An extension of the standard polynomial-time primal-dual path-following algorithm to the weighted determinant maximization problem with semidefinite constraints

Takashi Tsuchiya[*] Yu Xia[†]

February, 2006

Abstract

The problem of maximizing the sum of linear functional and several weighted logarithmic determinant (logdet) functions under semidefinite constraints is a generalization of the semidefinite programming (SDP) and has a number of applications in statistics and datamining, and other areas of informatics and mathematical sciences. In this paper, we extend the framework of standard primal-dual path-following algorithms for SDP to this problem. Employing this framework, we show that the long-step path-following algorithm analogous to the one in SDP has $\mathcal{O}(N \log(1/\varepsilon) + N)$ iteration-complexity to reduce the duality gap by a factor of ε, where $N = \sum N_i$, where N_i is the size of the i-th positive semidefinite matrix block which is assumed to be an $N_i \times N_i$ matrix.

1 Introduction

In this paper, we consider the problem of maximizing the sum of linear functional and several weighted logarithmic determinant (logdet) functions under semidefinite constraints. In the following, this problem is referred to as the weighted determinant maximization problem. We extend the framework of existing standard primal-dual interior-point algorithms for semidefinite programming (SDP) to this problem, and develop a polynomial-time long-step path-following algorithm.

The weighted determinant maximization problem is an extension of SDP problem. It includes the analytical central problem as well. Vandenberghe et. al in [22] focus on the (weighted) determinant maximization problem which involves one logdet function in the

[*]The Institute of Statistical Mathematics, Japan, tsuchiya@sun312.ism.ac.jp. This author is supported in part by Grants-in-Aid for Scientific Research (C) 15510144.

[†]The Institute of Statistical Mathematics, Japan, yuxia@ism.ac.jp. This author is supported by a postdoctoral fellowship for foreign researchers from JSPS (Japan Society for the Promotion of Science) and Grant-in-Aid for JSPS Fellows (16·04291)

objective and give many applications of this problem, including computational geometry, statistics, information and communication theory, etc. In [2], the weighted determinant maximization problem is applied to density estimation. In view of the important roles played by the logarithmic determinant function for positive definite matrices in various fields of informatics and mathematical sciences, there should be many potential applications to this problem. We also emphasize importance of considering the weighted version in this context. For example, if we apply the density estimation method developed in [2] to a histogram data, we need to solve a weighted determinant maximization problem.

In [22], a dual interior-point algorithm for the weighted determinant maximization problem with one logdet term is given with numerical examples. They outlined a polynomial-time convergence proof based on the self-concordance theory by Nesterov and Nemirovskii [17]. Toh presents a primal-dual interior-point algorithm for this problem (with one logdet term) in [20], which includes the search direction of [22]. Toh shows that the algorithms are efficient, robust and accurate through numerical results. No theoretical convergence analysis is given in the paper. The latest version of SDPT3 [21] provides an option to solve the weighted determinant maximization problem.

As was mentioned above, our main purpose in this paper is to extend the standard primal-dual path following algorithms to the weighted determinant maximization problem. To this end, we define an extended central trajectory and an extended normalized duality gap which play the same role as the central trajectory and the normalized duality gap in the primal-dual path-following algorithm for SDP, respectively. Then the neighborhood of the extended central trajectory is introduced analogous to the ordinary SDP case. With this setting, we solve the problem by generating points in the neighborhood of the extended central trajectory, reducing the extended duality gap to zero. As to the search direction, we consider the commutative class introduced in [12].

We prove polynomial-time complexity of the primal-dual path-following interior-point algorithm for the weighted determinant maximization problem. Specifically, we show that an analogue of the long-step path-following algorithm to SDP has $\mathcal{O}(N\log(1/\varepsilon)+N)$ iteration-complexity to reduce the duality gap by a factor of ε, where $N=\sum N_i$, where N_i is the size of the i-th positive semidefinite matrix block which is assumed to be an $N_i\times N_i$ matrix.

In the context of linear programming (LP), this problem is known as the weighted analytic center problem and polynomial-time primal-dual algorithm is first considered in [11] and later studied systematically in more detail by [6, 7, 18] under the name of target following method. Specifically, [18] considered the linear programming version of the extended central trajectory considered in this paper to obtain a short-step path-following algorithm with $\mathcal{O}(\sqrt{N}\log(1/\varepsilon))$ iteration-complexity (here N is the number of nonnegative variables in LP).

A main contribution of this paper is providing a complete analogue of the standard polynomial-time primal-dual path-following algorithms in [9, 8, 14] for linear programming and [10, 13] for semidefinite programming which are the basis for many practical implementations. In this respect, this is not an immediate extension of the results for LP described in the previous paragraph. We expect that on this point this paper serves as a recipe for how to systematically modify the existing practical implementations of the primal-dual path-following algorithms for LP and SDP to the weighted determinant maximization problem.

The remaining of the paper is organized as follows. In §2, we first present the weighted determinant maximization problem, and introduce extensions of the central trajectory, the normalized duality gap and the neighborhoods for SDP. Then we describe our path-following algorithm and the main result. Preliminary observation is also given. Complexity analysis for the algorithm is given in §3.

2 The weighted determinant maximization problem and the main result

2.1 The weighted determinant maximization problem

Let $\mathcal{S}^m$ denote the vector space of real $m \times m$ symmetric matrices. Note that the cone of positive semidefinite matrices induces a partial order. For $X \in \mathcal{S}^m$, we use $X \succeq 0$ ($\succ 0$) to represent that X is positive semidefinite (positive definite). The standard inner product on $\mathcal{S}^m$ is

$$A \bullet B = \mathrm{Tr}\, AB = \sum_{1 \le i,j \le m} A_{ij} B_{ij} \ .$$

We use $\det X$ to represent the determinant of a matrix X.

Let LD denote a subset of the index set $\{1, \dots, n\}$. We consider the following determinant optimization problem.

$$\begin{array}{ll} \min_{X_i \in \mathcal{S}^{N_i}} & \sum_{i=1}^n C_i \bullet X_i - \sum_{j \in \mathrm{LD}} \hat{c}_j \log \det X_j \\ \text{s.t.} & \sum_{i=1}^n A_{ki} \bullet X_i = b_k \quad (k = 1, \dots, m) \ , \\ & X_i \succeq 0 \quad (i = 1, \dots, n). \end{array} \tag{1}$$

Here, $C_i, A_{ki} \in \mathcal{S}^{N_i} (i = 1, \dots, n)$, $\hat{c}_j > 0$, $\mathbf{b} = (b_1, \dots, b_m)^T$ are given data. $X_i \in \mathcal{S}^{N_i}$ $(i = 1, \dots, n)$ are variables. The problem (1) is a convex program. Since the derivative of $(\log \det X)$ is X^{-1}, it is easily verifiable that the dual to (1) is

$$\begin{array}{ll} \max_{\mathbf{y} \in \mathbb{R}^m, Z_i \in \mathcal{S}^{N_i}} & \mathbf{b}^T \mathbf{y} + \sum_{j \in \mathrm{LD}} \hat{c}_j \log \det Z_j + \sum_{j \in \mathrm{LD}} \hat{c}_j N_j \\ \text{s.t.} & \sum_{k=1}^m y_k A_{ki} + Z_i = C_i \quad (i = 1, \dots, n) \ , \\ & Z_i \succeq 0 \ (i = 1, \dots, n). \end{array} \tag{2}$$

Let $N \stackrel{\text{def}}{=} \sum_{i=1}^n N_i$ and define the block diagonal matrices $X = \mathrm{Diag}(X_i)$. Also, define the operator $\mathcal{A}$: $\mathcal{S}^{N_1} \times \cdots \times \mathcal{S}^{N_n} \to \mathbb{R}^m$ as

$$(\mathcal{A}X)_i = \sum_{j=1}^n A_{ij} \bullet X_j \ , \quad (i = 1, \dots, m) \ .$$

We denote the adjoint of $\mathcal{A}$ by $\mathcal{A}^*$. Furthermore, let $Z = \mathrm{Diag}(Z_i)$ and $C = \mathrm{Diag}(C_i)$. Then the primal-dual feasible region, which will be denoted by $\mathcal{F}$ throughout this paper, is written as follows.

$$\mathcal{F} \stackrel{\text{def}}{=} \{(X, \mathbf{y}, Z) : \quad \mathcal{A}X = \mathbf{b}, \quad \mathcal{A}^* \mathbf{y} + Z = C, \quad X \succeq 0, \quad Z \succeq 0\}.$$

We assume that the rows of $\mathcal{A}$ is linearly independent and $\mathrm{int}(\mathcal{F})$ is not empty. Then the both problems have optimal solutions with the same optimal value. Under the setting above, $(X, \mathbf{y}, Z)$ is a solution to (1)–(2) iff it satisfies the following condition,

$$\mathcal{A}X = \mathbf{b}, \quad \mathcal{A}^*\mathbf{y} + Z = C, \quad X_iZ_i = 0 \ (i \notin \mathrm{LD}), \quad X_iZ_i = \hat{c}_iI \ (i \in \mathrm{LD}), \quad X \succeq 0, \quad Z \succeq 0, \tag{3}$$

or, equivalently,

$$X_iZ_i = 0 \ (i \notin \mathrm{LD}), \quad X_iZ_i = \hat{c}_iI \ (i \in \mathrm{LD}), \quad (X, \mathbf{y}, Z) \in \mathcal{F}.$$

The main purpose of this paper is to extend the standard polynomial-time primal-dual path-following algorithms for SDP to solve (3).

2.2 The central trajectory and its neighborhoods

In this part, we introduce an extended central trajectory and its neighborhoods together with an extended normalized duality gap.

In the following, we assume that the weight $\{\hat{c}_i\}$ is in decreasing order, i.e., $\hat{c}_1 \geq \hat{c}_2 \geq \ldots \geq \hat{c}_n$. Note that several blocks can have the same common weight, e.g., $\hat{c}_1 = \hat{c}_2$. Taking into account of this situation, let $\hat{c}_1, \ldots \hat{c}_n$ has D different values. Then we denote by $\hat{c}^d$ the d-th largest value among the D different values and denote by LD^d the index set of the blocks whose weight is greater than or equal to $\hat{c}^d$, the d-th largest weight. The relation between LD^d and LD^{d+1} is as follows.

$$LD^{d+1} = LD^d \cup \{\text{The set of } \hat{c}_i \ (i \in LD) \text{ such that } \hat{c}_i = \hat{c}^{d+1}\} \tag{4}$$

Now we introduce the extended central trajectory $\mathcal{C}$ as follows:

$$\mathcal{C} \equiv \{(X(\nu), Z(\nu), y(\nu)) : (X(\nu), Z(\nu), y(\nu)) \text{ is the solution of (5) for each } \nu \in (0, \infty]\},$$

$$X_iZ_i = \hat{c}_iI \quad (i \in LD^d), \quad X_iZ_i = \nu I \quad (i \notin LD^d), \quad (X, \mathbf{y}, Z) \in \mathcal{F} \quad (\text{for } \nu \in (\hat{c}^{d+1}, \hat{c}^d]). \tag{5}$$

If we denote by $\mathcal{C}_d$ the set of solutions of (5), then $\mathcal{C} = \cup_{d=0}^{D}\mathcal{C}_d$. Here we adopt $\hat{c}^0 = \infty$, $LD^0 = \emptyset$, $\hat{c}^{D+1} = 0$ and $LD^{D+1} = \{1, \ldots, n\}$ as a convention. We call $X_iZ_i = \hat{c}^dI$ $(i \in LD^d)$ in (5) as "the fixed blocks." When $\nu \geq \hat{c}^1$, $X_iZ_i = \nu I$ holds for every block. Therefore, the path coincides with the central trajectory for the ordinary dual pair of SDP obtained by setting all $\hat{c}_i$ to zero in (1) and (2). If $\hat{c}^1 \geq \nu \geq \hat{c}^2$, then we have $X_iZ_i = \hat{c}^1I$ for $i \in LD^1$ and $X_iZ_i = \nu I$ for other blocks. If $\hat{c}^2 \geq \nu \geq \hat{c}^3$, then we have $X_iZ_i = \hat{c}^1I$ for $i \in LD^1$ and $X_iZ_i = \hat{c}^2I$ for $i \in LD^2\backslash LD^1$ and $X_iZ_i = \nu I$ for $i \notin LD^2$. The number of fixed blocks strictly increases at $\nu = \hat{c}^d$ for $d = 1, \ldots, D$ when ν is decreased to zero.

In order to see that the set $\mathcal{C}$ indeed defines a path, we observe that

$$\lim_{\nu \downarrow \hat{c}^{d'+1}} (X(\nu), \mathbf{y}(\nu), Z(\nu)) = (X(\hat{c}^{d'+1}), \mathbf{y}(\hat{c}^{d'+1}), Z(\hat{c}^{d'+1}))$$

holds for $d' = 0, \ldots, D-1$. The existence of the left hand side is obvious from the standard argument since the limiting point must be in the interior of the feasible region.

Since $\hat{c}_i = \hat{c}^{d'+1}$ holds for all $i \in LD^{d'+1} \backslash LD^{d'}$ due to (4), the limit of the left hand side is the feasible point satisfying

$$X_i Z_i = \hat{c}_i \ \ (i \in LD^{d'}), \ \ X_i Z_i = \hat{c}^{d'+1} \ \ (i \in LD^{d'+1} \backslash LD^{d'}), \ \ X_i Z_i = \hat{c}^{d'+1} I \ \ (i \notin LD^{d'+1}),$$

which coincides with the right hand side defined by (5) with $d := d' + 1$.

We have the following proposition.

Proposition 1 *The following properties hold for $\mathcal{C}$.*

1. *$\mathcal{C}$ is a piecewise smooth path which can be nondifferentiable at $(X(\nu), \mathbf{y}(\nu), Z(\nu))$ with $\nu = \hat{c}^d$, $d = 1, \cdots, D$.*
2. *$X(\nu)$ and $(\mathbf{y}(\nu), Z(\nu))$ converge to optimal solutions of (1) and (2) respectively as $\nu \to 0$.*
3. *$X(\nu) \bullet Z(\nu) = \sum_{i \in LD^d} c_i N_i + \nu \sum_{i \notin LD^d} N_i$ for $\nu \in (\hat{c}^{d+1}, \hat{c}^d]$.*
4. *$X(\nu) \bullet Z(\nu)$ is a continuous strictly monotone increasing function of ν with $\lim_{\nu \downarrow 0} X(\nu) \bullet Z(\nu) = \sum_{i \in LD} \hat{c}_i N_i$.*

Proof: The first statement has been already shown by the discussion above. The second statement follows immediately by applying Theorem 9 of [3]. The third statement is easily seen by using (5). The fourth statement readily follows from the first and the third statement. ■

As is described in the above proposition, $X \bullet Z$ is monotonically decreasing along the extended central trajectory $\mathcal{C}$ when ν is decreased and it approaches $\sum_{i \in LD} \hat{c}_i N_i$ as $\nu \to 0$. In the following, we divide the feasible region $\mathcal{F}$ into D subregions SR^1 to SR^{D-1} according to the value of $X \bullet Z$.

$$\begin{aligned} SR^d &= \{(X, \mathbf{y}, Z) \in \mathcal{F} : X(\hat{c}^{d+1}) \bullet Z(\hat{c}^{d+1}) < X \bullet Z \leq X(\hat{c}^d) \bullet Z(\hat{c}^d)\} \\ &= \{(X, \mathbf{y}, Z) \in \mathcal{F} : \sum_{i \in LD^{d+1}} \hat{c}_i N_i + \hat{c}^{d+1} \sum_{i \notin LD^{d+1}} N_i \\ &\qquad < X \bullet Z \leq \sum_{i \in LD^d} \hat{c}_i N_i + \hat{c}^d \sum_{i \notin LD^d} N_i\} \end{aligned} \tag{6}$$

The subregions $SR^1, \ldots, SR^D$ play important roles in this paper, though it do not appear explicitly in the path-following algorithm introduced later. The extended central trajectory consists of D smooth piecses $\mathcal{C}_1, \ldots, \mathcal{C}_D$, and each SR^d exactly corresponds to each smooth piece $\mathcal{C}_d$, i.e., we have $\mathcal{C}_d = \mathcal{C} \cap SR^d$, $d = 1, \ldots, D$ (Figure 1). Later we define an extended normalized duality gap as a piecewise continuous smooth function in $\mathcal{F}$ in such a way that it is smooth in SR^d. The neighborhood of $\mathcal{C}$ will be defined by putting the neighborhood of $\mathcal{C}_d$ together, where the neighborhood of each $\mathcal{C}_d$ is defined as a subset of SR^d. Thus, SR^d is naturally associated with $\mathcal{C}_d$. On the other hand, $\mathcal{C}_d$ is defined with the index set LD^d by (5). Thus, it is natural to associate a feasible point $(X, \mathbf{y}, Z) \in SR^d$

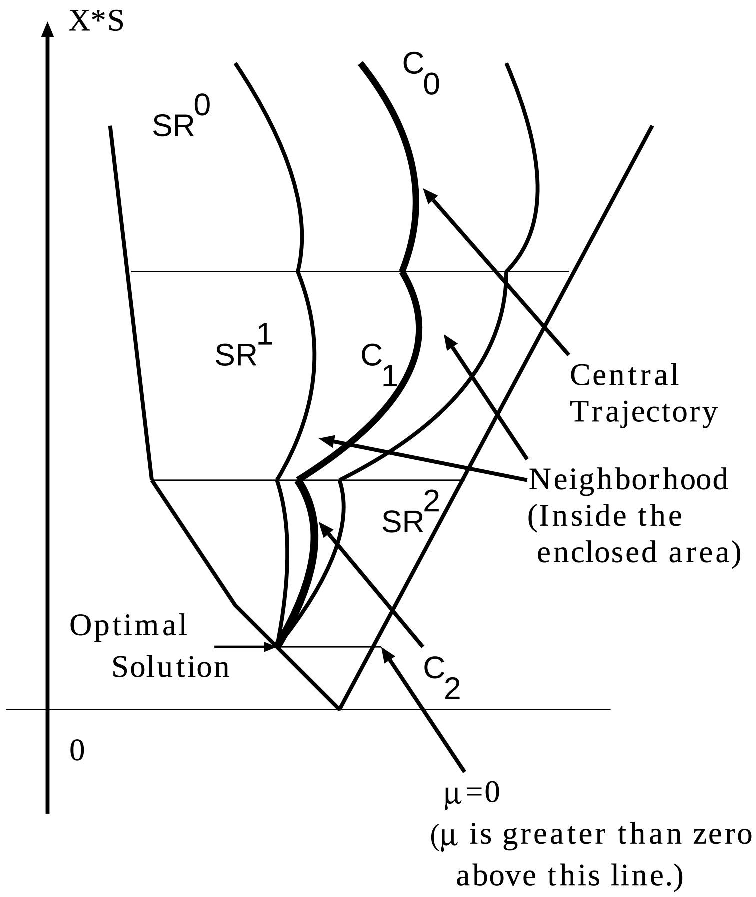

Figure 1: The extended central trajectory and its neighborhood. The subregions associated with the each smooth pieces of the trajectory are also shown.

to the index set LD^d. In view of this, we introduce the following mapping $LD^*(X,Z)$ to associate a point $(X,\mathbf{y},Z) \in SD^d$ to the index set LD^d:

$$LD^*(X,Z) \stackrel{\text{def}}{=} LD^d,$$

where "d is the number such that

$$X(\hat{c}^{d+1}) \bullet Z(\hat{c}^{d+1}) < X \bullet Z \le X(\hat{c}^d) \bullet Z(\hat{c}^d)$$

holds (see (6) and the third statement of Proposition 1)." Observe that

$$\big((X,\mathbf{y},Z) \in \mathcal{F} \text{ and } LD^*(X,Z) = LD^d\big) \Longleftrightarrow (X,\mathbf{y},Z) \in SR^d. \tag{7}$$

This relation plays an important role in this paper.

In the case of SDP, the quantity $X \bullet Z/N$ is referred to as the normalized duality gap. Now we introduce the extended normalized duality gap $\mu(X,Z)$ for our problem as follows:

$$\mu(X,Z) \equiv \frac{X \bullet Z - \sum_{i \in LD^*(X,Z)} \hat{c}_i N_i}{N - \sum_{i \in LD^*(X,Z)} N_i} = \frac{X \bullet Z - \sum_{i \in LD^*(X,Z)} \hat{c}_i N_i}{\sum_{i \notin LD^*(X,Z)} N_i}.$$

Note that in view of (7), $\mu(X,Z)$ have the following "subregion-wise" expression:

$$\begin{aligned}\mu(X,Z) &= \mu^d(X,Z) \quad \text{for } (X,\mathbf{y},Z)\in SR^d,\\ \mu^d(X,Z) &\stackrel{\text{def}}{=} \frac{X\bullet Z-\sum_{i\in LD^d}\hat{c}_iN_i}{N-\sum_{i\in LD^d}N_i}. = \frac{X\bullet Z-\sum_{i\in LD^d}\hat{c}_iN_i}{\sum_{i\notin LD^d}N_i}.\end{aligned}$$

Proposition 2 *The following properties hold.*

1. *For any $(X,\mathbf{y},Z)\in\mathcal{F}$ such that $X\bullet Z>\sum_{i\in LD}\hat{c}_iN_i$, $\mu(X,Z)$ is continuous.*
2. *Let $(X,\mathbf{y},Z)$ be a point on the extended central trajectory with parameter ν. Then we have $\mu(X,Z)=\nu$.*

Proof: To show the first statement, it is enough to check continuity of $\mu(X,Z)$ at the boundary between SD^{d-1} and SD^d, $d=1,\dots,D$. Let $(X,\mathbf{y},Z)$ be on the boundary between SD^{d-1} and SD^d. Observe that we have $X\bullet Z=X(\hat{c}^d)\bullet X(\hat{c}^d)$. Due to the definition of $\mu(X,Z)(=\mu_{d-1}(X,Z))$, for any sequence $(X^k,\mathbf{y}^k,Z^k)\in SR^{d-1}$ such that $\lim_{k=\infty}(X^k,Z^k)=(X,Z)$ and $X^k\bullet Z^k>X\bullet Z$, we readily see the limit of $\mu^{d-1}(X^k,Z^k)=\mu^d(X,Z)$. This proves continuity of $\mu(X,Z)$. The second statement readily follows from the third statement of Proposition 1 and the definition of $\mu(X,Z)$. ■

In the following, we denote by $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ the minimum and the largest eigenvalue of a matrix. For any feasible point such that $X\bullet Z>\sum_{i\in LD}\hat{c}_iN_i$, we define two distance functions to the central trajectory as follows:

$$\begin{aligned}d_F(X,Z) &\stackrel{\text{def}}{=} \sqrt{\sum_{i\in LD^*(X,Z)}\|X_i^{1/2}Z_iX_i^{1/2}-\hat{c}_iI\|^2+\sum_{i\notin LD^*(X,Z)}\|X_i^{1/2}Z_iX_i^{1/2}-\mu(X,Z)I\|^2},\\ d_{-\infty}(X,Z) &\stackrel{\text{def}}{=} \max\left[\max_{i\in LD^*(X,Z)}\hat{c}_i-\lambda_{\min}(X_iZ_i),\ \max_{i\notin LD^*(X,Z)}\mu(X,Z)-\lambda_{\min}(X_iZ_i)\right].\end{aligned}$$

Note that, as they contain $LD^*(X,Z)$ in its definition, these distance functions are written as follows for $(X,\mathbf{y},Z)\in SR^d$:

$$\begin{aligned}d_F(X,Z) &= \sqrt{\sum_{i\in LD^d}\|X_i^{1/2}Z_iX_i^{1/2}-\hat{c}_iI\|^2+\sum_{i\notin LD^d}\|X_i^{1/2}Z_iX_i^{1/2}-\mu^d(X,Z)I\|^2},\\ d_{-\infty}(X,Z) &= \max\left[\max_{i\in LD^d}\hat{c}_i-\lambda_{\min}(X_iZ_i),\ \max_{i\notin LD^d}\mu(X,Z)-\lambda_{\min}(X_iZ_i)\right].\end{aligned}$$

Now we are ready to introduce two neighborhoods of the extended central trajectory.

$$\begin{aligned}\mathcal{N}_F(\gamma) &\stackrel{\text{def}}{=} \{(X,\mathbf{y},Z)\in\mathcal{F}: d_F(X,Z)\le\gamma\mu(X,Z)\},\\ \mathcal{N}_{-\infty(\gamma)}(\mu) &\stackrel{\text{def}}{=} \{(X,\mathbf{y},Z)\in\mathcal{F}: d_{-\infty}(X,Z)\le\gamma\mu(X,Z)\}\end{aligned}$$

Observe that $d_{-\infty(X,Z)}(X,Z)\le\gamma\mu(X,Z)$ is equivalent to

$$\begin{aligned}\lambda_{\min}(XZ) &\ge \hat{c}_i-\gamma\mu(X,Z), \quad (i\in LD^*(X,Z)),\\ \lambda_{\min}(XZ) &\ge (1-\gamma)\mu(X,Z), \quad (i\notin LD^*(X,Z)).\end{aligned}$$

In the following, we show that $\lambda_{\max}(XZ)$ can be also bounded in terms of μ, N and γ. Let $i \in LD^*$. Then we have $\lambda_j(X_iZ_i) \geq \hat{c}_i - \gamma\mu$, where $\lambda_j(\cdot)$ $(i = 1, \ldots, n)$ denotes the jth eigenvalue of a matrix. Then let I_i^+ and I_i^- be the index sets consisting of j such that $\lambda_j(X_iZ_i) \geq \hat{c}_i$ and $\lambda_j(X_iZ_i) < \hat{c}_i$, respectively. Obviously, $I_i^+ \cap I_i^- = \emptyset$ and $I_i^+ \cup I_i^- = \{1, \ldots, N_i\}$. Furthermore, for $j \in I_i^-$, we have $\lambda_j(X_iZ_i) - \hat{c}_i \geq -\gamma\mu$. Based on this observation, we have

$$\begin{aligned}
N\mu &\geq (N - \sum_{i\in LD^*} N_i)\mu = X \bullet Z - \sum_{i\in LD^*} N_i\hat{c}_i \\
&= \sum_{i\in LD^*} \sum_{j=1}^{N_i} (\lambda_j(X_iZ_i) - \hat{c}_i) + \sum_{i\notin LD^*} \sum_{j=1}^{N_i} \lambda_j(X_iZ_i) \\
&= \sum_{i\in LD^*} \left\{ \sum_{j\in I_i^+} (\lambda_j(X_iZ_i) - \hat{c}_i) + \sum_{j\in I_i^-} (\lambda_j(X_iZ_i) - \hat{c}_i) \right\} + \sum_{i\notin LD^*} \sum_{j=1}^{N_i} \lambda_j(X_iZ_i) \\
&\geq \sum_{i\in LD^*} \left\{ \sum_{j\in I_i^+} (\lambda_j(X_iZ_i) - \hat{c}_i) + \sum_{j\in I_i^-} (-\gamma\mu) \right\} + \sum_{i\notin LD^*} \sum_{j=1}^{N_i} \lambda_j(X_iZ_i) \\
&\geq -\gamma\mu N + \sum_{i\in LD^*} \sum_{j\in I_i^+} (\lambda_j(X_iZ_i) - \hat{c}_i) + \sum_{i\notin LD^*} \sum_{j=1}^{N_i} \lambda_j(X_iZ_i).
\end{aligned}$$

From this inequality, we conclude that

$$(1+\gamma)\mu N \geq \sum_{i\in LD^*} \sum_{j\in I_i^+} (\lambda_j(X_iZ_i) - \hat{c}_i) + \sum_{i\notin LD^*} \sum_{j=1}^{N_i} \lambda_j(X_iZ_i).$$

Thus, we obtain the following proposition.

Proposition 3 *If $(X, \mathbf{y}, Z) \in \mathcal{N}_{-\infty}(\gamma)$ for $\gamma \in (0, 1)$, then we have, for any $i = 1, \ldots, n$ and $j = 1, \ldots, N_i$,*

$$\begin{aligned}
\hat{c}_i + (1+\gamma)N\mu(X, Z) \geq \lambda_j(X_iZ_i) \geq \hat{c}_i - \gamma\mu(X, Z) \quad (i \in LD^*(X, Z)) \\
(1+\gamma)N\mu(X, Z) \geq \lambda_j(X_iZ_i) \geq (1-\gamma)\mu(X, Z) \quad (i \notin LD^*(X, Z)).
\end{aligned}$$

Finally, the next proposition justifies solving the weighted determinant maximization problem by reducing the extended duality gap to zero staying inside the neighborhood.

Proposition 4 *Let $\{(X^k, \mathbf{y}^k, Z^k)\}$ be an infinite sequence such that $\mu(X^k, Z^k) \to 0$ and $(X^k, \mathbf{y}^k, Z^k) \in \mathcal{N}_{-\infty}(\gamma)$ for $\gamma \in (0, 1)$. Then any cluster point of X^k and $(\mathbf{y}^k, Z^k)$ are optimal solutions of (1) and (2), respectively.*

Proof: It is enough to consider the case when $\mu(X^k, Z^k)$ is sufficiently close to zero. Since $X^k \bullet Z^k = C \bullet X^k - b^Ty^k$ is bounded above and when $\mu(X^k, Z^k)$ converges to zero and since the level sets of the SDPs obtained by setting $\hat{c}_i$ to zero in (1) and (2) is bounded, the sequence has cluster points X^∞ and $(\mathbf{y}^\infty, Z^\infty)$. Then due to Proposition 3, we have $X_i^\infty Z_i^\infty = \hat{c}_iI$ for $i \in LD$ and $X_i^\infty Z_i^\infty = 0$ for $i \notin LD$. Since $(X^\infty, \mathbf{y}^\infty, Z^\infty)$ is a feasible point, the point satisfies (3) and hence optimal solutions of (1) and (2) as we desired. ■

2.3 The path-following algorithm

Now we introduce the scaled Newton direction for the path-following algorithm. We consider the scaled Newton direction for the following system of equations:

$$X_i Z_i = w_i I,\ i = 1, \dots, n, \quad (X, \mathbf{y}, Z) \in \mathcal{F},$$

where w is a nonnegative vector in $\mathbb{R}^n$. According to the standard symmetrization recipe, the scaled Newton direction for (9) is given as the solution of the following system of linear equations.

$$\mathcal{A}\,\Delta\, X = \mathbf{r}^p \overset{\text{def}}{=} \mathbf{b} - \mathcal{A}X, \tag{8a}$$

$$\mathcal{A}^*\,\Delta\,\mathbf{y} + \Delta\, Z = R^d \overset{\text{def}}{=} C - \mathcal{A}^*\mathbf{y} - Z, \tag{8b}$$

$$\mathcal{H}_{P_i}(X_i\,\Delta\, Z_i + \Delta\, X_i Z_i) = R^c_i \overset{\text{def}}{=} w_i I - \mathcal{H}_{P_i}(X_i Z_i) \quad (i = 1, \dots, n)\ . \tag{8c}$$

Here, $\mathcal{H}_{P_i}$ denotes the linear transformations on the ith block defined as, for $M \in \mathbb{R}^{n_i \times n_i}$,

$$\mathcal{H}_{P_i}(M) \overset{\text{def}}{=} \frac{1}{2}\left[P_i M P_i^{-1} + P_i^{-T} M^T P_i^T\right]$$

and $P_i (i = 1, \dots, n)$ is assumed to be a nonsingular matrix.

The scaled Newton directions with P_i satisfying the condition that $P_i X_i P_i^T$ and $P_i^{-T} Z_i P_i^{-1}$ commutes for each $i = 1, \dots, n$ constitute important class of search directions called the commutative class[12]. The search direction is ensured to exist at every interior-point of $\mathcal{F}$ for the commutative class. The commutative class contains the following two important search directions:

1. Nesterov-Todd direction: $P_i = \left[X_i^{1/2}\left(X_i^{1/2} Z_i X_i^{1/2}\right)^{1/2} X_i^{1/2}\right]^{1/2}$ $(i = 1, \dots, n)$.

2. HRVW/KSH/M direction: $P_i = Z_i^{1/2}$ $(i = 1, \dots, n)$.

Now we are ready to describe the primal-dual path-following algorithm. Below the notation $\mathcal{N}_*$ means either of $\mathcal{N}_F$ or $\mathcal{N}_{-\infty}$. The algorithm is exactly the same as the generic standard path-following algorithms for SDP, except for the definitions of the neighborhood and the (extended) normalized duality gap μ.

The path-following algorithm. Given an initial feasible interior point $(X^0, \mathbf{y}^0, Z^0)$ and accuracy threshold $\epsilon > 0$, our algorithm finds an ϵ-optimal solution $(\tilde{X}, \tilde{\mathbf{y}}, \tilde{Z})$ in the sense that $\mu(\tilde{X}, \tilde{Z}) \le \epsilon$, $\mathcal{A}\tilde{X} = \mathbf{b}$, $\mathcal{A}^*\tilde{\mathbf{y}} + \tilde{Z} = 0$.

1. Set the parameters $\gamma \in (0, 1)$ for specifying the neighborhood, $\sigma \in (0, 1)$ for updating the target point, and $\varepsilon > 0$ for accuracy.

2. Let $(X^0, \mathbf{y}^0, Z^0) \in \mathcal{N}_*(\gamma)$ be given as the initial point.

3. Compute $\mu^k = \mu(X^k, Z^k)$. If $\mu(X^k, Z^k) \le \varepsilon\mu(X^0, Z^0)$, then stop. Otherwise, proceed to Step 4.

4. Let $(X, \mathbf{y}, Z) := (X^k, \mathbf{y}^k, Z^k)$, and choose scaling matrices P_i, $i = 1, \ldots, n$, and compute the scaled Newton direction $(\Delta X, \Delta \mathbf{y}, \Delta Z) \equiv (\mathrm{Diag}(\Delta X_i), \Delta \mathbf{y}, \mathrm{Diag}(\Delta Z_i))$ for the target point

$$\widehat{X}_i \widehat{Z}_i = \hat{c}_i \quad (i \in LD^*(X,Z)), \quad \widehat{X}_i \widehat{Z}_i = \sigma\mu(X,Z) \quad (i \notin LD^*(X,Z)), \quad (\widehat{X}, \hat{\mathbf{y}}, \widehat{Z}) \in \mathcal{F} \tag{9}$$

by setting

$$w_i = \hat{c}_i \quad (i \in LD^*(X,Z)), \quad w_i = \sigma\mu(X,Z) \quad (i \notin LD^*(X,Z)) \tag{10}$$

in (8).

5. Let α^k be the largest $\alpha > 0$ such that $(X + \alpha\Delta X, \mathbf{y} + \alpha\Delta\mathbf{y}, Z + \alpha\Delta Z) \in \mathcal{N}_*(\gamma)$, and we let $(X^{k+1}, \mathbf{y}^{k+1}, Z^{k+1}) = (X + \alpha^k\Delta X, \mathbf{y} + \alpha^k\Delta\mathbf{y}, Z + \alpha^k\Delta Z)$.

Analogous to the ones in SDP, the short-step and long-step path-following algorithms are defined as follows:

1. Short-step path-following algorithm: We let $\mathcal{N}_* := \mathcal{N}_F$ and σ is set to $1 - \delta/\sqrt{N}$ for some $\delta \in (0, 1]$.

2. Long-step path-following algorithm: We let $\mathcal{N}_* := \mathcal{N}_{-\infty}$ (and $\sigma \in (0, 1)$).

In the following, we focus on the long-step path-following algorithm with the commutative class of search directions, and prove its polynomiality. Let

$$G_\infty \stackrel{\text{def}}{=} \sup\left\{ \max_{i=1}^{n} \ \mathrm{cond}(P_i^k X_i^k P_i^{k^T} P_i^k Z_i^{k^{-1}} P_i^{k^T}) \colon k = 0, 1, 2, \ldots \right\},$$

where, for a positive definite matrix K, we define $\mathrm{cond}(K) = \lambda_{\max}(K)/\lambda_{\min}(K)$. The main goal of this paper is to prove the following theorem.

Theorem 1 *The long-step path-following algorithm with the commutative class search direction terminates in $\mathcal{O}(\sqrt{G_\infty} N \log(1/\varepsilon) + N)$ iterations. Specifically, in the case of Nesterov-Todd direction [15, 16] and the HRVW/KSH/M direction [4, 10, 13], the algorithm terminates in $\mathcal{O}(N \log(1/\varepsilon) + N)$ iterations and $\mathcal{O}(N\sqrt{N} \log(1/\varepsilon) + N)$ iterations, respectively.*

2.4 Preliminary observations

Let $(X, \mathbf{y}, Z) \in \mathcal{F}$. Denote

$$(X(\alpha), \mathbf{y}(\alpha), Z(\alpha)) \stackrel{\text{def}}{=} (X, \mathbf{y}, Z) + \alpha\,(\Delta\,X, \Delta\,\mathbf{y}, \Delta\,Z) \ , \quad \mu(\alpha) \stackrel{\text{def}}{=} \mu\,(X(\alpha), Z(\alpha)) \ .$$

We will denote $\mu(0) = \mu(X, Z)$ by μ if it does not cause a confusion.

From the equations $\mathcal{A}\,\Delta\,X = 0$ and $\mathcal{A}^*\,\Delta\,\mathbf{y} + \Delta\,Z = 0$, it is easy to see that

$$\Delta\,X \bullet \Delta\,Z = 0 \ .$$

If $LD^*(X,Z) = LD^*(X(\alpha),Z(\alpha))$, i.e., $(X,\mathbf{y},Z)$ and $(X(\alpha),\mathbf{y}(\alpha),Z(\alpha))$ belong to the same subregion, we have

$$\begin{aligned}
\mu(\alpha)\sum_{i\notin LD^*} N_i + \sum_{i\in LD^*}\hat{c}_i N_i &= X(\alpha)\bullet Z(\alpha) = X\bullet Z + \alpha(X\bullet\Delta Z + Z\bullet\Delta X)\\
&= (1-\alpha)X\bullet Z + \alpha\sum_{i=1}^{n}\mathrm{Tr}\left[\mathcal{H}_{P_i}(X_i\,\Delta Z_i + \Delta X_i Z_i)\right]\\
&= (1-\alpha)\left(\sum_{i\in LD^*}\hat{c}_i N_i + \mu(X,Z)\sum_{i\notin LD^*} N_i\right) + \alpha\left(\sum_{i\in LD^*}\hat{c}_i N_i + \sigma\mu(X,Z)\sum_{i\notin LD^*} N_i\right)\\
&= \sum_{i\in LD^*}\hat{c}_i N_i + (1-\alpha+\alpha\sigma)\mu(X,Z)\sum_{i\notin LD^*} N_i\ .
\end{aligned}$$

The first equality is due to the definition of $\mu(\alpha)$. The second one is because of $\Delta X\bullet\Delta Z = 0$. The third one is due to $\mathrm{Tr}(PMP^{-1}) = \mathrm{Tr}(M)$ for any matrix M. The forth one is due to the definition of $\mu(X,Z)$ and (8). Therefore, we have the following proposition.

Proposition 5 *Let $(X,\mathbf{y},Z)\in\mathrm{int}(\mathcal{F})$ and let $X(\alpha)$ and $Z(\alpha)$ be defined as above with the solution $(\Delta X,\Delta\mathbf{y},\Delta Z)$ of (8), and suppose that $LD^*(X,Z) = LD^*(X(\alpha),Z(\alpha))$ holds, i.e., $(X,\mathbf{y},Z)$ and $(X(\alpha),\mathbf{y}(\alpha),Z(\alpha))$ belong to the same subregion. Then the following relation holds*

$$\mu(\alpha) = (1-\alpha+\sigma\alpha)\mu(X,Z).$$

In the proposition above, we assumed that $LD^*(X,Z) = LD^*(X(\alpha),Z(\alpha))$. Note that this condition always holds for sufficiently small $\alpha > 0$ since $LD^*(X,Z) = LD^*(X(\alpha),Z(\alpha))$ means that $(X,\mathbf{y},Z)$ and $(X(\alpha),\mathbf{y}(\alpha),Z(\alpha))$ belong to the same subregion. Generally, $\mu(\alpha)$ is not written in the simple manner as Proposition 5 because $(X,\mathbf{y},Z)$ and $(X(\alpha),\mathbf{y}(\alpha),Z(\alpha))$ may not belong to the same subregion. But we can still prove it is a monotone decreasing function. This makes sense when we recall that our purpose is to decrease $\mu(X,Z)$ to zero.

Proposition 6 *$\mu(\alpha)$ is a continuous monotone decreasing function of α.*

Proof: By definition, we have $\mu(\alpha) = \mu(X(\alpha),Z(\alpha))$. Furthermore, $\mu(X(\alpha),Z(\alpha)) = \mu_d(X(\alpha),Z(\alpha))$ if $(X(\alpha),\mathbf{y}(\alpha),Z(\alpha))\in SR^d$. As long as $(X(\alpha),\mathbf{y}(\alpha),Z(\alpha))$ stays in one subregion, $\mu(\alpha)$ is a monotone decreasing function of α, since $X(\alpha)\bullet Z(\alpha)$ is a monotone decreasing function of α. Then by continuity of $\mu(X,Z)$ in $\mathcal{F}$ the result readily follows. ■

Let

$$\hat{w}_i(\alpha) = \begin{cases} \hat{c}_i & (i\in LD^*(X,Z)) \\ \mu(\alpha) & (i\notin LD^*(X,Z)) \end{cases}.$$

By using (8) and Proposition 7, we readily obtain the following proposition.

Proposition 7 *Let* $(X, \mathbf{y}, Z) \in \text{int}(\mathcal{F})$ *and let* $X(\alpha)$ *and* $Z(\alpha)$ *be defined as above with the solution* $(\Delta X, \Delta \mathbf{y}, \Delta Z)$ *of (8), and suppose* $LD^*(X,Z) = LD^*(X(\alpha), Z(\alpha))$ *holds. Then we have, for* $i = 1, \ldots, n$,

$$\mathcal{H}_{P_i}(X_i(\alpha)Z_i(\alpha) - \hat{w}_i(\alpha)I) = (1-\alpha)\mathcal{H}_{P_i}(X_iZ_i - \hat{w}_i(\alpha)I) + \alpha^2\mathcal{H}_{P_i}(\Delta X_i \Delta Z_i).$$

Finally, we present the following result about $\|\mathcal{H}_{P_i}(\Delta X_i \Delta Z_i)\|_F$ $(i = 1, \ldots, n)$ which is a direct extension of the SDP case to bound the norm of the second order term.

Lemma 1 *Assume* $(X, \mathbf{y}, Z) \in \mathcal{N}_{-\infty}(\gamma)$, *let* $(\Delta X, \Delta \mathbf{y}, \Delta Z)$ *be the solution of* (8) *with* w_i *is set as in (10). Also suppose that* $P_iX_iP_i^T$ *and* $P_i^{-T}Z_iP_i^{-1}$ *commute for all* $i = 1, \ldots, n$. *Then*

$$\begin{aligned}&\|\mathcal{H}_{P_i}(\Delta X_i \Delta Z_i)\|_{\mathrm{F}} \le \left\|P \,\Delta X \,\Delta Z P^{-1}\right\|_{\mathrm{F}}\\ &\le \sqrt{\max_{i=1}^n\ \text{cond}(P_iX_iP_i^TP_iZ_i^{-1}P_i^T)}\left[\left(1 - 2\sigma + \frac{\sigma^2}{1-\gamma}\right)\sum_{i\notin LD^*(X,Z)} N_i + \sum_{i\in LD^*(X,Z)}\frac{\gamma\hat{c}_iN_i}{\hat{c}_i - \gamma\mu}\right]\mu\ ,\end{aligned} \tag{11}$$

where $P \stackrel{\text{def}}{=} \text{Diag}(P_i)$.

We put the proof in the appendix.

3 Proof of the main theorem

In this section, we prove the main theorem. In the following, we assume that $(X^k, \mathbf{y}^k, Z^k)$ is the kth iterate of the long-step path-following algorithm and P_i^k $(i = 1, \ldots, n)$ is the scaling matrix at the kth iteration.

Lemma 2 *Let* $(X, \mathbf{y}, Z) := (X^k, \mathbf{y}^k, Z^k) \in \mathcal{N}_\infty(\gamma)$ *and* $P_i := P_i^k$ $(i = 1, \ldots, n)$. *Assume that* $0 \le \alpha$ *satisfy*

$$\alpha \le \frac{\sigma\gamma}{\sqrt{\max_i\ \text{cond}(P_iX_iP_i^TP_iZ_i^{-1}P_i^T)}}\left[\left(1 - 2\sigma + \frac{\sigma^2}{1-\gamma}\right)\sum_{i\notin LD^*(X,Z)} N_i + \sum_{i\in LD^*(X,Z)}\frac{\gamma\hat{c}_iN_i}{\hat{c}_i - \gamma\mu}\right]^{-1},$$

and let $LD^*(X,Z) = LD^*(X(\alpha), Z(\alpha))$, *i.e.,* $(X, \mathbf{y}, Z)$ *and* $(X(\alpha), \mathbf{y}(\alpha), Z(\alpha))$ *belong to the same subregion. Then we have* $(X(\alpha), \mathbf{y}(\alpha), Z(\alpha)) \in \mathcal{N}_{-\infty}(\gamma)$, *i.e., we have* $X(\alpha) \succ 0$, $Z(\alpha) \succ 0$ *and*

$$\max_{i\in LD^*(X,Z)}\left[\hat{c}^i - \lambda_{\min}\left(X_i(\alpha)Z_i(\alpha)\right)\right] \ \le\ \gamma\mu(\alpha)\ , \tag{12}$$

$$\max_{i\notin LD^*(X,Z)}\left[\mu(\alpha) - \lambda_{\min}\left(X_i(\alpha)Z_i(\alpha)\right)\right] \ \le\ \gamma\mu(\alpha). \tag{13}$$

Proof: Since the real part of the spectrum of a real matrix is contained between the largest and the smallest eigenvalues of its Hermitian part (see [5, p. 187], for instance), we obtain

$$\lambda_{\max}\left\{\hat{w}_i(\alpha)I - X_i(\alpha)Z_i(\alpha)\right\} \le \lambda_{\max}\left\{\mathcal{H}_{P_i}\left(\hat{w}_i(\alpha)I - X_i(\alpha)Z_i(\alpha)\right)\right\} \ .$$

From the assumption that $P_iX_iP_i^T$ and $P_i^{-T}Z_iP_i^{-1}$ commute, we obtain $P_iX_iZ_iP_i^{-1} = P_i^{-T}Z_iX_iP_i^T$. Therefore,

$$\mathcal{H}_{P_i}[\hat{w}_i(\alpha) - X_iZ_i] = \frac{1}{2}[\hat{w}_i(\alpha) - P_iX_iZ_iP_i^{-1}] + \frac{1}{2}[\hat{w}_i(\alpha) - P_i^{-T}Z_iX_iP_i^T] = \hat{w}_i(\alpha) - P_iX_iZ_iP_i^{-1} \ .$$

Furthermore, $P_iX_iZ_iP_i^{-1}$ is similar to X_iZ_i, and $(X, \mathbf{y}, Z) \in \mathcal{N}_{-\infty}(\gamma)$. Then it follows that

$$\lambda_{\max}\{\mathcal{H}_{P_i}[\hat{w}_i(\alpha)I - X_iZ_i]\} \le \gamma\mu \ .$$

Noting that for all M, Q hermitian, we have $\lambda_{\max}(M+Q) \le \lambda_{\max}(M) + \lambda_{\max}(Q)$. Applying Proposition 7, we obtain, for each $i = 1, \ldots, n$,

$$\begin{aligned}
&\lambda_{\max}\left\{\hat{w}_i(\alpha)I - X_i(\alpha)Z_i(\alpha)\right\} \\
&\le \lambda_{\max}\left\{\mathcal{H}_{P_i}\left(\hat{w}_i(\alpha)I - X_i(\alpha)Z_i(\alpha)\right)\right\} \\
&\le (1-\alpha)\lambda_{\max}\left\{\mathcal{H}_{P_i}\left[\hat{w}_i(\alpha)I - X_iZ_i\right]\right\} + \alpha^2\lambda_{\max}\left[\mathcal{H}_{P_i}\left(\Delta X_i \, \Delta Z_i\right)\right] \\
&\le (1-\alpha)\gamma\mu + \alpha^2\left\|P\,\Delta X\,\Delta Z P^{-1}\right\|_{\mathrm{F}} \ .
\end{aligned}$$

Therefore, also by taking account of Proposition 5, (12) and (13) holds if

$$(1-\alpha)\gamma\mu + \alpha^2\left\|P\,\Delta X\,\Delta Z P^{-1}\right\|_{\mathrm{F}} \le \gamma\mu(\alpha) = \gamma((1-\alpha)\mu + \alpha\sigma\mu). \tag{14}$$

Furthermore, it is shown by the standard argument if this condition is satisfied then $X(\alpha) \succ 0$ and $Z(\alpha) \succ 0$. The lemma readily follows from (14) and Lemma 1. ■

Lemma 3 *Assume that $G_\infty < +\infty$. Let $(X, \mathbf{y}, Z) := (X^k, \mathbf{y}^k, Z^k) \in \mathcal{N}_\infty(\gamma)$ and $P_i := P_i^k$ $(i = 1, \ldots, n)$. If $LD^*(X, Z) = LD^*(X(\alpha^k), Z(\alpha^k))$ in the path-following algorithm, then we have*

$$\alpha^k \ \ge \ \frac{\sigma\gamma}{\sqrt{G_\infty}N}\left[\left((1-\sigma)^2 + \frac{\gamma\sigma^2}{1-\gamma}\right) + \frac{\gamma}{1-\gamma}\right]^{-1}$$

and

$$\mu^{k+1} \stackrel{\text{def}}{=} \mu\left(X^{k+1}, Z^{k+1}\right) = [1 - (1-\sigma)\alpha^k]\mu^k \ .$$

Proof: Since $\hat{c}_i \ge \mu(X, Z)$ and $\sum_{i=1}^n N_i = N$, we have

$$\begin{aligned}
&\frac{\sigma\gamma}{\sqrt{G_\infty}}\left[\left(1 - 2\sigma + \frac{\sigma^2}{1-\gamma}\right)\sum_{i \notin LD^*} N_i + \sum_{i \in LD^*}\frac{\gamma\hat{c}_iN_i}{\hat{c}_i - \gamma\mu}\right]^{-1} \\
&\ge \frac{\sigma\gamma}{\sqrt{G_\infty}N}\left[\left((1-\sigma)^2 + \frac{\gamma\sigma^2}{1-\gamma}\right) + \frac{\gamma}{1-\gamma}\right]^{-1} .
\end{aligned}$$

Then the lemma readily follows from Lemma 2. ■

Now we are ready to prove the main theorem.

Proof of Theorem 1

We divide the iterations into two cases:

1. $LD^*(X^k, Z^k) \neq LD^*(X^{k+1}, Z^{k+1})$, i.e., $(X^k, \mathbf{y}^k, Z^k)$ and $(X^{k+1}, \mathbf{y}^{k+1}, Z^{k+1})$ belong to different subregions.
2. $LD^*(X^k, Z^k) = LD^*(X^{k+1}, Z^{k+1})$, i.e., $(X^k, \mathbf{y}^k, Z^k)$ and $(X^{k+1}, \mathbf{y}^{k+1}, Z^{k+1})$ belong to the same subregion.

In the first case, we have no guarantee on reduction of the extended normalized duality gap μ, but since the number of subregions is bounded by $D(\leq N)$, this case occurs at most N times. Furthermore, μ^k does not increase at each such iteration.

On the other hand, due to Lemma 3 above, we see that μ decreases at least by a factor of $1 - \eta/(\sqrt{G_\infty}N)$ whenever the second case occurs, where η is a positive constant depending only on γ and σ. Therefore, the number of iterations to reduce the extended duality gap by a factor of ε is bounded by $\mathcal{O}(\sqrt{G_\infty}N(\log 1/\varepsilon) + N)$. Since G_∞ for the Nesterov-Todd direction is one and G_∞ for the HRVW/KSH/M direction is $\mathcal{O}(N)$ (See Appendix), the result readily follows.

4 Concluding remark

In this paper, we extended the standard primal-dual path-following algorithms for semidefinite programming to maximize the weighted sum of logdet functions under semidefinite constraints. Specifically, we presented an $\mathcal{O}(N\log(1/\varepsilon) + N)$ polynomial-time long-step path-following algorithm for this problem. We have not analyzed the short-step path following algorithm. The analysis of the short-step path-following algorithm is a bit more difficult if we focus on obtaining the (expected) best iteration-complexity bound of $\mathcal{O}(\sqrt{N}\log(1/\varepsilon))$ since the number of subregion can be $\mathcal{O}(N)$ and is larger than $\sqrt{N}$ in order. This means that we need to analyze behavior of the algorithm taking account of the situation when the iterate goes over several subregions in one iteration to establish such a bound. Nevertheless, we conjecture that it is possible to prove $\mathcal{O}(\sqrt{N}\log(1/\varepsilon))$ bound for the short-step algorithm. It is an interesting topic for further research.

5 Appendix

5.1 Proof of Lemma 1

To simplify notations, we use symmetric Kronecker product [1]. In the following, we define

$$W(\nu) = \mathrm{Diag}(W_i(\nu)), \quad W_i(\nu) = \hat{c}_i I \ (i \in LD^*), \quad W_i(\nu) = \nu I \ (i \notin LD^*).$$

We also let $R^c = \text{Diag}(R_i^c)$ (see (8)). Let svec represent the mapping from $\mathcal{S}^n$ to $\mathbb{R}^{\frac{n(n+1)}{2}}$:

$$\text{svec}(K) = \left[K_{11}, \sqrt{2}K_{12}, \ldots, \sqrt{2}K_{1n}, K_{22}, \ldots, \sqrt{2}K_{2n}, \ldots, K_{nn}\right]^T .$$

For $M, N \in \mathbb{R}^{n\times n}$, the symmetric Kronecker product $M \circledast N$ is defined as a linear operator on $\mathcal{S}^n$:

$$(M \circledast N)\,\text{svec}(K) = \text{svec}\left[\frac{1}{2}(NKM^T + MKN^T)\right] .$$

See the appendix of [19] for properties of the symmetric Kronecker product.

From $(X, \mathbf{y}, Z) \in \mathcal{N}_{-\infty}(\gamma)$, we get a lower bound on the eigenvalues of $X_i Z_i$ ($i = 1, \ldots, n$):

$$\lambda_j(X_i Z_i) \geq \hat{c}_i - \gamma\mu \;\; (i \in LD^*) , \quad \lambda_j(X_i Z_i) \geq (1-\gamma)\mu \;\; (i \notin LD^*) .$$

Therefore,

$$\begin{aligned}
&\left[W(\sigma\mu)X^{-1} - Z\right] \bullet \left[W(\sigma\mu)Z^{-1} - X\right] \\
&= Z \bullet X + W^2(\sigma\mu)X^{-1} \bullet Z^{-1} - 2\left(\sum_{i\in LD^*} \hat{c}_i N_i + \sigma\mu \sum_{i\notin LD^*} N_i\right) \\
&= \left(\sum_{i\in LD^*} \hat{c}_i N_i + \mu \sum_{i\notin LD^*} N_i\right) + \left(\sum_{i\in LD^*}\sum_{j=1}^{N_i} \frac{\hat{c}_i^2}{\lambda_j(X_i Z_i)} + \sum_{i\notin LD^*}\sum_{j=1}^{N_i} \frac{\sigma^2\mu^2}{\lambda_j(X_i Z_i)}\right) \\
&\quad -2\left(\sum_{i\in LD^*} \hat{c}_i N_i + \sigma\mu \sum_{i\notin LD^*} N_i\right) \\
&\leq \left[\left(1 - 2\sigma + \frac{\sigma^2}{1-\gamma}\right) \sum_{i\notin LD^*} N_i + \sum_{i\in LD^*} \frac{\gamma \hat{c}_i N_i}{\hat{c}_i - \gamma\mu}\right]\mu \qquad (15)
\end{aligned}$$

The second equality is because $X \bullet Z = \sum_{i\in LD^*} \hat{c}_i N_i + \mu \sum_{i\notin LD^*} N_i$; the inequality is due to the bound on the eigenvalues of $X_i Z_i$, and $-\hat{c}_i N_i + \frac{\hat{c}_i^2 N_i}{\hat{c}_i - \gamma\mu} = \frac{\gamma \hat{c}_i N_i \mu}{\hat{c}_i - \gamma\mu}$.

On the other hand, (8c) can be written as

$$\begin{aligned}
&\left[\left(P^{-T}ZP^{-1}\right) \circledast I\right](P \circledast P)\,\text{svec}\,\Delta X \\
&\qquad + \left[\left(PXP^T\right) \circledast I\right]\left(P^{-T} \circledast P^{-T}\right)\text{svec}\,\Delta Z = \text{svec}\,R^c. \qquad (16)
\end{aligned}$$

Also observe that

$$\begin{aligned}
\left[\left(P^{-T}ZP^{-1}\right) \circledast I\right]^{-1} \text{svec}\,R^c &= (P \circledast P)\,\text{svec}\left[W(\sigma\mu)Z^{-1} - X\right] , \\
\left[\left(PXP^T\right) \circledast I\right]^{-1} \text{svec}\,R^c &= \left(P^{-T} \circledast P^{-T}\right)\text{svec}\left[W(\sigma\mu)X^{-1} - Z\right] .
\end{aligned} \qquad (17)$$

Since for symmetric matrices M and Q, $[(P^{-T} \circledast P^{-T}) \operatorname{svec}(M)]^T [(P \circledast P) \operatorname{svec}(Q)] = M \bullet Q$, we get

$$
\begin{aligned}
&\left[W(\sigma\mu)X^{-1} - Z\right] \bullet \left[W(\sigma\mu)Z^{-1} - X\right] \\
&\quad = \left\{ \left(P^{-T} \circledast P^{-T}\right) \operatorname{svec}\left[W(\sigma\mu)X^{-1} - Z\right]\right\}^T \left\{(P \circledast P) \operatorname{svec}\left[W(\sigma\mu)Z^{-1} - X\right]\right\} \\
&\quad = \left\{ \left[\left(PXP^T\right) \circledast I\right]^{-1} \left[\left(P^{-T}ZP^{-1}\right) \circledast I\right] (P \circledast P) \operatorname{svec} \Delta X + \left(P^{-T} \circledast P^{-T}\right) \operatorname{svec} \Delta Z\right\}^T \\
&\qquad \left\{ (P \circledast P) \operatorname{svec} \Delta X + \left[\left(P^{-T}ZP^{-1}\right) \circledast I\right]^{-1} \left[\left(PXP^T\right) \circledast I\right] \left(P^{-T} \circledast P^{-T}\right) \operatorname{svec} \Delta Z\right\} .
\end{aligned} \tag{18}
$$

The second equality above is due to (16) and (17).

Because $\mathbf{r}^p = 0$, $R^d = 0$ in (8a) and (8b), it is easy to see that

$$
[(P \circledast P) \operatorname{svec} \Delta X]^T \left[\left(P^{-T} \circledast P^{-T}\right) \operatorname{svec} \Delta Z\right] = \Delta X \bullet \Delta Z = 0 .
$$

Observe that $C \circledast I$ and $D \circledast I$ commute if C and D commute. And the inverse of matrices C and D commute iff C and D commute. By assumption, (PXP^T) and $(P^{-T}ZP^{-1})$ commute. Denote $\mathcal{G}_i \stackrel{\text{def}}{=} \left[\left(P_iX_iP_i^T\right) \circledast I\right]^{-1} \left[\left(P_i^{-T}Z_iP_i^{-1}\right) \circledast I\right]$ $(i = 1, \dots, n)$. Further define $\operatorname{cond}_{\max}(\mathcal{G}) \stackrel{\text{def}}{=} \max_i[\operatorname{cond}(\mathcal{G}_i)]$. Then we obtain

$$
\begin{aligned}
(18) \quad = \quad & \sum_{i=1}^n [\mathcal{G}_i (P_i \circledast P_i) \operatorname{svec} \Delta X_i]^T [(P_i \circledast P_i) \operatorname{svec} \Delta X_i] \\
& \quad + \sum_{i=1}^n \left[\left(P_i^{-T} \circledast P_i^{-T}\right) \operatorname{svec} \Delta Z_i\right]^T \left[\mathcal{G}_i^{-1} \left(P_i^{-T} \circledast P_i^{-T}\right) \operatorname{svec} \Delta Z_i\right] \\
\geq \quad & \sum_{i=1}^n \lambda_{\min}(\mathcal{G}_i) \left\|(P_i \circledast P_i) \operatorname{svec} \Delta X_i\right\|_2^2 + \sum_{i=1}^n \lambda_{\min}(\mathcal{G}_i^{-1}) \left\|\left(P_i^{-T} \circledast P_i^{-T}\right) \operatorname{svec} \Delta Z_i\right\|_2^2 \\
\geq \quad & 2\sum_{i=1}^n \left[\lambda_{\min}(\mathcal{G}_i)\lambda_{\min}(\mathcal{G}_i^{-1})\right]^{-1/2} \left\|(P_i \circledast P_i) \operatorname{svec} \Delta X_i\right\|_2 \left\|\left(P_i^{-T} \circledast P_i^{-T}\right) \operatorname{svec} \Delta Z_i\right\|_2 \\
= \quad & 2\sum_{i=1}^n [\operatorname{cond}(\mathcal{G}_i)]^{-1/2} \left\|P_i \Delta X_i P_i^T\right\|_{\mathrm{F}} \left\|P_i^{-T} \Delta Z_i P_i^{-1}\right\|_{\mathrm{F}} \\
\geq \quad & 2\sum_{i=1}^n [\operatorname{cond}(\mathcal{G}_i)]^{-1/2} \left\|P_i \Delta X_i \Delta Z_i P_i^{-1}\right\|_{\mathrm{F}} \geq 2[\operatorname{cond}_{\max}(\mathcal{G})]^{-1/2} \left\|P \Delta X \Delta Z P^{-1}\right\|_{\mathrm{F}} .
\end{aligned} \tag{19}
$$

The second last inequality is also because the Frobenius norm is submultiplicaltive, i.e., for any matrices M and Q, $\|MQ\|_{\mathrm{F}} \leq \|M\|_{\mathrm{F}} \|Q\|_{\mathrm{F}}$.

It is proved in [12, Lemma 10.4.10] that for any two commuting symmetric positive definite matrices M and Q, $\operatorname{cond}\left[(M \circledast I)(Q \circledast I)^{-1}\right] < 4 \operatorname{cond}(MQ^{-1})$. Letting $M = P_iX_iP_i^T$ and $Q = P_i^{-T}Z_iP_i^{-1}$ in the above result; along with (15) and (19), we've proved the lemma.

5.2 Bounding G_∞ for the Nesterov-Todd direction and the HRVW/KSH/M direction

In this appendix we establish the bounds on G_∞ for the two directions mentioned in the end of the proof of Theorem 1. The proof of $G_\infty = 1$ for the Nesterov-Todd is exactly the same as the SDP case. In the following, we focus on the HRVW/KSH/M direftion.

For all k, the scaling matrix P^k for the HRVW/KSH/M direction is $P^k = {Z^k}^{1/2}$. Therefore, $G^k = {X^k}^{1/2} Z^k {X^k}^{1/2}$. Because $(X^k, \mathbf{y}^k, Z^k) \in \mathcal{N}_{-\infty}(\gamma)$, due to Proposition 3, we have

$$\hat{c}_i - \lambda_{\min}(X_i^k Z_i^k) \leq \gamma\mu^k \quad (i \in LD^*), \quad \mu^k - \lambda_{\min}(X_i^k Z_i^k) \leq \gamma\mu^k \quad (i \notin LD^*)$$

and

$$\hat{c}_i + (1+\gamma)N\mu^k \geq \lambda_{\max}(X_i^k Z_i^k) \;\; (i \in LD^*), \;\; (1+\gamma)N\mu^k \geq \lambda_{\max}(X_i^k Z_i^k) \;\; (i \notin LD^*).$$

Now, we have

$$\mathrm{cond}(G_i^k) = \mathrm{cond}(X_i^k Z_i^k) = \lambda_{\max}(X_i^k Z_i^k)\lambda_{\min}^{-1}(X_i^k Z_i^k).$$

For $i \in LD^*$, by $\hat{c}_i > \mu^k$ and Proposition 3, we obtain

$$\lambda_{\max}(X_i^k Z_i^k)\lambda_{\min}^{-1}(X_i^k Z_i^k) \leq \frac{\hat{c}_i + (1+\gamma)N\mu^k}{\hat{c}_i - \gamma\mu^k} \leq \frac{2N+1}{1-\gamma} \ .$$

For $i \notin LD^*$, in the similar manner,

$$\lambda_{\max}(X_i^k Z_i^k)\lambda_{\min}^{-1}(X_i^k Z_i^k) \leq \frac{2N}{1-\gamma} \ .$$

Hence

$$G_\infty \leq \frac{2N+1}{1-\gamma} \ .$$

References

[1] 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 J. Optim.*, 8(3):746–768 (electronic), 1998.

[2] Tadayoshi Fushiki, Shingo Horiuchi, and Takashi Tsuchiya. A new computational approach to density estimation with semidefinite programming. Research Memorandum 898, The Institute of Statistical Mathematics, Tokyo, Japan, November 2003.

[3] L. M. Graña Drummond and Y. Peterzil. The central path in smooth convex semidefinite programs. *Optimization*, 51(2):207–233, 2002.

[4] Christoph Helmberg, Franz Rendl, Robert J. Vanderbei, and Henry Wolkowicz. An interior-point method for semidefinite programming. *SIAM J. Optim.*, 6(2):342–361, 1996.

[5] Roger A. Horn and Charles R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1994.

[6] B. Jansen, C. Roos, T. Terlaky, and J.-Ph. Vial. Primal-dual target-following algorithms for linear programming. *Ann. Oper. Res.*, 62:197–231, 1996.

[7] Benjamin Jansen, Ccs Roos, Tamás Terlaky, and Jean-Philippe Vial. Long-step primal-dual target-following algorithms for linear programming. *Math. Methods Oper. Res.*, 44(1):11–30, 1996.

[8] Masakazu Kojima, Shinji Mizuno, and Akiko Yoshise. A polynomial-time algorithm for a class of linear complementarity problems. *Math. Programming*, 44(1, (Ser. A)):1–26, 1989.

[9] Masakazu Kojima, Shinji Mizuno, and Akiko Yoshise. A primal-dual interior point algorithm for linear programming. In *Progress in mathematical programming (Pacific Grove, CA, 1987)*, pages 29–47. Springer, New York, 1989.

[10] Masakazu Kojima, Susumu Shindoh, and Shinji Hara. Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM J. Optim.*, 7(1):86–125, 1997.

[11] Shinji Mizuno. An $O(n^3L)$ algorithm using a sequence for a linear complementarity problem. *J. Oper. Res. Soc. Japan*, 33(1):66–75, 1990.

[12] Renato Monteiro and Michael Todd. Path-following methods. In *Handbook of semidefinite programming*, volume 27 of *Internat. Ser. Oper. Res. Management Sci.*, pages 267–306. Kluwer Acad. Publ., Boston, MA, 2000.

[13] Renato D. C. Monteiro. Primal-dual path-following algorithms for semidefinite programming. *SIAM J. Optim.*, 7(3):663–678, 1997.

[14] Renato D. C. Monteiro and Ilan Adler. Interior path following primal-dual algorithms. I. Linear programming. *Math. Programming*, 44(1, (Ser. A)):27–41, 1989.

[15] Yu. E. Nesterov and M. J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Math. Oper. Res.*, 22(1):1–42, 1997.

[16] Yu. E. Nesterov and M. J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM J. Optim.*, 8(2):324–364 (electronic), 1998.

[17] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*, volume 13 of *SIAM Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

[18] C. Roos, T. Terlaky, and J.-Ph. Vial. *Theory and algorithms for linear optimization.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Ltd., Chichester, 1997.

[19] M. J. Todd, K. C. Toh, and R. H. Tütüncü. On the Nesterov-Todd direction in semidefinite programming. *SIAM J. Optim.*, 8(3):769–796 (electronic), 1998.

[20] Kim-Chuan Toh. Primal-dual path-following algorithms for determinant maximization problems with linear matrix inequalities. *Comput. Optim. Appl.*, 14(3):309–330, 1999.

[21] R. H. Tütüncü, K. C. Toh, and M. J. Todd. Solving semidefinite-quadratic-linear programs using SDPT3. *Math. Program.*, 95(2, Ser. B):189–217, 2003.

[22] Lieven Vandenberghe, Stephen Boyd, and Shao-Po Wu. Determinant maximization with linear matrix inequality constraints. *SIAM J. Matrix Anal. Appl.*, 19(2):499–533 (electronic), 1998.