_J)), \mathcal{Y} \in \text{int} \left(\mathbb{S}_+^{(n,k)} \right).$$

The barrier $f^{\text{FW}(k)}(\mathcal{Y})$ is self-concordant on $\text{int} \left(\mathbb{S}_+^{(n,k)} \right)$. Furthermore, if $X = \Psi(\mathcal{Y})$ then

$$\begin{aligned} f^{\text{FW}(k)}(\mathcal{Y}) &\geq - \binom{n-1}{k-1} \log(\det(X)) + n \binom{n-1}{k-1} \log \left(\binom{n-1}{k-1} \right) \\ &=: \binom{n-1}{k-1} f^{\text{SDP}}(X) + n \binom{n-1}{k-1} \log \left(\binom{n-1}{k-1} \right). \end{aligned}$$

Let us emphasize here that $f^{\text{FW}(k)}$ is a self-concordant barrier for $\mathbb{S}_+^{(n,k)}$ not $\text{FW}_n(k)$. Before proving Lemma 2 we need an auxiliary result which extends Lemma A.1 from [19] to general values of k such that $k|n$. To prove it we will make use of Theorem 1.

Lemma 3 Consider a $\mathcal{Y} = (Y_J) \in \mathbb{S}^{(n,k)}$ consisting of positive definite $k \times k$ matrices and let $\Psi(\mathcal{Y}) \in \text{FW}_n(k)$. Then there exists a set of $\binom{n-1}{k-1}$ matrices $Z_i \succ 0$ of size $n \times n$ such that $\Psi(\mathcal{Y}) = \sum_{i=1}^{\binom{n-1}{k-1}} Z_i$ and $f^{\text{FW}(k)}(\mathcal{Y}) = -\sum_{i=1}^{\binom{n-1}{k-1}} \log(\det(Z_i))$.

Proof Let K_k^n be the complete k -uniform hyper-graph on n vertices. We can identify each hyper-edge $\{i_1, i_2, \dots, i_k\} \subset [n]$ in K_k^n with exactly one element $Y_J \in \mathcal{Y}$, namely the one where $\{i_1, i_2, \dots, i_k\} = J$. Let $\{S_1, \dots, S_{\binom{n-1}{k-1}}\}$ be the color classes of a hyper-edge coloring of K_k^n . Define $\mathcal{Y}_i := \{Y_J : J \in S_i\}$ and set $Z_i := \Psi(\mathcal{Y}_i)$. Then $\Psi(\mathcal{Y}) = \sum_{i=1}^{\binom{n-1}{k-1}} Z_i$ since $S_i \cap S_j = \emptyset$ for $i \neq j$ and $\cup_i S_i = \mathcal{J}$. Moreover, since each S_i corresponds to disjoint index sets in \mathcal{J} , there exists a permutation matrix P_i for every $i = 1, \dots, \binom{n-1}{k-1}$ such that $P_i Z_i P_i^T$ is a block-diagonal matrix with blocks Y_J on the diagonal for $J \in S_i$. This shows that $Z_i \succ 0$.

From this we find

$$\log(\det(Z_i)) = \log(\det(P_i Z_i P_i^T)) = \sum_{J \in S_i} \log(\det(Y_J)).$$

Hence,

$$\begin{aligned} \sum_{i=1}^{\binom{n-1}{k-1}} \log(\det(Z_i)) &= \sum_{i=1}^{\binom{n-1}{k-1}} \sum_{J \in S_i} \log(\det(Y_J)) \\ &= \sum_{J \in \mathcal{J}} \log(\det(Y_J)) = -f^{\text{FW}(k)}(\mathcal{Y}), \end{aligned}$$

completing the proof. \square

We continue to prove Lemma 2.

Proof (Lemma 2) The self-concordance of $f^{\text{FW}(k)}$ on $\text{int}(\mathbb{S}_+^{(n,k)})$ follows immediately from the self-concordance of $-\log \det(\cdot)$ on $\text{int}(\mathbb{S}_+^k)$. By assumption $X = \Psi(\mathcal{Y}) = \sum_{i=1}^{\binom{n-1}{k-1}} Z_i \in \text{FW}_n(k)$. Therefore,

$$\begin{aligned} -\log(\det(X)) &= -\log \left(\det \left(\frac{1}{\binom{n-1}{k-1}} \sum_{i=1}^{\binom{n-1}{k-1}} \binom{n-1}{k-1} Z_i \right) \right) \\ &\leq -\sum_{i=1}^{\binom{n-1}{k-1}} \frac{1}{\binom{n-1}{k-1}} \log \left(\det \left(\binom{n-1}{k-1} Z_i \right) \right) \\ &= -\sum_{i=1}^{\binom{n-1}{k-1}} \frac{1}{\binom{n-1}{k-1}} \log \left(\left(\binom{n-1}{k-1} \right)^n \det(Z_i) \right), \end{aligned}$$

where the inequality follows by convexity of the function $-\log \det(\cdot)$ on $\text{int}(\mathbb{S}_+^n)$. Hence, we find

$$-\binom{n-1}{k-1} \log(\det(X)) \leq -\sum_{i=1}^{\binom{n-1}{k-1}} \left(n \log \left(\binom{n-1}{k-1} \right) + \log(\det(Z_i)) \right)$$

$$= - \sum_{i=1}^{\binom{n-1}{k-1}} \log(\det(Z_i)) - \binom{n-1}{k-1} n \log\left(\binom{n-1}{k-1}\right),$$

and the claim follows. \square

The following corollary is analogous to Corollary 4.5 from [19].

Corollary 4 *If*

$$\mathcal{Y}_0 = \left\{ Y_J = \binom{n-1}{k-1}^{-1} I_{k \times k} : J \subset [n], |J| = k \right\}$$

then

$$\begin{aligned} f^{\text{FW}(k)}(\mathcal{Y}_0) &= \binom{n-1}{k-1} f^{\text{SDP}}(X) + n \binom{n-1}{k-1} \log\left(\binom{n-1}{k-1}\right) \\ &= n \binom{n-1}{k-1} \log\left(\binom{n-1}{k-1}\right). \end{aligned}$$

Proof The first statement follows when noting that each $i \in [n]$ lies in exactly $\binom{n-1}{k-1}$ subsets of $[n]$ of size k . The reason is that when fixing i , there are $n-1$ elements left out of which we want to choose $k-1$ more elements to make a set of size k . For the second statement note that

$$\log\left(\det\left(\frac{1}{\binom{n-1}{k-1}} I_{k \times k}\right)\right) = \log\left(\binom{n-1}{k-1}^{-k}\right) = -k \log\left(\binom{n-1}{k-1}\right).$$

The result follows when noting that $k \binom{n}{k} = n \binom{n-1}{k-1}$. \square

4 Further properties of the barrier functions

To prove convergence of our algorithm we need two essential ingredients. First, we need to prove that the predictor step reduces the current objective value sufficiently, and secondly, we must prove that the corrector step converges to a point close to the central path. Moreover, we have to show that our criterion to decide which subroutine to call is valid. The issue here is that we compute the Newton decrement of $f^{\text{FW}(k)}$ at \mathcal{Y}_0 , but we need to be able to assert that the Newton decrement of f^{SDP} at X_ℓ is small enough.

The next result we present will allow us to lower bound the progress made by the corrector step. For this we need to be able to compare the barrier functions for \mathbb{S}_+^n and $\mathbb{S}_+^{(n,k)}$. We assume we have a given feasible solution X_ℓ such that $\langle A_0^{(\ell)}, I \rangle = v$. Define the vector $b(v) := (v, b_1, \dots, b_m)^T$. For further reference, consider

$$\min \left\{ f^{\text{SDP}}(X) : \langle A_i^{(\ell)}, X \rangle = b(v)_i \ \forall i = 0, 1, \dots, m, X \in \mathbb{S}_+^n \right\}, \quad (15)$$

which we would like to compare to

$$\min \left\{ f^{\text{FW}(k)}(\mathcal{Y}) : \mathcal{Y} \in L_\ell^\Psi(v) \cap \mathbb{S}_+^{(n,k)} \right\}. \quad (16)$$

Suppose \mathcal{Y}^* is an approximate solution to (16). Defining

$$X_{\ell+1} = X_\ell^{1/2} \Psi(\mathcal{Y}^*) X_\ell^{1/2},$$

we find that $X_\ell \in \mathcal{F}_{\text{SDP}}$ for all ℓ . In other words, the points X_ℓ we obtain via this procedure are all feasible for the original SDP (2). The following lemma allows us to lower bound the decrease achieved by one corrector step in terms of an element in $\mathbb{S}_+^{(n,k)}$.

Lemma 5 *Let \mathcal{Y}^* be a feasible solution to (16) and \mathcal{Y}_0 as in (11). Further, let $X_{\ell+1} = X_\ell^{1/2} \Psi(\mathcal{Y}^*) X_\ell^{1/2}$ for X_ℓ a feasible solution. Then*

$$\binom{n-1}{k-1} \left(f^{\text{SDP}}(X_\ell) - f^{\text{SDP}}(X_{\ell+1}) \right) \geq f^{\text{FW}(k)}(\mathcal{Y}_0) - f^{\text{FW}(k)}(\mathcal{Y}^*).$$

Proof The proof follows immediately when noting that

$$\begin{aligned} & \binom{n-1}{k-1} \left(f^{\text{SDP}}(X_\ell) - f^{\text{SDP}}(X_{\ell+1}) \right) \\ &= \binom{n-1}{k-1} \left(f^{\text{SDP}}(X_\ell) - f^{\text{SDP}}(X_\ell^{1/2} \Psi(\mathcal{Y}^*) X_\ell^{1/2}) \right) \\ &= n \underbrace{\binom{n-1}{k-1} \log\left(\binom{n-1}{k-1}\right)}_{= f^{\text{FW}(k)}(\mathcal{Y}_0) \text{ by Cor. 4}} - f^{\text{SDP}}(\Psi(\mathcal{Y}^*)) - n \underbrace{\binom{n-1}{k-1} \log\left(\binom{n-1}{k-1}\right)}_{\geq -f^{\text{FW}(k)}(\mathcal{Y}^*) \text{ by Lemma 2}} \end{aligned}$$

□

4.1 Relation of the Newton decrements

In this subsection we will prove that we can upper bound the Newton decrement of f^{SDP} at the identity in terms of the Newton decrement of $f^{\text{FW}(k)}$ at \mathcal{Y}_0 . We now define the following operator

$$\Psi^\dagger : \mathbb{S}^n \rightarrow \mathbb{S}^{(n,k)}$$

via

$$(\Psi^\dagger(X))_J = \left(\frac{1}{\binom{n-1}{k-1}} I + \frac{1}{\binom{n-2}{k-2}} (ee^T - I) \right) \circ X_{J,J} \quad \text{for } J \subset [n], |J| = k,$$

where \circ denotes the Hadamard product. See Figure 2 for a visualization of the surjection from $\mathbb{S}_+^{(n,k)}$ to $\text{FW}_n(k)$.

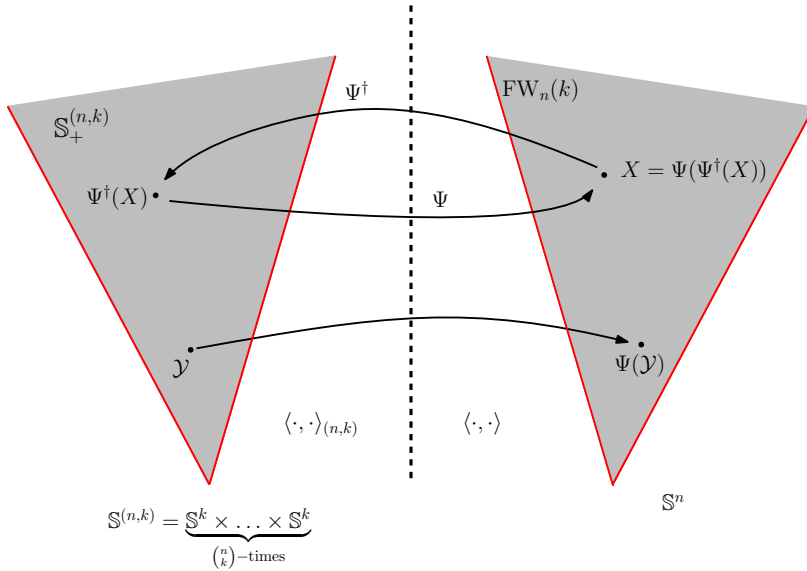


Fig. 2 Visualization the surjection from $\mathbb{S}_+^{(n,k)}$ to $\text{FW}_n(k)$

This operator satisfies

$$\Psi(\Psi^\dagger(X)) = X \text{ for all } X \in \mathbb{S}^n.$$

An inner product on $\mathbb{S}^{(n,k)}$ is given by

$$\langle \mathcal{X}, \mathcal{Y} \rangle_{(n,k)} := \sum_{J \in \mathcal{J}} \langle X_J, Y_J \rangle, \quad (\mathcal{X} = (X_J) \in \mathbb{S}^{(n,k)}, \mathcal{Y} = (Y_J) \in \mathbb{S}^{(n,k)}).$$

It is straightforward to verify the following relation between the norms induced by this inner product, and the Frobenius norm on \mathbb{S}^n .

Lemma 6 For any $X \in \mathbb{S}^n$ we have

$$\|\Psi^\dagger(X)\|_{(n,k)} \leq \|X\|.$$

Suppose now X_ℓ is a feasible solution to (5) such that $\langle A_0, X_\ell \rangle = v$. We define the two subspaces

$$L_\ell^\Psi = \{\mathcal{Y} \in \mathbb{S}^{(n,k)} : (\mathcal{A}^{(\ell)} \circ \Psi)(\mathcal{Y}) = b\}$$

and

$$L_\ell = \{X \in \mathbb{S}^n : \mathcal{A}^{(\ell)}(X) = b\}.$$

We may also add an equality for the objective, in which case we will refer to the following operator

$$\mathcal{A}_0^{(\ell)}(X) = (\langle A_0^{(\ell)}, X \rangle, \langle A_1^{(\ell)}, X \rangle, \dots, \langle A_m^{(\ell)}, X \rangle) \in \mathbb{R}^{m+1}.$$

The respective subspaces will be denoted as follows

$$L_\ell^\Psi(v) = \{\mathcal{Y} \in \mathbb{S}^{(n,k)} : (\mathcal{A}_0^{(\ell)} \circ \Psi)(\mathcal{Y}) = b(v)\} \quad (17)$$

and

$$L_\ell(v) = \{X \in \mathbb{S}^n : \mathcal{A}_0^{(\ell)}(X) = b(v)\}, \quad (18)$$

where $b(v) := (v, b_1, \dots, b_m)^T$.

When we consider the subspaces defined via the operator with respect to the initial data matrices, we omit the subscript ℓ , e.g.,

$$L^\Psi = \{\mathcal{Y} \in \mathbb{S}^{(n,k)} : \langle A_i, \Psi(\mathcal{Y}) \rangle = b_i, \forall i \in [m]\}.$$

The following lemma corresponds to Lemma A.2 in [19], and allows us to bound the Newton decrement of $f_{|L}^{\text{SDP}}$ in terms of $f_{|L}^{\text{FW}(k)}$.

Lemma 7 Assume $\mathcal{Y}_0 \in L^\Psi$ and $I \in L$. At \mathcal{Y}_0 one has

$$\Delta\left(f_{|L^\Psi}^{\text{FW}(k)}, \mathcal{Y}_0\right) \geq \frac{\Delta\left(\binom{n-1}{k-1} f_{|L}^{\text{SDP}}, I\right)}{\sqrt{\binom{n-1}{k-1}}} = \sqrt{\binom{n-1}{k-1}} \Delta\left(f_{|L}^{\text{SDP}}, I\right).$$

Proof Following (10) we have

$$\Delta\left(f_{|L^\Psi}^{\text{FW}(k)}, \mathcal{Y}\right) \geq \frac{\langle d, n^{\text{FW}}(\mathcal{Y}) \rangle_{(n,k), \mathcal{Y}}}{\|d\|_{(n,k), \mathcal{Y}}} \text{ for all } d \in L \setminus \{0\}.$$

Choosing $d = \Psi^\dagger(n_L^{\text{SDP}}(X)) \in L$ leads to

$$\Delta\left(f_{|L^\Psi}^{\text{FW}(k)}, \mathcal{Y}\right) \geq \frac{\langle \Psi^\dagger(n_L^{\text{SDP}}(X)), n^{\text{FW}}(\mathcal{Y}) \rangle_{(n,k), \mathcal{Y}}}{\|\Psi^\dagger(n_L^{\text{SDP}}(X))\|_{(n,k), \mathcal{Y}}},$$

and evaluating the expression at \mathcal{Y}_0 we find

$$\begin{aligned} \Delta\left(f_{|L^\Psi}^{\text{FW}(k)}, \mathcal{Y}_0\right) &\geq \frac{\langle \Psi^\dagger(n_L^{\text{SDP}}(X)), n^{\text{FW}}(\mathcal{Y}_0) \rangle_{(n,k), \mathcal{Y}_0}}{\|\Psi^\dagger(n_L^{\text{SDP}}(X))\|_{(n,k), \mathcal{Y}_0}} \\ &= \frac{\langle \Psi^\dagger(n_L^{\text{SDP}}(X)), -g^{\text{FW}}(\mathcal{Y}_0) \rangle_{(n,k)}}{\binom{n-1}{k-1} \|\Psi^\dagger(n_L^{\text{SDP}}(X))\|_{(n,k)}} \\ &\geq \frac{\langle \Psi^\dagger(n_L^{\text{SDP}}(X)), (I, I, \dots, I) \rangle_{(n,k)}}{\|n_L^{\text{SDP}}(X)\|} \\ &= \frac{\text{tr}(n_L^{\text{SDP}}(X))}{\|n_L^{\text{SDP}}(X)\|}, \end{aligned}$$

where the second inequality follows from Lemma 6. Setting $X = I$ and noting

$$\text{tr}(n_L^{\text{SDP}}(I)) = \langle I, n_L^{\text{SDP}}(I) \rangle = \frac{1}{\binom{n-1}{k-1}} \langle g^{\text{SDP}}(I), -n_L^{\text{SDP}}(I) \rangle$$

18 *An IPM for SDP using the factor width cone*

$$= \frac{1}{\binom{n-1}{k-1}} \left(\Delta \left(\binom{n-1}{k-1} f_{|L}^{\text{SDP}}, I \right) \right)^2$$

we conclude

$$\Delta \left(f_{|L^\Psi}^{\text{FW}(k)}, \mathcal{Y}_0 \right) \geq \frac{1}{\binom{n-1}{k-1}} \frac{\Delta \left(\binom{n-1}{k-1} f_{|L}^{\text{SDP}}, I \right)^2}{\|n_L^{\text{SDP}}(I)\|} = \frac{\Delta \left(\binom{n-1}{k-1} f_{|L}^{\text{SDP}}, I \right)}{\sqrt{\binom{n-1}{k-1}}},$$

because

$$\|n_L^{\text{SDP}}(I)\| = \frac{\Delta \left(\binom{n-1}{k-1} f_{|L}^{\text{SDP}}, I \right)}{\sqrt{\binom{n-1}{k-1}}} = \sqrt{\binom{n-1}{k-1}} \Delta \left(f_{|L}^{\text{SDP}}, I \right).$$

□

5 Complexity analysis

We begin the complexity analysis with the following lemma, which helps us to check whether the current point is close enough to the central path of the SDP.

Lemma 8 *Let X_ℓ be a feasible iterate for the SDP (15) and let the objective value at X_ℓ be v , i.e., $\langle A_0, X_\ell \rangle = v$. Define the two subspaces $L_\ell^\Psi(v)$, L_ℓ as in (17), (12) respectively. Then, if*

$$\Delta \left(f_{|L_\ell^\Psi(v)}^{\text{FW}(k)}, \mathcal{Y}_0 \right) \leq \frac{1}{14},$$

one has

$$\Delta \left(f_{\eta_v|L_\ell}^{\text{SDP}}, I \right) \leq \frac{1}{9},$$

where

$$f_{\eta_v}^{\text{SDP}}(X) = \eta_v \langle A_0, X \rangle - \log \det(X),$$

and η_v is the value of the central path parameter that corresponds to the objective value v .

Proof By Lemma 7 we know that

$$\frac{1}{14} \geq \Delta \left(f_{|L_\ell^\Psi(v)}^{\text{FW}(k)}, \mathcal{Y}_0 \right) \geq \Delta \left(f_{|L_\ell(v)}^{\text{SDP}}, I \right).$$

Let now $z(v)$ be the point on the central path of the rescaled SDP with objective value v and let the corresponding parameter be η_v . By Theorem 2.2.5 from [18] we have

$$\|z(v) - I\|_I \leq \Delta \left(f_{|L_\ell(v)}^{\text{SDP}}, I \right) + \frac{3\Delta \left(f_{|L_\ell(v)}^{\text{SDP}}, I \right)^2}{\left(1 - \Delta \left(f_{|L_\ell(v)}^{\text{SDP}}, I \right)\right)^3} \leq \frac{1}{11}. \quad (19)$$

Let X_+ be the point returned by taking a Newton step at $X = I$ with respect to the function $f_{\eta_v}^{\text{SDP}}$ restricted to L_ℓ . By Theorem 2.2.3 in [18] we have

$$\frac{\|z(v) - I\|_I^2}{1 - \|z(v) - I\|_I} \geq \|X_+ - z(v)\|_I$$

and hence

$$\begin{aligned} \Delta \left(f_{\eta_v|_{L_\ell}}^{\text{SDP}}, I \right) &= \|X_+ - I\|_I \leq \|X_+ - z(v)\|_I + \|z(v) - I\|_I \\ &\leq \frac{\|z(v) - I\|_I^2}{1 - \|z(v) - I\|_I} + \|z(v) - I\|_I \leq \frac{1}{9}. \end{aligned}$$

□

The Newton decrement of the rescaled SDP being smaller than $1/9$ means we can safely perform the next predictor step. If the current point is too far away from the central path and one were to perform the predictor step the direction may not be approximately tangential to the central path. Hence, once the Newton decrement of the factor width program is small enough, so is the one of the SDP and we can perform the next predictor step, knowing the direction will be approximately tangential to the central path. After each predictor step we may have to take several corrector steps, to get back close to the central path.

Corrector step

We will now find an upper bound on the number of corrector steps needed to get close to the central path. We know from Lemma 5 that a decrease in the barrier for the factor width cone will lead to a decrease in the barrier function for our original SDP, meaning we made progress towards its central path. The following lemma asserts that if we are too far away from the central path we can attain at least a constant reduction in the barrier of the factor width cone and therefore obtain a constant reduction in the SDP barrier as well.

Lemma 9 *Let X_ℓ be a feasible iterate for the SDP (15) and let the objective value at X_ℓ be v . Define the subspace $L_\ell^\Psi(v)$ as in (17). If*

$$\Delta \left(f_{|_{L_\ell^\Psi(v)}}^{\text{FW}(k)}, \mathcal{Y}_0 \right) > \frac{1}{14}$$

then

$$f_{|_{L_\ell^\Psi(v)}}^{\text{FW}(k)}(\mathcal{Y}_0) - f_{|_{L_\ell^\Psi(v)}}^{\text{FW}(k)}(\mathcal{Y}^*) \geq \frac{1}{2688}.$$

Proof If $\Delta \left(f_{|_{L_\ell^\Psi(v)}}^{\text{FW}(k)}, \mathcal{Y}_0 \right) > \frac{1}{14}$ the corrector step will employ a line search to find \mathcal{Y}^* , i.e. the point in $L_\ell^\Psi(v)$ that minimizes $f^{\text{FW}(k)}$. Let $n_{L_\ell^\Psi(v)}(\mathcal{Y}_0)$ be the Newton step taken from \mathcal{Y}_0 and let $t = \frac{1}{8\|n_{L_\ell^\Psi(v)}(\mathcal{Y}_0)\|_{(n,k),\mathcal{Y}_0}}$, where the norm in the denominator is the local norm at \mathcal{Y}_0 induced by $\langle \cdot, \cdot \rangle_{(n,k)}$. Then, for

$$\tilde{\mathcal{Y}} = \mathcal{Y}_0 + t n_{L_\ell^\Psi(v)}(\mathcal{Y}_0)$$

we find by Theorem 2.2.2 in [18]

$$f^{\text{FW}(k)}(\tilde{\mathcal{Y}}) \leq f^{\text{FW}(k)}(\mathcal{Y}_0) - \frac{1}{14} \frac{1}{8} + \frac{1}{2} \left(\frac{1}{8} \right)^2 + \frac{(1/8)^3}{3(1-1/8)}$$

20 *An IPM for SDP using the factor width cone*

$$\leq f^{\text{FW}(k)}(\mathcal{Y}_0) - \frac{1}{2688}.$$

□

Note that this implies together with Lemma 5 that

$$\begin{aligned} \frac{1}{2688} &\leq f^{\text{FW}(k)}(\mathcal{Y}_0) - f^{\text{FW}(k)}(\tilde{\mathcal{Y}}) \leq f^{\text{FW}(k)}(\mathcal{Y}_0) - f^{\text{FW}(k)}(\mathcal{Y}^*) \\ &\leq \binom{n-1}{k-1} (f^{\text{SDP}}(X_\ell) - f^{\text{SDP}}(X_{\ell+1})). \end{aligned} \tag{20}$$

Knowing each line search reduces the distance to the targeted point on the central path at least by a constant amount will allow us to bound the number of line searches we need to get close enough if we have an upper bound on the distance of the result of the predictor step and the corresponding point on the central path of the SDP.

Lemma 10 *Let X_1 be close to a point $z(v_1)$ on the central path of the SDP in the sense that $\Delta(f_{L_\ell(v_1)}^{\text{SDP}}, X_1) \leq \frac{1}{9}$. Further, let X_2 be the result of the predictor step and $z(v_2)$ be the point on the central path with the same objective value as X_2 . Then*

$$f^{\text{SDP}}(X_2) - f^{\text{SDP}}(z(v_2)) \leq n \left(\log \frac{1}{1-\sigma} \right) + \frac{1}{154}.$$

Proof A proof of this statement for generic self-concordant barriers may be found on page 54 of [18]. We have used that the barrier parameter for the barrier of the psd cone is given by $\vartheta_{f^{\text{SDP}}} = n$. □

Lemma 11 *Let v_2 be the objective value of the result X_2 of the predictor step. The maximum number K of line searches needed to find a point X_{K+2} which is close enough to $z(v_2)$ in the sense that $\Delta(f_{L_\ell(v_2)}^{\text{SDP}}, X_{K+2}) \leq \frac{1}{9}$ is*

$$K = \left\lceil 2688 \binom{n-1}{k-1} \left(n \log \left(\frac{1}{1-\sigma} \right) + \frac{1}{154} \right) \right\rceil,$$

where $z(v_2)$ is the point on the central path with objective value v_2 .

Proof We know that the distance between the result of the predictor phase and the targeted point on the central path is at most $n \left(\log \frac{1}{1-\sigma} \right) + \frac{1}{154}$ by Lemma 10. Moreover, using Lemma 9 and 5 we find that in each corrector step we reduce this distance by at least $\frac{1}{2688 \binom{n-1}{k-1}}$, unless the SDP Newton decrement at I is already small enough to perform the next predictor step. If after rescaling the Newton decrement of the factor width program satisfies

$$\Delta(f_{L_\ell^\Psi(v)}^{\text{FW}(k)}, \mathcal{Y}_0) > \frac{1}{14},$$

thereby implying by Lemma 8 that I is not close to the central path of the SDP we can perform another corrector step yielding at least a constant decrease of $\frac{1}{2688\binom{n-1}{k-1}}$ of the distance to the central path, and rescale again. This process can be continued until we do not get such a constant decrease anymore at which point we know we must be close enough to the central path, in the sense of Lemma 8. This is because if the decrease is not greater than $\frac{1}{2688\binom{n-1}{k-1}}$ we know that the Newton decrement cannot satisfy

$$\Delta\left(f_{|L_\ell^\Psi(v)}^{\text{FW}(k)}, \mathcal{Y}_0\right) > \frac{1}{14},$$

from which follows by Lemma 8 that

$$\Delta\left(f_{L_\ell(v)}^{\text{SDP}}, I\right) \leq \frac{1}{9}.$$

This implies we are close enough to the central path to perform the next predictor step. Hence, after at most

$$K = \left\lceil 2688 \binom{n-1}{k-1} \left(n \log\left(\frac{1}{1-\sigma}\right) + \frac{1}{154} \right) \right\rceil$$

corrector steps we are close enough to the central path so that we can perform the next predictor step. \square

Predictor step

We will make use of the analysis of the short step interior point method discussed in Section 2.4.2 in [18]. We will show that each predictor step reduces the objective value by an amount at least as large as the objective decrease by the short-step interior point method. This will allow us to conclude the maximum number of predictor steps needed to obtain an ε optimal solution of the given SDP. Note that the decrease in objective value obtained by our predictor method is as follows. Let X be the point from where the predictor method starts and $-(A_0)_X := -H(X)A_0$ be the direction. Then for $\sigma \geq \frac{1}{4}$ and s^* as in (14) we find

$$\begin{aligned} \langle A_0, X - s^* \sigma (A_0)_X \rangle &= \langle (A_0)_X, X \rangle - s^* \sigma \langle A_0, (A_0)_X \rangle \\ &\leq \langle A_0, X \rangle - \frac{1}{4} \|(A_0)_X\|_X. \end{aligned}$$

This implies the decrease is at least as large the one obtained in one iteration of the short-step method, as discussed in [18, § 2.4.2]. Renegar's analysis shows that short-step method leads to an ε optimal solution in at most

$$K = 10\sqrt{\vartheta_f} \log(\vartheta_f/(\varepsilon\eta_0))$$

steps, where η_0 is such that our starting point X_0 is close to z_{η_0} . By an ε optimal solution we mean a feasible solution X such that

$$v_{\text{SDP}}^* \leq \langle A_0, X \rangle \leq v_{\text{SDP}}^* + \varepsilon.$$

Predictor and corrector steps combined

Combining the complexity analysis of predictor and corrector steps we arrive at the following theorem.

Theorem 12 *Let X_0 be a feasible solution of the SDP (2) and assume it is close to some point z_{η_0} on the corresponding central path in the sense that $\Delta\left(f_{|L(v)}^{\text{SDP}}, X_0\right) < 1/14$, where L is as in (18) for $v = \langle A_0, X_0 \rangle$. Algorithm 1 converges to an ε optimal solution in at most*

$$\begin{aligned} K &= \left\lceil 2688 \binom{n-1}{k-1} \left(n \log \left(\frac{1}{1-\sigma} \right) + \frac{1}{154} \right) \right\rceil 10\sqrt{n} \log(n/(\varepsilon \eta_0)) \\ &= O \left(\binom{n-1}{k-1} n^{3/2} \log \left(\frac{1}{1-\sigma} \right) \log \left(\frac{n}{\varepsilon \eta_0} \right) \right). \end{aligned}$$

steps.

The assumption of a starting point “close to the central path” may be satisfied by the self-dual embedding strategy [11]. Alternatively, one may first solve an auxiliary SDP problem, as in [18, § 2.4.2], by using the algorithm we have presented. The solution of this auxiliary problem then yields a point close to the central path of the original SDP problem.

6 Discussion and future prospects

We finish with a brief discussion on various topics surrounding Algorithm 1.

Replacing the predictor step

In their paper [19], the authors propose to perform a fixed number of decrease steps, where a decrease step consists of solving (7) and rescaling with respect to the optimal solution. In our algorithm we considered a different method to decrease the objective value, i.e., the predictor method, where we use the traditional SDP affine scaling direction.

Tractability of factor width cones

Some recent ideas regarding factor width cones that could influence future research in this area are:

- the idea to optimize over the dual cone of $\text{FW}_n(k)$ by utilizing clique trees [24]
- a variation on the factor width cone involving fewer blocks [25].

In addition, it would be very helpful to know a computable self-concordant barrier functional for the cone $\text{FW}_n(k)$, as well as its complexity parameter.

Acknowledgements. The authors would like to thank Georgina Hall for

insightful discussions on the topic on multiple occasions. Moreover, the authors thank Michaël Gabay and Arefeh Kavand for fruitful conversations on different angles of the subject matter.

Funding. This work is supported by the European Union’s Framework Programme for Research and Innovation Horizon 2020 under the Marie Skłodowska-Curie grant agreement N. 813211 (POEMA).

References

- [1] Amir Ali Ahmadi, Sanjeeb Dash, and Georgina Hall. Optimization over structured subsets of positive semidefinite matrices via column generation. *Discrete Optimization*, 24:129–151, 2017. Conic Discrete Optimization.
- [2] Amir Ali Ahmadi and Georgina Hall. *Sum of squares basis pursuit with linear and second order cone*, Contemporary Mathematics. American Mathematical Society, United States, 2017.
- [3] Farid Alizadeh, Jean-Pierre A. Haeberly, and Michael L. Overton. Primal-dual interior-point methods for semidefinite programming: Convergence rates, stability and numerical results. *SIAM Journal on Optimization*, 8(3):746–768, 1998.
- [4] Amir Ali Ahmadi and Anirudha Majumdar. DSOS and SDSOS optimization: LP and SOCP-based alternatives to sum of squares optimization. *2014 48th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–5, 2014.
- [5] Amir Ali Ahmadi and Anirudha Majumdar. DSOS and SDSOS optimization: More tractable alternatives to sum of squares and semidefinite optimization. *SIAM Journal on Applied Algebra and Geometry*, 3(2):193–230, 2019.
- [6] Zsolt Baranyai. On the factorization of the complete uniform hypergraph. *Infinite and Finite Sets*, 1:91–108, 1975. Proceedings of a Colloquium held at Keszthely, June 25-July 1, 1973. Dedicated to Paul Erdős on his 60th Birthday.
- [7] Erik G. Boman, Doron Chen, Ojas Parekh, and Sivan Toledo. On factor width and symmetric H-matrices. *Linear Algebra and its Applications*, 405:239–248, 2005.
- [8] Grigoriy Blekherman, Santanu S. Dey, Marco Molinaro, and Shengding Sun. Sparse psd approximation of the psd cone. *Mathematical Programming*, 191(2):981–1004, 2022.

- [9] Brian Borchers. CSDP, A C library for semidefinite programming. *Optimization Methods and Software*, 11(1-4):613–623, 1999.
- [10] Etienne de Klerk and Frank Vallentin. On the Turing model complexity of interior point methods for semidefinite programming. *SIAM Journal on Optimization*, 26(3):1944–1961, 2016.
- [11] E. de Klerk, C. Roos, and T. Terlaky. Initialization in semidefinite programming via a self-dual skew-symmetric embedding. *Operations Research Letters*, 20(5):213–221, 1997.
- [12] Katsuki Fujisawa, Masakazu Kojima, Kazuhide Nakata, and Makoto Yamashita. SDPA (semidefinite programming algorithm) user’s manual — version 6.00. *Math. Comp. Sci. Series B: Oper. Res.*, 12 2002.
- [13] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.
- [14] Marvin Marcus and Henryk Minc. *A survey of matrix theory and matrix inequalities*, volume 14. Allyn and Bacon, Inc., 1964. p. 115.
- [15] Renato D. C. Monteiro and Paulo Zanjácomo. Implementation of primal-dual methods for semidefinite programming based on Monteiro and Tsuchiya Newton directions and their variants. *Optimization Methods and Software*, 11(1-4):91–140, 1999.
- [16] MOSEK, ApS. MOSEK Optimization Software. Technical report, Version 9.1.9, 2019. <http://docs.mosek.com/9.1/toolbox/index.html>.
- [17] Yurii Nesterov and Arkadii Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics, 1994.
- [18] J. Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization*. MPS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, 2001.
- [19] Biel Roig-Solvas and Mario Sznaiier. A globally convergent LP and SOCP-based algorithm for semidefinite programming. 2022. preprint, <https://arxiv.org/abs/2202.12374>.
- [20] Jos F Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11(1-4):625–653, 1999.
- [21] K.C. Toh, M.J. Todd, and R.H. Tütüncü. SDPT3 - a MATLAB software package for semidefinite programming, version 1.3. *Optimization Methods*

and Software, 11(1):545–581, 1999.

- [22] Makoto Yamashita, Katsuki Fujisawa, Mituhiro Fukuda, Kazuhiro Kobayashi, Kazuhide Nakata, and Maho Nakata. *Latest Developments in the SDPA Family for Solving Large-Scale SDPs*, pages 687–713. Springer US, Boston, MA, 2012.
- [23] Fariba Zohrizadeh, Cedric Jozs, Ming Jin, Ramtin Madani, Javad Lavaei, and Somayeh Sojoudi. A survey on conic relaxations of optimal power flow problem. *European Journal of Operational Research*, 287(2):391–409, 2020.
- [24] Richard Y. Zhang and Javad Lavaei. Sparse semidefinite programs with guaranteed near-linear time complexity via dualized clique tree conversion. *Mathematical Programming*, 188(1):351–393, 2021.
- [25] Yang Zheng, Aivar Sootla, and Antonis Papachristodoulou. Block factor-width-two matrices and their applications to semidefinite and sum-of-squares optimization. *IEEE Transactions on Automatic Control*, pages 1–1, 2022.