TARGET FOLLOWING ALGORITHMS FOR SEMIDEFINITE PROGRAMMING

CHEK BENG CHUA

Abstract. We present a target-following framework for semidefinite programming, which generalizes the target-following framework for linear programming. We use this framework to build weighted path-following interior-point algorithms of three distinct flavors: short-step, predictor-corrector, and large-update. These algorithms have worst-case iteration bounds that parallel their counterparts in linear programming. We further consider the problem of finding analytic centers given a pair of primal-dual strictly feasible solutions. An algorithm that moves towards the analytic center prior to reducing the duality gap has a better iteration bound than the weighted path-following algorithms. In the case of linear programming, this bound is also an improvement over existing similar algorithms.

2000 *Mathematics Subject Classification.* 90C22; 90C25.

Key words and phrases. Semidefinite programming; Weighted analytic centers; Weighted central path; Target-following algorithm; Weighted path-following algorithm; Target space; Cholesky search directions; Efficient centering; Local Lipschitz constant of Cholesky factorization.

This research was supported in part by a grant from the Faculty of Mathematics, University of Waterloo and by a Discovery Grant from NSERC.

Contents

1. Introduction

The target-following framework was first introduced by Mizuno [8] for linear complementarity problems and Jansen, Roos, Terlaky and Vial [6] for linear programming as a unifying framework for various primal-dual path-following algorithms and algorithms that find analytic centers. The essential ingredient of this framework is the *target map* $(\mathbf{x}, \mathbf{s}) \mapsto [\mathbf{x}_1\mathbf{s}_1, \ldots, \mathbf{x}_n\mathbf{s}_n]^T$, defined for each pair of positive n-vectors $(\mathbf{x}, \mathbf{s})$. An important feature of the target map is its bijection between the primal-dual strictly feasible region and the cone of positive n-vectors $\mathbb{R}^n_{++}$ [6, 7], whence identifying the primal-dual strictly feasible region with the relatively simple cone $\mathbb{R}^n_{++}$ known as the *target space* (or v-space). Interior-point algorithms based on the target map are known as *target-following algorithms*, which are conceptually simple when viewed as following a sequence of targets in the target space.

Various attempts were made to generalize the concept of target maps to semidefinite programming (SDP) [10, 11, 17], symmetric cone programming [5, 21] and general convex conic programming [22]. We present a target map and a target-following framework for SDP, from which we derive weighted path-following algorithms and target-following algorithms with provable polynomial worst-case iteration bounds. Our target map is based on the notion of *Cholesky weighted analytic centers* first introduced by the author in [3].

In recent reports [2, 3], the author analyzed the convergence behavior of the *weighted central paths* corresponding to the Cholesky weighted centers. In these reports, the study of Cholesky weighted centers were mainly motivated by homogeneous cone programming: the central path for a homogeneous cone programming problem coincide with certain weighted central path of a particular SDP-representation of the problem.

In this paper, we explore a different aspect of Cholesky weighted centers: the target map derived from these weighted centers. We present a generic target-following framework based on this target map, and analyze the iteration complexity of target-following algorithms based on two distinct choices of search directions, and weighted path-following algorithms of three distinct flavors.

1.1. **Organization of material.** This paper is organized as follows.

We begin section 2 with a generic target-following framework based on the target map derived from Cholesky weighted centers. These weighted centers were first introduced by the author in [3], and are related to a notion of weighted centers studied by Monteiro and Zanjácomo [12] in a general framework. We present a different perspective on these weighted centers that relates them with analytic centers of larger SDP problems which we called *expanded SDP problems*. We define a measure of proximity to the Cholesky weighted centers based on the l_2-proximity measure for the expanded SDP problems. We also show that search directions for the expanded SDP problems, which translate naturally to search directions for the original SDP problems, can be efficiently computed. However the computation of search directions in each iteration may require the solving for $\Theta(n^3)$ real variables, in contrast with $O(n^2)$ variables in a regular path-following algorithm.

This issue is addressed in section 3, where we reduce the size of the Newton system to match that of a typical Newton system in a regular path-following algorithm. This is achieved with a specific choice of search directions, which we called the *Cholesky search directions*. We then use these search directions in our target-following framework to produce a target-following algorithm. We further consider three weighted path-following algorithms: a short-step algorithm, a predictor-corrector algorithm and a large-update algorithm. Our analyses on these algorithms show that the first two take $O(\sqrt{n\rho})$ iterations to improve the duality gap by a fixed fraction, while the last algorithm takes $O(n\sqrt{\rho})$ iterations. Here ρ denotes the ratio of the average weight to the smallest weight. These bounds parallel their counterparts in linear programming. These search directions were discussed in a general framework by Burer and Monteiro [1], with which they built a long-step path-following algorithm. Their analysis was based on the derivatives of the map $(\mathbf{X}, \mathbf{S}) \mapsto \mathbf{L}_\mathbf{S}^T \mathbf{X} \mathbf{L}_\mathbf{S}$, where $\mathbf{L}_\mathbf{S}$ denotes the Cholesky factor of $\mathbf{S}$. In contrast, we use the local Lipschitz property of the Cholesky factorization $\mathbf{X} \mapsto \mathbf{L}_\mathbf{X}$.

In section 4, we investigate the application of our target-following framework in the approximation of analytic centers. We work in a subset of the target space containing only diagonal matrices, hence our investigation is very closely related, and directly applicable, to the work of Mizuno [8] on linear complementarity problems and the work of Jansen et. al. [6] on linear programming. From a given pair of primal-dual strictly feasible solutions, we generate a finite sequence of targets towards the pair of solution on the central path with the same duality gap as the given pair. Using a technique first developed by Todd [18] for linear programming, and subsequently used by Nesterov and Todd [15], and Nemirovski and Nesterov [14] for general convex conic programming, we derive an upper bound on the number of targets in the sequence. For SDP problems, we obtained the improved worst-case iteration bound $O(\sqrt{n}\log\rho)$. For linear programming problems, this bound is an improvement over the existing best bound $O(\sqrt{n}(\log\rho + \log\widetilde{\rho}))$, where $\widetilde{\rho} \in [1, n]$ denotes the ratio of the largest weight to the average weight (see [6, 8]).

1.2. **Notations and conventions.** Throughout this paper, we use the following notations and conventions.

We use uppercase bold letters (e.g., $\mathbf{X}, \mathbf{L}$, etc.) to denote matrices, and use lowercase bold letters (e.g., $\mathbf{y}, \mathbf{b}$, etc.) to denote vectors.

The space of real n-vectors is denoted by $\mathbb{R}^n$, and the cone of vectors in $\mathbb{R}^n$ with nonnegative (resp., positive) entries is denoted by $\mathbb{R}^n_+$ (resp., $\mathbb{R}^n_{++}$). The cone of vectors in $\mathbb{R}^n_{++}$ with entries in nonincreasing order is denoted by $\mathbb{R}^n_{\downarrow,++}$.

The space of real n-by-n matrices is denoted by $\mathbb{M}^n$. We equip $\mathbb{M}^n$ with the inner product $\bullet : (\mathbf{A}, \mathbf{B}) \in \mathbb{M}^n \times \mathbb{M}^n \mapsto \operatorname{tr}(\mathbf{A}^T\mathbf{B})$. The induced norm $\|\cdot\|_F$ is the Frobenius norm.

Cartesian product of matrix spaces $\mathbb{M}^{n_1} \times \cdots \times \mathbb{M}^{n_k}$ is equipped with the inner product

$$(\mathfrak{A}, \mathfrak{B}) \mapsto \sum_{l=1}^{k} \mathfrak{A}_l \bullet \mathfrak{B}_l.$$

The transposes, inverses, products and Cholesky factors of tuples in the Cartesian product are defined componentwise.

The subspace of lower triangular (resp., upper triangular) matrices in $\mathbb{M}^n$ is denoted by $\mathbb{L}^n$ (resp., $\mathbb{U}^n$).

For any matrix $\mathbf{M} \in \mathbb{M}^n$, the unique lower triangular matrix $\mathbf{L}$ satisfying $\mathbf{M} - \mathbf{L} \in \mathbb{U}^n$ and $\mathbf{L}_{ii} = \mathbf{M}_{ii}/2$ for $i = 1, \ldots, n$, is denoted by $\langle\!\langle \mathbf{M} \rangle\!\rangle$. For any matrix $\mathbf{M} \in \mathbb{M}^n$, we denote by $\mathbf{M}_H$ the symmetric matrix $\mathbf{M} + \mathbf{M}^T$. Consequently $\langle\!\langle \mathbf{M} \rangle\!\rangle_H$ denotes the unique symmetric matrix whose entries in the lower triangular part coincide with those of $\mathbf{M}$.

For any symmetric, positive definite matrix $\mathbf{X} \in \mathbb{S}^n_{++}$, its unique Cholesky factor (i.e., the unique lower triangular matrix $\mathbf{L} \in \mathbb{L}^n$ with positive diagonal entries satisfying $\mathbf{L}\mathbf{L}^T = \mathbf{X}$) is denoted by $\mathbf{L}_\mathbf{X}$.

The group of orthonormal matrices in $\mathbb{M}^n$ is denoted by $\mathbb{O}^n$.

The space of symmetric matrices of order n is denoted by $\mathbb{S}^n$, and the cone of symmetric, positive semidefinite (resp., positive definite) matrices of order n is denoted by $\mathbb{S}^n_+$ (resp., $\mathbb{S}^n_{++}$).

The subspace of diagonal matrices in $\mathbb{S}^n$ is denoted by $\mathbb{D}^n$, and its intersection with $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$ are, respectively, denoted by $\mathbb{D}^n_+$ and $\mathbb{D}^n_{++}$. The cone of matrices in $\mathbb{D}^n_{++}$ with diagonal entries in nonincreasing order is denoted by $\mathbb{D}^n_{\downarrow,++}$.

For each diagonalizable matrix $\mathbf{M} \in \mathbb{M}^n$, we denote by $\boldsymbol{\lambda}(\mathbf{M})$ the vector of eigenvalues of $\mathbf{M}$ in nonincreasing order.

For any m-by-n matrix $\mathbf{M}$ and any subsets of indices $I \subseteq \{1, \ldots, m\}$ and $J \subseteq \{1, \ldots, n\}$, the sub-matrix of $\mathbf{M}$ with row indices in I and column indices in J is denoted by $\mathbf{M}_{IJ}$. If $I = \{i\}$ (resp., $J = \{j\}$) is a singleton, we may also write i (resp., j) in place of $\{i\}$ (resp., $\{j\}$). For any matrix $\mathbf{M}$, we denote by $[\mathbf{M}]_i$ the square sub-matrix $\mathbf{M}_{\{1,\ldots,i\},\{1,\ldots,i\}}$.

The zero matrix and the identity matrix of appropriate size (in the context used) are denoted, respectively, by $\mathbf{0}$ and $\mathbf{I}$. The vector of ones of appropriate size (in the context used) is denoted by $\mathbf{1}$.

For each linear map $\boldsymbol{\mathcal{A}} : \mathbb{E} \to \mathbb{F}$ between two Euclidean spaces, $\boldsymbol{\mathcal{A}}^H : \mathbb{F} \to \mathbb{E}$ denotes its adjoint map.

For each sequence $x_1, \ldots, x_n$ of real numbers, $\operatorname{Diag}(x_1, \ldots, x_n)$ denotes the diagonal matrix in $\mathbb{D}^n$ with $x_1, \ldots, x_n$ on its diagonal. For each matrix $\mathbf{M} \in \mathbb{M}^n$, we denote by $\operatorname{diag}(\mathbf{M})$ the vector $[\mathbf{M}_{11}, \ldots, \mathbf{M}_{nn}]^T \in \mathbb{R}^n$.

For each pair of real numbers (x, y), we denote by $x \vee y$ the greater of the two.

2. Target-Following Framework

We consider the following pair of primal-dual SDP problems:

$$\begin{array}{lll} \inf_{\mathbf{X}} & \mathbf{C} \bullet \mathbf{X} & \\ \text{subject to} & \mathbf{A}_k \bullet \mathbf{X} = \mathbf{b}_k \ (1 \leq k \leq m), \quad \mathbf{X} \in \mathbb{S}^n_+, \end{array} \tag{SDP}$$

and

$$\begin{array}{ll} \sup_{\mathbf{S},\mathbf{y}} & \mathbf{b}^T\mathbf{y} \\ \text{subject to} & \sum_{k=1}^{m} \mathbf{y}_k \mathbf{A}_k + \mathbf{S} = \mathbf{C}, \quad \mathbf{S} \in \mathbb{S}^n_+, \end{array} \tag{SDD}$$

where $\mathbf{A}_1, \ldots, \mathbf{A}_m, \mathbf{C} \in \mathbb{S}^n$ and $\mathbf{b} \in \mathbb{R}^m$ are given.

We assume there exists primal-dual strictly feasible solutions $(\widehat{\mathbf{X}}, \widehat{\mathbf{S}})$; i.e., a pair of primal-dual feasible solutions in $\mathbb{S}^n_{++} \times \mathbb{S}^n_{++}$.

Consider the target map $\boldsymbol{\mathcal{T}} : \mathbb{S}^n_{++} \times \mathbb{S}^n_{++} \to \mathbb{S}^n_{++}$ defined by

$$(\mathbf{X}, \mathbf{S}) \mapsto \mathbf{Q}\mathbf{D}\mathbf{Q}^T,$$

where $\mathbf{Q}^T\mathbf{X}\mathbf{S}\mathbf{Q} = \mathbf{D} + \mathbf{U} \in \mathbb{U}^n$ is a Schur-decomposition of $\mathbf{X}\mathbf{S}$ with $\operatorname{diag}(\mathbf{U}) = \mathbf{0}$, and $\mathbf{D} \in \mathbb{D}^n_{\downarrow,++}$.

Theorem 1. *The map $\boldsymbol{\mathcal{T}}$ is well defined. Moreover, it is a bijection between the cone $\mathbb{S}^n_{++}$ and the set of primal-dual strictly feasible solutions of the primal-dual pair (SDP, SDD).*

Proof. See [3, Theorem 10]. □

Using the target map $\boldsymbol{\mathcal{T}}$, we propose the following general framework for target-following algorithms:

Algorithm 1. (Target-following framework for SDP)
Given a pair of primal-dual strictly feasible solutions $(\mathbf{X}_{in}, \mathbf{S}_{in})$.

1. *Find a target* $\mathbf{W}_+ \in \mathbb{S}^n_{++}$ *close to* $\boldsymbol{\mathcal{T}}(\mathbf{X}_{in}, \mathbf{S}_{in})$. *Set* $(\mathbf{X}_+, \mathbf{S}_+) = (\mathbf{X}_{in}, \mathbf{S}_{in})$.
2. *Repeat the following:*
   - (a) *Pick* $\mathbf{W}_{++} \in \mathbb{S}^n_{++}$ *close to* $\mathbf{W}_+$.
   - (b) *Compute a pair of primal-dual strictly feasible solutions* $(\mathbf{X}_{++}, \mathbf{S}_{++})$ *that approximates* $\boldsymbol{\mathcal{T}}^{-1}(\mathbf{W}_{++})$.
   - (c) *Update* $(\mathbf{X}_+, \mathbf{S}_+) \leftarrow (\mathbf{X}_{++}, \mathbf{S}_{++})$ *and* $\mathbf{W}_+ \leftarrow \mathbf{W}_{++}$.

The two main steps in this framework are the choosing of $\mathbf{W}_{++}$ and the computation of approximate solutions $(\mathbf{X}_{++}, \mathbf{S}_{++})$. The sequence of $\mathbf{W}_{++}$ chosen is called the *sequence of targets*, and the sequence approximate solutions $(\mathbf{X}_{++}, \mathbf{S}_{++})$ computed is called the *sequence of iterates.*

In the next section, we consider the problem of computing the next pair of iterates. Following that, we address the issue of choosing the next target $\mathbf{W}_{++}$.

2.1. **Expanded semidefinite programming problems.** For the sake of clarity, we assume that $\mathbf{W}_{++}$ is the diagonal matrix $\mathbf{D}_{++} \in \mathbb{D}^n_{\downarrow,++}$. This is without any loss of generality as we can transform our primal-dual SDP problems via the orthonormal similarity transformation

$$(\mathbf{X}, \mathbf{S}) \mapsto (\mathbf{Q}^T\mathbf{X}\mathbf{Q}, \mathbf{Q}^T\mathbf{S}\mathbf{Q}),$$

where $\mathbf{Q} \in \mathbb{O}^n$ is such that $\mathbf{Q}^T\mathbf{W}_{++}\mathbf{Q} = \mathbf{D}_{++}$ is a diagonalization of $\mathbf{W}_{++}$.

Consider the pair of Cholesky weighted centers $\boldsymbol{\mathcal{T}}^{-1}(\mathbf{D}_{++})$: the unique pair of matrices $(\mathbf{X}, \mathbf{S})$ satisfying

$$\begin{aligned} \mathbf{A}_k \bullet \mathbf{X} = \mathbf{b}_k \ (1 \le k \le m), &\quad \mathbf{X} \in \mathbb{S}^n_{++}, \\ \sum_{k=1}^m \mathbf{y}_k\mathbf{A}_k + \mathbf{S} = \mathbf{C}, &\quad \mathbf{S} \in \mathbb{S}^n_{++}, \\ \mathbf{L}_\mathbf{S}^T\mathbf{X}\mathbf{L}_\mathbf{S} = \mathbf{D}_{++}. & \end{aligned} \tag{$CP_{\mathbf{D}_{++}}$}$$

Suppose further that all entries in $\mathbf{D}_{++}$ are rational numbers. Then there exists a positive real number κ and positive integers $w_1, \ldots, w_n$ such that $\mathbf{D}_{++} = \kappa \operatorname{Diag}(w_1, \ldots, w_n)$. Recall that $w_1 \ge \cdots \ge w_n$. For each $l \in \{1, \ldots, n-1\}$, let π_l denote the difference $w_l - w_{l+1}$, and let $\pi_n = w_n$. Let $\mathcal{L}$ denote the index set $\{l : \pi_l > 0\}$. Note that $\mathcal{L} \supseteq \{n\}$ is nonempty.

Let $\mathcal{S}$ denote the Cartesian product

$$\underbrace{\mathbb{S}^1 \times \cdots \times \mathbb{S}^1}_{\pi_1 \text{ copies}} \times \underbrace{\mathbb{S}^2 \times \cdots \times \mathbb{S}^2}_{\pi_2 \text{ copies}} \times \cdots \times \underbrace{\mathbb{S}^n \times \cdots \times \mathbb{S}^n}_{\pi_n \text{ copies}},$$

and let $\mathcal{P}_l$ denote the index set $\left\{1 + \sum_{j=1}^{l-1} \pi_j, \ldots, \sum_{j=1}^{l} \pi_j\right\}$ for each $l \in \{1, \ldots, n\}$ so that the p-th component $\mathfrak{X}_p$ of any element $\mathfrak{X} \in \mathcal{S}$ is a matrix in $\mathbb{S}^l$ whenever $p \in \mathcal{P}_l$. Note that $\mathcal{P}_l$ is empty when $l \notin \mathcal{L}$. Let $\mathcal{S}_+$ and $\mathcal{S}_{++}$ denote cones of $\mathcal{S}$ containing elements with positive semidefinite and positive definite components, respectively.

Define the injective linear map $\boldsymbol{\mathcal{E}} : \mathbb{S}^n \to \mathcal{S}$ by

$$\mathbf{X} \mapsto (\underbrace{[\mathbf{X}]_1, \ldots, [\mathbf{X}]_1}_{\pi_1 \text{ copies}}, \underbrace{[\mathbf{X}]_2, \ldots, [\mathbf{X}]_2}_{\pi_2 \text{ copies}}, \ldots, \underbrace{[\mathbf{X}]_n, \ldots, [\mathbf{X}]_n}_{\pi_n \text{ copies}}).$$

Its adjoint map $\boldsymbol{\mathcal{E}}^H$ satisfies

$$(\boldsymbol{\mathcal{E}}^H(\mathfrak{X}))_{ij} = \sum_{l=i\vee j}^n \sum_{p\in\mathcal{P}_l} (\mathfrak{X}_p)_{ij} \ (1 \le i, j \le n).$$

Let $\mathfrak{A}_1, \ldots, \mathfrak{A}_m$ and $\mathfrak{C}$ denote, respectively, $\boldsymbol{\mathcal{E}}(\mathbf{A}_1), \ldots, \boldsymbol{\mathcal{E}}(\mathbf{A}_m)$ and $\boldsymbol{\mathcal{E}}(\mathbf{C})$. Consider the expanded primal-dual SDP problems

$$\begin{aligned} \inf_{\mathfrak{X}} &\quad \sum_{p=1}^{w_1} \mathfrak{C}_p \bullet \mathfrak{X}_p \\ \text{subject to} &\quad \sum_{p=1}^{w_1} (\mathfrak{A}_k)_p \bullet \mathfrak{X}_p = \mathbf{b}_k \ (1 \le k \le m), \ \mathfrak{X} \in \mathcal{S}_+, \end{aligned} \tag{$\widetilde{SDP}$}$$

and

$$\begin{array}{ll}\sup_{\mathfrak{S},\mathbf{y}} & \mathbf{b}^T\mathbf{y} \\ \text{subject to} & \sum_{k=1}^{m}\mathbf{y}_k(\mathfrak{A}_k)_p + \mathfrak{S}_p = \mathfrak{C}_p \ (1 \le p \le w_1), \ \mathfrak{S} \in \mathcal{S}_+. \end{array} \qquad (\widetilde{SDD})$$

Let us look at the pair of analytic centers $(\mathfrak{X}(\kappa\mathbf{I}), \mathfrak{S}(\kappa\mathbf{I}))$ of the pair of problems $(\widetilde{SDP}, \widetilde{SDD})$: the unique pair of tuples $(\mathfrak{X}, \mathfrak{S})$ satisfying the central path equations

$$\sum_{p=1}^{w_1}(\mathfrak{A}_k)_p \bullet \mathfrak{X}_p = \mathbf{b}_k \ (1 \le k \le m), \quad \mathfrak{X} \in \mathcal{S}_{++},$$

$$\sum_{k=1}^{m}\mathbf{y}_k(\mathfrak{A}_k)_p + \mathfrak{S}_p = \mathfrak{C}_p \ (1 \le p \le w_1), \quad \mathfrak{S} \in \mathcal{S}_{++},$$

$$\langle\!\langle \mathfrak{X}_p\mathfrak{S}_p \rangle\!\rangle_H = \kappa\mathbf{I} \ (1 \le p \le w_1).$$

Let $\widehat{\mathbf{X}}$ and $\widehat{\mathbf{S}}$ denote, respectively, $\boldsymbol{\mathcal{E}}^H(\mathfrak{X}(\kappa\mathbf{I}))$ and $\boldsymbol{\mathcal{E}}^{-1}(\mathfrak{S}(\kappa\mathbf{I}))$. It is straightforward to check that $\mathbf{v} \in \mathbb{R}^n \mapsto \mathbf{v}^T\widehat{\mathbf{X}}\mathbf{v}$, whence $\widehat{\mathbf{X}}$, is positive definite, and that $\mathbf{A}_k \bullet \widehat{\mathbf{X}} = \mathbf{b}_k$ for each $k \in \{1, \ldots, m\}$. Thus $\widehat{\mathbf{X}}$ is strictly feasible for (SDP). Also, $\widehat{\mathbf{S}} = \mathfrak{S}(\kappa\mathbf{I})_{w_1} \in \mathbb{S}^n_{++}$ and $\sum_{k=1}^{m}\mathbf{y}_k(\mathfrak{A}_k)_{w_1} + \mathfrak{S}(\kappa\mathbf{I})_{w_1} = \mathfrak{C}_{w_1}$ shows that $\widehat{\mathbf{S}}$ is strictly feasible for (SDD). Moreover, the bilinear equations in the central path equations imply that $\langle\!\langle \widehat{\mathbf{X}}\widehat{\mathbf{S}} \rangle\!\rangle_H = \mathbf{D}_{++}$, or equivalently,

$$\mathbf{L}_{\widehat{\mathbf{S}}}^T\widehat{\mathbf{X}}\mathbf{L}_{\widehat{\mathbf{S}}} = \mathbf{D}_{++}.$$

Thus $(\boldsymbol{\mathcal{E}}^H(\mathfrak{X}(\kappa\mathbf{I})), \boldsymbol{\mathcal{E}}^{-1}(\mathfrak{S}(\kappa\mathbf{I})))$ is the pair of Cholesky weighted centers $\boldsymbol{\mathcal{T}}^{-1}(\mathbf{D}_{++})$.

This observation allows us to view Cholesky weighted centers as (unweighted) analytic centers of a pair of larger primal-dual SDP problems. Moreover, all existing path-following algorithms and their analyses apply directly to Cholesky weighted centers via this observation.

It is immediately clear that without further exploitation of the special structures of the expanded problems, this approach is ill-advised as $\dim(\mathcal{S}) = \sum_{l=1}^{n} w_l l$, the size of the expanded pair, is (generally) much larger than $\dim(\mathbb{S}^n) = \sum_{l=1}^{n} l$, the size of the original pair. This much larger size affects computational complexity of the resulting algorithm in two ways:

(1) the step size at each iteration, hence the worst-case iteration bound, and
(2) the complexity of the computation of search directions.

2.1.1. *Proximity measure.* We shall use the following measure of proximity to analytic centers of the expanded SDP problems:

$$\begin{aligned}\widetilde{d}_2 : (\mathfrak{X}, \mathfrak{S}; \mu) \in \mathcal{S}_{++} \times \mathcal{S}_{++} \times \mathbb{R}_{++} \mapsto \ & \mu^{-1}\left(\sum_{p=1}^{w_1}\|\boldsymbol{\lambda}(\mathfrak{X}_p\mathfrak{S}_p) - \mu\mathbf{1}\|_2^2\right)^{\frac{1}{2}} \\ = \ & \mu^{-1}\left(\sum_{p=1}^{w_1}\|\mathbf{L}_{\mathfrak{S}_p}^T\mathfrak{X}_p\mathbf{L}_{\mathfrak{S}_p} - \mu\mathbf{I}\|_F^2\right)^{\frac{1}{2}}.\end{aligned}$$

The definition of $\widetilde{d}_2$ only requires $\mathfrak{S} \in \mathcal{S}_{++}$. Moreover, we can extend its definition continuously to include all $\mathfrak{S} \in \mathcal{S}_+ \setminus \mathcal{S}_{++}$. Thus $\widetilde{d}_2$ is well defined over $\mathcal{S} \times \mathcal{S}_+ \times \mathbb{R}_{++}$.

This leads to the following measure of proximity to $\boldsymbol{\mathcal{T}}^{-1}(\mathbf{D}_{++})$:

$$(\mathbf{X}, \mathbf{S}) \mapsto \inf_{\mathfrak{X}} \left\{ \widetilde{d}_2(\mathfrak{X}, \boldsymbol{\mathcal{E}}(\mathbf{S}); \kappa) : \boldsymbol{\mathcal{E}}^H(\mathfrak{X}) = \mathbf{X} \right\}.$$

We compute the infimum in Lemma 2 using the following lemma.

Lemma 1. *Suppose that $u_1 \geq \cdots \geq u_n > u_{n+1} = 0$, $(\mathbf{X}, \mathbf{S}) \in \mathbb{S}^n \times \mathbb{S}^n_{++}$, and $\mu > 0$. Then for every sequence of symmetric matrices $\{\mathfrak{X}_l \in \mathbb{S}^l\}_{l=1}^n$ satisfying*

$$\mathbf{X}_{ij} = \sum_{l=i\vee j}^{n} (u_l - u_{l+1})(\mathfrak{X}_l)_{ij} \ (1 \leq i, j \leq n), \tag{2.1}$$

it holds

$$\begin{aligned}\sum_{l=1}^{n}(u_l - u_{l+1})\|[\mathbf{L_S}]_l^T \mathfrak{X}_l [\mathbf{L_S}]_l - \mu \mathbf{I}\|_F^2 &\geq \sum_{i,j=1}^{n} u_{i\vee j}^{-1}((\mathbf{L_S^T X L_S})_{ij} - \mu u_i \mathbf{I}_{ij})^2 \\ &= \left\| (\mathbf{D}^{-\frac{1}{2}} \langle\!\langle \mathbf{L_S^T X L_S} - \mu \mathbf{D} \rangle\!\rangle)_H \right\|_F^2,\end{aligned}$$

where $\mathbf{D}$ denotes the diagonal matrix $\operatorname{Diag}(u_1, \ldots, u_n)$*. Moreover, equality holds if and only if*

$$\mathfrak{X}_l = [\mathbf{L_S}]_l^{-T} [(\mathbf{D}^{-1} \langle\!\langle \mathbf{L_S^T X L_S} \rangle\!\rangle)_H]_l [\mathbf{L_S}]_l^{-1} \quad \forall l \in \mathcal{L}, \tag{2.2}$$

where $\mathcal{L}$ denotes the set $\{l : u_l > u_{l+1}\}$.

Proof. For each $l \in \mathcal{L}$, let $\mathfrak{Z}_l = [\mathbf{L_S}]_l^T \mathfrak{X}_l [\mathbf{L_S}]_l$. In terms of $\mathfrak{Z}_l$,

$$\begin{aligned}&\sum_{l=1}^{n}(u_l - u_{l+1})\|[\mathbf{L_S}]_l^T \mathfrak{X}_l [\mathbf{L_S}]_l - \mu \mathbf{I}\|_F^2 \\ &= \sum_{l\in\mathcal{L}}(u_l - u_{l+1}) \sum_{i,j=1}^{l} ((\mathfrak{Z}_l)_{ij} - \mu \mathbf{I}_{ij})^2 \\ &= \sum_{i,j=1}^{n} \sum_{l\in\mathcal{L},\ l\geq i\vee j} (u_l - u_{l+1})((\mathfrak{Z}_l)_{ij} - \mu \mathbf{I}_{ij})^2.\end{aligned}$$

Using Cauchy's inequality, we bound, for each $i, j \in \{1, \ldots, n\}$,

$$\begin{aligned}&\sum_{l\in\mathcal{L},\ l\geq i\vee j} (u_l - u_{l+1})((\mathfrak{Z}_l)_{ij} - \mu \mathbf{I}_{ij})^2 \\ &\geq \left(\sum_{l\in\mathcal{L},\ l\geq i\vee j} (u_l - u_{l+1}) \right)^{-1} \left(\sum_{l\in\mathcal{L},\ l\geq i\vee j} (u_l - u_{l+1})((\mathfrak{Z}_l)_{ij} - \mu \mathbf{I}_{ij}) \right)^2.\end{aligned} \tag{2.3}$$

The first part of the lemma then follow from

$$\sum_{l\in\mathcal{L},\ l\geq i\vee j}(u_l-u_{l+1})(\mathfrak{Z}_l)_{ij}=\sum_{l=i\vee j}^{n}(u_l-u_{l+1})\sum_{\tilde{\imath}=i}^{l}\sum_{\tilde{\jmath}=j}^{l}(\mathbf{L_S})_{\tilde{\imath}i}(\overline{\mathfrak{X}}_l)_{\tilde{\imath}\tilde{\jmath}}(\mathbf{L_S})_{\tilde{\jmath}j}$$
$$=\sum_{\tilde{\imath}=i}^{n}\sum_{\tilde{\jmath}=j}^{n}\sum_{l=i\vee j}^{n}(u_l-u_{l+1})(\mathbf{L_S})_{\tilde{\imath}i}(\overline{\mathfrak{X}}_l)_{\tilde{\imath}\tilde{\jmath}}(\mathbf{L_S})_{\tilde{\jmath}j}$$
$$=\sum_{\tilde{\imath}=i}^{n}\sum_{\tilde{\jmath}=j}^{n}(\mathbf{L_S})_{\tilde{\imath}i}\mathbf{X}_{\tilde{\imath}\tilde{\jmath}}(\mathbf{L_S})_{\tilde{\jmath}j}=(\mathbf{L_S}^T\mathbf{X}\mathbf{L_S})_{ij},$$

for all $i,j\in\{1,\dots,n\}$.

Equality in (2.3) holds if and only if

$$(\mathfrak{Z}_l)_{ij}=(\mathfrak{Z}_{\tilde{l}})_{ij}\quad\forall l,\tilde{l}\in\mathcal{L}\cap\{i\vee j,\dots,n\};$$

i.e., there exists $\mathbf{Z}\in\mathbb{S}^n$ such that

$$\mathfrak{Z}_l=[\mathbf{Z}]_l\quad\forall l\in\mathcal{L}. \tag{2.4}$$

By (2.1), any $\mathbf{Z}\in\mathbb{S}^n$ satisfying the above set of equations must also satisfy

$$\mathbf{X}_{ij}=\sum_{l\in\mathcal{L},\ l\geq i\vee j}(u_l-u_{l+1})([\mathbf{L_S}]_l^{-T}[\mathbf{Z}]_l[\mathbf{L_S}]_l^{-1})_{ij}\ (1\leq i,j\leq n). \tag{2.5}$$

Let $\boldsymbol{\mathcal{F}}:\mathbb{S}^n\to\bigoplus_{l\in\mathcal{L}}\mathbb{S}^l$ be restriction of the map $\mathbf{X}\mapsto([\mathbf{X}]_1,\dots,[\mathbf{X}]_n)$ to the subspace $\bigoplus_{l\in\mathcal{L}}\mathbb{S}^l$. Consider the following inner product on $\bigoplus_{l\in\mathcal{L}}\mathbb{S}^l$:

$$(\mathfrak{A},\mathfrak{B})\mapsto\sum_{l\in\mathcal{L}}(u_l-u_{l+1})\mathfrak{A}_l\bullet\mathfrak{B}_l.$$

Under this inner product, the adjoint $\boldsymbol{\mathcal{F}}^H$ of $\boldsymbol{\mathcal{F}}$ satisfies

$$(\boldsymbol{\mathcal{F}}^H(\mathfrak{X}))_{ij}=\sum_{l=i\vee j}^{n}(u_l-u_{l+1})(\mathfrak{X}_l)_{ij}\ (1\leq i,j\leq n),$$

hence (2.5) is equivalent to $\mathbf{X}=\boldsymbol{\mathcal{F}}^H(\boldsymbol{\mathcal{F}}(\mathbf{L_S})^{-T}\boldsymbol{\mathcal{F}}(\mathbf{Z})\boldsymbol{\mathcal{F}}(\mathbf{L_S})^{-1})$. For all $\mathbf{W}\in\mathbb{S}^n$,

$$\boldsymbol{\mathcal{F}}^H(\boldsymbol{\mathcal{F}}(\mathbf{L_S})^T\boldsymbol{\mathcal{F}}(\mathbf{Z})\boldsymbol{\mathcal{F}}(\mathbf{L_S}))\bullet\mathbf{W}=\sum_{l\in\mathcal{L}}(u_l-u_{l+1})[\mathbf{Z}]_l\bullet[\mathbf{L_S}]_l[\mathbf{W}]_l[\mathbf{L_S}]_l^T$$
$$=\sum_{l\in\mathcal{L}}(u_l-u_{l+1})[\mathbf{Z}]_l\bullet[\mathbf{L_S}\mathbf{W}\mathbf{L_S}^T]_l$$
$$=(\mathbf{L_S}^T\boldsymbol{\mathcal{F}}^H(\boldsymbol{\mathcal{F}}(\mathbf{Z}))\mathbf{L_S})\bullet\mathbf{W},$$

hence (2.5) is equivalent to $\mathbf{X}=\mathbf{L_S}^{-T}\boldsymbol{\mathcal{F}}^H(\boldsymbol{\mathcal{F}}(\mathbf{Z}))\mathbf{L_S}^{-1}$. The map $\boldsymbol{\mathcal{F}}$ is clearly injective, hence $(\boldsymbol{\mathcal{F}}^H\circ\boldsymbol{\mathcal{F}})^{-1}$ is bijective. Subsequently the only $\mathbf{Z}\in\mathbb{S}^n$ satisfying (2.5) is

$$(\boldsymbol{\mathcal{F}}^H\circ\boldsymbol{\mathcal{F}})^{-1}(\mathbf{L_S}^T\mathbf{X}\mathbf{L_S})=(\mathbf{D}^{-1}\langle\!\langle\mathbf{L_S}^T\mathbf{X}\mathbf{L_S}\rangle\!\rangle)_H,$$

where the equality follows from $\boldsymbol{\mathcal{F}}^H\circ\boldsymbol{\mathcal{F}}:\mathbf{V}\in\mathbb{S}^n\mapsto(\mathbf{D}\langle\!\langle\mathbf{V}\rangle\!\rangle)_H$. Consequently equality in (2.3) holds if and only if (2.2) holds. $\square$

Lemma 2. *Suppose* $(\mathbf{X}, \mathbf{S}) \in \mathbb{S}^n \times \mathbb{S}^n_{++}$ *and* $\mu > 0$*. Then*

$$\begin{aligned}&\inf_{\mathfrak{X}} \left\{ \widetilde{d}_2(\mathfrak{X}, \boldsymbol{\mathcal{E}}(\mathbf{S}); \mu) : \boldsymbol{\mathcal{E}}^H(\mathfrak{X}) = \mathbf{X} \right\} \\ &= \widetilde{d}_2\left(\boldsymbol{\mathcal{E}}(\mathbf{L_S})^{-T} \boldsymbol{\mathcal{E}}((\kappa \mathbf{D}_{++}^{-1} \langle\!\langle \mathbf{L_S}^T \mathbf{X} \mathbf{L_S} \rangle\!\rangle)_H) \boldsymbol{\mathcal{E}}(\mathbf{L_S})^{-1}, \boldsymbol{\mathcal{E}}(\mathbf{S}); \mu \right) \\ &= \mu^{-1} \left(\sum_{i,j=1}^{n} w_{i\vee j}^{-1} ((\mathbf{L_S}^T \mathbf{X} \mathbf{L_S})_{ij} - \mu w_i \mathbf{I}_{ij})^2 \right)^{\frac{1}{2}} \\ &= \mu^{-1} \left\| (\sqrt{\kappa} \mathbf{D}_{++}^{-\frac{1}{2}} \langle\!\langle \mathbf{L_S}^T \mathbf{X} \mathbf{L_S} - \mu \kappa^{-1} \mathbf{D}_{++} \rangle\!\rangle)_H \right\|_F .\end{aligned}$$

Proof. Let $\mathfrak{S} = \boldsymbol{\mathcal{E}}(\mathbf{S})$ and let $\mathfrak{X} \in (\boldsymbol{\mathcal{E}}^H)^{-1}(\mathbf{X})$ be arbitrary. For each $l \in \mathcal{L}$, it follows from the convexity of the square of the Frobenius norm that

$$\frac{1}{\pi_l} \sum_{p \in \mathcal{P}_l} \left\| \mathbf{L}_{\mathfrak{S}_p}^T \mathfrak{X}_p \mathbf{L}_{\mathfrak{S}_p} - \mu \mathbf{I} \right\|_F^2 = \frac{1}{\pi_l} \sum_{p \in \mathcal{P}_l} \left\| \mathbf{L}_{[\mathbf{S}]_l}^T \mathfrak{X}_p \mathbf{L}_{[\mathbf{S}]_l} - \mu \mathbf{I} \right\|_F^2 \geq \left\| \mathbf{L}_{[\mathbf{S}]_l}^T \overline{\mathfrak{X}}_l \mathbf{L}_{[\mathbf{S}]_l} - \mu \mathbf{I} \right\|_F^2 ,$$

where $\overline{\mathfrak{X}}_l$ denotes the average $\pi_l^{-1} \sum_{p \in \mathcal{P}_l} \mathfrak{X}_p$. Since $\mathfrak{X}$ is arbitrary, it follows

$$\begin{aligned}&\inf_{\mathfrak{X}} \left\{ \widetilde{d}_2(\mathfrak{X}, \boldsymbol{\mathcal{E}}(\mathbf{S}); \mu) : \boldsymbol{\mathcal{E}}^H(\mathfrak{X}) = \mathbf{X} \right\} \\ &= \inf_{\mathfrak{X}} \left\{ \widetilde{d}_2(\mathfrak{X}, \boldsymbol{\mathcal{E}}(\mathbf{S}); \mu) : \boldsymbol{\mathcal{E}}^H(\mathfrak{X}) = \mathbf{X},\ \mathfrak{X}_p = \mathfrak{X}_q\ \forall p, q \in \mathcal{P}_l,\ \forall l \in \mathcal{L} \right\}.\end{aligned}$$

Thus we may assume without loss of generality that for each $l \in \mathcal{L}$ and all $p \in \mathcal{P}_l$, $\mathfrak{X}_p = \overline{\mathfrak{X}}_l$. The lemma then follows from Lemma 1. □

The following lemma shows that under the proximity measure

$$\begin{aligned}(\mathbf{X}, \mathbf{S}) \in \mathbb{S}^n \times \mathbb{S}^n_+ \mapsto &\inf_{\mathfrak{X}} \left\{ \widetilde{d}_2(\mathfrak{X}, \boldsymbol{\mathcal{E}}(\mathbf{S}); \kappa) : \boldsymbol{\mathcal{E}}^H(\mathfrak{X}) = \mathbf{X} \right\} \\ &= \kappa^{-1} \left(\sum_{i,j=1}^{n} w_{i\vee j}^{-1} ((\mathbf{L_S}^T \mathbf{X} \mathbf{L_S})_{ij} - \kappa w_i \mathbf{I}_{ij})^2 \right)^{\frac{1}{2}} ,\end{aligned} \tag{2.6}$$

solutions on the boundary of the primal-dual feasible regions are at a distance at least $\sqrt{w_n}$ from $\boldsymbol{\mathcal{T}}^{-1}(\mathbf{D}_{++})$. This suggests scaling the measure (2.6) by $w_n^{-\frac{1}{2}}$.

Lemma 3. *If* $u_1 \geq \cdots \geq u_n > 0$*,* $\mu > 0$*, and* $\mathbf{Z} \in \mathbb{S}^n$*, then*

$$\sum_{i,j=1}^{n} u_{i\vee j}^{-1} (\mathbf{Z}_{ij} - \mu u_i \mathbf{I}_{ij})^2 \geq \sum_{i=1}^{n} u_i^{-1} \left(\boldsymbol{\lambda}(\mathbf{Z})_i - \mu u_i \right)^2 .$$

Consequently

$$\inf \left\{ \mu^{-1} \sum_{i,j=1}^{n} u_{i\vee j}^{-1} (\mathbf{Z}_{ij} - \mu u_i \mathbf{I}_{ij})^2 : \mathbf{Z} \in \mathbb{S}^n_+ \setminus \mathbb{S}^n_{++} \right\} = \mu u_n.$$

Proof. By expanding both sides of the desired inequality, it is clear that we only need to bound the sum $\sum_{i,j=1}^n u_{i\vee j}^{-1}\mathbf{Z}_{ij}^2$ from below by $\sum_{i=1}^n u_i^{-1}\boldsymbol{\lambda}(\mathbf{Z})_i^2$. Since $\mathbf{Z}^2$ is symmetric, there exists an orthogonal matrix $\mathbf{Q}\in\mathbb{O}^n$ such that $\mathbf{Z}^2=\mathbf{Q}\mathrm{Diag}(\boldsymbol{\lambda}(\mathbf{Z}))^2\mathbf{Q}^T$, which leads to

$$\begin{bmatrix}(\mathbf{Z}^2)_{11}\\ \vdots\\ (\mathbf{Z}^2)_{nn}\end{bmatrix}=\begin{bmatrix}\mathbf{Q}_{11}^2 & \cdots & \mathbf{Q}_{1n}^2\\ \vdots & \ddots & \vdots\\ \mathbf{Q}_{n1}^2 & \cdots & \mathbf{Q}_{nn}^2\end{bmatrix}\begin{bmatrix}\boldsymbol{\lambda}(\mathbf{Z})_1^2\\ \vdots\\ \boldsymbol{\lambda}(\mathbf{Z})_n^2\end{bmatrix},$$

where the matrix on the right side of the above equation is doubly-stochastic. By the Hardy, Littlewood and Pólya theorem [4], we have

$$\sum_{i=1}^l(\mathbf{Z}^2)_{ii}\le\sum_{i=1}^l\boldsymbol{\lambda}(\mathbf{Z})_i^2$$

for all $l\in\{1,\dots,n\}$. Consequently by writing

$$\sum_{i,j=1}^n u_{i\vee j}^{-1}\mathbf{Z}_{ij}^2=u_n^{-1}\sum_{i,j=1}^n\mathbf{Z}_{ij}^2-\sum_{l=1}^{n-1}(u_{l+1}^{-1}-u_l^{-1})\sum_{i,j=1}^l\mathbf{Z}_{ij}^2,$$

and using the upper bounds

$$\sum_{i,j=1}^l\mathbf{Z}_{ij}^2\le\sum_{i=1}^l(\mathbf{Z}^2)_{ii}\le\sum_{i=1}^l\boldsymbol{\lambda}(\mathbf{Z})_i^2\quad(1\le l\le n),$$

we conclude the desired inequality

$$\sum_{i,j=1}^n u_{i\vee j}^{-1}\mathbf{Z}_{ij}^2\ge\sum_{i=1}^n u_i^{-1}\boldsymbol{\lambda}(\mathbf{Z})_i^2,$$

hence proving the theorem. □

We shall use the scaled proximity measure $d_2:\mathbb{S}^n\times\mathbb{S}^n_+\times\mathbb{D}^n_{\downarrow,++}\to\mathbb{R}$ defined by

$$\begin{aligned}(\mathbf{X},\mathbf{S};\mathbf{D})\mapsto\mathbf{D}_{nn}^{-\frac{1}{2}}&\left(\sum_{i,j=1}^n\mathbf{D}_{i\vee j,i\vee j}^{-1}\left((\mathbf{L}_\mathbf{S}^T\mathbf{X}\mathbf{L}_\mathbf{S})_{ij}-\mathbf{D}_{ij}\right)^2\right)^{\frac{1}{2}}\\ &=\mathbf{D}_{nn}^{-\frac{1}{2}}\left\|(\mathbf{D}^{-\frac{1}{2}}\langle\!\langle\mathbf{L}_\mathbf{S}^T\mathbf{X}\mathbf{L}_\mathbf{S}-\mathbf{D}\rangle\!\rangle)_H\right\|_F.\end{aligned}\tag{2.7}$$

Note that we do not restrict $\mathbf{D}$ to have rational entries only. When we restrict $\mathbf{D}$ to be a positive multiple of the identity matrix $\mathbf{I}$, the proximity measure naturally reduces to the standard l_2-proximity measure.

When $\mathbf{D}$ has rational entries with $\mathbf{D}=\kappa\,\mathrm{Diag}(w_1,\dots,w_n)$, κ positive and $w_1,\dots,w_n$ integers, this proximity measure can be written

$$\begin{aligned}d_2(\mathbf{X},\mathbf{S};\mathbf{D})&=w_n^{-\frac{1}{2}}\inf_{\mathfrak{X}}\left\{\widetilde{d}_2(\mathfrak{X},\boldsymbol{\mathcal{E}}(\mathbf{S});\kappa):\boldsymbol{\mathcal{E}}^H(\mathfrak{X})=\mathbf{X}\right\}\\ &=w_n^{-\frac{1}{2}}\widetilde{d}_2(\boldsymbol{\mathcal{E}}(\mathbf{L}_\mathbf{S})^{-T}\boldsymbol{\mathcal{E}}((\kappa\mathbf{D}^{-1}\langle\!\langle\mathbf{L}_\mathbf{S}^T\mathbf{X}\mathbf{L}_\mathbf{S}\rangle\!\rangle)_H)\boldsymbol{\mathcal{E}}(\mathbf{L}_\mathbf{S})^{-1},\boldsymbol{\mathcal{E}}(\mathbf{S});\kappa).\end{aligned}\tag{2.8}$$

In the context of the expanded SDP problems, the scaling factor $w_n^{-\frac{1}{2}}$ corresponds to the use of a larger neighborhood size, hence allowing for larger step sizes, or full steps with farther targets. This scaling factor is further justified in the following theorem.

Theorem 2. *Suppose that* $\alpha \in [0,1] \mapsto (\mathbf{X}_\alpha, \mathbf{S}_\alpha, \mu_\alpha) \in \mathbb{S}^n \times \mathbb{S}^n \times \mathbb{R}_{++}$ *is continuous with* $(\mathbf{X}_0, \mathbf{S}_0) \in \mathbb{S}^n_{++} \times \mathbb{S}^n_{++}$*. If there exist* $\mathbf{D} \in \mathbb{D}^n_{\downarrow,++}$ *and* $\beta < 1$ *such that*

$$\mu_\alpha^{-1} \|(\mathbf{D}^{-\frac{1}{2}} \langle\!\langle \mathbf{L}_{\mathbf{S}_\alpha}^T \mathbf{X}_\alpha \mathbf{L}_{\mathbf{S}_\alpha} - \mu_\alpha \mathbf{D} \rangle\!\rangle)_H\|_F \le \beta \sqrt{\mathbf{D}_{nn}} \tag{2.9}$$

whenever $\mathbf{S}_\alpha \in \mathbb{S}^n_{++}$*, then* $(\mathbf{X}_\alpha, \mathbf{S}_\alpha) \in \mathbb{S}^n_{++} \times \mathbb{S}^n_{++}$ *for all* $\alpha \in [0,1]$.

Proof. Let $\widehat{\alpha} = \inf\{\alpha \in [0,1] : (\mathbf{X}_\alpha, \mathbf{S}_\alpha) \notin \mathbb{S}^n_{++} \times \mathbb{S}^n_{++}\}$. Suppose on the contrary $\widehat{\alpha} \le 1$. By the continuity assumption, $\widehat{\alpha} > 0$. Under the hypothesis (2.9),

$$\sum_{i,j=1}^n \mathbf{D}_{i\vee j, i\vee j}^{-1} \left((\mathbf{L}_{\mathbf{S}_\alpha}^T \mathbf{X}_\alpha \mathbf{L}_{\mathbf{S}_\alpha})_{ij} - \mu_\alpha \mathbf{D}_{ij} \right)^2 \le \beta^2 \mathbf{D}_{nn} \mu_\alpha^2$$

for all $0 \le \alpha < \widehat{\alpha}$, and subsequently

$$\liminf_{\alpha \to \widehat{\alpha}} \sum_{i,j=1}^n \mu_\alpha^{-1} \mathbf{D}_{i\vee j, i\vee j}^{-1} \left((\mathbf{L}_{\mathbf{S}_\alpha}^T \mathbf{X}_\alpha \mathbf{L}_{\mathbf{S}_\alpha})_{ij} - \mu_\alpha \mathbf{D}_{ij} \right)^2 \le \beta^2 \mathbf{D}_{nn} \mu_{\widehat{\alpha}} < \mathbf{D}_{nn} \mu_{\widehat{\alpha}}.$$

On the other hand, Lemma 3 implies

$$\liminf_{\alpha \to \widehat{\alpha}} \sum_{i,j=1}^n \mu_\alpha^{-1} \mathbf{D}_{i\vee j, i\vee j}^{-1} \left((\mathbf{L}_{\mathbf{S}_\alpha}^T \mathbf{X}_\alpha \mathbf{L}_{\mathbf{S}_\alpha})_{ij} - \mu_\alpha \mathbf{D}_{ij} \right)^2 \ge \mathbf{D}_{nn} \mu_{\widehat{\alpha}},$$

a contradiction. □

2.1.2. *Search directions.* In this section, we discuss the computation of search directions. Once again, we use the pair of expanded SDP problems ($\widetilde{SDP}$, $\widetilde{SDD}$) for our purpose. As discussed in the preceding section, we may (and should) use

$$\begin{aligned}(\mathfrak{X}_+, \mathfrak{S}_+) &= (\boldsymbol{\mathcal{E}}(\mathbf{L}_{\mathbf{S}_+})^{-T} \boldsymbol{\mathcal{E}}((\kappa \mathbf{D}_{++}^{-1} \langle\!\langle \mathbf{L}_{\mathbf{S}_+}^T \mathbf{X}_+ \mathbf{L}_{\mathbf{S}_+} \rangle\!\rangle)_H) \boldsymbol{\mathcal{E}}(\mathbf{L}_{\mathbf{S}_+})^{-1}, \boldsymbol{\mathcal{E}}(\mathbf{S}_+)) \\ &= \operatorname*{argmin}_{\mathfrak{X}, \mathfrak{S}} \left\{ \widetilde{d}_2(\mathfrak{X}, \mathfrak{S}; \kappa) : (\mathbf{X}_+, \mathbf{S}_+) = (\boldsymbol{\mathcal{E}}^H(\mathfrak{X}), \boldsymbol{\mathcal{E}}^{-1}(\mathfrak{S})) \right\}\end{aligned}$$

as the pair of current iterates for the expanded SDP problems.

The pair of search directions $(\boldsymbol{\Delta}_{\mathfrak{X}}, \boldsymbol{\Delta}_{\mathfrak{S}})$ for the expanded SDP problems is obtained by linearizing the constraints

$$\boldsymbol{\mathcal{H}}_l(\mathfrak{X}_p, \mathfrak{S}_p) = \kappa \mathbf{I} \ (l \in \mathcal{L},\ p \in \mathcal{P}_l)$$

at $(\mathfrak{X}_+, \mathfrak{S}_+)$, for some maps $\{\boldsymbol{\mathcal{H}}_l : \mathbb{S}^l_{++} \times \mathbb{S}^l_{++} \mapsto \mathbb{S}^l : l \in \mathcal{L}\}$ satisfying

$$\boldsymbol{\mathcal{H}}_l(\mathbf{X}, \mathbf{S}) = \mu \mathbf{I} \iff \mathbf{X}\mathbf{S} = \mu \mathbf{I}$$

for all $\mu > 0$. In other words, $(\boldsymbol{\Delta}_{\mathfrak{X}}, \boldsymbol{\Delta}_{\mathfrak{S}})$ solves

$$\sum_{p=1}^{w_1} (\mathfrak{A}_k)_p \bullet (\boldsymbol{\Delta}_{\mathfrak{X}})_p = 0 \ (1 \le k \le m), \tag{2.10a}$$

$$\sum_{k=1}^{m} (\boldsymbol{\Delta}_{\mathbf{y}})_k (\mathfrak{A}_k)_p + (\boldsymbol{\Delta}_{\mathfrak{S}})_p = \mathbf{0} \ (1 \le p \le w_1), \tag{2.10b}$$

$$D\boldsymbol{\mathcal{H}}_l((\mathfrak{X}_+)_p, (\mathfrak{S}_+)_p)((\boldsymbol{\Delta}_{\mathfrak{X}})_p, (\boldsymbol{\Delta}_{\mathfrak{S}})_p) = \kappa \mathbf{I} - \boldsymbol{\mathcal{H}}_l((\mathfrak{X}_+)_p, (\mathfrak{S}_+)_p) \ (l \in \mathcal{L}, \ p \in \mathcal{P}_l), \tag{2.10c}$$

where $D\boldsymbol{\mathcal{H}}_l((\mathfrak{X}_+)_p, (\mathfrak{S}_+)_p)$ denotes the gradient of $\boldsymbol{\mathcal{H}}_l$ at $((\mathfrak{X}_+)_p, (\mathfrak{S}_+)_p)$; i.e., the linear map

$$(\mathbf{U}, \mathbf{V}) \in \mathbb{S}^l \times \mathbb{S}^l \mapsto \partial_{\mathbf{X}} \boldsymbol{\mathcal{H}}_l((\mathfrak{X}_+)_p, (\mathfrak{S}_+)_p)[\mathbf{U}] + \partial_{\mathbf{S}} \boldsymbol{\mathcal{H}}_l((\mathfrak{X}_+)_p, (\mathfrak{S}_+)_p)[\mathbf{V}].$$

This linear system has $2\sum_{l=1}^n w_l l + m = \Theta(\sum_{l \in L} \pi_l l^2)$ variables. The pair of search directions for the original SDP problems is then given by

$$(\boldsymbol{\Delta}_{\mathbf{X}}, \boldsymbol{\Delta}_{\mathbf{S}}) = (\boldsymbol{\mathcal{E}}^H(\boldsymbol{\Delta}_{\mathfrak{X}}), \boldsymbol{\mathcal{E}}^{-1}(\boldsymbol{\Delta}_{\mathfrak{S}})).$$

By the choice of $\mathfrak{X}_+$, there exists, for each $l \in \mathcal{L}$, $(\overline{\mathfrak{X}_+})_l \in \mathbb{S}^l$ such that

$$(\overline{\mathfrak{X}_+})_l = (\mathfrak{X}_+)_p \ (p \in \mathcal{P}_l).$$

By symmetry, it follows that there is a search direction $\boldsymbol{\Delta}_{\mathfrak{X}}$ such that for each $l \in \mathcal{L}$,

$$(\boldsymbol{\Delta}_{\mathfrak{X}})_p = \boldsymbol{\Delta}_{\mathbf{X}(l)} \ (p \in \mathcal{P}_l)$$

for some $\boldsymbol{\Delta}_{\mathbf{X}(l)} \in \mathbb{S}^l$. Thus we may compute the search directions by solving the smaller system

$$\sum_{l=1}^{n} \pi_l [\mathbf{A}_k]_l \bullet \boldsymbol{\Delta}_{\mathbf{X}(l)} = 0 \ (1 \le k \le m), \tag{2.11a}$$

$$\sum_{k=1}^{m} (\boldsymbol{\Delta}_{\mathbf{y}})_k \mathbf{A}_k + \boldsymbol{\Delta}_{\mathbf{S}} = \mathbf{0}, \tag{2.11b}$$

$$D\boldsymbol{\mathcal{H}}_l((\overline{\mathfrak{X}_+})_l, (\overline{\mathfrak{S}_+})_l)(\boldsymbol{\Delta}_{\mathbf{X}(l)}, [\boldsymbol{\Delta}_{\mathbf{S}}]_l) = \kappa \mathbf{I} - \boldsymbol{\mathcal{H}}_l((\overline{\mathfrak{X}_+})_l, (\overline{\mathfrak{S}_+})_l) \ (l \in \mathcal{L}), \tag{2.11c}$$

where $(\overline{\mathfrak{S}_+})_l$ denotes $[\mathbf{S}_+]_l$, and set

$$(\boldsymbol{\Delta}_{\mathbf{X}})_{ij} = \sum_{l \in \mathcal{L}, \ l \ge i \vee j} \pi_l (\boldsymbol{\Delta}_{\mathbf{X}(l)})_{ij} \tag{2.12}$$

for each $i, j \in \{1, \ldots, n\}$. This system has $\Theta(\sum_{l \in \mathcal{L}} l^2) = O(n^3)$ variables. The number of variables is actually $\Omega(n^3)$ in certain cases; e.g., when $\mathcal{L} = \{1, \ldots, n\}$.

It is necessary for the above system to have unique solution so that the search directions are well defined. Since distinct solutions of the above system give distinct solutions to (2.10), this requirement is satisfied by (2.10) having unique solution. Typically, sufficient conditions for this is given by $(\mathfrak{X}_+, \mathfrak{S}_+) \in \mathcal{S}_{++} \times \mathcal{S}_{++}$, and at times together with the existence of some $\mu > 0$ such that

$$\|\boldsymbol{\lambda}((\mathfrak{X}_+)_p (\mathfrak{S}_+)_p) - \mu \mathbf{1}\|_\infty \le \gamma \mu \ \forall p \in \{1, \ldots, w_1\},$$

where $\gamma \in (0,1)$ is given. As $\mathbf{S}_+ \in \mathbb{S}^n_{++}$ implies $\mathfrak{S}_+ \in \mathcal{S}_{++}$, the above condition is sufficient for $\mathfrak{X}_+ \in \mathcal{S}_{++}$. The following lemma shows that this condition is satisfied when $(\mathbf{X}_+, \mathbf{S}_+)$ is sufficiently close to $\boldsymbol{\mathcal{T}}^{-1}(\mathbf{D}_{++})$.

Lemma 4. *It holds*

$$\|\boldsymbol{\lambda}((\mathfrak{X}_+)_p(\mathfrak{S}_+)_p) - \kappa\mathbf{1}\|_2 \leq \kappa d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_{++})$$

for all $p \in \{1, \ldots, w_1\}$.

Proof. By definition,

$$\begin{aligned}
\kappa d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_{++}) &= w_n^{-\frac{1}{2}} \left(\sum_{p=1}^{w_1} \|\mathbf{L}^T_{(\mathfrak{S}_+)_p} (\mathfrak{X}_+)_p \mathbf{L}_{(\mathfrak{S}_+)_p} - \kappa\mathbf{I}\|_F^2 \right)^{\frac{1}{2}} \\
&\geq w_n^{-\frac{1}{2}} \left(\sum_{p\in\mathcal{P}_n} \|\mathbf{L}^T_{(\mathfrak{S}_+)_p} (\mathfrak{X}_+)_p \mathbf{L}_{(\mathfrak{S}_+)_p} - \kappa\mathbf{I}\|_F^2 \right)^{\frac{1}{2}} \\
&= \|\mathbf{V} - \kappa\mathbf{I}\|_F,
\end{aligned}$$

where $\mathbf{V}$ denotes the matrix $(\kappa\mathbf{D}_{++}^{-1}\langle\!\langle \mathbf{L}_\mathbf{S}^T \mathbf{X} \mathbf{L}_\mathbf{S} \rangle\!\rangle)_H$. Consequently for any $p \in \{1, \ldots, w_1\}$,

$$\begin{aligned}
\kappa d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_{++}) \geq \|\mathbf{V} - \kappa\mathbf{I}\|_F &\geq \|[\mathbf{V}]_l - \kappa\mathbf{I}\|_F \\
&= \|\mathbf{L}^T_{(\mathfrak{S}_+)_p} (\mathfrak{X}_+)_p \mathbf{L}_{(\mathfrak{S}_+)_p} - \kappa\mathbf{I}\|_F \\
&= \|\boldsymbol{\lambda}((\mathfrak{X}_+)_p(\mathfrak{S}_+)_p) - \kappa\mathbf{1}\|_2,
\end{aligned}$$

where $l \in \mathcal{L}$ is such that $p \in \mathcal{P}_l$. □

2.2. **Choice of targets.** Suppose that the pair of input matrices $(\mathbf{X}_{\text{in}}, \mathbf{S}_{\text{in}})$ satisfies

$$\mathbf{L}^T_{\mathbf{S}_{\text{in}}} \mathbf{X}_{\text{in}} \mathbf{L}_{\mathbf{S}_{\text{in}}} \in \mathbb{D}^n_{\downarrow,++}.$$

This is without loss of generality if we apply the orthonormal similarity transformation defined by the orthogonal matrix that upper-triangularizes the product $\mathbf{X}_{\text{in}}\mathbf{S}_{\text{in}}$ to both primal and dual SDP problems.

We first consider the task of picking the initial target $\mathbf{W}_+$. Using the proximity measure d_2, the proximity of $\boldsymbol{\mathcal{T}}(\mathbf{X}_{\text{in}}, \mathbf{S}_{\text{in}})$ to $\mathbf{W}_+$ can be quantified by

$$d_2(\mathbf{Q}_+^T \mathbf{X}_{\text{in}} \mathbf{Q}_+, \mathbf{Q}_+^T \mathbf{S}_{\text{in}} \mathbf{Q}_+; \mathbf{D}_+),$$

where $\mathbf{Q}_+ \in \mathbb{O}^n$ and $\mathbf{D}_+ \in \mathbb{D}^n_{\downarrow,++}$ are such that $\mathbf{Q}_+^T \mathbf{W}_+ \mathbf{Q}_+ = \mathbf{D}_+$ is a diagonalization of $\mathbf{W}_+$. By Lemma 3, for each fixed $\mathbf{D}_+ \in \mathbb{D}^n_{\downarrow,++}$, the above measure is minimized at $\mathbf{Q}_+ = \mathbf{I}$. Thus it makes sense to pick $\mathbf{W}_+ \in \mathbb{D}^n_{\downarrow,++}$. Henceforth, we shall assume that $\mathbf{W}_+$ is the diagonal matrix $\mathbf{D}_+ \in \mathbb{D}^n_{\downarrow,++}$.

We now consider the task of picking the next target $\mathbf{W}_{++}$. Once again, the next target $\mathbf{W}_{++}$ should thus be chosen so that $d_2(\mathbf{Q}_{++}^T \mathbf{X}_+ \mathbf{Q}_{++}, \mathbf{Q}_{++}^T \mathbf{S}_+ \mathbf{Q}_{++}; \mathbf{D}_{++})$ can be readily bounded, where $\mathbf{Q}_{++} \in \mathbb{O}^n$ and $\mathbf{D}_{++} \in \mathbb{D}^n_{\downarrow,++}$ are such that $\mathbf{Q}_{++}^T \mathbf{W}_{++} \mathbf{Q}_{++} = \mathbf{D}_{++}$ is a diagonalization of $\mathbf{W}_{++}$. An natural criterion would be the size of $d_2(\mathbf{W}_{++}, \mathbf{I}; \mathbf{D}_+)$. Once again, since Lemma 3 implies that

$$\begin{aligned}\inf\{d_2(\mathbf{Q}^T\mathbf{D}_{++}\mathbf{Q}, \mathbf{I}; \mathbf{D}_+) : \mathbf{Q} \in \mathbb{O}^n\} &= d_2(\mathbf{D}_{++}, \mathbf{I}; \mathbf{D}_+) \\ &= (\mathbf{D}_+)_{nn}^{-\frac{1}{2}} \|\mathbf{D}_{++}\mathbf{D}_+^{-\frac{1}{2}} - \mathbf{D}_+^{\frac{1}{2}}\|_F,\end{aligned}$$

it makes sense to choose $\mathbf{W}_{++} \in \mathbb{D}^n_{\downarrow,++}$.

With these choices of targets, we can use the following lemma to get an upper bound on $d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_{++})$ in terms of $d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_+)$ and $d_2(\mathbf{D}_{++}, \mathbf{I}; \mathbf{D}_+)$.

Lemma 5. *If $d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_+) \leq \beta$ and $(\mathbf{D}_+)_{nn}^{-\frac{1}{2}} \|\mathbf{D}_{++}\mathbf{D}_+^{-\frac{1}{2}} - \mathbf{D}_+^{\frac{1}{2}}\|_F \leq \delta$ for some $\beta, \delta \in (0, 1)$, then*

$$d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_{++}) \leq \frac{\beta + \delta}{1 - \delta}.$$

Proof. For simplicity of notation, let $\mathbf{Z}$ denote the product $\mathbf{L}_{\mathbf{S}_+}^T \mathbf{X}_+ \mathbf{L}_{\mathbf{S}_+}$. From definition,

$$\begin{aligned}&d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_{++}) \\ &= (\mathbf{D}_{++})_{nn}^{-\frac{1}{2}} \left(\sum_{i,j=1}^n (\mathbf{D}_{++})_{i\vee j, i\vee j}^{-1} \left(\mathbf{Z}_{ij} - (\mathbf{D}_{++})_{ij}\right)^2 \right)^{\frac{1}{2}} \\ &\leq (\mathbf{D}_{++})_{nn}^{-\frac{1}{2}} \left(\sum_{i,j=1}^n (\mathbf{D}_{+})_{i\vee j, i\vee j}^{-1} \left(\mathbf{Z}_{ij} - (\mathbf{D}_{++})_{ij}\right)^2 \right)^{\frac{1}{2}} \max_{i=1,\ldots,n} \frac{\sqrt{(\mathbf{D}_+)_{ii}}}{\sqrt{(\mathbf{D}_{++})_{ii}}}\end{aligned}$$

with

$$\begin{aligned}&\left(\sum_{i,j=1}^n (\mathbf{D}_{+})_{i\vee j, i\vee j}^{-1} \left(\mathbf{Z}_{ij} - (\mathbf{D}_{++})_{ij}\right)^2 \right)^{\frac{1}{2}} \\ &\leq \left(\sum_{i,j=1}^n (\mathbf{D}_{+})_{i\vee j, i\vee j}^{-1} \left(\mathbf{Z}_{ij} - (\mathbf{D}_{+})_{ij}\right)^2 \right)^{\frac{1}{2}} + \|\mathbf{D}_{++}\mathbf{D}_+^{-\frac{1}{2}} - \mathbf{D}_+^{\frac{1}{2}}\|_F \\ &= \sqrt{(\mathbf{D}_+)_{nn}} \left(d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_+) + (\mathbf{D}_+)_{nn}^{-\frac{1}{2}} \|\mathbf{D}_{++}\mathbf{D}_+^{-\frac{1}{2}} - \mathbf{D}_+^{\frac{1}{2}}\|_F \right).\end{aligned}$$

If $(\mathbf{D}_+)_{nn}^{-\frac{1}{2}} \|\mathbf{D}_{++}\mathbf{D}_+^{-\frac{1}{2}} - \mathbf{D}_+^{\frac{1}{2}}\|_F \leq \delta$, then

$$\delta \geq (\mathbf{D}_+)_{nn}^{-\frac{1}{2}} \left(\sum_{i=1}^n \left(\frac{(\mathbf{D}_{++})_{ii}}{\sqrt{(\mathbf{D}_+)_{ii}}} - \sqrt{(\mathbf{D}_+)_{ii}} \right)^2 \right)^{\frac{1}{2}} \geq \left| \frac{(\mathbf{D}_{++})_{ii}}{(\mathbf{D}_+)_{ii}} - 1 \right|$$

for all $i \in \{1, \ldots, n\}$, and hence

$$\min_{i=1,\ldots,n} \frac{\sqrt{(\mathbf{D}_{++})_{ii}}}{\sqrt{(\mathbf{D}_+)_{ii}}} \geq \sqrt{1 - \delta}.$$

Consequently,

$$d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_{++}) \leq (\mathbf{D}_{++})_{nn}^{-\frac{1}{2}} \sqrt{(\mathbf{D}_+)_{nn}}(\beta + \delta) \max_{i=1,\ldots,n} \frac{\sqrt{(\mathbf{D}_+)_{ii}}}{\sqrt{(\mathbf{D}_{++})_{ii}}} \leq \frac{\beta + \delta}{1 - \delta}$$

under the hypotheses of the lemma. □

3. Target-Following Framework Based on Cholesky Search Directions

In this section, we highlight a choice of search directions whose Newton system can be further reduced in size.

Consider the maps $\boldsymbol{\mathcal{H}}_l : \mathbb{S}^l_{++} \times \mathbb{S}^l_{++} \mapsto \mathbb{S}^l$ defined by

$$\boldsymbol{\mathcal{H}}_l : (\mathbf{X}, \mathbf{S}) \mapsto \mathbf{L}_{\mathbf{S}}^T \mathbf{X} \mathbf{L}_{\mathbf{S}}.$$

Note that $\boldsymbol{\mathcal{H}}_n(\mathbf{X}, \mathbf{S}) = \mathbf{D}$ is precisely the defining equation for the weighted centers $\boldsymbol{\mathcal{T}}^{-1}(\mathbf{D})$ for each $\mathbf{D} \in \mathbb{D}^n_{\downarrow,++}$. The gradient $D\boldsymbol{\mathcal{H}}_l(\mathbf{X}, \mathbf{S})$ of $\boldsymbol{\mathcal{H}}_l$ at $(\mathbf{X}, \mathbf{S})$ is given by

$$(\mathbf{U}, \mathbf{V}) \mapsto \mathbf{L}_{\mathbf{S}}^T \mathbf{U} \mathbf{L}_{\mathbf{S}} + \left(\mathbf{L}_{\mathbf{S}}^T \mathbf{X} \mathbf{L}_{\mathbf{S}} \langle\!\langle \mathbf{L}_{\mathbf{S}}^{-1} \mathbf{V} \mathbf{L}_{\mathbf{S}}^{-T} \rangle\!\rangle\right)_H.$$

We shall use this choice of $\boldsymbol{\mathcal{H}}_l$ in the linear system (2.10) with

$$\begin{aligned}(\mathfrak{X}_+, \mathfrak{S}_+) &= (\boldsymbol{\mathcal{E}}(\mathbf{L}_{\mathbf{S}_+})^{-T} \boldsymbol{\mathcal{E}}((\kappa \mathbf{D}_{++}^{-1} \langle\!\langle \mathbf{L}_{\mathbf{S}_+}^T \mathbf{X}_+ \mathbf{L}_{\mathbf{S}_+} \rangle\!\rangle)_H) \boldsymbol{\mathcal{E}}(\mathbf{L}_{\mathbf{S}_+})^{-1}, \boldsymbol{\mathcal{E}}(\mathbf{S}_+)) \\ &= \operatorname*{argmin}_{\mathfrak{X},\mathfrak{S}} \left\{ \widetilde{d}_2(\mathfrak{X}, \mathfrak{S}; \kappa) : (\mathbf{X}_+, \mathbf{S}_+) = (\boldsymbol{\mathcal{E}}^H(\mathfrak{X}), \boldsymbol{\mathcal{E}}^{-1}(\mathfrak{S})) \right\}.\end{aligned}$$

This gives

$$\sum_{p=1}^{w_1} (\mathfrak{A}_k)_p \bullet (\boldsymbol{\Delta}_{\mathfrak{X}})_p = 0 \ (1 \leq k \leq m), \tag{3.1a}$$

$$\sum_{k=1}^{m} (\boldsymbol{\Delta}_{\mathbf{y}})_k (\mathfrak{A}_k)_p + (\boldsymbol{\Delta}_{\mathfrak{S}})_p = \mathbf{0} \ (1 \leq p \leq w_1), \tag{3.1b}$$

$$\begin{aligned}&\kappa[\mathbf{Z}]_l + \mathbf{L}_{(\mathfrak{S}_+)_p}^T (\boldsymbol{\Delta}_{\mathfrak{X}})_p \mathbf{L}_{(\mathfrak{S}_+)_p} \\ &+ \left(\kappa[\mathbf{Z}]_l \langle\!\langle \mathbf{L}_{(\mathfrak{S}_+)_p}^{-1} (\boldsymbol{\Delta}_{\mathfrak{S}})_p \mathbf{L}_{(\mathfrak{S}_+)_p}^{-T} \rangle\!\rangle\right)_H = \kappa \mathbf{I} \ (l \in \mathcal{L}, p \in \mathcal{P}_l),\end{aligned} \tag{3.1c}$$

where $\mathbf{Z}$ denotes the matrix $(\mathbf{D}_{++}^{-1} \langle\!\langle \mathbf{L}_{\mathbf{S}_+}^T \mathbf{X}_+ \mathbf{L}_{\mathbf{S}_+} \rangle\!\rangle)_H$, so that

$$\kappa[\mathbf{Z}]_l = \mathbf{L}_{(\mathfrak{S}_+)_p}^T (\mathfrak{X}_+)_p \mathbf{L}_{(\mathfrak{S}_+)_p}$$

for each $l \in \mathcal{L}$ and each $p \in \mathcal{P}_l$. The corresponding pair of search directions for the original SDP problems is called the pair of *Cholesky search directions*, and is given by

$$(\boldsymbol{\Delta}_{\mathbf{X}}, \boldsymbol{\Delta}_{\mathbf{S}}) = (\boldsymbol{\mathcal{E}}^H(\boldsymbol{\Delta}_{\mathfrak{X}}), \boldsymbol{\mathcal{E}}^{-1}(\boldsymbol{\Delta}_{\mathfrak{S}})).$$

Adding up (3.1c) over all $l \in \mathcal{L}$ and all $p \in \mathcal{P}_l$ gives

$$\mathbf{A}_k \bullet \mathbf{\Delta_X} = 0 \; (1 \le k \le m), \tag{3.2a}$$

$$\sum_{k=1}^{m} (\mathbf{\Delta_y})_k \mathbf{A}_k + \mathbf{\Delta_S} = \mathbf{0}, \tag{3.2b}$$

$$\mathbf{V} + \mathbf{L}_{\mathbf{S}_+}^T \mathbf{\Delta_X} \mathbf{L}_{\mathbf{S}_+} + \left(\mathbf{V} \langle\!\langle \mathbf{L}_{\mathbf{S}_+}^{-1} \mathbf{\Delta_S} \mathbf{L}_{\mathbf{S}_+}^{-T} \rangle\!\rangle \right)_H = \mathbf{D}_{++}, \tag{3.2c}$$

where $\mathbf{V}$ denotes $\mathbf{L}_{\mathbf{S}_+}^T \mathbf{X}_+ \mathbf{L}_{\mathbf{S}_+}$. Thus we may compute the pair of search directions $(\mathbf{\Delta_X}, \mathbf{\Delta_S})$ by solving a linear system with only $O(n^2)$ variables. Moreover, we do not require that $\mathbf{D}_{++}$ has rational entries for this system to be well defined. Not surprisingly, this system is the linearization of $(CP_{\mathbf{D}_{++}})$.

The algorithm based on the Cholesky search directions is the following:

Algorithm 2. (Target-following algorithm based on Cholesky search directions)
Given a pair of primal-dual strictly feasible solutions $(\mathbf{X}_{in}, \mathbf{S}_{in})$ *with* $\boldsymbol{\mathcal{T}}(\mathbf{X}_{in}, \mathbf{S}_{in}) \in \mathbb{D}^n_{\downarrow,++}$*, and the required accuracy* $\varepsilon > 0$*.*

(1) *Find a target* $\mathbf{D}_+ \in \mathbb{D}^n_{\downarrow,++}$ *satisfying* $d_2(\mathbf{X}_{in}, \mathbf{S}_{in}; \mathbf{D}_+) \le \beta$ *for some* $\beta \in (0,1)$*. Set* $(\mathbf{X}_+, \mathbf{S}_+) = (\mathbf{X}_{in}, \mathbf{S}_{in})$*.*
(2) *While* $\mathbf{X}_+ \bullet \mathbf{S}_+ > \varepsilon(\mathbf{X}_{in} \bullet \mathbf{S}_{in})$*,*
 (a) *Pick target* $\mathbf{D}_{++} \in \mathbb{D}^n_{\downarrow,++}$ *satisfying*

$$(\mathbf{D}_+)_{nn}^{-\frac{1}{2}} \left\| \mathbf{D}_{++} \mathbf{D}_+^{-\frac{1}{2}} - \mathbf{D}_+^{\frac{1}{2}} \right\|_F \le \delta$$

 for some $\delta \in (0,1)$*.*
 (b) *Solve* (3.2) *and set* $(\mathbf{X}_{++}, \mathbf{S}_{++}) = (\mathbf{X}_+ + \mathbf{\Delta_X}, \mathbf{S}_+ + \mathbf{\Delta_S})$*.*
 (c) *Update* $(\mathbf{X}_+, \mathbf{S}_+) \leftarrow (\mathbf{X}_{++}, \mathbf{S}_{++})$ *and* $\mathbf{D}_+ \leftarrow \mathbf{D}_{++}$*.*
(3) *Output* $(\mathbf{X}_{out}, \mathbf{S}_{out}) = (\mathbf{X}_+, \mathbf{S}_+)$*.*

3.1. **Analysis of algorithm.** For the analysis of this algorithm, we focus on each iteration of the algorithm.

We write $\mathbf{D}_{++} = \mathrm{Diag}(w_1, \ldots, w_n)$, where $w_1, \ldots, w_n \in \mathbb{R}_{++}$. Note that we no longer require the w_i's to be integers. Let π_l denote $w_l - w_{l+1}$ for $l \in \{1, \ldots, n-1\}$, let π_n denote w_n, and let $\mathcal{L}$ denote $\{l : \pi_l > 0\}$. For each $l \in \mathcal{L}$, let $\mathfrak{X}_l$ and $\mathfrak{S}_l$ denote, respectively, $[\mathbf{L}_{\mathbf{S}_+}]_l^{-T} [(\mathbf{D}_{++}^{-1} \langle\!\langle \mathbf{L}_{\mathbf{S}_+}^T \mathbf{X}_+ \mathbf{L}_{\mathbf{S}_+} \rangle\!\rangle)_H]_l [\mathbf{L}_{\mathbf{S}_+}]_l^{-1}$ and $[\mathbf{S}_+]_l$.

We further simplify (3.2) to

$$\widetilde{\mathbf{A}}_k \bullet \widetilde{\mathbf{\Delta}}_\mathbf{X} = 0 \; (1 \le k \le m), \tag{3.3a}$$

$$\sum_{k=1}^{m} (\mathbf{\Delta_y})_k \widetilde{\mathbf{A}}_k + \widetilde{\mathbf{\Delta}}_\mathbf{S} = \mathbf{0}, \tag{3.3b}$$

$$\mathbf{V} + \widetilde{\mathbf{\Delta}}_\mathbf{X} + \left(\mathbf{V} \langle\!\langle \widetilde{\mathbf{\Delta}}_\mathbf{S} \rangle\!\rangle \right)_H = \mathbf{D}_{++}, \tag{3.3c}$$

where $\widetilde{\mathbf{\Delta}}_\mathbf{X}$ and $\widetilde{\mathbf{\Delta}}_\mathbf{S}$ denote, respectively, $\mathbf{L}_{\mathbf{S}_+}^T \mathbf{\Delta_X} \mathbf{L}_{\mathbf{S}_+}$ and $\mathbf{L}_{\mathbf{S}_+}^{-1} \mathbf{\Delta_S} \mathbf{L}_{\mathbf{S}_+}^{-T}$, and $\widetilde{\mathbf{A}}_k$ denotes $\mathbf{L}_{\mathbf{S}_+}^{-1} \mathbf{A}_k \mathbf{L}_{\mathbf{S}_+}^{-T}$ for each $k \in \{1, \ldots, m\}$. For each $\alpha \in \mathbb{R}$, let $\widetilde{\mathbf{X}}_\alpha$ and $\widetilde{\mathbf{S}}_\alpha$ denote, respectively, the sums $\mathbf{V}+\alpha\widetilde{\boldsymbol{\Delta}}_{\mathbf{X}}$ and $\mathbf{I}+\alpha\widetilde{\boldsymbol{\Delta}}_{\mathbf{S}}$. It is easy to check that for each α satisfying $\widetilde{\mathbf{S}}_\alpha \in \mathbb{S}^n_{++}$, it holds $d_2(\mathbf{X}_\alpha, \mathbf{S}_\alpha; \mathbf{D}) = d_2(\widetilde{\mathbf{X}}_\alpha, \widetilde{\mathbf{S}}_\alpha; \mathbf{D})$.

Consider the following linear system:

$$\sum_{l=1}^{n} \pi_l [\widetilde{\mathbf{A}}_k]_l \bullet \widetilde{\boldsymbol{\Delta}}_{\mathbf{X}(l)} = 0 \ (k=1,\ldots,m), \tag{3.4a}$$

$$\sum_{k=1}^{m} (\boldsymbol{\Delta}_{\mathbf{y}})_k \widetilde{\mathbf{A}}_k + \widetilde{\boldsymbol{\Delta}}_{\mathbf{S}} = \mathbf{0}, \tag{3.4b}$$

$$[\mathbf{Z}]_l + \widetilde{\boldsymbol{\Delta}}_{\mathbf{X}(l)} + \Big([\mathbf{Z}]_l \langle\!\langle [\widetilde{\boldsymbol{\Delta}}_{\mathbf{S}}]_l \rangle\!\rangle\Big)_H = \mathbf{I} \ (l \in \mathcal{L}), \tag{3.4c}$$

which is actually (2.11) with $\mathbf{A}_k = \widetilde{\mathbf{A}}_k$ and $((\overline{\mathfrak{X}_+})_l, (\overline{\mathfrak{S}_+})_l) = (\mathfrak{X}_l, \mathfrak{S}_l)$. Thus the solution of this system is related to the solution of (3.3) via (2.12).

We shall now derive an upper bound on the error in the linearization (3.1). The following bound on the Newton step is useful.

Lemma 6. *If $d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_{++}) \le \gamma$ for some $\gamma \in (0, 1/\sqrt{2})$, and $\widehat{\boldsymbol{\Delta}}_{\mathbf{X}(l)} \in \mathbb{S}^l$ $(l \in \mathcal{L})$ and $\widehat{\boldsymbol{\Delta}}_{\mathbf{S}} \in \mathbb{S}^n$ satisfy*

$$\sum_{l\in\mathcal{L}} \pi_l \operatorname{tr} \widehat{\boldsymbol{\Delta}}_{\mathbf{X}(l)} [\widehat{\boldsymbol{\Delta}}_{\mathbf{S}}]_l \ge 0$$

and

$$\widehat{\boldsymbol{\Delta}}_{\mathbf{X}(l)} + \Big([\mathbf{Z}]_l \langle\!\langle [\widehat{\boldsymbol{\Delta}}_{\mathbf{S}}]_l \rangle\!\rangle\Big)_H = \mathbf{M}_l \quad (l \in \mathcal{L})$$

for $\mathbf{Z} = (\mathbf{D}_{++}^{-1} \langle\!\langle \mathbf{L}_{\mathbf{S}_+}^T \mathbf{X}_+ \mathbf{L}_{\mathbf{S}_+} \rangle\!\rangle)_H$, and some $\mathbf{M}_l \in \mathbb{S}^l$ $(l \in \mathcal{L})$, then

$$\max\left\{ \sum_{l\in\mathcal{L}} \pi_l \|\widehat{\boldsymbol{\Delta}}_{\mathbf{X}(l)}\|_F^2, \sum_{l\in\mathcal{L}} \pi_l \|[\widehat{\boldsymbol{\Delta}}_{\mathbf{S}}]_l\|_F^2 \right\} \le \frac{1}{(1-\sqrt{2}\gamma)^2} \sum_{l\in\mathcal{L}} \pi_l \|\mathbf{M}_l\|_F^2.$$

Proof. Since

$$w_n^{-1} \sum_{i,j=1}^{n} w_{i\vee j} \left(w_{i\vee j}^{-1} \mathbf{V}_{ij} - \mathbf{I}_{ij} \right)^2 \ge \sum_{i,j=1}^{n} \left(w_{i\vee j}^{-1} \mathbf{V}_{ij} - \mathbf{I}_{ij} \right)^2 = \|\mathbf{Z} - \mathbf{I}\|_F^2$$

it follows that

$$\|\mathbf{Z} - \mathbf{I}\|_F \le \gamma. \tag{3.5}$$

By summing the following inequalities

$$\max\{\|\widehat{\boldsymbol{\Delta}}_{\mathbf{X}(l)}\|_F^2, \|[\widehat{\boldsymbol{\Delta}}_{\mathbf{S}}]_l\|_F^2\} \le \|\widehat{\boldsymbol{\Delta}}_{\mathbf{X}(l)} + [\widehat{\boldsymbol{\Delta}}_{\mathbf{S}}]_l\|_F^2 - \operatorname{tr} \widehat{\boldsymbol{\Delta}}_{\mathbf{X}(l)} [\widehat{\boldsymbol{\Delta}}_{\mathbf{S}}]_l \quad (l \in \mathcal{L}),$$

we deduce, using $\sum_{l\in\mathcal{L}} \pi_l \operatorname{tr} \widehat{\boldsymbol{\Delta}}_{\mathbf{X}(l)} [\widehat{\boldsymbol{\Delta}}_{\mathbf{S}}]_l \ge 0$, that

$$\max\left\{ \sum_{l\in\mathcal{L}} \pi_l \|\widehat{\boldsymbol{\Delta}}_{\mathbf{X}(l)}\|_F^2, \sum_{l\in\mathcal{L}} \pi_l \|[\widehat{\boldsymbol{\Delta}}_{\mathbf{S}}]_l\|_F^2 \right\} \le \sum_{l\in\mathcal{L}} \pi_l \|\widehat{\boldsymbol{\Delta}}_{\mathbf{X}(l)} + [\widehat{\boldsymbol{\Delta}}_{\mathbf{S}}]_l\|_F^2.$$

It then follows from $\widehat{\mathbf{\Delta}}_{\mathbf{X}(l)} + \left([\mathbf{Z}]_l \langle\!\langle [\widehat{\mathbf{\Delta}}_{\mathbf{S}}]_l \rangle\!\rangle\right)_H = \mathbf{M}_l$ $(l \in \mathcal{L})$ and the triangle inequality on the 2-norm of $\mathbb{R}^n$ that

$$\left(\max\left\{\sum_{l\in\mathcal{L}} \pi_l \|\widehat{\mathbf{\Delta}}_{\mathbf{X}(l)}\|_F^2, \sum_{l\in\mathcal{L}} \pi_l \|[\widehat{\mathbf{\Delta}}_{\mathbf{S}}]_l\|_F^2\right\}\right)^{\frac{1}{2}}$$
$$\leq \left(\sum_{l\in\mathcal{L}} \pi_l \|\mathbf{M}_l\|_F^2\right)^{\frac{1}{2}} + \left(\sum_{l\in\mathcal{L}} \pi_l \left\|\left(([\mathbf{Z}]_l - \mathbf{I})\langle\!\langle [\widehat{\mathbf{\Delta}}_{\mathbf{S}}]_l \rangle\!\rangle\right)_H\right\|_F^2\right)^{\frac{1}{2}}.$$

Using (3.5) we estimate

$$\left\|\left(([\mathbf{Z}]_l - \mathbf{I})\langle\!\langle [\widehat{\mathbf{\Delta}}_{\mathbf{S}}]_l \rangle\!\rangle\right)_H\right\|_F \leq 2\left\|([\mathbf{Z}]_l - \mathbf{I})\langle\!\langle [\widehat{\mathbf{\Delta}}_{\mathbf{S}}]_l \rangle\!\rangle\right\|_F$$
$$\leq 2\gamma \left\|\langle\!\langle [\widehat{\mathbf{\Delta}}_{\mathbf{S}}]_l \rangle\!\rangle\right\|_F \leq \sqrt{2}\gamma \left\|[\widehat{\mathbf{\Delta}}_{\mathbf{S}}]_l\right\|_F.$$

Consequently

$$\left(\max\left\{\sum_{l\in\mathcal{L}} \pi_l \|\widehat{\mathbf{\Delta}}_{\mathbf{X}(l)}\|_F^2, \sum_{l\in\mathcal{L}} \pi_l \|[\widehat{\mathbf{\Delta}}_{\mathbf{S}}]_l\|_F^2\right\}\right)^{\frac{1}{2}}$$
$$\leq \left(\sum_{l\in\mathcal{L}} \pi_l \|\mathbf{M}_l\|_F^2\right)^{\frac{1}{2}} + \sqrt{2}\gamma \left(\sum_{l\in\mathcal{L}} \pi_l \left\|[\widehat{\mathbf{\Delta}}_{\mathbf{S}}]_l\right\|_F^2\right)^{\frac{1}{2}}$$

proves the lemma. □

In addition, we require the following local Lipschitz constant of Cholesky factorization.

Lemma 7. *If $\mathbf{\Delta} \in \mathbb{S}^n$ satisfies $\|\mathbf{\Delta}\|_F \leq 1/2$, then*

$$\|\mathbf{L}_{\mathbf{I}+\mathbf{\Delta}} - \mathbf{I}\|_F \leq \sqrt{2}\|\mathbf{\Delta}\|_F.$$

Proof. Let $\mathbf{\Delta_L}(t)$ denote the lower triangular matrix $\mathbf{L}_{\mathbf{I}+t\mathbf{\Delta}} - \mathbf{I}$. Note that

$$\mathbf{\Delta_L}(t)_H + \mathbf{\Delta_L}(t)\mathbf{\Delta_L}(t)^T = t\mathbf{\Delta}.$$

For $t \in [0,1]$, we have

$$\begin{aligned} t\|\mathbf{\Delta}\|_F &= \|\mathbf{\Delta_L}(t)_H + \mathbf{\Delta_L}(t)\mathbf{\Delta_L}(t)^T\|_F \\ &\geq \|\mathbf{\Delta_L}(t)_H\|_F - \|\mathbf{\Delta_L}(t)\mathbf{\Delta_L}(t)^T\|_F \\ &\geq \sqrt{2}\|\mathbf{\Delta_L}(t)\|_F - \|\mathbf{\Delta_L}(t)\|_F^2. \end{aligned} \tag{3.6}$$

Solving this quadratic in $\|\mathbf{\Delta_L}(t)\|_F$ gives

$$\|\mathbf{\Delta_L}(t)\|_F \leq \frac{1}{\sqrt{2}} - \sqrt{\frac{1}{2} - t\|\mathbf{\Delta}\|_F} \quad \text{or} \quad \|\mathbf{\Delta_L}(t)\|_F \geq \frac{1}{\sqrt{2}} + \sqrt{\frac{1}{2} - t\|\mathbf{\Delta}\|_F}.$$

Since $\mathbf{\Delta_L}(t)$, whence $\|\mathbf{\Delta_L}(t)\|_F$, is continuous in t, it follows that

$$\|\mathbf{\Delta_L}(t)\|_F \leq \frac{1}{\sqrt{2}} - \sqrt{\frac{1}{2} - t\|\mathbf{\Delta}\|_F}$$

whenever $t\|\mathbf{\Delta}\|_F \le 1/2$. Under the hypothesis $\|\mathbf{\Delta}\|_F \le 1/2$, this indeed hold for $t = 1$, thus

$$\|\mathbf{L}_{\mathbf{I}+\mathbf{\Delta}} - \mathbf{I}\|_F \le \frac{1}{\sqrt{2}} - \sqrt{\frac{1}{2} - \|\mathbf{\Delta}\|_F} \le \frac{1}{\sqrt{2}}.$$

Finally, applying this upper bound in (3.6) with $t = 1$ gives

$$\|\mathbf{\Delta}\|_F \ge \sqrt{2}\|\mathbf{L}_{\mathbf{I}+\mathbf{\Delta}} - \mathbf{I}\|_F - \frac{1}{\sqrt{2}}\|\mathbf{L}_{\mathbf{I}+\mathbf{\Delta}} - \mathbf{I}\|_F = \frac{1}{\sqrt{2}}\|\mathbf{L}_{\mathbf{I}+\mathbf{\Delta}} - \mathbf{I}\|_F$$

as required. □

Lemma 8. *If $d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_{++}) \le \gamma$ for some $\gamma \in (0, 1/\sqrt{2})$, then the linear system (3.4), with the matrix $\mathbf{I}$ in (3.4c) replaced by $\sigma\mathbf{I}$ for some $\sigma \in [0, 1]$, has a unique solution $(\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}, \widetilde{\mathbf{\Delta}}_{\mathbf{S}}, \mathbf{\Delta}_{\mathbf{y}})$. Moreover for every $\alpha \in [0, \min\{1, (1-\sqrt{2}\gamma)/(2\chi)\}]$, it holds $\widetilde{\mathfrak{S}}_{l,\alpha} \in \mathbb{S}^l_{++}$ for all $l \in \mathcal{L}$, and*

$$\begin{aligned} &w_n^{-\frac{1}{2}}\left(\sum_{l\in\mathcal{L}} \pi_l \left\|\mathbf{L}^T_{\widetilde{\mathfrak{S}}_{l,\alpha}} \widetilde{\mathfrak{X}}_{l,\alpha} \mathbf{L}_{\widetilde{\mathfrak{S}}_{l,\alpha}} - \mu_\alpha \mathbf{I}\right\|_F^2\right)^{\frac{1}{2}} \\ &\le (1-\alpha)\gamma + \alpha^2\frac{\chi^2(7+5\gamma)}{(1-\sqrt{2}\gamma)^2} + 2\alpha^3\frac{\chi^3}{(1-\sqrt{2}\gamma)^3}, \end{aligned} \tag{3.7}$$

where $\widetilde{\mathfrak{X}}_{l,\alpha}$, $\widetilde{\mathfrak{S}}_{l,\alpha}$ and μ_α denote, respectively, the sums $[\mathbf{Z}]_l + \alpha\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}$, $\mathbf{I} + \alpha[\widetilde{\mathbf{\Delta}}_{\mathbf{S}}]_l$ and $(1 - \alpha + \alpha\sigma)$, and

$$\chi = \begin{cases} \sigma d_2(\mathbf{X}_+, \mathbf{S}_+; \sigma\mathbf{D}_{++}) & \text{if } \sigma > 0, \\ \left(w_n^{-1}\sum_{l\in\mathcal{L}} \pi_l \|[\mathbf{Z}]_l\|_F^2\right)^{\frac{1}{2}} & \text{if } \sigma = 0. \end{cases}$$

Proof. Since the system (3.4) is square, Lemma 6 shows that it has unique solution whenever $d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_{++}) < 1/\sqrt{2}$.

For $\mathbf{M}_l = \sigma\mathbf{I} - [\mathbf{Z}]_l$ ($l \in \mathcal{L}$), we have

$$\begin{aligned} \sum_{l\in\mathcal{L}} \pi_l\|\mathbf{M}_l\|_F^2 &= \sum_{l\in\mathcal{L}} \pi_l \sum_{i,j=1}^{l} (w_{i\vee j}^{-1}\mathbf{V}_{ij} - \sigma\mathbf{I}_{ij})^2 \\ &= \sum_{i,j=1}^{n} \sum_{l=i\vee j}^{n} \pi_l (w_{i\vee j}^{-1}\mathbf{V}_{ij} - \sigma\mathbf{I}_{ij})^2 \\ &= \sum_{i,j=1}^{n} w_{i\vee j}\left(w_{i\vee j}^{-1}\mathbf{V}_{ij} - \sigma\mathbf{I}_{ij}\right)^2 \le w_n\chi^2. \end{aligned}$$

It thus follows from Lemma 6 that

$$w_n^{-1}\max\left\{\sum_{l=1}^{n} \pi_l\|\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\|_F^2, \sum_{l=1}^{n} \pi_l\|[\widetilde{\mathbf{\Delta}}_{\mathbf{S}}]_l\|_F^2\right\} \le \frac{\chi^2}{(1-\sqrt{2}\gamma)^2}. \tag{3.8}$$

A useful consequence of this bound is

$$\|[\widetilde{\mathbf{\Delta}}_{\mathbf{S}}]_l\|_F^2 \le \|\widetilde{\mathbf{\Delta}}_{\mathbf{S}}\|_F^2 \le w_n^{-1}\sum_{l=1}^{n} \pi_l\|[\widetilde{\mathbf{\Delta}}_{\mathbf{S}}]_l\|_F^2 \le \frac{\chi^2}{(1-\sqrt{2}\gamma)^2}, \tag{3.9}$$

which implies that $\widetilde{\mathfrak{S}}_{n,\alpha} = \mathbf{I} + \alpha\widetilde{\mathbf{\Delta}}_{\mathbf{S}}$, whence $\widetilde{\mathfrak{S}}_{l,\alpha} = \mathbf{I} + \alpha[\widetilde{\mathbf{\Delta}}_{\mathbf{S}}]_l$, is positive definite whenever $|\alpha| < (1-\sqrt{2}\gamma)/\chi$.

For each $l \in \mathcal{L}$, let $\mathfrak{L}_{l,\alpha}$ denote the lower triangular matrix $\mathbf{L}_{\widetilde{\mathfrak{S}}_{l,\alpha}} - \mathbf{I} = \mathbf{L}_{\mathbf{I}+\alpha[\widetilde{\mathbf{\Delta}}_{\mathbf{S}}]_l} - \mathbf{I}$. In terms of $\mathfrak{L}_{l,\alpha}$, the difference $\left(\mathbf{L}^T_{\widetilde{\mathfrak{S}}_{l,\alpha}} \widetilde{\mathfrak{X}}_{l,\alpha} \mathbf{L}_{\widetilde{\mathfrak{S}}_{l,\alpha}} - \mu_\alpha \mathbf{I}\right)$ is

$$\begin{aligned}&[\mathbf{Z}]_l - (1-\alpha)\mathbf{I} - \alpha\sigma\mathbf{I} + \alpha\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)} + ([\mathbf{Z}]_l\mathfrak{L}_{l,\alpha})_H \\ &\quad + \alpha(\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\mathfrak{L}_{l,\alpha})_H + \mathfrak{L}^T_{l,\alpha}[\mathbf{Z}]_l\mathfrak{L}_{l,\alpha} + \alpha\mathfrak{L}^T_{l,\alpha}\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\mathfrak{L}_{l,\alpha}.\end{aligned}$$

Using (3.4c) with $\sigma\mathbf{I}$ replacing $\mathbf{I}$, and $\alpha[\widetilde{\mathbf{\Delta}}_{\mathbf{S}}]_l = (\mathfrak{L}_{l,\alpha})_H + \mathfrak{L}_{l,\alpha}\mathfrak{L}^T_{l,\alpha}$, this reduces to

$$\begin{aligned}&(1-\alpha)([\mathbf{Z}]_l - \mathbf{I}) + \Big([\mathbf{Z}]_l(\mathfrak{L}_{l,\alpha} - \alpha\langle\!\langle[\widetilde{\mathbf{\Delta}}_{\mathbf{S}}]_l\rangle\!\rangle)\Big)_H \\ &+ \alpha(\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\mathfrak{L}_{l,\alpha})_H + \mathfrak{L}^T_{l,\alpha}\mathfrak{L}_{l,\alpha} + \mathfrak{L}^T_{l,\alpha}([\mathbf{Z}]_l - \mathbf{I})\mathfrak{L}_{l,\alpha} + \alpha\mathfrak{L}^T_{l,\alpha}\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\mathfrak{L}_{l,\alpha}, \\ &= (1-\alpha)([\mathbf{Z}]_l - \mathbf{I}) - \mathfrak{L}_{l,\alpha}\mathfrak{L}^T_{l,\alpha} - \big(([\mathbf{Z}]_l - \mathbf{I})\langle\!\langle\mathfrak{L}_{l,\alpha}\mathfrak{L}^T_{l,\alpha}\rangle\!\rangle\big)_H \\ &\quad + \alpha(\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\mathfrak{L}_{l,\alpha})_H + \mathfrak{L}^T_{l,\alpha}\mathfrak{L}_{l,\alpha} + \mathfrak{L}^T_{l,\alpha}([\mathbf{Z}]_l - \mathbf{I})\mathfrak{L}_{l,\alpha} + \alpha\mathfrak{L}^T_{l,\alpha}\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\mathfrak{L}_{l,\alpha}.\end{aligned}$$

Using Lemma 7, we bound for all $\alpha \in [0, (1-\sqrt{2}\gamma)/(2\chi)]$,

$$\|\mathfrak{L}_{l,\alpha}\mathfrak{L}^T_{l,\alpha}\|_F = \|\mathfrak{L}^T_{l,\alpha}\mathfrak{L}_{l,\alpha}\|_F \le \|\mathfrak{L}_{l,\alpha}\|_F^2 \le 2\alpha^2\|[\widetilde{\mathbf{\Delta}}_{\mathbf{S}}]_l\|_F^2,$$

$$\begin{aligned}\|\big(([\mathbf{Z}]_l - \mathbf{I})\langle\!\langle\mathfrak{L}_{l,\alpha}\mathfrak{L}^T_{l,\alpha}\rangle\!\rangle\big)_H\|_F &\le 2\|[\mathbf{Z}]_l - \mathbf{I}\|_F\|\langle\!\langle\mathfrak{L}_{l,\alpha}\mathfrak{L}^T_{l,\alpha}\rangle\!\rangle\|_F \\ &\le \sqrt{2}\|[\mathbf{Z}]_l - \mathbf{I}\|_F\|\mathfrak{L}_{l,\alpha}\mathfrak{L}^T_{l,\alpha}\|_F \\ &\le 2\sqrt{2}\alpha^2\|[\mathbf{Z}]_l - \mathbf{I}\|_F\|[\widetilde{\mathbf{\Delta}}_{\mathbf{S}}]_l\|_F^2,\end{aligned}$$

$$\|\alpha(\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\mathfrak{L}_{l,\alpha})_H\| \le 2\alpha\|\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\|_F\|\mathfrak{L}_{l,\alpha}\|_F \le 2\sqrt{2}\alpha^2\|\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\|_F\|[\widetilde{\mathbf{\Delta}}_{\mathbf{S}}]_l\|_F,$$

$$\|\mathfrak{L}^T_{l,\alpha}([\mathbf{Z}]_l - \mathbf{I})\mathfrak{L}_{l,\alpha}\|_F \le \|[\mathbf{Z}]_l - \mathbf{I}\|_F\|\mathfrak{L}_{l,\alpha}\|_F^2 \le 2\alpha^2\|[\mathbf{Z}]_l - \mathbf{I}\|_F\|[\widetilde{\mathbf{\Delta}}_{\mathbf{S}}]_l\|_F^2,$$

and

$$\|\alpha\mathfrak{L}^T_{l,\alpha}\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\mathfrak{L}_{l,\alpha}\|_F \le \alpha\|\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\|_F\|\mathfrak{L}_{l,\alpha}\|_F^2 \le 2\alpha^3\|\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\|_F\|[\widetilde{\mathbf{\Delta}}_{\mathbf{S}}]_l\|_F^2.$$

Thus if $d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_{++}) \le \gamma$ for some $\gamma \in (0, 1/\sqrt{2})$, then for all $\alpha \in [0, \min\{1, (1-\sqrt{2}\gamma)/(2\chi)\}]$, $\left(\sum_{l\in\mathcal{L}} \pi_l\|\mathbf{L}^T_{\widetilde{\mathfrak{S}}_{l,\alpha}}\widetilde{\mathfrak{X}}_{l,\alpha}\mathbf{L}_{\widetilde{\mathfrak{S}}_{l,\alpha}} - \mathbf{I}\|_F^2\right)^{\frac{1}{2}}$ is bounded above by the sum of

$$(1-\alpha)\left(\sum_{l\in\mathcal{L}} \pi_l\|[\mathbf{Z}]_l - \mathbf{I}\|_F^2\right)^{\frac{1}{2}} = (1-\alpha)\sqrt{w_n}d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_{++}) \le \sqrt{w_n}(1-\alpha)\gamma,$$

$$4\alpha^2\left(\sum_{l\in\mathcal{L}} \pi_l\|[\widetilde{\mathbf{\Delta}}_{\mathbf{S}}]_l\|_F^4\right)^{\frac{1}{2}} \le 4\alpha^2\|\widetilde{\mathbf{\Delta}}_{\mathbf{S}}\|_F\left(\sum_{l\in\mathcal{L}} \pi_l\|[\widetilde{\mathbf{\Delta}}_{\mathbf{S}}]_l\|_F^2\right)^{\frac{1}{2}} \le \frac{4\sqrt{w_n}\alpha^2\chi^2}{(1-\sqrt{2}\gamma)^2},$$

$$2(1+\sqrt{2})\alpha^2\left(\sum_{l\in\mathcal{L}}\pi_l\|[\mathbf{Z}]_l-\mathbf{I}\|_F^2\|[\widetilde{\mathbf{\Delta}}_\mathbf{S}]_l\|_F^4\right)^{\frac{1}{2}} \leq 2(1+\sqrt{2})\alpha^2\|\widetilde{\mathbf{\Delta}}_\mathbf{S}\|_F^2\left(\sum_{l\in\mathcal{L}}\pi_l\|[\mathbf{Z}]_l-\mathbf{I}\|_F^2\right)^{\frac{1}{2}}$$
$$\leq \frac{2(1+\sqrt{2})\sqrt{w_n}\alpha^2\chi^2\gamma}{(1-\sqrt{2}\gamma)^2} \leq \frac{5\sqrt{w_n}\alpha^2\chi^2\gamma}{(1-\sqrt{2}\gamma)^2},$$

$$2\sqrt{2}\alpha^2\left(\sum_{l\in\mathcal{L}}\pi_l\|\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\|_F^2\|[\widetilde{\mathbf{\Delta}}_\mathbf{S}]_l\|_F^2\right)^{\frac{1}{2}} \leq 2\sqrt{2}\alpha^2\|\widetilde{\mathbf{\Delta}}_\mathbf{S}\|_F\left(\sum_{l\in\mathcal{L}}\pi_l\|\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\|_F^2\right)^{\frac{1}{2}}$$
$$\leq \frac{2\sqrt{2}\sqrt{w_n}\alpha^2\chi^2}{(1-\sqrt{2}\gamma)^2} \leq \frac{3\sqrt{w_n}\alpha^2\chi^2}{(1-\sqrt{2}\gamma)^2},$$

and

$$2\alpha^3\left(\sum_{l\in\mathcal{L}}\pi_l\|\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\|_F^2\|[\widetilde{\mathbf{\Delta}}_\mathbf{S}]_l\|_F^4\right)^{\frac{1}{2}} \leq 2\alpha^3\|\widetilde{\mathbf{\Delta}}_\mathbf{S}\|_F^2\left(\sum_{l\in\mathcal{L}}\pi_l\|\widetilde{\mathbf{\Delta}}_{\mathbf{X}(l)}\|_F^2\right)^{\frac{1}{2}}$$
$$\leq \frac{2\sqrt{w_n}\alpha^3\chi^3}{(1-\sqrt{2}\gamma)^3},$$

where we have used (3.8) and (3.9) to bound the last four terms. □

We are ready to give the main theorem of this section.

Theorem 3. *If $\beta,\delta\in(0,1)$ satisfies*

$$\frac{\gamma^2(7+5\gamma)}{(1-\sqrt{2}\gamma)^2}+2\frac{\gamma^3}{(1-\sqrt{2}\gamma)^3}<\beta, \tag{3.10}$$

where $\gamma=(\beta+\delta)/(1-\delta)$, then in each iteration of Algorithm 2, the search directions are well defined. Moreover, in each iteration, the iterates are primal-dual strictly feasible solutions satisfying $d_2(\mathbf{X}_+,\mathbf{S}_+;\mathbf{D}_+)\leq\beta$.

Proof. We shall prove the theorem by induction on the iterations. Suppose that at the beginning of an iteration, the iterates $(\mathbf{X}_+,\mathbf{S}_+)$ are strictly feasible and $d_2(\mathbf{X}_+,\mathbf{S}_+;\mathbf{D}_+)$ is at most β. This is certainly true for the first iteration. By the choice of $\mathbf{D}_{++}$ and Lemma 5, we have

$$d_2(\mathbf{X}_+,\mathbf{S}_+;\mathbf{D}_{++})\leq\gamma,$$

where $\gamma=(\beta+\delta)/(1-\delta)$. If (3.10) holds with $\beta<1$, then it is straightforward to check that $\gamma<1/\sqrt{2}$ and $(1-\sqrt{2}\gamma)/(2\gamma)>1$. Thus we may apply Lemma 8 with $\sigma=1$ to deduce that the search directions $\mathbf{\Delta}_\mathbf{X}$ and $\mathbf{\Delta}_\mathbf{S}$ are well defined, and that for all $\alpha\in[0,1]$,

$$w_n^{-\frac{1}{2}}\left(\sum_{l\in\mathcal{L}}\pi_l\left\|\mathbf{L}_{\widetilde{\mathfrak{S}}_{l,\alpha}}^T\widetilde{\mathfrak{X}}_{l,\alpha}\mathbf{L}_{\widetilde{\mathfrak{S}}_{l,\alpha}}-\mathbf{I}\right\|_2^2\right)^{\frac{1}{2}}$$
$$\leq(1-\alpha)\gamma+\alpha^2\frac{\gamma^2(7+5\gamma)}{(1-\sqrt{2}\gamma)^2}+2\alpha^3\frac{\gamma^3}{(1-\sqrt{2}\gamma)^3},$$

and $\mathbf{S}_+ + \alpha\mathbf{\Delta}_\mathbf{S} = \mathbf{L}_{\mathbf{S}_+}\widetilde{\mathfrak{S}}_{n,\alpha}\mathbf{L}_{\mathbf{S}_+}^T \in \mathbb{S}^n_{++}$, where $\widetilde{\mathfrak{X}}_{l,\alpha}$ and $\widetilde{\mathfrak{S}}_{l,\alpha}$ denote, respectively, the sums $\mathfrak{X}_l + \alpha\mathbf{\Delta}_{\mathbf{X}(l)}$ and $\mathfrak{S}_l + \alpha[\mathbf{\Delta}_\mathbf{S}]_l$. By Lemma 1, we have

$$
\begin{aligned}
d_2(\mathbf{X}_+ + \alpha\mathbf{\Delta}_\mathbf{X}, \mathbf{S}_+ + \alpha\mathbf{\Delta}_\mathbf{S}; \mathbf{D}_{++}) &\leq w_n^{-\frac{1}{2}} \left(\sum_{l\in\mathcal{L}} \pi_l \left\| \mathbf{L}_{\widetilde{\mathfrak{S}}_{l,\alpha}}^T \widetilde{\mathfrak{X}}_{l,\alpha} \mathbf{L}_{\widetilde{\mathfrak{S}}_{l,\alpha}} - \mathbf{I} \right\|_2^2 \right)^{\frac{1}{2}} \\
&\leq (1-\alpha)\gamma + \alpha^2 \frac{\gamma^2(7+5\gamma)}{(1-\sqrt{2}\gamma)^2} + 2\alpha^3 \frac{\gamma^3}{(1-\sqrt{2}\gamma)^3}.
\end{aligned}
$$

Under the hypothesis (3.10), the above upper bound is at most $(1-\alpha)\gamma + \alpha\beta < 1$ for all $\alpha \in [0,1]$. We then conclude from Theorem 2 that the next pair of iterates $(\mathbf{X}_{++}, \mathbf{S}_{++}) = (\mathbf{X}_+ + \mathbf{\Delta}_\mathbf{X}, \mathbf{S}_+ + \mathbf{\Delta}_\mathbf{S})$ are positive definite, whence strictly feasible as they clearly satisfy the linear equations in their respective SDP problems. Finally, the induction is completed by observing that the upper bound $((1-\alpha)\gamma + \alpha\beta)$ is precisely β when $\alpha = 1$. □

3.2. **Weighted path-following algorithms.** In this section, we describe three weighted path-following algorithms using the above target-following framework based on the Cholesky search directions. The first is a short-step algorithm that is actually a special case of the above target-following framework. The next is a weighted path-following version of the Mizuno-Todd-Ye (MTY) predictor-corrector algorithm. Finally, we present a weighted path-following algorithm that attempts to take a large step towards optimality in each iteration. The analyses of the first two algorithms demonstrate the same worst-case iteration bound of $O(\sqrt{n\rho}\log(\varepsilon^{-1}))$, while the analysis of the last algorithm gives a bound of $O(n\rho\log(\varepsilon^{-1}))$ to obtain a pair of primal-dual feasible solutions $(\mathbf{X}_\text{out}, \mathbf{S}_\text{out})$ satisfying $\mathbf{X}_\text{out} \bullet \mathbf{S}_\text{out} \leq \varepsilon \mathbf{X}_\text{in} \bullet \mathbf{S}_\text{in}$, where ρ denotes the ratio

$$
\frac{\mathbf{X}_\text{in} \bullet \mathbf{S}_\text{in}}{n\boldsymbol{\lambda}(\mathbf{X}_\text{in}\mathbf{S}_\text{in})_n}. \tag{3.11}
$$

We begin with the following generic weighted path-following algorithm:

Algorithm 3. (Cholesky weighted path-following algorithm)
Given a pair of primal-dual strictly feasible solutions $(\mathbf{X}_{in}, \mathbf{S}_{in})$ *with* $\boldsymbol{\mathcal{T}}(\mathbf{X}_{in}, \mathbf{S}_{in}) \in \mathbb{D}^n_{\downarrow,++}$, *and the required accuracy* $\varepsilon > 0$.

(1) *Find a target* $\mathbf{D}_+ \in \mathbb{D}^n_{\downarrow,++}$ *satisfying* $d_2(\mathbf{X}_{in}, \mathbf{S}_{in}; \mathbf{D}_+) \leq \beta$ *for some* $\beta \in (0,1)$. *Set* $(\mathbf{X}_+, \mathbf{S}_+) = (\mathbf{X}_{in}, \mathbf{S}_{in})$.
(2) *While* $\mathbf{X}_+ \bullet \mathbf{S}_+ > \varepsilon(\mathbf{X}_{in} \bullet \mathbf{S}_{in})$,
  (a) *Pick* $\sigma \in [0,1]$.
  (b) *Solve* (3.2) *with* $\mathbf{D}_{++}$ *replaced by* $\sigma\mathbf{D}_+$. *For each* $\alpha \in [0,1]$, *let* $(\mathbf{X}_\alpha, \mathbf{S}_\alpha) = (\mathbf{X}_+ + \alpha\mathbf{\Delta}_\mathbf{X}, \mathbf{S}_+ + \alpha\mathbf{\Delta}_\mathbf{S})$, *and let* $\mu_\alpha = 1 - \alpha + \alpha\sigma$. *Pick* $\widetilde{\beta} \in (0,1)$. *Pick* $\widehat{\alpha} \in [0,1]$ *such that* $\mathbf{S}_{\widehat{\alpha}} \in \mathbb{S}^n_{++}$ *and*

$$
d_2(\mathbf{X}_{\widehat{\alpha}}, \mathbf{S}_{\widehat{\alpha}}; \mu_{\widehat{\alpha}}\mathbf{D}_+) \leq \widetilde{\beta}.
$$

  (c) *Update* $(\mathbf{X}_+, \mathbf{S}_+) \leftarrow (\mathbf{X}_{\widehat{\alpha}}, \mathbf{S}_{\widehat{\alpha}})$ *and* $\mathbf{D}_+ \leftarrow \mu_{\widehat{\alpha}}\mathbf{D}_+$.
(3) *Output* $(\mathbf{X}_{out}, \mathbf{S}_{out}) = (\mathbf{X}_+, \mathbf{S}_+)$.

3.2.1. *Short-step algorithm.* Let ρ denote the ratio (3.11). The short-step algorithm uses the choices $\sigma = 1 - \delta(n\rho)^{-\frac{1}{2}}$ for a fixed constant $\delta \in (0, 1)$, and $\widetilde{\beta} = \beta$ throughout all iterations.

Theorem 4. *If $\beta, \delta \in (0, 1)$ satisfies the hypothesis of Theorem 3, then in each iteration of Algorithm 3, with $\sigma = 1-\delta(n\rho)^{-\frac{1}{2}}$, the search directions are well defined and we may use $\widehat{\alpha} = 1$. Moreover, with this choice of $\widehat{\alpha}$, the algorithm terminates after at most $O(\sqrt{n\rho}\log(\varepsilon^{-1}))$ iterations.*

Proof. The proof of Theorem 3 shows that we may use $\widehat{\alpha} = 1$ in each iteration. Therefore Algorithm 3, with $\sigma = 1 - \delta(n\rho)^{-\frac{1}{2}}$ in each iteration, is precisely Algorithm 2 with $\mathbf{D}_{++} = \sigma\mathbf{D}_+$. Consequently the first part of the theorem holds. Moreover, the duality gap of the iterates decreases by a factor of $1 - \delta(n\rho)^{-\frac{1}{2}}$ in each iteration, whence the iteration bound holds. $\square$

3.2.2. *Predictor-corrector algorithm.* The MTY predictor-corrector algorithm alternates between $(\sigma, \widetilde{\beta}) = (0, 2\beta)$ and $(\sigma, \widetilde{\beta}) = (1, \beta)$. The iterations in the former case are called the *predictor steps*, and those in the latter the *corrected steps*. As before, ρ denotes the ratio (3.11).

Theorem 5. *If $\beta \in (0, 1/(2\sqrt{2}))$ satisfies*

$$\frac{4\beta^2(7 + 10\beta)}{(1 - 2\sqrt{2}\beta)^2} + 2\frac{8\beta^3}{(1 - 2\sqrt{2}\beta)^3} < \beta, \tag{3.12}$$

then in each iteration of Algorithm 3, with $(\sigma, \widetilde{\beta})$ alternating between $(0, 2\beta)$ and $(1, \beta)$, the search directions are well defined and we may take $\widehat{\alpha}$ to be the positive real root of

$$\alpha \mapsto \alpha^2\frac{(\beta + \sqrt{n\rho})^2(7 + 5\beta)}{(1 - \sqrt{2}\beta)^2} + 2\alpha^3\frac{(\beta + \sqrt{n\rho})^3}{(1 - \sqrt{2}\beta)^3} - (1 - \alpha)\beta \tag{3.13}$$

in the predictor steps, and $\widehat{\alpha} = 1$ in the corrector steps. Moreover, with these choices of $\widehat{\alpha}$, the algorithm terminates after at most $O(\sqrt{n\rho}\log(\varepsilon^{-1}))$ iterations.

Proof. We shall prove by induction on the iterations that under the hypothesis of the theorem,

- all search directions are well defined and all iterates are strictly feasible,
- $d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_+) \le \beta$ in all predictor steps, and
- $d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_+) \le 2\beta$ in all corrector steps.

Suppose that at the beginning of a predictor step, we have strictly feasible $(\mathbf{X}_+, \mathbf{S}_+)$ satisfying $d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_+) \le \beta$. This is certainly true for the first predictor step, which happens to be the very first iteration. Under the hypothesis of the theorem, Lemma 8 with $\mathbf{D}_{++} = \mathbf{D}_+$ shows that the search directions $\mathbf{\Delta_X}$ and $\mathbf{\Delta_S}$ are well defined, and for each $\alpha \in [0, (1 - \sqrt{2}\beta)/(2\chi)]$,

$$\begin{aligned} &w_n^{-\frac{1}{2}}\left(\sum_{l\in\mathcal{L}}\pi_l\left\|\mathbf{L}_{\widetilde{\mathfrak{S}}_{l,\alpha}}^T\widetilde{\mathfrak{X}}_{l,\alpha}\mathbf{L}_{\widetilde{\mathfrak{S}}_{l,\alpha}} - (1 - \alpha)\mathbf{I}\right\|_2^2\right)^{\frac{1}{2}} \\ &\le (1 - \alpha)\beta + \alpha^2\frac{\chi^2(7 + 5\beta)}{(1 - \sqrt{2}\beta)^2} + 2\alpha^3\frac{\chi^3}{(1 - \sqrt{2}\beta)^3}, \end{aligned} \tag{3.14}$$

where $\widetilde{\mathfrak{X}}_{l,\alpha}$ and $\widetilde{\mathfrak{S}}_{l,\alpha}$ denote, respectively, the sums $\mathfrak{X}_l + \alpha \mathbf{\Delta}_{\mathbf{X}(l)}$ and $\mathfrak{S}_l + \alpha[\mathbf{\Delta}_{\mathbf{S}}]_l$, and $\chi = \left(w_n^{-1} \sum_{l \in \mathcal{L}} \pi_l \left\|\boldsymbol{\lambda}(\mathfrak{X}_l \mathfrak{S}_l)\right\|_F^2\right)^{\frac{1}{2}}$. Using the triangle inequality on the 2-norm of $\mathbb{R}^n$, we bound

$$\begin{aligned} \sqrt{w_n}\chi &\leq \left(\sum_{l \in \mathcal{L}} \pi_l \left\|\boldsymbol{\lambda}(\mathfrak{X}_l \mathfrak{S}_l) - \mathbf{I}\right\|_2^2\right)^{\frac{1}{2}} + \left(\sum_{l \in \mathcal{L}} \pi_l \left\|\mathbf{I}\right\|_2^2\right)^{\frac{1}{2}} \\ &\leq \sqrt{w_n}(\beta + \sqrt{n\rho}). \end{aligned} \tag{3.15}$$

By the definition (2.7) and Lemma 1, we have

$$\begin{aligned} & d_2(\mathbf{X}_\alpha, \mathbf{S}_\alpha; \mu_\alpha \mathbf{D}_+) \\ &= (1-\alpha)^{-\frac{1}{2}} w_n^{-\frac{1}{2}} \left\| \left((1-\alpha)^{-\frac{1}{2}} \mathbf{D}_+^{-\frac{1}{2}} \langle\!\langle \mathbf{L}_{\mathbf{S}_\alpha}^T \mathbf{X}_\alpha \mathbf{L}_{\mathbf{S}_\alpha} - (1-\alpha)\mathbf{D}_+ \rangle\!\rangle \right)_H \right\|_F \\ &\leq (1-\alpha)^{-1} w_n^{-\frac{1}{2}} \left(\sum_{l \in \mathcal{L}} \pi_l \left\| \mathbf{L}_{\widetilde{\mathfrak{S}}_{l,\alpha}}^T \widetilde{\mathfrak{X}}_{l,\alpha} \mathbf{L}_{\widetilde{\mathfrak{S}}_{l,\alpha}} - (1-\alpha)\mathbf{I} \right\|_2^2 \right)^{\frac{1}{2}}. \end{aligned}$$

This, together with (3.14) and (3.15), leads to the bound

$$\begin{aligned} & (1-\alpha)\left(d_2(\mathbf{X}_\alpha, \mathbf{S}_\alpha; \mu_\alpha \mathbf{D}_+) - 2\beta\right) \\ &\leq \alpha^2 \frac{(\beta + \sqrt{n\rho})^2 (7 + 5\beta)}{(1 - \sqrt{2}\beta)^2} + 2\alpha^3 \frac{(\beta + \sqrt{n\rho})^3}{(1 - \sqrt{2}\beta)^3} - (1-\alpha)\beta. \end{aligned}$$

This upper bound is not only cubic in α, it is in fact increasing in α whenever α is nonnegative. Thus if we take $\widehat{\alpha}$ to be the positive real root of (3.13), then we have

$$d_2(\mathbf{X}_\alpha, \mathbf{S}_\alpha; \mu_\alpha \mathbf{D}_+) \leq 2\beta$$

for all $\alpha \in [0, \widehat{\alpha}]$. Thus we conclude from Theorem 2 that the next iterates $(\mathbf{X}_{\widehat{\alpha}}, \mathbf{S}_{\widehat{\alpha}})$ are positive definite, whence strictly feasible as they clearly satisfy the linear equations in their respective SDP problems. Furthermore, $d_2(\mathbf{X}_{\widehat{\alpha}}, \mathbf{S}_{\widehat{\alpha}}; \mu_{\widehat{\alpha}} \mathbf{D}_+) \leq 2\beta$, whence in the next iteration we have $d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_+) \leq 2\beta$, which is a corrector step.

Now consider a corrector step. Suppose that at the beginning of the corrector step, we have with strictly feasible $(\mathbf{X}_+, \mathbf{S}_+)$ satisfying $d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_+) \leq 2\beta$. This is shown above to be true for the first corrector step. As before, we conclude from Lemma 8 that the search directions $\mathbf{\Delta}_{\mathbf{X}}$ and $\mathbf{\Delta}_{\mathbf{S}}$ are well defined. The proof of Theorem 3 shows that we may use $\widehat{\alpha} = 1$ in this iteration, and that

$$d_2(\mathbf{X}_1, \mathbf{S}_1; \mu_1 \mathbf{D}_+) \leq \frac{4\beta^2(7 + 10\beta)}{(1 - 2\sqrt{2}\beta)^2} + 2\frac{8\beta^3}{(1 - 2\sqrt{2}\beta)^3} \leq \beta$$

under the hypothesis (3.12), whence $d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_+) \leq \beta$ in the next iteration, which is a predictor step. This completes the induction.

Finally, since $\widehat{\alpha} = \Omega((n\rho)^{-\frac{1}{2}})$ for each predictor step, the duality gap decreases by a factor of $1 - \Omega((n\rho)^{-\frac{1}{2}})$ every two iterations. Thus the iteration bound holds. □

3.2.3. *Large-update algorithm.* Rather aiming for conservatively close targets, the large-update algorithm aims at weighted analytic centers with duality gap that is a constant fraction σ of the current duality gap. As Newton's method is not guaranteed to perform well when not in a neighborhood of the target, we shall use damped Newton steps instead. As before, ρ denotes the ratio (3.11).

Theorem 6. *If $\beta \in (0, 1/\sqrt{2})$ and $\sigma \in (0, 1)$, then in each iteration of Algorithm 3, with $\widetilde{\beta} = \beta$, the search directions are well defined and we may take $\widehat{\alpha}$ in each step to be the positive real root of*

$$\alpha \mapsto \alpha \frac{(\beta + \sqrt{(1-\sigma)n\rho})^2 (7 + 5\beta)}{(1 - \sqrt{2}\beta)^2} + 2\alpha^2 \frac{(\beta + \sqrt{(1-\sigma)n\rho})^3}{(1 - \sqrt{2}\beta)^3} - \sigma\beta.$$

Moreover, with this choice of $\widehat{\alpha}$, the algorithm terminates after at most $O(n\rho \log(\varepsilon^{-1}))$ iterations.

Proof. We shall prove by induction on the iterations, under the hypothesis of the theorem, that all search directions are well defined, all iterates are strictly feasible, and that $d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_+) \le \beta$. Suppose that at the beginning of an iteration, we have with strictly feasible $(\mathbf{X}_+, \mathbf{S}_+)$ satisfying $d_2(\mathbf{X}_+, \mathbf{S}_+; \mathbf{D}_+) \le \beta$. This is certainly true for the first iteration. Under the hypothesis of the theorem, Lemma 8 with $\mathbf{D}_{++} = \mathbf{D}_+$ shows that the search directions $\mathbf{\Delta_X}$ and $\mathbf{\Delta_S}$ are well defined, and for each $\alpha \in [0, (1 - \sqrt{2}\beta)/(2\chi)]$,

$$\begin{aligned} &w_n^{-\frac{1}{2}} \left(\sum_{l \in \mathcal{L}} \pi_l \left\| \mathbf{L}^T_{\widetilde{\mathfrak{S}}_{l,\alpha}} \widetilde{\mathfrak{X}}_{l,\alpha} \mathbf{L}_{\widetilde{\mathfrak{S}}_{l,\alpha}} - \mu_\alpha \mathbf{I} \right\|_2^2 \right)^{\frac{1}{2}} \\ &\le (1-\alpha)\beta + \alpha^2 \frac{\chi^2 (7 + 5\beta)}{(1 - \sqrt{2}\beta)^2} + 2\alpha^3 \frac{\chi^3}{(1 - \sqrt{2}\beta)^3}, \end{aligned} \tag{3.16}$$

where $\widetilde{\mathfrak{X}}_{l,\alpha}$ and $\widetilde{\mathfrak{S}}_{l,\alpha}$ denote, respectively, the sums $\mathfrak{X}_l + \alpha \mathbf{\Delta}_{\mathbf{X}(l)}$ and $\mathfrak{S}_l + \alpha[\mathbf{\Delta_S}]_l$, and $\chi = \sigma d_2(\mathbf{X}_+, \mathbf{S}_+; \sigma\mathbf{D}_+)$. Using the triangle inequality on the 2-norm of $\mathbb{R}^n$, we bound

$$\chi = \sigma d_2(\mathbf{X}_+, \mathbf{S}_+; \sigma\mathbf{D}_{++}) \le \beta + \sqrt{(1-\sigma)n\rho}. \tag{3.17}$$

By the definition (2.7) of d_2, Lemma 1 and the bounds (3.16) and (3.17), we have

$$\begin{aligned} &\mu_\alpha d_2(\mathbf{X}_\alpha, \mathbf{S}_\alpha; \mu_\alpha \mathbf{D}_+) \\ &\le w_n^{-\frac{1}{2}} \left(\sum_{l \in \mathcal{L}} \pi_l \left\| \mathbf{L}^T_{\widetilde{\mathfrak{S}}_{l,\alpha}} \widetilde{\mathfrak{X}}_{l,\alpha} \mathbf{L}_{\widetilde{\mathfrak{S}}_{l,\alpha}} - \mu_\alpha \mathbf{I} \right\|_2^2 \right)^{\frac{1}{2}} \\ &\le (1-\alpha)\beta + \alpha^2 \frac{(\beta + \sqrt{(1-\sigma)n\rho})^2 (7 + 5\beta)}{(1 - \sqrt{2}\beta)^2} + 2\alpha^3 \frac{(\beta + \sqrt{(1-\sigma)n\rho})^3}{(1 - \sqrt{2}\beta)^3}. \end{aligned}$$

Subsequently $d_2(\mathbf{X}_\alpha, \mathbf{S}_\alpha; \mu_\alpha \mathbf{D}_+) \le \beta$ whenever

$$\alpha^2 \frac{(\beta + \sqrt{(1-\sigma)n\rho})^2 (7 + 5\beta)}{(1 - \sqrt{2}\beta)^2} + 2\alpha^3 \frac{(\beta + \sqrt{(1-\sigma)n\rho})^3}{(1 - \sqrt{2}\beta)^3} \le \alpha\sigma\beta.$$

The left hand side of the above inequality is increasing for nonnegative α. Hence the inequality holds for all $\alpha \in [0, \widehat{\alpha}]$, where $\widehat{\mathbf{a}}lpha$ is the positive real root of

$$\alpha \mapsto \alpha \frac{(\beta + \sqrt{(1-\sigma)n\rho})^2(7+5\beta)}{(1-\sqrt{2}\beta)^2} + 2\alpha^2 \frac{(\beta + \sqrt{(1-\sigma)n\rho})^3}{(1-\sqrt{2}\beta)^3} - \sigma\beta.$$

Thus we conclude from Theorem 2 that the next iterates $(\mathbf{X}_{\widehat{\alpha}}, \mathbf{S}_{\widehat{\alpha}})$ are positive definite, whence strictly feasible as they clearly satisfy the linear equations in their respective SDP problems. Furthermore, $d_2(\mathbf{X}_{\widehat{\alpha}}, \mathbf{S}_{\widehat{\alpha}}; \mu_{\widehat{\alpha}}\mathbf{D}_+) \leq \beta$.

Finally, since $\widehat{\alpha} = \Omega((n\rho)^{-1})$ in each iteration, the duality gap decreases by a factor of $1 - \Omega((n\rho)^{-1})$ every iteration. Thus the iteration bound holds. □

4. Finding Analytic Centers

Consider the problem of finding an analytic center $\boldsymbol{\mathcal{T}}^{-1}(\widehat{\mu}\mathbf{I})$ for some given $\widehat{\mu} > 0$. Given a pair primal-dual strictly feasible solutions $(\widehat{\mathbf{X}}, \widehat{\mathbf{S}})$ with $\mathbf{L}_{\widehat{\mathbf{S}}}^T \widehat{\mathbf{X}} \mathbf{L}_{\widehat{\mathbf{S}}} \in \mathbb{D}^n_{\downarrow,++}$, we shall construct a finite sequence of targets $\{\mathbf{D}_k\}_{k=0}^N$ such that

$$d_2(\widehat{\mathbf{X}}, \widehat{\mathbf{S}}; \mathbf{D}_0) \leq \beta,$$
$$(\mathbf{D}_{k-1})_{nn}^{-\frac{1}{2}} \left\| \mathbf{D}_k \mathbf{D}_{k-1}^{-\frac{1}{2}} - \mathbf{D}_{k-1}^{\frac{1}{2}} \right\|_F \leq \delta \quad (1 \leq k \leq N) \tag{4.1}$$

and $\mathbf{D}_N = \mu\mathbf{I}$, with β and δ satisfying the hypothesis of Theorem 3, thus allowing us to apply Algorithm 2 to approximate $\boldsymbol{\mathcal{T}}^{-1}(\widehat{\mu}\mathbf{I})$.

Of course, if $\mathbf{L}_{\widehat{\mathbf{S}}}^T \widehat{\mathbf{X}} \mathbf{L}_{\widehat{\mathbf{S}}} \in \mathbb{D}^n_{\downarrow,++}$ is a positive multiple of $\mathbf{I}$, then we need simply to follow the central path to approximate $\boldsymbol{\mathcal{T}}^{-1}(\widehat{\mu}\mathbf{I})$. Henceforth, we assume that $\mathbf{L}_{\widehat{\mathbf{S}}}^T \widehat{\mathbf{X}} \mathbf{L}_{\widehat{\mathbf{S}}} \in \mathbb{D}^n_{\downarrow,++}$ is not a positive multiple of $\mathbf{I}$.

Since the targets are diagonal matrices $\mathbf{D}_k \in \mathbb{D}^n_{\downarrow,++}$, we may restrict our attention to the diagonal entries $\mathbf{x}^k = \operatorname{diag}(\mathbf{D}_k)$ and work in $\mathbb{R}^n_{\downarrow,++}$ instead. Under this restriction, the condition (4.1) becomes

$$\sqrt{\frac{1}{\mathbf{x}_n^{k-1}} \sum_{i=1}^n \frac{\left(\mathbf{x}_i^k - \mathbf{x}_i^{k-1}\right)^2}{\mathbf{x}_i^{k-1}}} \leq \delta \quad (1 \leq k \leq N).$$

Such sequence $\{\mathbf{x}^k\}_{k=0}^N$ is called a δ-*sequence*; see [8]. We first give an upper bound on the length N of a δ-sequence.

Consider the local metric defined by the inner product

$$\langle \cdot, \cdot \rangle_{\mathbf{x}} : (\mathbf{u}, \mathbf{v}) \in \mathbb{R}^n \times \mathbb{R}^n \mapsto \frac{1}{\mathbf{x}_n} \sum_{i=1}^n \frac{\mathbf{u}_i \mathbf{v}_i}{\mathbf{x}_i}$$

at each $\mathbf{x} \in \mathbb{R}^n_{\downarrow,++}$. We denote by $\|\cdot\|_{\mathbf{x}}$ the norm induced by the above inner product. In terms of this local metric, a δ-sequence $\{\mathbf{x}^k\}_{k=0}^N$ is one that satisfies

$$\left\|\mathbf{x}^k - \mathbf{x}^{k-1}\right\|_{\mathbf{x}^{k-1}} \leq \delta \quad (1 \leq k \leq N).$$

The length of a piecewise smooth curve $\boldsymbol{\xi} : [0,1] \to \mathbb{R}^n_{\downarrow,++}$ is defined to be

$$\int_0^1 \left\| \frac{d\boldsymbol{\xi}(t)}{dt} \right\|_{\boldsymbol{\xi}(t)} dt = \int_0^1 \left(\frac{1}{\boldsymbol{\xi}_n} \sum_{i=1}^n \frac{\dot{\boldsymbol{\xi}}_i^2}{\boldsymbol{\xi}_i} \right)^{\frac{1}{2}} dt,$$

and denoted by $l(\boldsymbol{\xi})$.

Lemma 9 (c.f. Lemma 3.1 of [15])**.** *Suppose $\boldsymbol{\xi} : [0,1] \to \mathbb{R}^n_{\downarrow,++}$ is a piecewise smooth curve. If $\|\boldsymbol{\xi}(1) - \boldsymbol{\xi}(0)\|_{\boldsymbol{\xi}(0)} < 1$, then*

$$l(\boldsymbol{\xi}) \geq r - \frac{1}{2}r^2,$$

where r denotes $\|\boldsymbol{\xi}(1) - \boldsymbol{\xi}(0)\|_{\boldsymbol{\xi}(0)}$.

Proof. Let $\mathbf{x}$ denote the vector $\boldsymbol{\xi}(0)$. Let $\eta : [0,1] \to \mathbb{R}$ denote the map

$$t \mapsto \|\boldsymbol{\xi}(t) - \mathbf{x}\|_{\mathbf{x}}.$$

Using Cauchy-Schwarz inequality, we have, for any $t \in (0,1)$,

$$\frac{d}{dt}\eta(t) = \frac{1}{\eta(t)}(\boldsymbol{\xi}(t) - \mathbf{x})^T \frac{d}{dt}\boldsymbol{\xi}(t) \leq \left\| \frac{d}{dt}\boldsymbol{\xi}(t) \right\|_{\mathbf{x}}.$$

Moreover, if $t \in (0,1)$ is such that $\eta(t) < 1$, then we deduce from $\|\boldsymbol{\xi}(t) - \mathbf{x}\|_{\mathbf{x}} = \eta(t)$ that

$$\boldsymbol{\xi}(t)_i \leq \frac{\mathbf{x}(t)_i}{1 - \eta(t)}$$

for each $i \in \{1, \ldots, n\}$. Subsequently we may bound

$$(1 - \eta(t)) \left\| \frac{d}{dt}\boldsymbol{\xi}(t) \right\|_{\mathbf{x}} \leq \left\| \frac{d}{dt}\boldsymbol{\xi}(t) \right\|_{\boldsymbol{\xi}(t)}.$$

Thus, if we let $\widehat{t}$ denote the least $t \in [0,1]$ satisfying $\eta(t) = r$, then

$$l(\boldsymbol{\xi}) \geq \int_0^{\widehat{t}} \left\| \frac{d}{dt}\boldsymbol{\xi} \right\|_{\boldsymbol{\xi}} dt \geq \int_0^{\widehat{t}} (1 - \eta) \left\| \frac{d}{dt}\boldsymbol{\xi} \right\|_{\mathbf{x}} dt \geq \int_0^{\widehat{t}} (1 - \eta) \frac{d}{dt}\eta dt = \left[\eta - \frac{1}{2}\eta^2 \right]_0^r$$

proves the lemma. □

Lemma 10 (c.f. Lemma 3.3 of [15])**.** *For every piecewise smooth curve $\boldsymbol{\xi} : [0,1] \to \mathbb{R}^n_{\downarrow,++}$ and every $\delta \in (0,1)$, there exists a δ-sequence $\{\mathbf{x}^k\}_{k=0}^N$ with $\mathbf{x}^0 = \boldsymbol{\xi}(0)$, $\mathbf{x}^1 = \boldsymbol{\xi}(1)$ and length*

$$N \leq \left\lceil \frac{l(\boldsymbol{\xi})}{\delta - \frac{1}{2}\delta^2} \right\rceil.$$

Proof. Consider the sequence $\{\boldsymbol{\xi}(t_k)\}_{k=0}^{\infty}$, where $t_0 = 0$, and t_k is defined recursively to be the least $t \in [t_{k-1}, 1]$ satisfying

$$\|\boldsymbol{\xi}(t) - \boldsymbol{\xi}(t_{k-1})\|_{\boldsymbol{\xi}(t_{k-1})} = \delta$$

whenever $\|\boldsymbol{\xi}(1) - \boldsymbol{\xi}(t_{k-1})\|_{\boldsymbol{\xi}(t_{k-1})} > \delta$, or $t_k = 1$ otherwise. By the preceding lemma, each segment $\{\boldsymbol{\xi}(t) : t \in [t_{k-1}, t_k]\}$ of the curve $\boldsymbol{\xi}$ has length greater than $(\delta - \frac{1}{2}\delta^2)$ whenever $\|\boldsymbol{\xi}(1) - \boldsymbol{\xi}(t_{k-1})\|_{\boldsymbol{\xi}(t_{k-1})} > \delta$. Subsequently, we have $\|\boldsymbol{\xi}(1) - \boldsymbol{\xi}(t_{k-1})\|_{\boldsymbol{\xi}(t_{k-1})} \leq \delta$ when

$$k \geq \frac{l(\boldsymbol{\xi})}{\delta - \frac{1}{2}\delta^2}.$$

Thus $\{\mathbf{x}^k := \boldsymbol{\xi}(t_k)\}_{k=0}^N$ with

$$N = \left\lceil \frac{l(\boldsymbol{\xi})}{\delta - \frac{1}{2}\delta^2} \right\rceil$$

is the required δ-sequence. □

4.1. **Approximation of analytic centers.** We now construct a piecewise linear curve $\boldsymbol{\xi}$ joining $\operatorname{diag}(\mathbf{L}_{\widehat{\mathbf{S}}}^T \widehat{\mathbf{X}} \mathbf{L}_{\widehat{\mathbf{S}}})$ and $\mu\mathbf{1}$, where $\mu = (\widehat{\mathbf{X}} \bullet \widehat{\mathbf{S}})/n$, and shall demonstrate a good upper bound on the length of $\boldsymbol{\xi}$. Each linear piece of the curve $\boldsymbol{\xi}$ raises all entries with the least value at the same rate and reduces the remainder at another rate, while keeping the sum of all entries constant throughout. Each linear piece ends when the entries with the least value coincide with some other entries.

Let $\widehat{\mathbf{x}}$ denote the vector $\operatorname{diag}(\mathbf{L}_{\widehat{\mathbf{S}}}^T \widehat{\mathbf{X}} \mathbf{L}_{\widehat{\mathbf{S}}})$. Let K denote the number distinct entries in $\widehat{\mathbf{x}}$. Since we assumed that $\mathbf{L}_{\widehat{\mathbf{S}}}^T \widehat{\mathbf{X}} \mathbf{L}_{\widehat{\mathbf{S}}}$ is not a multiple of the identity matrix, we necessarily have $K > 1$. Let $y_1 > \cdots > y_K$ denote the values of the distinct entries of $\widehat{\mathbf{x}}$. For each $p \in \{1, \ldots, K\}$, let J_p denote the index set $\{i : \widehat{\mathbf{x}}_i = y_p\}$, and let n_p denote the number of indices in J_p.

For each $p \in \{1, \ldots, K\}$, let $\widehat{\mathbf{x}}^p$ denote the vector in $\mathbb{R}^n_{\downarrow,++}$ satisfying

$$\widehat{\mathbf{x}}_i^p = \begin{cases} \alpha_p \widehat{\mathbf{x}}_i & \text{if } i \in \mathcal{J}_1 \cup \cdots \cup \mathcal{J}_{K-p+1}, \\ \alpha_p y_{K-p+1} & \text{if } i \in \mathcal{J}_{K-p+2} \cup \cdots \cup \mathcal{J}_K, \end{cases} \tag{4.2}$$

where $\alpha_p \in \mathbb{R}_{++}$ is such that

$$\sum_{i=1}^n \widehat{\mathbf{x}}_i^p = \sum_{i=1}^n \widehat{\mathbf{x}}_i. \tag{4.3}$$

Since $\widehat{\mathbf{x}}_i = y_{K-p+1}$ when $i \in J_{K-p+1}$, we may alternatively write

$$\widehat{\mathbf{x}}_i^p = \begin{cases} \alpha_p \widehat{\mathbf{x}}_i & \text{if } i \in \mathcal{J}_1 \cup \cdots \cup \mathcal{J}_{K-p}, \\ \alpha_p y_{K-p+1} & \text{if } i \in \mathcal{J}_{K-p+1} \cup \cdots \cup \mathcal{J}_K. \end{cases} \tag{4.4}$$

The curve $\boldsymbol{\xi}$ consists of $(K-1)$ pieces of linear segments $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{K-1}$, where the p-th segment $\boldsymbol{\xi}_p$ joins $\widehat{\mathbf{x}}^k$ and $\widehat{\mathbf{x}}^{k+1}$.

Lemma 11. *In the curve $\boldsymbol{\xi}$, each line segment $\boldsymbol{\xi}_p$ has length*

$$l(\boldsymbol{\xi}_p) \leq \sqrt{n} \log\left(\frac{4 y_{K-p} \sigma_{K-p+1}}{y_{K-p+1} \sigma_{K-p}}\right),$$

where σ_p *denotes* $\sum_{q=1}^{p} n_q y_q + y_p \sum_{q=p+1}^{K} n_q$. *Consequently,*

$$l(\boldsymbol{\xi}) \leq \sqrt{n} \log\left(\frac{4y_1\sigma_K}{y_K\sigma_1}\right) = \sqrt{n}\log\left(\frac{4\widehat{\mathbf{X}} \bullet \widehat{\mathbf{S}}}{n\boldsymbol{\lambda}(\widehat{\mathbf{X}}\widehat{\mathbf{S}})_n}\right).$$

Proof. We fix a $p \in \{1, \ldots, K-1\}$ and consider the p-th linear segment $\boldsymbol{\xi}_p$. Recall that $\boldsymbol{\xi}_p : [0,1] \to \mathbb{R}^n_{\downarrow,++}$ is defined by

$$\boldsymbol{\xi}_p(t) = (1-t)\widehat{\mathbf{x}}^p + t\widehat{\mathbf{x}}^{p+1}.$$

Thus its length is

$$\int_0^1 \left(\frac{1}{\widehat{\mathbf{x}}^p_n + t(\widehat{\mathbf{x}}^{p+1}_n - \widehat{\mathbf{x}}^p_n)} \sum_{i=1}^{n} \frac{(\widehat{\mathbf{x}}^{p+1}_i - \widehat{\mathbf{x}}^p_i)^2}{\widehat{\mathbf{x}}^p_i + t(\widehat{\mathbf{x}}^{p+1}_i - \widehat{\mathbf{x}}^p_i)}\right)^{\frac{1}{2}} dt.$$

Using (4.4) for $\widehat{\mathbf{x}}^p$ and (4.2) for $\widehat{\mathbf{x}}^{p+1}$, the integrand is

$$\left(\begin{array}{c}\frac{1}{\alpha_p y_{K-p+1} + t\beta_p} \sum_{i \in \mathcal{J}_1 \cup \cdots \cup \mathcal{J}_{K-p}} \frac{(\alpha_{p+1}\widehat{\mathbf{x}}_i - \alpha_p\widehat{\mathbf{x}}_i)^2}{\alpha_p\widehat{\mathbf{x}}_i + t(\alpha_{p+1}\widehat{\mathbf{x}}_i - \alpha_p\widehat{\mathbf{x}}_i)} \\ + \frac{1}{\alpha_p y_{K-p+1} + t\beta_p} \sum_{i \in \mathcal{J}_{K-p+1} \cup \cdots \cup \mathcal{J}_K} \frac{\beta_p^2}{\alpha_p y_{K-p+1} + t\beta_p}\end{array}\right)^{\frac{1}{2}},$$

where β_p denotes $(\alpha_{p+1}y_{K-p} - \alpha_p y_{K-p+1})$. This simplifies to

$$\left(\begin{array}{c}\frac{\gamma_p^2}{(\alpha_p y_{K-p+1} + t\beta_p)(\alpha_p - t\gamma_p)} \sum_{i \in \mathcal{J}_1 \cup \cdots \cup \mathcal{J}_{K-p}} \widehat{\mathbf{x}}_i \\ + \frac{\beta_p^2}{(\alpha_p y_{K-p+1} + t\beta_p)^2} \sum_{q=K-p+1}^{K} n_q\end{array}\right)^{\frac{1}{2}},$$

where γ_p denotes $(\alpha_p - \alpha_{p+1})$. The condition (4.3) for $\widehat{\mathbf{x}}^p$ and $\widehat{\mathbf{x}}^{p+1}$ implies that

$$\begin{aligned}\alpha_p \sum_{i \in \mathcal{J}_1 \cup \cdots \cup \mathcal{J}_{K-p}} \widehat{\mathbf{x}}_i + \alpha_p y_{K-p+1} \sum_{q=K-p+1}^{K} n_q &= \sum_{i=1}^{n} \widehat{\mathbf{x}}_i \\ = \alpha_{p+1} \sum_{i \in \mathcal{J}_1 \cup \cdots \cup \mathcal{J}_{K-p}} \widehat{\mathbf{x}}_i + \alpha_{p+1} y_{K-p} \sum_{q=K-p+1}^{K} n_q&, \end{aligned} \tag{4.5}$$

and subsequently

$$\gamma_p \sum_{i \in \mathcal{J}_1 \cup \cdots \cup \mathcal{J}_{K-p}} \widehat{\mathbf{x}}_i = \beta_p \sum_{q=K-p+1}^{K} n_q.$$

Thus we may further simplify the integrand to

$$\left(\sum_{q=K-p+1}^{K} n_q\right)^{\frac{1}{2}} \left(\frac{\gamma_p\beta_p}{(\alpha_p y_{K-p+1} + t\beta_p)(\alpha_p - t\gamma_p)} + \frac{\beta_p^2}{(\alpha_p y_{K-p+1} + t\beta_p)^2}\right)^{\frac{1}{2}}.$$

The length of $\boldsymbol{\xi}_p$ is then

$$\left(\sum_{q=K-p+1}^{K} n_q\right)^{\frac{1}{2}} \left[\log\left(\frac{\sqrt{\alpha_p y_{K-p+1}\gamma_p + \alpha_p\beta_p} - \sqrt{\beta_p(\alpha_p - t\gamma_p)}}{\sqrt{\alpha_p y_{K-p+1}\gamma_p + \alpha_p\beta_p} + \sqrt{\beta_p(\alpha_p - t\gamma_p)}}\right)\right]_0^1$$
$$= \left(\sum_{q=K-p+1}^{K} n_q\right)^{\frac{1}{2}} \log\left(\frac{\sqrt{\alpha_p y_{K-p+1}\gamma_p + \alpha_p\beta_p} - \sqrt{\beta_p(\alpha_p - \gamma_p)}}{\sqrt{\alpha_p y_{K-p+1}\gamma_p + \alpha_p\beta_p} + \sqrt{\beta_p(\alpha_p - \gamma_p)}}\right)$$
$$- \left(\sum_{q=K-p+1}^{K} n_q\right)^{\frac{1}{2}} \log\left(\frac{\sqrt{\alpha_p y_{K-p+1}\gamma_p + \alpha_p\beta_p} - \sqrt{\beta_p\alpha_p}}{\sqrt{\alpha_p y_{K-p+1}\gamma_p + \alpha_p\beta_p} + \sqrt{\beta_p\alpha_p}}\right).$$

An upper bound may be obtained using the following inequalities:

$$\frac{4}{u} \le \frac{1-\sqrt{1-u}}{1+\sqrt{1-u}} \le u \quad (0 < u \le 1).$$

The first inequality follows from $\sqrt{1-u} \le 1 - \frac{1}{2}u$, while the second from the convexity of the ratio as a function of u on $[0,1]$. These inequalities imply

$$\frac{\sqrt{\alpha_p y_{K-p+1}\gamma_p + \alpha_p\beta_p} - \sqrt{\beta_p(\alpha_p - \gamma_p)}}{\sqrt{\alpha_p y_{K-p+1}\gamma_p + \alpha_p\beta_p} + \sqrt{\beta_p(\alpha_p - \gamma_p)}} \le 1 - \frac{\beta_p(\alpha_p - \gamma_p)}{\alpha_p y_{K-p+1}\gamma_p + \alpha_p\beta_p}$$
$$= \frac{\alpha_p y_{K-p+1}\gamma_p + \beta_p\gamma_p}{\alpha_p y_{K-p+1}\gamma_p + \alpha_p\beta_p},$$

and

$$\frac{\sqrt{\alpha_p y_{K-p+1}\gamma_p + \alpha_p\beta_p} - \sqrt{\beta_p\alpha_p}}{\sqrt{\alpha_p y_{K-p+1}\gamma_p + \alpha_p\beta_p} + \sqrt{\beta_p\alpha_p}} \ge 4\left(1 - \frac{\beta_p\alpha_p}{\alpha_p y_{K-p+1}\gamma_p + \alpha_p\beta_p}\right)^{-1}$$
$$= 4\frac{\alpha_p y_{K-p+1}\gamma_p + \alpha_p\beta_p}{\alpha_p y_{K-p+1}\gamma_p},$$

and thus the length of $\boldsymbol{\xi}_p$ is bounded from above by

$$\sqrt{n}\log\left(4\frac{\alpha_p y_{K-p+1} + \beta_p}{\alpha_p y_{K-p+1}}\right) = \sqrt{n}\log\left(4\frac{\alpha_{p+1} y_{K-p}}{\alpha_p y_{K-p+1}}\right).$$

From (4.5) we deduce the ratio

$$\frac{\alpha_{p+1}}{\alpha_p} = \frac{\sum_{i\in\mathcal{J}_1\cup\cdots\cup\mathcal{J}_{K-p}} \widehat{\mathbf{x}}_i + y_{K-p+1}\sum_{q=K-p+1}^{K} n_q}{\sum_{i\in\mathcal{J}_1\cup\cdots\cup\mathcal{J}_{K-p}} \widehat{\mathbf{x}}_i + y_{K-p}\sum_{q=K-p+1}^{K} n_q}$$
$$= \frac{\sum_{q=1}^{K-p} n_q y_q + y_{K-p+1}\sum_{q=K-p+1}^{K} n_q}{\sum_{q=1}^{K-p} n_q y_q + y_{K-p}\sum_{q=K-p+1}^{K} n_q}.$$

The numerator equals $\sum_{q=1}^{K-p+1} n_q y_q + y_{K-p+1}\sum_{q=K-p+2}^{K} n_q$, hence the lemma is proved. □

Combining Lemmas 10 and 11 with a short-step path-following sequence of targets, we have the following theorem.

Theorem 7. *Suppose $\beta \in (0,1)$ is fixed. Given any pair of primal-dual strictly feasible solutions $(\widehat{\mathbf{X}}, \widehat{\mathbf{S}})$, any positive real number $\widehat{\mu}$, there is a sequence of at most*

$$O\left(\sqrt{n}\left(\log \frac{\widehat{\mathbf{X}} \bullet \widehat{\mathbf{S}}}{n\boldsymbol{\lambda}(\widehat{\mathbf{X}}\widehat{\mathbf{S}})_n} + \left|\log \frac{\widehat{\mathbf{X}} \bullet \widehat{\mathbf{S}}}{n\widehat{\mu}}\right|\right)\right)$$

targets such that Algorithm 2 find a pair of primal-dual feasible solutions $(\mathbf{X}, \mathbf{S})$ satisfying $\|\boldsymbol{\lambda}(\mathbf{XS}) - \widehat{\mu}\mathbf{1}\|_2 \le \beta$.

As an immediate corollary, we have an improved worst-case iteration bound on solving SDP problems using our target-following framework.

Corollary 1. *Given any pair of primal-dual strictly feasible solutions $(\widehat{\mathbf{X}}, \widehat{\mathbf{S}})$ and any $\varepsilon > 0$, there is a sequence of at most*

$$O\left(\sqrt{n}\left(\log \frac{\widehat{\mathbf{X}} \bullet \widehat{\mathbf{S}}}{n\boldsymbol{\lambda}(\widehat{\mathbf{X}}\widehat{\mathbf{S}})_n} + \left|\log \varepsilon^{-1}\right|\right)\right)$$

targets such that Algorithm 2 find a pair of primal-dual feasible solutions $(\mathbf{X}, \mathbf{S})$ satisfying $\mathbf{X} \bullet \mathbf{S} \le \varepsilon \widehat{\mathbf{X}} \bullet \widehat{\mathbf{S}}$.

5. Conclusion

A target-following framework for SDP is presented. This framework is an extension of similar frameworks for linear programming [6] and linear complementarity problems [8]. Within this framework, we designed three primal-dual weighted path-following algorithms for SDP, and proved that their iteration complexity parallels their counterparts in linear programming.

In addition, we showed that with a specific choice of search directions, which we called *Cholesky search directions*, the computational efforts in each iteration is comparable with a regular path-following algorithm.

Finally, we showed that the target-following framework can be used to efficiently approximate points on the central path, when given any pair of primal-dual strictly feasible solutions. This, followed by any path-following algorithm, provides an algorithm for SDP. If we use a path-following algorithm with the best known iteration complexity, then we obtain an improved worst-case iteration bound for solving SDPs with a given pair of primal-dual strictly feasible solutions. It should be noted that there are existing initialization schemes, such as the homogeneous self-dual embedding used in SeDuMi [16], that solve the SDP by solving a larger problem with an easily obtained pair of primal-dual solutions near (or on) the central path. However, such methods are not known to have similar iteration complexity for obtaining approximately optimal solutions in general.

In addition to solving SDPs, the target-following framework can also be used to solve max-det problems. A *max-det* problem is the minimization of the sum of a linear function and a log-determinant term subjected to linear matrix inequalities. The many applications of max-det problems are discussed in [23]. The primal-dual central path of a max-det problem is the path of Cholesky weighted centers $\{\boldsymbol{\mathcal{T}}^{-1}(\left[\begin{smallmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mu\mathbf{I} \end{smallmatrix}\right]) : \mu \in (0,1]\}$ of the SDP obtained by dropping the log-determinant term. If we start close to a pair $\boldsymbol{\mathcal{T}}^{-1}(\left[\begin{smallmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \bar{\mu}\mathbf{I} \end{smallmatrix}\right])$, and follow a $\delta-$sequence along this path of Cholesky weighted centers, then within $O(\sqrt{n}|\log\varepsilon^{-1}|)$ iterations of Algorithm 2, where n denotes the order of the matrix in the log-determinant term, we will obtain a pair of primal-dual feasible solutions with duality gap $\varepsilon\widehat{\mu}$. This iteration bound is comparable with the one obtained in [23] for a dual short-step path-following algorithm. Primal-dual path-following algorithms for the max-det problem were proposed by Toh [19]. Although no theoretical convergence analysis was given in [19], the numerical results in the paper shows that the algorithms are efficient and able to obtain highly accuracy.

Recently, a primal-dual long-step path-following algorithm was proposed by Tsuchiya and Xia [20] for a generalization of the max-det problem where more than one log-determinant term is allowed. The authors showed that the algorithm takes $O(n\log\varepsilon^{-1}+n)$ iterations to reduce the duality gap by a factor of ε, where n denotes the sum of the orders of the matrices in the log-determinant terms and the order of the matrix in the linear term. This assumes that the initial primal-dual pair is near the "extended central trajectory" defined in [20]. The "extended central trajectory" is actually the path of Cholesky weighted centers

$$\left\{ \boldsymbol{\mathcal{T}}^{-1}\left(\begin{bmatrix} (\mu\vee c_1)\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & (\mu\vee c_p)\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mu\mathbf{I} \end{bmatrix}\right) : \mu > 0 \right\}$$

of the corresponding SDP after dropping all log-determinant terms, where $c_1 \geq c_2 \geq \cdots \geq c_p > 0$ are the weights of the log-determinant terms. While long-step algorithms are usually more efficient than short-step algorithms in practice, the short-step algorithms usually have better iteration complexity in theory. Indeed, if Algorithm 2 is used to solve the weighted max-det problem by following a $\delta-$sequence along the above path of Cholesky weighted centers, then we can show that the algorithm takes $O(\sqrt{n}\log\varepsilon^{-1})$ iterations to reduce the duality gap by a factor of ε. As the proof of this iteration complexity uses precisely the technique in the last section, we leave this as an exercise for the readers.

References

[1] S. Burer and R. D. C. Monteiro, *A general framework for establishing polynomial convergence of long-step methods for semidefinite programming*, Optim. Methods Softw. **18** (2003), 1–38.

[2] C. B. Chua, *Analyticity of weighted central path and error bound for semidefinite programming*, CORR 2005-15, Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Canada, July 2005.

[3] ———, *A new notion of weighted centers for semidefinite programming*, SIAM J. Optim. **16** (2006), no. 4, 1092–1109.

[4] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Some simple inequalities satisfied by convex functions*, Messenger of Math. **58** (1929), 145–152.

[5] R. Hauser, *Square-root fields and the v-space approach to primal-dual interior-point methods for self-scaled conic programming*, Numerical Analysis Report DAMTP 1999/NA 14, Department of Applied Mathematics and Theoretical Physics, Cambridge, England, 1999.

[6] B. Jansen, C. Roos, T. Terlaky, and J.-P. Vial, *Primal-dual target-following algorithms for linear programming*, Annals of Oper. Res. **62** (1996), 197–231.

[7] M. Kojima, N. Megiddo, T. Noma, and A. Yoshise, *A unified approach to interior point algorithms for linear complementarity problems*, Lecture Notes Comput. Sci., vol. 538, Springer-Verlag, Berlin-Heidelberg-New York, 1991.

[8] S. Mizuno, *A new polynomial time method for a linear complementarity problem*, Math. Program. **56** (1992), 31–43.

[9] R. D. C. Monteiro, *Polynomial convergence of primal-dual algorithms for semidefinite programming based on the Monteiro and Zhang family of directions*, SIAM J. Optim. **8** (1998), no. 3, 797–812.

[10] R. D. C. Monteiro and J.-S. Pang, *On two interior-point mappings for nonlinear semidefinite complementarity problems*, Math. Oper. Res. **23** (1998), 39–60.

[11] R. D. C. Monteiro and P. Zanjácomo, *Implementation of primal-dual methods for semidefinite programming based on Monteiro and Tsuchiya Newton directions and their variants*, Optim. Methods Softw. **11/12** (1999), 91–140.

[12] ———, *General interior-point maps and existence of weighted paths for nonlinear semidefinite complementarity problems*, Math. Oper. Res. **25** (2000), no. 3, 382–399.

[13] R. D. C. Monteiro and Y. Zhang, *A unified analysis for a class of long-step primal-dual path-following interior-point algorithms for semidefinite programming*, Math. Program. **81** (1981), 281–299.

[14] A. Nemirovski and Yu. E. Nesterov, *Central path and Riemannian distances*, CORE DP 2003/51, Université catholique de Louvain, Center for Operations Research and Econometrics, Belgium, 2003.

[15] Yu. E. Nesterov and M. J. Todd, *On the Riemannian geometry defined by self-concordant barriers and interior-point methods*, Found. Comput. Math. **2** (2002), 333–361.

[16] J. F. Sturm, *Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones*, Optim. Methods Softw. **11-12** (1999), 625–653.

[17] J. F. Sturm and S. Zhang, *On weighted centers for semidefinite programming*, Eur. J. Oper. Res. **126** (2000), 391–407.

[18] M. J. Todd, *On adjusting parameters in homotopy methods for linear programming*, Approximation Theory and Optimization (M. Buhmann and A. Iserles, eds.), Cambridge University Press, Cambridge, UK, 1997, pp. 201–220.

[19] K.-C. Toh, *Primal-dual path-following algorithms for determinant maximization problems with linear matrix inequalities*, Comput. Optim. Appl. **14** (1999), 309–330.

[20] T. Tsuchiya and Y. Xia, *An extension of the standard polynomial-time primal-dual path-following algorithm to the weighted determinant maximization problem with semidefinite constraints*, Pac. J. Optim. **3** (2007), 165–182.

[21] L. Tunçel, *Primal-dual symmetry and scale invariance of interior-point algorithms for convex optimization*, Math. Oper. Res. **23** (1998), 708–718.

[22] ———, *Generalization of primal-dual interior-point methods to convex optimization problems in conic form*, Found. Comput. Math. **1** (2001), 229–254.

[23] L. Vandenberghe, S. Boyd and S-P WU, *Determinant Maximization With Linear Matrix Inequality Constraints*, SIAM J. Matrix Anal. Appl. **19** (1998), 499–533.

Division of Mathematical Sciences, School of Physical & Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore

E-mail address: `cbchua@ntu.edu.sg`