A Polynomial Time Constraint-Reduced Algorithm for Semidefinite Optimization Problems, with Convergence Proofs. *

Sungwoo Park[†]and Dianne P. O'Leary[‡]

February 18, 2015

Abstract

We present an infeasible *primal-dual* interior point method for semidefinite optimization problems, making use of constraint reduction. We show that the algorithm is globally convergent and has polynomial complexity, the first such complexity result for *primal-dual* constraint reduction algorithms for any class of problems. Our algorithm is a modification of one with no constraint reduction due to Potra and Sheng (1998) and can be applied whenever the data matrices are block diagonal. It thus solves as special cases any optimization problem that is a linear, convex quadratic, convex quadratically constrained, or second-order cone problem.

Keywords: semidefinite programming, semidefinite optimization, interior point methods, constraint reduction, primal-dual interior point method, primal dual infeasible, polynomial complexity, linear programming, linear optimization, quadratic programming, quadratic optimization, second-order cone optimization, second-order cone programming.
AMS Classification: 90C22, 65K05, 90C51

1 Introduction

In this work, we propose a new infeasible *primal-dual predictor-corrector* interior point method (IPM) with adaptive criteria for constraint reduction. The algorithm is a modification of one with no constraint reduction, due to Potra and Sheng [34]. Our algorithm can be applied when the data matrices are block

*This work was supported by the US Department of Energy under grants DESC0002218 and DESC0001862. We are very grateful to André Tits for careful reading of the manuscript, many suggestions, and insightful comments that helped shape the choice of active and inactive blocks, to Florian Potra for helpful discussions.
[†]KCG holdings, 545 Washington BLVD, Jersey City, NJ 07310, USA. swpark81@gmail.com
[‡]Computer Science Department and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA. oleary@cs.umd.edu

diagonal. We verify its validity by proving global convergence. We also prove its polynomial complexity for a given convergence tolerance and an initial residual.

IPMs tend to require fewer iterations than active set methods like the simplex algorithm but require many more computations per iteration. Thus constraint reduction, which can substantially reduce the work per iteration, is an important tool in IPMs.

Prior work on constraint reduction begins with linear programming (LP). For example, Dantzig and Ye [5], Tone [39], Kaliski and Ye [20], and Hertog et al. [13] have applied different constraint reduction schemes to LP. More recently, Tits, Absil, and Woessner [37] introduced new constraint-reduced LP versions of a *primal-dual affine-scaling* method (rPDAS) and of Mehrotra's *predictor-corrector* method (rMPC). In contrast to previous constraint reduction schemes, their method adaptively updates the working set without any backtracking. They proved global convergence and quadratic local convergence of rPDAS under a nondegeneracy assumption, but polynomial complexity was not proved. Later, Winternitz et al. [41] proved global convergence of an rMPC, relaxing the assumptions of [37].

Adaptive constraint reduction also has been applied to other optimization problems. Jung, O'Leary, and Tits [18] proposed constraint reduction for training support vector machines (SVM), and Williams [40] improved the efficiency of the SVM training by applying a preconditioner. Jung, O'Leary, and Tits [19] developed a constraint-reduced affine-scaling method for convex quadratic programming (QP), and verified its global convergence and local quadratic convergence.

For semidefinite programming (SDP), many studies focused on *primal-dual* IPM. Applying Newton's method to the central path equation results in nonsymmetric directions. To symmetrize, different search directions have been proposed. Helmberg et al. [12], Kojima et al. [23], and Monteiro [24] suggested the HKM direction. Alizadeh, Haeberly, and Overton [1] introduced the AHO direction, and Nesterov and Todd [27, 28] proposed the NT direction. Later, Monteiro and Zhang [24, 26, 42] noticed that these methods all had the form

$$\operatorname{symm}\left(\boldsymbol{P}\boldsymbol{X}\boldsymbol{Z}\boldsymbol{P}^{-1}\right) = \mu\boldsymbol{I},$$

where $\boldsymbol{P} = \boldsymbol{Z}^{1/2}$ for the HKM direction, $\boldsymbol{P} = \boldsymbol{I}$ for the AHO direction, and $\boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{Z}^{1/2}(\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2})^{-1/2}\boldsymbol{Z}^{1/2}$ for the NT direction. This has become known as the Monteiro-Zhang (MZ) family of directions. Alizadeh et al. [2] also investigated various directions in a unified framework.

Convergence of primal-dual IPMs for SDP has also been intensely studied. Monteiro [25] developed a *short-step path-following* algorithm using an MZ search direction in a *predictor-corrector* algorithm, establishing polynomial convergence. Potra and Sheng [34] proposed a *primal-dual infeasible predictor-corrector* algorithm using $\boldsymbol{P} = \boldsymbol{X}^{-1/2}$ among the MZ family search directions, and proved polynomial global convergence and local superlinear convergence. Later, Kojima et al. [21] relaxed one of the assumptions of [34] and suggested a modified algorithm with superlinear convergence, repeating corrector steps until the iterate converges tangentially to the central path. Potra et al. [33] proved superlinear convergence of the modified algorithm without the nondegeneracy assumption needed by [21]. Kojima et al. [22] proposed a *predictor-corrector* algorithm using the AHO direction, with quadratic convergence under a nondegeneracy assumption but without repeating corrector steps. Inspired by the good centering effect of the AHO direction, Potra and Sheng [32] and Ji et al. [17] proposed a superlinearly converging algorithm using the HKM direction in the predictor step and a MZ direction with a bounded scaling matrix.

In this study, we extend constraint reduction to a *primal-dual infeasible predictor-corrector* method of Potra and Sheng [34] for block diagonal SDP problems. The most computationally intensive step in an IPM for SDP is the calculation of the Newton direction. By ignoring unnecessary constraints, we show how to reduce the computational load for constructing the Schur complement matrix that determines this direction, thus reducing the cost of each iteration.

Many important classes of problems have block diagonal form: LP, QP, quadratically constrained quadratic programming (QCQP), second-order cone programming (SOCP), etc. See, for example, [4]. From this point of view, our study generalizes [37, 41, 19].

Block diagonal SDPs also arise from relaxations of many important problems with discrete variables. These problems include the maximum binary code problem [35], the traveling salesman's problem [7], the kissing number problem [3], and the quadratic assignment problem [43].

Our work is motivated by numerical experiments [31, Chap. 4] in which we modified SDPT3 4.0,[1] written by Toh, Todd, and Tütüncü [38], to incorporate constraint reduction. Constraint reduction can be quite effective, saving up to 25% of the computational work on a binary code problem, but convergence was quite sensitive to the choice of the parameter controlling the reduction. Therefore, here, modifying and refining an algorithm in [31, Chap. 5], we develop rigorous criteria for reduction, guaranteeing convergence.

Our paper is structured as follows. In Section 2 we present our algorithm. We prove global convergence and polynomial complexity in Section 3. Section 4 summarizes results and open questions.

2 Interior Point Methods for SDP

We discuss how standard IPMs find an optimal solution of SDP. For more details, see, for example, [6, 16]. We make use of the definitions in Table 1. The primal and dual SDP problems are as follows:

$$\text{Primal SDP:} \quad \min_{\boldsymbol{X}} \boldsymbol{C} \bullet \boldsymbol{X} \quad \text{s.t.} \quad \boldsymbol{A}_i \bullet \boldsymbol{X} = b_i \text{ for } i = 1, \ldots, m, \quad \boldsymbol{X} \succeq \boldsymbol{0}, \tag{1}$$

$$\text{Dual SDP:} \quad \max_{\boldsymbol{y}, \boldsymbol{Z}} \boldsymbol{b}^T \boldsymbol{y} \quad \text{s.t.} \quad \sum_{i=1}^{m} y_i \boldsymbol{A}_i + \boldsymbol{Z} = \boldsymbol{C}, \quad \boldsymbol{Z} \succeq \boldsymbol{0}, \tag{2}$$

[1]The Matlab package is available in `http://www.math.nus.edu.sg/~mattohkc/sdpt3.html`.

$\mathcal{S}^n$	the set of $n \times n$ symmetric matrices
$\mathcal{S}^n_+$	the set of $n \times n$ symmetric positive semidefinite matrices
$\mathcal{S}^n_{++}$	the set of $n \times n$ symmetric positive definite matrices
$\boldsymbol{X} \succ \mathbf{0}$	a positive definite matrix
$\boldsymbol{X} \succeq \mathbf{0}$	a positive semidefinite matrix
$\boldsymbol{A} \bullet \boldsymbol{B} = \mathrm{tr}\left(\boldsymbol{A}\boldsymbol{B}^T\right)$	the dot-product of matrices
$\mu = (\boldsymbol{X} \bullet \boldsymbol{Z})/n$	the duality gap
$\boldsymbol{x} = \mathrm{vec}\,(\boldsymbol{X})$	the vectorization of a given matrix $\boldsymbol{X}$, a stack of columns of $\boldsymbol{X}^T$
$\mathrm{mat}\,(\boldsymbol{x})$	the inverse of $\mathrm{vec}\,(\boldsymbol{X})$
$\mathrm{symm}\,(\boldsymbol{X}) = \frac{1}{2}(\boldsymbol{X} + \boldsymbol{X}^T)$	the symmetric part of $\boldsymbol{X}$
$\boldsymbol{G} \otimes \boldsymbol{H}$	Kronecker product of matrices $\boldsymbol{G}$ and $\boldsymbol{H}$
$\lVert\boldsymbol{A}\rVert$	the 2-norm of a matrix $\boldsymbol{A}$
$\lVert\boldsymbol{A}\rVert_F = (\sum_{ij} a_{ij}^2)^{1/2}$	the Frobenius norm of a matrix $\boldsymbol{A}$

Table 1: Notation for the SDP.

where $\boldsymbol{C} \in \mathcal{S}^n$, $\boldsymbol{A}_i \in \mathcal{S}^n$, $\boldsymbol{X} \in \mathcal{S}^n$, and $\boldsymbol{Z} \in \mathcal{S}^n$.
We focus on problems in which the matrices $\boldsymbol{A}_i$ and $\boldsymbol{C}$ are block diagonal:

$$\boldsymbol{A}_i = \begin{bmatrix} \boldsymbol{A}_{i1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{A}_{ip} \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{C}_p \end{bmatrix},$$

where $\boldsymbol{A}_{ij}, \boldsymbol{C}_j \in \mathcal{S}^{n_j}$ for $i = 1, \ldots, m$ and $j = 1, \ldots, p$. By the block structure of $\boldsymbol{A}_i$ and $\boldsymbol{C}$, we can also partition the parameters $\boldsymbol{X}$ and $\boldsymbol{Z}$ into $\boldsymbol{X}_j$'s and $\boldsymbol{Z}_j$'s. This is because nonzero elements outside of the diagonal blocks of $\boldsymbol{Z}$ violate the dual constraint of (2), and nonzero elements outside of the diagonal blocks in $\boldsymbol{X}$ do not make any contribution to minimize the primal objective value $\boldsymbol{C} \bullet \boldsymbol{X}$. Throughout our work, we assume the Slater condition.

Assumption 2.1 (Slater condition)**.** *There exists a primal and dual feasible point* $(\mathbf{X}, \mathbf{y}, \mathbf{Z})$ *such that* $\mathbf{X} \succ \mathbf{0}$ *and* $\mathbf{Z} \succ \mathbf{0}$.

Under Assumption 2.1, the primal and dual SDP problems have optimal solutions with equal optimal values; see, for example, de Klerk [6, Theorem 2.6 in p.33].

2.1 The HKM Newton Direction

Under Assumption 2.1, $(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z})$ is an optimal solution if and only if

$$\boldsymbol{A}_i \bullet \boldsymbol{X} = \boldsymbol{b}_i \quad \text{for } i = 1, \ldots, m, \tag{3}$$

$$(\sum_{i=1}^{m} y_i \boldsymbol{A}_i) + \boldsymbol{Z} = \boldsymbol{C}, \tag{4}$$

$$\boldsymbol{X}\boldsymbol{Z} = \boldsymbol{0}, \tag{5}$$

$$\boldsymbol{X} \succeq 0, \quad \boldsymbol{Z} \succeq 0. \tag{6}$$

So, given a current iterate $(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z})$, we solve the following equations

$$\boldsymbol{A}_i \bullet \Delta\dot{\boldsymbol{X}} = r_{pi} \quad \text{for } i = 1, \ldots, m, \tag{7}$$

$$(\sum_{i=1}^{m} \Delta y_i \boldsymbol{A}_i) + \Delta\boldsymbol{Z} = \boldsymbol{R}_d, \tag{8}$$

$$\boldsymbol{X}\Delta\boldsymbol{Z} + \Delta\dot{\boldsymbol{X}}\boldsymbol{Z} = \boldsymbol{R}_c, \tag{9}$$

$$\Delta\boldsymbol{X} = \text{ symm}\left(\Delta\dot{\boldsymbol{X}}\right), \tag{10}$$

where the primal, dual, and complementarity residuals are defined by

$$r_{pi} \quad := \quad b_i - \boldsymbol{A}_i \bullet \boldsymbol{X} \quad \text{for } i = 1, \ldots, m, \tag{11}$$

$$\boldsymbol{R}_d \quad := \quad \boldsymbol{C} - \boldsymbol{Z} - \sum_{i=1}^{m} y_i \boldsymbol{A}_i, \tag{12}$$

$$\boldsymbol{R}_c \quad := \quad \overline{\mu}\boldsymbol{I} - \boldsymbol{X}\boldsymbol{Z}, \tag{13}$$

and $\overline{\mu}$ defines the current target point on the central path. The solution of the equations is called the HKM direction, named after Helmberg et al., Kojima et al., and Monteiro [12, 23, 24].

Kojima et al.[23, Theorem 4.2] proved that the equations above have a unique solution for $(\boldsymbol{X}, \boldsymbol{Z}) \in \mathcal{S}^n_{++} \times \mathcal{S}^n_{++}$. Monteiro [24] showed that we can obtain the same direction without the symmetrization step (10) by solving

$$\boldsymbol{A}_i \bullet \Delta\boldsymbol{X} = r_{pi} \quad \text{for } i = 1, \ldots, m, \tag{14}$$

$$(\sum_{i=1}^{m} \Delta y_i \boldsymbol{A}_i) + \Delta\boldsymbol{Z} = \boldsymbol{R}_d, \tag{15}$$

$$\text{symm}\left(\boldsymbol{Z}^{1/2}(\boldsymbol{X}\Delta\boldsymbol{Z} + \Delta\boldsymbol{X}\boldsymbol{Z})\boldsymbol{Z}^{-1/2}\right) = \overline{\mu}\boldsymbol{I} - \boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2}. \tag{16}$$

Monteiro [24, Lemma 2.1ff] proved that the solution of (7)-(10) is the unique solution of (14)-(16). So, we frequently refer to (16) for convergence analysis later.

2.2 Constraint Reduction and Inactive Blocks

In the following equations, we make frequent use of identities involving vectorization and the Kronecker product, in particular,

$$\mathrm{vec}\,(\boldsymbol{KNL}) = (\boldsymbol{K} \otimes \boldsymbol{L}^T)\ \mathrm{vec}\,(\boldsymbol{N}),$$
$$(\boldsymbol{K} \otimes \boldsymbol{L})(\boldsymbol{N} \otimes \boldsymbol{Q}) = (\boldsymbol{KN}) \otimes (\boldsymbol{LQ}).$$

Let $\mathcal{A} \in \mathbb{R}^{m \times n^2}$ be the matrix with ith row equal to $\mathrm{vec}\,(\boldsymbol{A}_i)^T$, $i = 1, \ldots, m$, and recall from Table 1 that we use a lowercase letter for a row-ordered vectorization of a given matrix. Using Gauss elimination, equations (7)-(10) can be reduced to

$$\boldsymbol{M}\Delta\boldsymbol{y} = \boldsymbol{g},$$

where the Schur complement matrix $\boldsymbol{M} = \mathcal{A}(\boldsymbol{X} \otimes \boldsymbol{Z}^{-1})\mathcal{A}^T$ and $\boldsymbol{g} = \boldsymbol{r}_p + \mathcal{A}(\boldsymbol{X} \otimes \boldsymbol{Z}^{-1})\boldsymbol{r}_d - \mathcal{A}(\boldsymbol{I} \otimes \boldsymbol{Z}^{-1})\boldsymbol{r}_c$. After solving the Schur complement equation, we compute

$$\Delta\boldsymbol{z} = \boldsymbol{r}_d - \mathcal{A}^T\Delta\boldsymbol{y}, \quad (17)$$
$$\Delta\dot{\boldsymbol{x}} = (\boldsymbol{X} \otimes \boldsymbol{Z}^{-1})(\mathcal{A}^T\Delta\boldsymbol{y} - \boldsymbol{r}_d) + (\boldsymbol{I} \otimes \boldsymbol{Z}^{-1})\boldsymbol{r}_c. \quad (18)$$

Using the block structure of $\boldsymbol{A}_i$ and $\boldsymbol{C}$, the Schur complement matrix $\boldsymbol{M}$ can be computed as

$$\boldsymbol{M} = \sum_{j=1}^{p} \boldsymbol{M}_j,$$

where

$$\boldsymbol{M}_j := \mathcal{A}_j(\boldsymbol{X}_j \otimes \boldsymbol{Z}_j^{-1})\mathcal{A}_j^T, \quad \text{and} \quad \mathcal{A}_j := \begin{bmatrix} \mathrm{vec}\,(\boldsymbol{A}_{1j})^T \\ \vdots \\ \mathrm{vec}\,(\boldsymbol{A}_{mj})^T \end{bmatrix} \in \mathbb{R}^{m \times n_j^2}.$$

Hence, each element $(\boldsymbol{M}_j)_{lh}$ of $\boldsymbol{M}_j$, where $1 \le l \le m,\ l \le h \le m,\ 1 \le j \le p$, can be computed as

$$(\boldsymbol{M}_j)_{lh} = (\boldsymbol{X}_j\ \boldsymbol{A}_{lj}\ \boldsymbol{Z}_j^{-1}) \bullet \boldsymbol{A}_{hj}. \quad (19)$$

If $\boldsymbol{A}_{ij}$ is dense[2], then the cost of computing the entire Schur complement matrix $\boldsymbol{M}$, including Cholesky factorization of $\boldsymbol{Z}_j$, is

$$\sum_{j=1}^{p}(4m + 1/3)n_j^3 + 2m^2 n_j^2 \text{ operations.} \quad (20)$$

This is the most expensive computation in the algorithm. Our goal is to drop matrices $\boldsymbol{M}_j$ that do not play important roles in $\boldsymbol{M}$, reducing the computational

[2]Refer to Fujisawa, Kojima, and Nakata [9] to see how to exploit the sparsity of $\boldsymbol{A}_{ij}$.

cost. We classify the blocks into *active* and *inactive* blocks and discuss why the latter can be dropped.

From the optimality condition (5), we see that $r_x + r_z \le n$, where r_x and r_z are the ranks of an optimal solution $\boldsymbol{X}^*$ and $\boldsymbol{Z}^*$. This implies that there may exist blocks $\boldsymbol{X}_j^*$ and $\boldsymbol{Z}_j^*$ such that $\boldsymbol{X}_j^* = \boldsymbol{0}$ and $\boldsymbol{Z}_j^*$ has full rank, so $\boldsymbol{Z}_j^* \succ \boldsymbol{0}$ and $\boldsymbol{Z}_j^*$ is in the interior of the semidefinite cone. For such a block, $(\boldsymbol{X}_j^* \otimes \boldsymbol{Z}_j^{*-1}) = \boldsymbol{0}$. Given that the algorithm converges to an optimal solution, $(\boldsymbol{X}_j \otimes \boldsymbol{Z}_j^{-1})$ will approach $\boldsymbol{0}$ as the iteration proceeds. In Section 2.3, we define thresholds on $\|\boldsymbol{X}_j \otimes \boldsymbol{Z}_j^{-1}\|_F$ allowing us to drop blocks in $\boldsymbol{M}$ while guaranteeing convergence of the algorithm. At a given iteration, we will say that blocks larger than the threshold are *active*, and that the other blocks are *inactive*. Without loss of generality, we assume that the first $\widehat{p}$ blocks are active and the remaining $\tilde{p}$ blocks are inactive. We let $\widehat{\boldsymbol{A}}_i$ and $\widetilde{\boldsymbol{A}}_i$ denote the active and inactive blocks of $\boldsymbol{A}_i$, so

$$\widehat{\boldsymbol{A}}_i = \begin{bmatrix} \boldsymbol{A}_{i1} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{A}_{i\widehat{p}} \end{bmatrix}, \quad \widetilde{\boldsymbol{A}}_i = \begin{bmatrix} \boldsymbol{A}_{i(\widehat{p}+1)} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{A}_{ip} \end{bmatrix},$$

where $\widehat{\boldsymbol{A}}_i \in \mathbb{R}^{\widehat{n}\times\widehat{n}}$, $\widetilde{\boldsymbol{A}}_i \in \mathbb{R}^{\widetilde{n}\times\widetilde{n}}$, and $n = \widehat{n} + \widetilde{n}$. Furthermore, let $\widehat{n}_j$ and $\widetilde{n}_j$ denote the size of active and inactive blocks, so that

$$\widehat{n} := \sum_{j=1}^{\widehat{p}} \widehat{n}_j, \qquad \widetilde{n} := \sum_{j=1}^{\tilde{p}} \widetilde{n}_j.$$

Block matrices $(\widehat{\boldsymbol{X}}, \widetilde{\boldsymbol{X}})$, $(\widehat{\boldsymbol{Z}}, \widetilde{\boldsymbol{Z}})$, $(\widehat{\boldsymbol{R}}_d, \widetilde{\boldsymbol{R}}_d)$, and $(\widehat{\boldsymbol{R}}_c, \widetilde{\boldsymbol{R}}_c)$ are defined similarly. We also define $\widehat{\mathcal{A}} \in \mathbb{R}^{m\times\widehat{n}^2}$, with rows equal to $\operatorname{vec}\left(\widehat{\boldsymbol{A}}_i\right)^T$, $i = 1, \ldots, \widehat{p}$, and $\widetilde{\mathcal{A}} \in \mathbb{R}^{m\times\widetilde{n}^2}$, with rows equal to $\operatorname{vec}\left(\widetilde{\boldsymbol{A}}_i\right)^T$, $i = 1, \ldots, \tilde{p}$.

Then we can expand $\boldsymbol{M}$ into active and inactive parts as $\boldsymbol{M} = \widehat{\boldsymbol{M}} + \widetilde{\boldsymbol{M}}$, where

$$\widehat{\boldsymbol{M}} = \widehat{\mathcal{A}}(\widehat{\boldsymbol{X}} \otimes \widehat{\boldsymbol{Z}}^{-1})\widehat{\mathcal{A}}^T, \;\; \widetilde{\boldsymbol{M}} = \widetilde{\mathcal{A}}(\widetilde{\boldsymbol{X}} \otimes \widetilde{\boldsymbol{Z}}^{-1})\widetilde{\mathcal{A}}^T.$$

If $\|(\widetilde{\boldsymbol{X}} \otimes \widetilde{\boldsymbol{Z}}^{-1})\|$ is small, we expect that $\widetilde{\boldsymbol{M}}$ is also negligible and we can omit it when we solve the linear system.

Constraint reduction generates an extra term $\Delta \boldsymbol{X}_\epsilon$ in the primal direction, perturbing the complementarity equation. Due to this, a series of lemmas for global convergence by Potra and Sheng [34] need to be modified. The proposed adaptive criteria restrain the magnitude of $\Delta \boldsymbol{X}_\epsilon$ so that we can guarantee that we can take a step long enough to ensure convergence.

The constraint-reduced equation is

$$\widehat{\boldsymbol{M}} \Delta \boldsymbol{y} = \boldsymbol{g}, \tag{21}$$

or equivalently,

$$\left(\widehat{\mathcal{A}}(\widehat{\boldsymbol{X}} \otimes \widehat{\boldsymbol{Z}}^{-1})\widehat{\mathcal{A}}^T\right) \Delta \boldsymbol{y} = \boldsymbol{r}_p + \mathcal{A}(\boldsymbol{X} \otimes \boldsymbol{Z}^{-1})\boldsymbol{r}_d - \mathcal{A}(\boldsymbol{I} \otimes \boldsymbol{Z}^{-1})\boldsymbol{r}_c. \tag{22}$$

We select $\widehat{\boldsymbol{X}}$ and $\widehat{\boldsymbol{Z}}$ so that the resulting $\widehat{\boldsymbol{M}}$ has full rank, which is always possible because the equation above has a unique solution if no reduction is done [23, Theorem 4.2]. We discuss checking the rank of $\widehat{\boldsymbol{M}}$ at the end of this section. After solving the equation above, we then compute $\Delta\dot{\boldsymbol{x}}$ and $\Delta\boldsymbol{z}$ from

$$\Delta\dot{\boldsymbol{x}} = \operatorname{vec}\left(\begin{bmatrix} \Delta\dot{\widehat{\boldsymbol{X}}} & \mathbf{0} \\ \mathbf{0} & \Delta\dot{\widetilde{\boldsymbol{X}}} \end{bmatrix}\right), \tag{23}$$

$$\Delta\boldsymbol{z} = \boldsymbol{r}_d - \mathcal{A}^T\Delta\boldsymbol{y}, \tag{24}$$

where

$$\Delta\dot{\widehat{\boldsymbol{X}}} := \operatorname{mat}\left((\widehat{\boldsymbol{X}}\otimes\widehat{\boldsymbol{Z}}^{-1})\widehat{\mathcal{A}}^T\Delta\boldsymbol{y} - (\widehat{\boldsymbol{X}}\otimes\widehat{\boldsymbol{Z}}^{-1})\widehat{\boldsymbol{r}}_d + (\boldsymbol{I}\otimes\widehat{\boldsymbol{Z}}^{-1})\widehat{\boldsymbol{r}}_c\right), \tag{25}$$

$$\Delta\dot{\widetilde{\boldsymbol{X}}} := \operatorname{mat}\left(-(\widetilde{\boldsymbol{X}}\otimes\widetilde{\boldsymbol{Z}}^{-1})\widetilde{\boldsymbol{r}}_d + (\boldsymbol{I}\otimes\widetilde{\boldsymbol{Z}}^{-1})\widetilde{\boldsymbol{r}}_c\right). \tag{26}$$

Note that while (18) contains $(\boldsymbol{X}\otimes\boldsymbol{Z}^{-1})\mathcal{A}^T\Delta\boldsymbol{y}$ as its first term, $\Delta\dot{\widetilde{\boldsymbol{X}}}$ in (26) does not have the corresponding term $(\widetilde{\boldsymbol{X}}\otimes\widetilde{\boldsymbol{Z}}^{-1})\widetilde{\mathcal{A}}^T\Delta\boldsymbol{y}$, which will cause a perturbation $\Delta\dot{\boldsymbol{X}}_\epsilon$ in the primal direction. In the following lemma, we show that $\Delta\dot{\boldsymbol{x}}$, $\Delta\boldsymbol{z}$, and $\Delta\boldsymbol{y}$ from equations (21), (23), and (24), are a solution of the perturbed equations

$$\mathcal{A}\Delta\dot{\boldsymbol{x}} = \boldsymbol{r}_p, \tag{27}$$

$$\mathcal{A}^T\Delta\boldsymbol{y} + \Delta\boldsymbol{z} = \boldsymbol{r}_d, \tag{28}$$

$$(\boldsymbol{X}\otimes\boldsymbol{I})\Delta\boldsymbol{z} + (\boldsymbol{I}\otimes\boldsymbol{Z})(\Delta\dot{\boldsymbol{x}} + \Delta\dot{\boldsymbol{x}}_\epsilon) = \boldsymbol{r}_c, \tag{29}$$

where

$$\Delta\dot{\boldsymbol{X}}_\epsilon := \operatorname{mat}(\Delta\dot{\boldsymbol{x}}_\epsilon) = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \operatorname{mat}\left((\widetilde{\boldsymbol{X}}\otimes\widetilde{\boldsymbol{Z}}^{-1})\widetilde{\mathcal{A}}^T\Delta\boldsymbol{y}\right) \end{bmatrix}. \tag{30}$$

Note the new vector $\Delta\dot{\boldsymbol{x}}_\epsilon$ in the second term of (29).

Lemma 2.1 (Perturbed Newton equations)**.** *The solution* $(\Delta\dot{\mathbf{x}}, \Delta\mathbf{y}, \Delta\mathbf{z})$ *of (21), (23), and (24) satisfies equations (27)-(29).*

Proof. First, we note the primal equation (27) is satisfied since, by (23),

$$\begin{aligned}
\mathcal{A}\Delta\dot{\boldsymbol{x}} &= \widehat{\mathcal{A}}\Delta\dot{\widehat{\boldsymbol{x}}} + \widetilde{\mathcal{A}}\Delta\dot{\widetilde{\boldsymbol{x}}} \\
&= \widehat{\mathcal{A}}(\widehat{\boldsymbol{X}}\otimes\widehat{\boldsymbol{Z}}^{-1})\widehat{\mathcal{A}}^T\Delta\boldsymbol{y} - \widehat{\mathcal{A}}(\widehat{\boldsymbol{X}}\otimes\widehat{\boldsymbol{Z}}^{-1})\widehat{\boldsymbol{r}}_d + \widehat{\mathcal{A}}(\boldsymbol{I}\otimes\widehat{\boldsymbol{Z}}^{-1})\widehat{\boldsymbol{r}}_c \\
&\quad - \widetilde{\mathcal{A}}(\widetilde{\boldsymbol{X}}\otimes\widetilde{\boldsymbol{Z}}^{-1})\widetilde{\boldsymbol{r}}_d + \widetilde{\mathcal{A}}(\boldsymbol{I}\otimes\widetilde{\boldsymbol{Z}}^{-1})\widetilde{\boldsymbol{r}}_c \quad \text{(by (25) and (26))} \\
&= \widehat{\mathcal{A}}(\widehat{\boldsymbol{X}}\otimes\widehat{\boldsymbol{Z}}^{-1})\widehat{\mathcal{A}}^T\Delta\boldsymbol{y} - \mathcal{A}(\boldsymbol{X}\otimes\boldsymbol{Z}^{-1})\boldsymbol{r}_d + \mathcal{A}(\boldsymbol{I}\otimes\boldsymbol{Z}^{-1})\boldsymbol{r}_c \\
&= \left(\boldsymbol{r}_p + \mathcal{A}(\boldsymbol{X}\otimes\boldsymbol{Z}^{-1})\boldsymbol{r}_d - \mathcal{A}(\boldsymbol{I}\otimes\boldsymbol{Z}^{-1})\boldsymbol{r}_c\right) \\
&\quad - \mathcal{A}(\boldsymbol{X}\otimes\boldsymbol{Z}^{-1})\boldsymbol{r}_d + \mathcal{A}(\boldsymbol{I}\otimes\boldsymbol{Z}^{-1})\boldsymbol{r}_c \quad \text{(by (22))} \\
&= \boldsymbol{r}_p.
\end{aligned}$$

In addition, (28) is immediately satisfied by (24).

To see that (29) is satisfied, we first calculate $(\Delta\dot{\boldsymbol{x}} + \Delta\boldsymbol{x}_\epsilon)$. By (23), (25), (26), and (30),

$$\begin{aligned}\Delta\dot{\boldsymbol{X}} + \Delta\dot{\boldsymbol{X}}_\epsilon &= \begin{bmatrix} \Delta\dot{\widehat{\boldsymbol{X}}} & \mathbf{0} \\ \mathbf{0} & \Delta\dot{\widetilde{\boldsymbol{X}}} + \operatorname{mat}\left((\widetilde{\boldsymbol{X}} \otimes \widetilde{\boldsymbol{Z}}^{-1})\widetilde{\mathcal{A}}^T\Delta\boldsymbol{y}\right) \end{bmatrix} \\ &= \operatorname{mat}\left((\boldsymbol{X} \otimes \boldsymbol{Z}^{-1})\mathcal{A}^T\Delta\boldsymbol{y} - (\boldsymbol{X} \otimes \boldsymbol{Z}^{-1})\boldsymbol{r}_d + (\boldsymbol{I} \otimes \boldsymbol{Z}^{-1})\boldsymbol{r}_c\right),\end{aligned}$$

so

$$\Delta\dot{\boldsymbol{x}} + \Delta\dot{\boldsymbol{x}}_\epsilon = (\boldsymbol{X} \otimes \boldsymbol{Z}^{-1})\mathcal{A}^T\Delta\boldsymbol{y} - (\boldsymbol{X} \otimes \boldsymbol{Z}^{-1})\boldsymbol{r}_d + (\boldsymbol{I} \otimes \boldsymbol{Z}^{-1})\boldsymbol{r}_c.$$

Thus,

$$\begin{aligned}(\boldsymbol{I} \otimes \boldsymbol{Z})(\Delta\dot{\boldsymbol{x}} + \Delta\dot{\boldsymbol{x}}_\epsilon) &= (\boldsymbol{I} \otimes \boldsymbol{Z})(\boldsymbol{X} \otimes \boldsymbol{Z}^{-1})\mathcal{A}^T\Delta\boldsymbol{y} - (\boldsymbol{I} \otimes \boldsymbol{Z})(\boldsymbol{X} \otimes \boldsymbol{Z}^{-1})\boldsymbol{r}_d \\ &\quad + (\boldsymbol{I} \otimes \boldsymbol{Z})(\boldsymbol{I} \otimes \boldsymbol{Z}^{-1})\boldsymbol{r}_c \\ &= (\boldsymbol{X} \otimes \boldsymbol{I})\mathcal{A}^T\Delta\boldsymbol{y} - (\boldsymbol{X} \otimes \boldsymbol{I})\boldsymbol{r}_d + (\boldsymbol{I} \otimes \boldsymbol{I})\boldsymbol{r}_c = (\boldsymbol{X} \otimes \boldsymbol{I})(\mathcal{A}^T\Delta\boldsymbol{y} - \boldsymbol{r}_d) + \boldsymbol{r}_c \\ &= -(\boldsymbol{X} \otimes \boldsymbol{I})\Delta\boldsymbol{z} + \boldsymbol{r}_c \quad \text{(by (24))}.\end{aligned}$$

Therefore,

$$(\boldsymbol{X} \otimes \boldsymbol{I})\Delta\boldsymbol{z} + (\boldsymbol{I} \otimes \boldsymbol{Z})(\Delta\dot{\boldsymbol{x}} + \Delta\dot{\boldsymbol{x}}_\epsilon) = \boldsymbol{r}_c.$$

□

From equations (27)-(30) and Lemma 2.1, we can see that constraint reduction does not affect the primal and dual equations (7) and (8) but only the complementarity equation (9). Furthermore, considering the relations between (7)-(10) and (14)-(16), the solution $(\Delta\dot{\boldsymbol{x}}, \Delta\boldsymbol{y}, \Delta\boldsymbol{z})$ of (27)-(29) also satisfies the following equations by the symmetrization of $\Delta\boldsymbol{X} = \operatorname{symm}\left(\Delta\dot{\boldsymbol{X}}\right)$ and $\Delta\boldsymbol{X}_\epsilon = \operatorname{symm}\left(\dot{\boldsymbol{X}}_\epsilon\right)$,

$$\mathcal{A}\Delta\boldsymbol{x} = \boldsymbol{r}_p, \tag{31}$$

$$\mathcal{A}^T\Delta\boldsymbol{y} + \Delta\boldsymbol{z} = \boldsymbol{r}_d, \tag{32}$$

$$\boldsymbol{Z}^{1/2}(\boldsymbol{X} + \Delta\boldsymbol{X}_\epsilon)\boldsymbol{Z}^{1/2} + \operatorname{symm}\left(\boldsymbol{Z}^{1/2}(\boldsymbol{X}\Delta\boldsymbol{Z} + \Delta\boldsymbol{X}\boldsymbol{Z})\boldsymbol{Z}^{-1/2}\right) = \overline{\mu}\boldsymbol{I}. \tag{33}$$

2.3 Algorithm SDP:Reduced

In this section, we introduce an interior point method, similar to that of Potra and Sheng [34], but including constraint reduction. We define a set $\mathcal{F}$ of feasible solutions and a set $\mathcal{F}^*$ of optimal solutions as

$$\begin{aligned}\mathcal{F} &:= \{(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z}) \in \mathcal{S}_+^n \times \mathbb{R}^m \times \mathcal{S}_+^n : (\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z}) \text{ satisfies (3) and (4).} \}, \\ \mathcal{F}^* &:= \{(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z}) \in \mathcal{F} : \boldsymbol{X} \bullet \boldsymbol{Z} = 0\}.\end{aligned}$$

We also define the neighborhood $\mathcal{N}(\gamma, \tau)$ of the central path as

$$\mathcal{N}(\gamma, \tau) := \{(\boldsymbol{X}, \boldsymbol{Z}) \in \mathcal{S}_{++}^n \times \mathcal{S}_{++}^n : \|\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2} - \tau\boldsymbol{I}\|_F \leq \gamma\tau\}.$$

In the predictor step, given the current (possibly infeasible) iterate $(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z})$ and inactive blocks $(\widetilde{\boldsymbol{X}}, \widetilde{\boldsymbol{Z}})$ of $(\boldsymbol{X}, \boldsymbol{Z})$, we find a solution $(\Delta\boldsymbol{X}, \Delta\boldsymbol{y}, \Delta\boldsymbol{Z})$ of (31)-(33), setting $\overline{\mu} = 0$. We refer to the resulting equations as (31P)-(33P) and set

$$\Delta\boldsymbol{X}_\epsilon = \operatorname{symm}\left(\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \operatorname{mat}\left((\widetilde{\boldsymbol{X}} \otimes \widetilde{\boldsymbol{Z}}^{-1})\widetilde{\mathcal{A}}^T \Delta\boldsymbol{y}\right) \end{bmatrix}\right). \tag{34}$$

We then compute an updated point $(\overline{\boldsymbol{X}}, \overline{\boldsymbol{y}}, \overline{\boldsymbol{Z}})$ by taking a step of length $\overline{\theta} < 1$ in this direction.

In the corrector step, we set the target duality gap $\overline{\mu} = (1 - \overline{\theta})\tau$, where the parameter τ decreases at each iteration. Then, with inactive blocks $(\widetilde{\boldsymbol{X}}, \widetilde{\boldsymbol{Z}})$ of $(\overline{\boldsymbol{X}}, \overline{\boldsymbol{Z}})$, we find a solution $(\Delta\overline{\boldsymbol{X}}, \Delta\overline{\boldsymbol{y}}, \Delta\overline{\boldsymbol{Z}})$ of (31)-(33) with $\boldsymbol{r}_p = \boldsymbol{0}$ and $\boldsymbol{r}_d = \boldsymbol{0}$. We refer to the resulting equations as (31C)-(33C) and set

$$\Delta\overline{\boldsymbol{X}}_\epsilon = \operatorname{symm}\left(\begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \operatorname{mat}\left((\widetilde{\boldsymbol{X}} \otimes \widetilde{\boldsymbol{Z}}^{-1})\widetilde{\mathcal{A}}^T \Delta\overline{\boldsymbol{y}}\right) \end{bmatrix}\right). \tag{35}$$

We denote the directions' norms as

$$\delta := \frac{1}{\tau}\|\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}\Delta\boldsymbol{Z}\boldsymbol{Z}^{-1/2}\|_F, \tag{36}$$

$$\delta_x := \|\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}\boldsymbol{Z}^{1/2}\|_F, \tag{37}$$

$$\delta_z := \tau\|\boldsymbol{Z}^{-1/2}\Delta\boldsymbol{Z}\boldsymbol{Z}^{-1/2}\|_F, \tag{38}$$

$$\delta_\epsilon := \frac{1}{\tau}\|\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}_\epsilon\boldsymbol{Z}^{1/2}\|_F, \tag{39}$$

$$\overline{\delta}_\epsilon := \frac{1}{\tau}\|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}_\epsilon\overline{\boldsymbol{Z}}^{1/2}\|_F. \tag{40}$$

We use two fixed positive parameters α and β with the property

$$\frac{\beta^2}{2(1-\beta)^2} < \alpha < \beta \le \frac{\beta}{1-\beta} < 1. \tag{41}$$

This inequality restrains the ranges of α and β as

$$0 < \alpha < \beta < 0.5. \tag{42}$$

For example, we can choose $(\alpha, \beta) = (0.17, 0.3)$. Based on these parameters, we define $\widehat{\theta}$ and $\breve{\theta}$ (which change at each iteration) as

$$\begin{aligned} \widehat{\theta} &:= \frac{(\alpha - \beta - \delta_\epsilon) + \sqrt{(\alpha - \beta - \delta_\epsilon)^2 + 4\delta(\beta - \alpha)}}{2\delta} \\ &= \frac{2(\beta - \alpha)}{\sqrt{(\beta - \alpha + \delta_\epsilon)^2 + 4\delta(\beta - \alpha)} + (\beta - \alpha + \delta_\epsilon)}, \end{aligned} \tag{43}$$

$$\breve{\theta} := \max\{\tilde{\theta} \in [0,1] : (\boldsymbol{X} + \theta\Delta\boldsymbol{X}, \boldsymbol{y} + \theta\Delta\boldsymbol{y}, \boldsymbol{Z} + \theta\Delta\boldsymbol{Z}) \in \mathcal{N}(\beta, (1-\theta)\tau),\ \forall\theta \in [0, \tilde{\theta}]\}. \tag{44}$$

In our *predictor-corrector* algorithm, specified below, at each step we incrementally build a set of active blocks that satisfies one of the following two conditions, the first for a predictor step and the second for a corrector step.

Condition 2.1.

$$\delta_\epsilon \leq \frac{q}{\tau}\delta_x, \tag{45}$$

or equivalently

$$\|\mathbf{Z}^{1/2}\Delta\mathbf{X}_\epsilon\mathbf{Z}^{1/2}\|_F \leq q\|\mathbf{Z}^{1/2}\Delta\mathbf{X}\mathbf{Z}^{1/2}\|_F,$$

where the input parameter q of the algorithm has a range of

$$0 \leq q < 0.5. \tag{46}$$

Condition 2.2.

$$\overline{\delta}_\epsilon < (1-\overline{\theta})(\sqrt{s^2+t}-s), \tag{47}$$

where

$$s := \beta^2 - \beta + 1, \quad t := 2\alpha(1-\beta)^2 - \beta^2. \tag{48}$$

Based on these parameters and conditions, we define our *primal-dual predictor-corrector* algorithm, Algorithm SDP:Reduced. Before analyzing the algorithm, the following overview is useful.

1. For step 2, an appropriate choice of ρ is discussed by Toh, Todd and Tütüncü [38, Section 3.4]. Convergence holds for any $\rho > 0$, but we follow [34, Thm. 3.8] in our constraint on ρ. A smaller ρ increases the size of w in Lemma 3.9 below, slowing convergence. In practice, if a bound on $\max(||\boldsymbol{X}^*||, ||\boldsymbol{Z}^*||)$ is not known and convergence is slow, the iteration can be restarted with a larger ρ.

2. In step 3(d), the choice of step length in the predictor step is valid only when $\widehat{\theta} \leq \breve{\theta}$, which will be proved in Lemma 3.2.

3. It may not be practical to compute $\breve{\theta}$. Instead, $\overline{\theta}$ can be defined by (43) or computed by line search.

4. In step 3(e), the algorithm terminates since $(\overline{\boldsymbol{X}}, \overline{\boldsymbol{y}}, \overline{\boldsymbol{Z}})$ is an optimal solution, which will be shown in Lemma 3.2.

5. Since $\boldsymbol{r}_p = 0$ and $\boldsymbol{r}_d = 0$ in the corrector step, the corrector step's only purpose is to move the point toward the central path.

6. By (43), $\widehat{\theta}$ is a decreasing function of δ_ϵ. Thus, there is a trade-off between the allowance for constraint reduction and the step length in the predictor step.

7. The predictor step moves the point from $\mathcal{N}(\alpha,\tau)$ into $\mathcal{N}(\beta,(1-\overline{\theta})\tau)$ (Lemma 3.2), and the corrector moves it into $\mathcal{N}(\alpha,(1-\overline{\theta})\tau)$ (Lemma 3.3).

Algorithm SDP:Reduced: Primal-Dual Infeasible Constraint-Reduced Predictor-Corrector Algorithm for Block Diagonal SDP

1. Input : $\mathcal{A}$, $\boldsymbol{b}$, $\boldsymbol{C}$; α and β satisfying (41); convergence tolerance τ^*; q for the perturbation bound of the primal direction in the predictor step, satisfying (46).
2. Set $\boldsymbol{X}^0 = \boldsymbol{Z}^0 = \rho \boldsymbol{I}$ for $\rho > \max(||\boldsymbol{X}^*||, ||\boldsymbol{Z}^*||)$, and set $\tau = \tau_0 = \mu_0 = (\boldsymbol{X}^0 \bullet \boldsymbol{Z}^0)/n$ so that $(\boldsymbol{X}^0, \boldsymbol{Z}^0) \in \mathcal{N}(\alpha, \tau_0)$.
3. Repeat until $\tau < \tau^*$: For $k = 0, 1, \ldots,$
    (a) Set $(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z}) = (\boldsymbol{X}^k, \boldsymbol{y}^k, \boldsymbol{Z}^k)$ and $\tau = \tau_k$.
    (b) Sort the constraint blocks in decreasing order of $\|\boldsymbol{X}_j \otimes \boldsymbol{Z}_j^{-1}\|$.
    (c) Initially, $\widehat{\boldsymbol{M}}_P = \boldsymbol{0}$. For $j = 1, \ldots, p$, until $\widehat{\boldsymbol{M}}_P$ is full-rank and Condition 2.1 (above) is satisfied, replace $\widehat{\boldsymbol{M}}_P$ by $\widehat{\boldsymbol{M}}_P + \mathcal{A}_j(\boldsymbol{X}_j \otimes \boldsymbol{Z}_j^{-1})\mathcal{A}_j^T$. Set $\widehat{p} = j$.
    (d) *Predictor step:* Solve (22) with $\widehat{\boldsymbol{M}} = \widehat{\boldsymbol{M}}_P$ and $\boldsymbol{r}_c = \operatorname{vec}(-\boldsymbol{X}\boldsymbol{Z})$ for $(\Delta\boldsymbol{X}, \Delta\boldsymbol{y}, \Delta\boldsymbol{Z})$ satisfying (31P)–(33P). Choose a step length $\overline{\theta} \in [\widehat{\theta}, \breve{\theta}]$ defined by (43) and (44),
    $\overline{\boldsymbol{X}} = \boldsymbol{X} + \overline{\theta}\Delta\boldsymbol{X}, \quad \overline{\boldsymbol{y}} = \boldsymbol{y} + \overline{\theta}\Delta\boldsymbol{y}, \quad \overline{\boldsymbol{Z}} = \boldsymbol{Z} + \overline{\theta}\Delta\boldsymbol{Z}.$
    (e) If $\overline{\theta} = 1$, terminate the iteration with optimal solution $(\overline{\boldsymbol{X}}, \overline{\boldsymbol{y}}, \overline{\boldsymbol{Z}})$.
    (f) Sort the constraint blocks in decreasing order of $\|\overline{\boldsymbol{X}}_j \otimes \overline{\boldsymbol{Z}}_j^{-1}\|$.
    (g) Initially, $\widehat{\boldsymbol{M}}_C = \boldsymbol{0}$. For $j = 1, \ldots, p$, until $\widehat{\boldsymbol{M}}_C$ is full-rank and Condition 2.2 (above) is satisfied, replace $\widehat{\boldsymbol{M}}_C$ by $\widehat{\boldsymbol{M}}_C + \mathcal{A}_j(\overline{\boldsymbol{X}}_j \otimes \overline{\boldsymbol{Z}}_j^{-1})\mathcal{A}_j^T$. Set $\widehat{p} = j$.
    (h) *Corrector step:* Solve (22) with $\widehat{\boldsymbol{M}} = \widehat{\boldsymbol{M}}_C$, $\boldsymbol{r}_p = \boldsymbol{0}$, $\boldsymbol{r}_d = \boldsymbol{0}$, and $\boldsymbol{r}_c = \operatorname{vec}\big((1-\overline{\theta})\tau\boldsymbol{I} - \overline{\boldsymbol{X}}\,\overline{\boldsymbol{Z}}\big)$, for $(\Delta\overline{\boldsymbol{X}}, \Delta\overline{\boldsymbol{y}}, \Delta\overline{\boldsymbol{Z}})$ satisfying (31C)-(33C). Take a full step as
    $\boldsymbol{X}^{(k+1)} = \boldsymbol{X}^+ = \overline{\boldsymbol{X}} + \Delta\overline{\boldsymbol{X}}, \quad \boldsymbol{y}^{(k+1)} = \boldsymbol{y}^+ = \overline{\boldsymbol{y}} + \Delta\overline{\boldsymbol{y}}, \quad \boldsymbol{Z}^{(k+1)} = \boldsymbol{Z}^+ = \overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{Z}}.$
    (i) Set $\tau_{k+1} = (1-\overline{\theta})\tau$.
    (j) Update $\boldsymbol{r}_p = \boldsymbol{b} - \mathcal{A}\boldsymbol{x}$ and $\boldsymbol{r}_d = \boldsymbol{c} - \boldsymbol{z} - \mathcal{A}^T\boldsymbol{y}$.

8. Condition 2.1 and Condition 2.2 restrict the magnitude of $\Delta\boldsymbol{X}_\epsilon$ and $\Delta\overline{\boldsymbol{X}}_\epsilon$, which are the perturbations caused by constraint reduction.
9. $\widetilde{\boldsymbol{X}} \otimes \widetilde{\boldsymbol{Z}}^{-1} = \boldsymbol{0}$ when we use the full Schur complement matrix, so by (34) and (35), Condition 2.1 and Condition 2.2 can always be satisfied by making all blocks active. The thresholds are updated dynamically every iteration.

In the process of incremental construction of $\widehat{\boldsymbol{M}}_P$, we check that it is full rank and satisfies Condition 2.1. To do this, we can solve for $\Delta \boldsymbol{y}$ and calculate $\Delta \boldsymbol{X}_\epsilon$, which may require the Cholesky factor of $\widehat{\boldsymbol{M}}_P$ to compute $\Delta \boldsymbol{y} = \widehat{\boldsymbol{M}}_P^{-1} \boldsymbol{g}$. In some cases it is obvious that $\widehat{\boldsymbol{M}}_P$ is rank deficient (e.g, if $\sum_{l=1}^{j} n_l^2 < m$ for the blocks included) or full-rank (e.g., when $\widehat{\boldsymbol{M}}_P$ is already full-rank before a block is added). In other cases we verify the full-rank condition by checking that the Cholesky factorization exists and then applying an inexpensive condition number estimator [11, 14, 29, 36]. Once $\widehat{\boldsymbol{M}}_P$ has full rank, we can use rank-1 updating of the Cholesky factor[3] for $\widehat{\boldsymbol{M}}_P$ depending on the size of m and n_j. Similar comments hold for $\widehat{\boldsymbol{M}}_C$.

We now discuss this updating. Let $\boldsymbol{R}_{X_j}$ and $\boldsymbol{R}_{Z_j}$ be Cholesky factors of $\boldsymbol{X}_j$ and $\boldsymbol{Z}_j$. Note that the factor $\boldsymbol{R}_{Z_j}$ is required to compute $\boldsymbol{M}_j$ by (19), regardless of constraint reduction, unless $\boldsymbol{Z}_j^{-1}$ is computed explicitly. Then, the partial Schur complement $\boldsymbol{M}_j$ can be written as

$$\begin{aligned}\boldsymbol{M}_j &= \mathcal{A}_j(\boldsymbol{X}_j \otimes \boldsymbol{Z}_j^{-1})\mathcal{A}_j^T = \mathcal{A}_j \left((\boldsymbol{R}_{X_j}^T \boldsymbol{R}_{X_j}) \otimes (\boldsymbol{R}_{Z_j}^T \boldsymbol{R}_{Z_j})^{-1} \right) \mathcal{A}_j^T \\ &= \mathcal{A}_j \left((\boldsymbol{R}_{X_j}^T \otimes \boldsymbol{R}_{Z_j}^{-1})(\boldsymbol{R}_{X_j} \otimes \boldsymbol{R}_{Z_j}^{-T}) \right) \mathcal{A}_j^T = \boldsymbol{H}_j \boldsymbol{H}_j^T, \end{aligned} \tag{49}$$

where

$$\boldsymbol{H}_j = \mathcal{A}_j(\boldsymbol{R}_{X_j}^T \otimes \boldsymbol{R}_{Z_j}^{-1}) \in \mathbb{R}^{m \times n_j^2}.$$

Thus, $\boldsymbol{h}_l^T$, the l-th row of $\boldsymbol{H}_j$, can be computed as

$$\boldsymbol{h}_l = \operatorname{vec}\left(\boldsymbol{R}_{X_j} \; \boldsymbol{A}_{lj} \; \boldsymbol{R}_{Z_j}^{-1} \right).$$

Furthermore, we can rewrite $(\boldsymbol{M}_j)_{lh}$ in (19) as

$$\begin{aligned}(\boldsymbol{M}_j)_{lh} &= (\boldsymbol{X}_j \boldsymbol{A}_{lj} \boldsymbol{Z}_j^{-1}) \bullet \boldsymbol{A}_{hj} = \left((\boldsymbol{R}_{X_j}^T \boldsymbol{R}_{X_j}) \boldsymbol{A}_{lj} (\boldsymbol{R}_{Z_j}^{-1} \boldsymbol{R}_{Z_j}^{-T}) \right) \bullet \boldsymbol{A}_{hj} \\ &= \left(\boldsymbol{R}_{X_j}^T (\boldsymbol{R}_{X_j} \boldsymbol{A}_{lj} \boldsymbol{R}_{Z_j}^{-1}) \boldsymbol{R}_{Z_j}^{-T} \right) \bullet \boldsymbol{A}_{hj} = \left(\boldsymbol{R}_{X_j}^T \operatorname{mat}(\boldsymbol{h}_l) \, \boldsymbol{R}_{Z_j}^{-T} \right) \bullet \boldsymbol{A}_{hj}. \end{aligned}$$

Therefore, $\boldsymbol{H}_j$ can be obtained in the process of computing $\boldsymbol{M}_j$ with additional computation for the factor $\boldsymbol{R}_{X_j}$ of $\boldsymbol{X}_j$.

From (49), we can write the j-th update of $\widehat{\boldsymbol{M}}$ in step 3.(c) and 3.(f) in the algorithm as

$$\widehat{\boldsymbol{M}}^{(j)} = \widehat{\boldsymbol{M}}^{(j-1)} + \boldsymbol{M}_j = \widehat{\boldsymbol{M}}^{(j-1)} + \boldsymbol{H}_j \boldsymbol{H}_j^T.$$

If we already have the Cholesky factor $\boldsymbol{R}_{\widehat{M}}^{(j-1)}$ of $\widehat{\boldsymbol{M}}^{(j-1)}$, the Cholesky factor $\boldsymbol{R}_{\widehat{M}}^{(j)}$ of $\widehat{\boldsymbol{M}}^{(j)}$ can be computed by n_j^2 rank-1 Cholesky updates [10]. The term $\boldsymbol{R}_{Z_j}^{-1}$ in $\boldsymbol{H}_j$ can be computed column-by-column if space is at a premium, but since the matrices $\mathcal{A}_j$ require $m \sum_{j=1}^{p} n_j^2$ space, the additional $m \times n_j^2$

[3] Rank-1 modification of Cholesky factor is implemented by *"schud.f"* and *"dchud.f"* in LINPACK. See Gill et al. [10] and LINPACK documentation [8].

memory for $\boldsymbol{H}_j$ is not burdensome. Rank-1 update of a Cholesky factor requires $2m^2 + O(m)$ flops. Using the updated factor of $\widehat{\boldsymbol{M}}$, we can compute $\Delta \boldsymbol{y} = \widehat{\boldsymbol{M}}^{-1} \boldsymbol{g} = \boldsymbol{R}_{\widehat{M}}^{-1} (\boldsymbol{R}_{\widehat{M}}^{-T} \boldsymbol{g})$ in $2m^2$ flops. Since we do not need a very accurate $\Delta \boldsymbol{y}$ for determining constraint reduction, iterative refinement may not be necessary. Once we finish updating $\widehat{\boldsymbol{M}}$, the factor $\boldsymbol{R}_{\widehat{M}}$ can be reused as a preconditioner for an iterative method like *SYMMLQ* [30] to compute $\Delta \boldsymbol{y}$ to a high accuracy.

In summary, for each update of $\widehat{\boldsymbol{M}}$, the extra cost is Cholesky factorization of $\boldsymbol{X}_j$ ($n_j^3/3$ flops), update of Cholesky factor of $\widehat{\boldsymbol{M}}$ ($n_j^2(2m^2 + O(m))$ flops), and computation of $\Delta \boldsymbol{y} = \widehat{\boldsymbol{M}}^{-1} \boldsymbol{g}$ ($2m^2$ flops). The total, $\frac{1}{3} {n_j}^3 + 2m^2(n_j^2 + 1) + O(mn_j^2)$, is a reasonable cost for the constraint reduction decision, considering that it takes $(4m+1/3)n_j^3 + 2m^2 n_j^2$ to compute $\boldsymbol{M}_i$ by (19) and (20). An analysis based on memory access rather than floating point operations yields a similar conclusion. If $m^3/3 < (n_j^3/3 + 2m^2 n_j^2)$, then we can compute the Cholesky factor $\boldsymbol{R}_{\widehat{M}}$ of $\widehat{\boldsymbol{M}}$ explicitly with no Cholesky factorization of $\boldsymbol{X}_j$ and no updating of the factor $\boldsymbol{R}_{\widehat{M}}$. In that case, it costs $m^3/3 + 2m^2$.

3 Global Convergence of Algorithm SDP:Reduced

3.1 Primal and Dual Residuals and Closeness to Central Path

Convergence analysis for our constraint-reduced algorithm SDP:Reduced is based on a series of lemmas similar to those presented by Monteiro [24] and Potra and Sheng [34]. Note that in their lemmas the roles of $\boldsymbol{X}$ and $\boldsymbol{Z}$ in (16) are switched:

$$\operatorname{symm}\left(\boldsymbol{X}^{-1/2}(\boldsymbol{X}\Delta\boldsymbol{Z} + \Delta\boldsymbol{X}\boldsymbol{Z})\boldsymbol{X}^{1/2}\right) = \overline{\mu}\boldsymbol{I} - \boldsymbol{X}^{1/2}\boldsymbol{Z}\boldsymbol{X}^{1/2}. \tag{50}$$

We use (16) rather than (50) because, as Zhang [42] noted, (16) is computationally easy to solve. It also explains how the active and inactive blocks are involved in the Schur complement matrix computation as we described in Section 2.2. However, all of our results can be extended to the algorithm with (50) replacing (16).

For Algorithm SDP:Reduced, each component of the primal and dual residuals moves toward zero at each iteration, bringing us closer to feasibility.

Lemma 3.1. *In Algorithm SDP:Reduced,*

$$\mathbf{r}_d^+ = (1-\overline{\theta})\mathbf{r}_d \text{ and } \mathbf{r}_p^+ = (1-\overline{\theta})\mathbf{r}_p.$$

Proof. First, let us see how the dual residual changes. By (32P) and (32C), $\Delta \boldsymbol{z} = \boldsymbol{r}_d - \mathcal{A}^T \Delta \boldsymbol{y}$ and $\Delta \overline{\boldsymbol{z}} = -\mathcal{A}^T \Delta \overline{\boldsymbol{y}}$, so

$$\begin{aligned}\boldsymbol{r}_d^+ &= \boldsymbol{c} - \boldsymbol{z}^+ - \mathcal{A}^T \boldsymbol{y}^+ = \boldsymbol{c} - (\boldsymbol{z} + \overline{\theta}\Delta\boldsymbol{z} + \Delta\overline{\boldsymbol{z}}) - \mathcal{A}^T(\boldsymbol{y} + \overline{\theta}\Delta\boldsymbol{y} + \Delta\overline{\boldsymbol{y}}) \\ &= (\boldsymbol{c} - \boldsymbol{z} - \mathcal{A}^T\boldsymbol{y}) - \overline{\theta}(\Delta\boldsymbol{z} + \mathcal{A}^T\Delta\boldsymbol{y}) - (\Delta\overline{\boldsymbol{z}} + \mathcal{A}^T\Delta\overline{\boldsymbol{y}}) = \boldsymbol{r}_d - \overline{\theta}\ \boldsymbol{r}_d = (1-\overline{\theta})\boldsymbol{r}_d.\end{aligned}$$

Next, we consider the primal residual. By (31P) and (31C), $\mathcal{A}\Delta\boldsymbol{x} = \boldsymbol{r}_p$ and $\mathcal{A}\Delta\overline{\boldsymbol{x}} = 0$, so

$$\begin{aligned} \boldsymbol{r}_p^+ &= \boldsymbol{b} - \mathcal{A}(\boldsymbol{x}^+) = \boldsymbol{b} - \mathcal{A}(\boldsymbol{x} + \overline{\theta}\Delta\boldsymbol{x} + \Delta\overline{\boldsymbol{x}}) \\ &= \boldsymbol{r}_p - \overline{\theta}\mathcal{A}\Delta\boldsymbol{x} - \mathcal{A}\Delta\overline{\boldsymbol{x}} = \boldsymbol{r}_p - \overline{\theta}\boldsymbol{r}_p = (1-\overline{\theta})\boldsymbol{r}_p. \end{aligned}$$

□

Next we analyze how the iterate moves relative to the central path during the predictor and corrector steps. Assume that the current point $(\boldsymbol{X}, \boldsymbol{Z}) \in \mathcal{N}(\alpha, \tau)$. The initial point $(\boldsymbol{X}^0, \boldsymbol{y}^0, \boldsymbol{Z}^0)$ in the algorithm is perfectly placed on the central path, so this assumption is satisfied. With this assumption, we show in Lemma 3.2 that

$$(\overline{\boldsymbol{X}}, \overline{\boldsymbol{Z}}) \in \mathcal{N}(\beta, (1-\overline{\theta})\tau), \tag{51}$$

after the predictor step, and in Lemma 3.3 that

$$(\boldsymbol{X}^+, \boldsymbol{Z}^+) \in \mathcal{N}(\alpha, (1-\overline{\theta})\tau), \tag{52}$$

after the corrector step. Some proofs are omitted in this section but supplied in the Appendix, to keep the outline of the argument clear.

To show (51), recall that we know from (44) that

$$(\boldsymbol{X} + \theta\Delta\boldsymbol{X}, \boldsymbol{y} + \theta\Delta\boldsymbol{y}, \boldsymbol{Z} + \theta\Delta\boldsymbol{Z}) \in \mathcal{N}(\beta, (1-\theta)\tau),$$

for any $\theta \in [0, \breve{\theta}]$. We show in Lemma 3.2 that this relation holds for any $\theta \in [0, \widehat{\theta}]$, so we conclude that $\widehat{\theta} \le \breve{\theta}$ and therefore the predictor step exists.

Lemma 3.2. *(Similar to* [34, Lemma 2.5]*.) If* $(\mathbf{X}, \mathbf{Z}) \in \mathcal{N}(\alpha, \tau)$ *then* $\widehat{\theta} \le \breve{\theta}$*. In particular,*

1. *if* $\overline{\theta} < 1$, *then* $(\overline{\mathbf{X}}, \overline{\mathbf{Z}}) \in \mathcal{N}(\beta, (1-\overline{\theta})\tau)$, *so* $\overline{\mathbf{X}} \succ \mathbf{0}$ *and* $\overline{\mathbf{Z}} \succ \mathbf{0}$.
2. *if* $\overline{\theta} = 1$, *then* $\overline{\mathbf{X}}\,\overline{\mathbf{Z}} = \mathbf{0}$.

Proof. See Appendix. □

Lemma 3.3. *(Similar to* [34, Theorem 2.6]*.) Suppose that* $(\overline{\mathbf{X}}, \overline{\mathbf{Z}}) \in \mathcal{N}(\beta, (1-\overline{\theta})\tau)$ *in Algorithm SDP:Reduced. Let* $\tau^+ = (1-\overline{\theta})\tau$*. Then, after the corrector step,*

$$\begin{aligned} (\mathbf{X}^+, \mathbf{Z}^+) &\in \mathcal{N}(\alpha, \tau^+), \quad \text{and} \\ (\mathbf{Z}^+)^{1/2}\mathbf{X}^+(\mathbf{Z}^+)^{1/2} &\succeq (1-\alpha)\tau^+\mathbf{I}. \end{aligned}$$

Proof. See Appendix. □

Next, we quantify the bound on the duality gap $\mu = (\boldsymbol{X} \bullet \boldsymbol{Z})/n$. For the analysis, the following properties of the Frobenius norm and the trace of a matrix are useful. For a matrix $\boldsymbol{E} \in \mathcal{S}^n$,

$$|\mathrm{tr}\,(\boldsymbol{E})| = \left|\sum_{i=1}^{m} \lambda_i(\boldsymbol{E})\right| \leq \left|\sum_{i=1}^{m} \sigma_i(\boldsymbol{E})\right|,$$

where $\lambda_i(\boldsymbol{E})$ is the i-th eigenvalue and $\sigma_i(\boldsymbol{E})$ is the i-th singular value of $\boldsymbol{E}$. By the Cauchy-Schwarz inequality, for $\boldsymbol{E} \in \mathcal{S}^n$,

$$n\|\boldsymbol{E}\|_F^2 = n\sum_{i=1}^{n} \sigma_i^2(\boldsymbol{E}) \geq \left(\sum_{i=1}^{n} \sigma_i(\boldsymbol{E})\right)^2 \geq (\mathrm{tr}\,(\boldsymbol{E}))^2,$$

so

$$n\|\boldsymbol{E}\|_F^2 \geq (\mathrm{tr}\,(\boldsymbol{E}))^2 \tag{53}$$

Lemma 3.4. *If* $(\mathbf{X}, \mathbf{Z}) \in \mathcal{N}(\alpha, \tau)$, *then*

$$(1 - \frac{\alpha}{\sqrt{n}})\tau \leq \mu = \frac{1}{n}(\mathbf{X} \bullet \mathbf{Z}) \leq (1 + \frac{\alpha}{\sqrt{n}})\tau.$$

Proof. Since $(\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2} - \tau\boldsymbol{I})$ is symmetric, by (53),

$$\begin{aligned} n\|\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2} - \tau\boldsymbol{I}\|_F^2 &\geq \left(\mathrm{tr}\left(\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2} - \tau\boldsymbol{I}\right)\right)^2 \\ &= \left(\mathrm{tr}\left(\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2}\right) - n\tau\right)^2 \\ &= (\mathrm{tr}\,(\boldsymbol{X}\boldsymbol{Z}) - n\tau)^2 = (\boldsymbol{X} \bullet \boldsymbol{Z} - n\tau)^2 . \end{aligned}$$

Thus, since $(\boldsymbol{X}, \boldsymbol{Z}) \in \mathcal{N}(\alpha, \tau)$,

$$(\boldsymbol{X} \bullet \boldsymbol{Z} - n\tau)^2 \leq n\|\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2} - \tau\boldsymbol{I}\|_F^2 \leq n\alpha^2\tau^2,$$

i.e.,

$$\left(\frac{1}{n}(\boldsymbol{X} \bullet \boldsymbol{Z}) - \tau\right)^2 \leq \frac{1}{n}\alpha^2\tau^2,$$

and the rest of the proof is straightforward. □

The theorem below summarizes the convergence properties of SDP:Reduced.

Theorem 3.1. *At the k-th iteration of SDP:Reduced,* τ_k, $\mathbf{r}_p^k$, $\mathbf{r}_d^k$, *and* $(\mathbf{X}^k, \mathbf{Z}^k)$ *satisfy*

$$\tau_k = \psi_k\tau_0, \tag{54}$$

$$\mathbf{r}_p^k = \psi_k\mathbf{r}_p^0, \tag{55}$$

$$\mathbf{r}_d^k = \psi_k\mathbf{r}_d^0 \qquad (\mathbf{R}_d^k = \psi_k\mathbf{R}_d^0), \tag{56}$$

$$(\mathbf{X}^k, \mathbf{Z}^k) \in \mathcal{N}(\alpha, \tau_k), \tag{57}$$

$$(1 - \frac{\alpha}{\sqrt{n}})\tau_k \leq \mu_k = \frac{1}{n}(\mathbf{X}^k \bullet \mathbf{Z}^k) \leq (1 + \frac{\alpha}{\sqrt{n}})\tau_k. \tag{58}$$

where

$$\psi_k := \prod_{i=1}^{k}(1-\overline{\theta}_i).$$

Proof. From Lemmas 3.1 – Lemma 3.4, we can obtain the results. □

In order to prove the convergence of $\boldsymbol{r}_p^k$, $\boldsymbol{r}_d^k$, and μ_k to zero, all that remains is to show that the step lengths $\overline{\theta}_i$ are bounded away from zero.

3.2 Lower Bound on Step Length

In this section, we omit the k in ψ_k, $\boldsymbol{r}_p^k$, and $\boldsymbol{r}_d^k$ whenever it is evident in the context, and let $(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z})$ denote the k-th iterate of our algorithm.

To show that the step lengths are bounded away from zero (Theorem 3.2), we need to show that quantities related to $\Delta\boldsymbol{X}$, $\Delta\boldsymbol{Z}$, and δ are bounded (Lemmas 3.7 and 3.8). Two preliminary lemmas lead us to these bounds.

Lemma 3.5. *(Similar to* [34, Lemma 3.2]*.) For an initial point* $(\mathbf{X}^0, \mathbf{y}^0, \mathbf{Z}^0)$ *and an optimal solution* $(\mathbf{X}^*, \mathbf{y}^*, \mathbf{Z}^*) \in \mathcal{F}^*$*, define*

$$\zeta = \frac{\mathbf{X}^0 \bullet \mathbf{Z}^* + \mathbf{X}^* \bullet \mathbf{Z}^0}{\mathbf{X}^0 \bullet \mathbf{Z}^0}.$$

Then

$$\mathbf{X} \bullet \mathbf{Z}^0 + \mathbf{X}^0 \bullet \mathbf{Z} \leq n\tau_0 \left(2 + \zeta + \frac{\alpha}{\sqrt{n}}\right), \quad (59)$$

$$\mathbf{X} \bullet \mathbf{Z}^* + \mathbf{X}^* \bullet \mathbf{Z} \leq \left(\frac{1+\alpha+\psi}{1-\psi} + \zeta\right) n\tau. \quad (60)$$

Proof. Let us define

$$\begin{aligned} \boldsymbol{X}' &= \boldsymbol{X} - \psi\boldsymbol{X}^0 - (1-\psi)\boldsymbol{X}^*, \\ \boldsymbol{y}' &= \boldsymbol{y} - \psi\boldsymbol{y}^0 - (1-\psi)\boldsymbol{y}^*, \\ \boldsymbol{Z}' &= \boldsymbol{Z} - \psi\boldsymbol{Z}^0 - (1-\psi)\boldsymbol{Z}^*. \end{aligned}$$

By (11), (55) and the primal feasibility of $\boldsymbol{X}^*$,

$$\begin{aligned} \boldsymbol{A}_i \bullet \boldsymbol{X} &= b_i - r_{pi}, \\ \psi\boldsymbol{A}_i \bullet \boldsymbol{X}^0 &= \psi(b_i - r_{pi}^0) = \psi b_i - r_{pi}, \\ (1-\psi)\boldsymbol{A}_i \bullet \boldsymbol{X}^* &= (1-\psi)b_i, \end{aligned}$$

for $i = 1, \ldots, m$, and by (12), (56), and the dual feasibility of $(\boldsymbol{y}^*, \boldsymbol{Z}^*)$

$$\begin{aligned}
\sum_{i=1}^{m} y_i \boldsymbol{A}_i + \boldsymbol{Z} &= \boldsymbol{C} - \boldsymbol{R}_d, \\
\psi(\sum_{i=1}^{m} y_i^0 \boldsymbol{A}_i + \boldsymbol{Z}^0) &= \psi(\boldsymbol{C} - \boldsymbol{R}_d^0) = \psi\boldsymbol{C} - \boldsymbol{R}_d, \\
(1-\psi)\left(\sum_{i=1}^{m} y_i^* \boldsymbol{A}_i + \boldsymbol{Z}^*\right) &= (1-\psi)\boldsymbol{C}.
\end{aligned}$$

Thus, $(\boldsymbol{X}', \boldsymbol{y}', \boldsymbol{Z}')$ satisfies

$$\begin{aligned}
\boldsymbol{A}_i \bullet \boldsymbol{X}' = 0 \text{ for } i = 1, \ldots, m, \\
\sum_{i=1}^{m} y_i' \boldsymbol{A}_i + \boldsymbol{Z}' = 0.
\end{aligned}$$

Therefore, $\boldsymbol{X}' \bullet \boldsymbol{Z}' = -\sum_{i=1}^{m} y_i'(\boldsymbol{A}_i \bullet \boldsymbol{X}) = 0$, so

$$[\boldsymbol{X} - \psi\boldsymbol{X}^0 - (1-\psi)\boldsymbol{X}^*] \bullet [\boldsymbol{Z} - \psi\boldsymbol{Z}^0 - (1-\psi)\boldsymbol{Z}^*] = 0.$$

By expanding this equation using $\boldsymbol{X}^* \bullet \boldsymbol{Z}^* = 0$, we obtain

$$\begin{aligned}
\psi(\boldsymbol{X} \bullet \boldsymbol{Z}^0 + \boldsymbol{X}^0 \bullet \boldsymbol{Z}) &+ (1-\psi)(\boldsymbol{X} \bullet \boldsymbol{Z}^* + \boldsymbol{X}^* \bullet \boldsymbol{Z}) \\
&= \boldsymbol{X} \bullet \boldsymbol{Z} + \psi^2 \boldsymbol{X}^0 \bullet \boldsymbol{Z}^0 + \psi(1-\psi)(\boldsymbol{X}^0 \bullet \boldsymbol{Z}^* + \boldsymbol{X}^* \bullet \boldsymbol{Z}^0).
\end{aligned}$$

Since $\boldsymbol{X} \in \mathcal{S}_+^n, \boldsymbol{Z} \in \mathcal{S}_+^n, \boldsymbol{X}^* \in \mathcal{S}_+^n, \boldsymbol{Z}^* \in \mathcal{S}_+^n$ and $\psi \in [0,1]$, this equation gives us two inequalities:

$$\begin{aligned}
\psi(\boldsymbol{X} \bullet \boldsymbol{Z}^0 + \boldsymbol{X}^0 \bullet \boldsymbol{Z}) &\leq \boldsymbol{X} \bullet \boldsymbol{Z} + \psi^2 \boldsymbol{X}^0 \bullet \boldsymbol{Z}^0 + \psi(1-\psi)(\boldsymbol{X}^0 \bullet \boldsymbol{Z}^* + \boldsymbol{X}^* \bullet \boldsymbol{Z}^0), \\
(1-\psi)(\boldsymbol{X} \bullet \boldsymbol{Z}^* + \boldsymbol{X}^* \bullet \boldsymbol{Z}) &\leq \boldsymbol{X} \bullet \boldsymbol{Z} + \psi^2 \boldsymbol{X}^0 \bullet \boldsymbol{Z}^0 + \psi(1-\psi)(\boldsymbol{X}^0 \bullet \boldsymbol{Z}^* + \boldsymbol{X}^* \bullet \boldsymbol{Z}^0).
\end{aligned}$$

Because $\boldsymbol{X}^0 \bullet \boldsymbol{Z}^0 = n\tau_0$, $\boldsymbol{X} \bullet \boldsymbol{Z} \leq (1 + \alpha/\sqrt{n})\psi n\tau_0$ by (58), and $\psi(1-\psi) \leq \psi$, we can bound the right hand sides of the inequalities above,

$$\begin{aligned}
\boldsymbol{X} \bullet \boldsymbol{Z} + \psi^2 \boldsymbol{X}^0 \bullet \boldsymbol{Z}^0 &+ \psi(1-\psi)(\boldsymbol{X}^0 \bullet \boldsymbol{Z}^* + \boldsymbol{X}^* \bullet \boldsymbol{Z}^0) \\
&\leq (1 + \alpha/\sqrt{n})\psi n\tau_0 + \psi^2 n\tau_0 + \psi(1-\psi)\zeta n\tau_0 \\
&= \psi\left(1 + \alpha/\sqrt{n} + \psi + (1-\psi)\zeta\right) n\tau_0.
\end{aligned}$$

With this upper bound, we can rewrite the inequalities as

$$\begin{aligned}
\psi(\boldsymbol{X} \bullet \boldsymbol{Z}^0 + \boldsymbol{X}^0 \bullet \boldsymbol{Z}) &\leq \psi\left(1 + \alpha/\sqrt{n} + \psi + (1-\psi)\zeta\right) n\tau_0, \\
(1-\psi)(\boldsymbol{X} \bullet \boldsymbol{Z}^* + \boldsymbol{X}^* \bullet \boldsymbol{Z}) &\leq \psi\left(1 + \alpha/\sqrt{n} + \psi + (1-\psi)\zeta\right) n\tau_0.
\end{aligned}$$

Therefore, by using $\psi \in [0,1]$, $n \geq 1$, and $\tau = \psi\tau_0$, we can derive (59) and (60).

$\square$

Lemma 3.6. *(Similar to* [34, Corollary 3.3]*.)*

$$\|\mathbf{X}^{1/2}(\mathbf{Z}^0)^{1/2}\|_F \le (n\tau_0)^{1/2}\left(2+\zeta+\frac{\alpha}{\sqrt{n}}\right)^{1/2}, \tag{61}$$

$$\|\mathbf{Z}^{1/2}(\mathbf{X}^0)^{1/2}\|_F \le (n\tau_0)^{1/2}\left(2+\zeta+\frac{\alpha}{\sqrt{n}}\right)^{1/2}, \tag{62}$$

$$\|\mathbf{X}^{1/2}\|_F \le \|(\mathbf{Z}^0)^{-1/2}\|\,(n\tau_0)^{1/2}\ \left(2+\zeta+\frac{\alpha}{\sqrt{n}}\right)^{1/2}, \tag{63}$$

$$\|\mathbf{Z}^{1/2}\|_F \le \|(\mathbf{X}^0)^{-1/2}\|\,(n\tau_0)^{1/2}\ \left(2+\zeta+\frac{\alpha}{\sqrt{n}}\right)^{1/2}, \tag{64}$$

$$\|\mathbf{X}^{1/2}\mathbf{Z}^{1/2}\|^2 = \|\mathbf{Z}^{1/2}\mathbf{X}\mathbf{Z}^{1/2}\| \le (1+\alpha)\tau, \tag{65}$$

$$\|\mathbf{X}^{-1/2}\mathbf{Z}^{-1/2}\|^2 = \|\mathbf{Z}^{-1/2}\mathbf{X}^{-1}\mathbf{Z}^{-1/2}\| \le \frac{1}{(1-\alpha)\tau}. \tag{66}$$

Proof. See Appendix. ☐

Recalling the definition of δ, δ_x, and δ_z in (36)–(38), δ is bounded by

$$\begin{aligned}\delta &= \frac{1}{\tau}\|\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}\Delta\boldsymbol{Z}\boldsymbol{Z}^{-1/2}\|_F\\ &\le \frac{1}{\tau}\|\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}\boldsymbol{Z}^{1/2}\|_F\|\boldsymbol{Z}^{-1/2}\Delta\boldsymbol{Z}\boldsymbol{Z}^{-1/2}\|_F = \frac{1}{\tau^2}\delta_x\delta_z.\end{aligned} \tag{67}$$

Lemma 3.7. *(Similar to* [34, Lemma 3.4]*.) For* $(\breve{\mathbf{X}}, \breve{\mathbf{y}}, \breve{\mathbf{Z}}) \in \mathcal{F}$, *denote*

$$\begin{aligned}\mathbf{T} &:= \psi\left[\mathbf{Z}^{1/2}(\mathbf{X}^0-\breve{\mathbf{X}})\mathbf{Z}^{1/2} + \ \mathrm{symm}\left(\mathbf{Z}^{1/2}\mathbf{X}(\mathbf{Z}^0-\breve{\mathbf{Z}})\mathbf{Z}^{-1/2}\right)\right] - \mathbf{Z}^{1/2}(\mathbf{X}+\Delta\mathbf{X}_\epsilon)\mathbf{Z}^{1/2},\\ \mathbf{T}_x &:= \psi\mathbf{Z}^{1/2}(\mathbf{X}^0-\breve{\mathbf{X}})\mathbf{Z}^{1/2},\\ \mathbf{T}_z &:= \psi\mathbf{Z}^{-1/2}(\mathbf{Z}^0-\breve{\mathbf{Z}})\mathbf{Z}^{-1/2}.\end{aligned}$$

Then

$$\begin{aligned}\delta_x &= \|\mathbf{Z}^{1/2}\Delta\mathbf{X}\mathbf{Z}^{1/2}\|_F \le \|\mathbf{T}_x\|_F + \frac{\|\mathbf{T}\|_F}{1-\alpha},\\ \delta_z &= \tau\|\mathbf{Z}^{-1/2}\Delta\mathbf{Z}\mathbf{Z}^{-1/2}\|_F \le \tau\|\mathbf{T}_z\|_F + \frac{\|\mathbf{T}\|_F}{1-\alpha}.\end{aligned}$$

Proof. See Appendix. ☐

Lemma 3.8. *For any given* $(\breve{\mathbf{X}}, \breve{\mathbf{y}}, \breve{\mathbf{Z}}) \in \mathcal{F}$, *we have*

$$\begin{aligned}\delta_x &\le \left(\frac{1-\alpha}{1-\alpha-q}\right)\left(C_x+\frac{C_0}{1-\alpha}\right)\tau,\\ \delta_z &\le \left(\frac{1-\alpha}{1-\alpha-q}\right)\left(C_z+\frac{C_0}{1-\alpha}\right)\tau,\\ \delta &\le \left(\frac{1-\alpha}{1-\alpha-q}\right)^2\left(C_z+\frac{C_0}{1-\alpha}\right)^2,\end{aligned}$$

where

$$C_x := nd_0(2+\zeta+\alpha/\sqrt{n}), \tag{68}$$

$$C_z := \frac{nd_0(2+\zeta+\alpha/\sqrt{n})}{1-\alpha}, \tag{69}$$

$$C_0 := \frac{2nd_0(2+\zeta+\alpha/\sqrt{n})}{1-\alpha} + \sqrt{n}(1+\alpha), \text{ and} \tag{70}$$

$$d_0 := \max\left(\|(\mathbf{X}^0)^{-1/2}(\mathbf{X}^0-\breve{\mathbf{X}})(\mathbf{X}^0)^{-1/2}\|_F, \|(\mathbf{Z}^0)^{-1/2}(\mathbf{Z}^0-\breve{\mathbf{Z}})(\mathbf{Z}^0)^{-1/2}\|_F\right).$$

Proof. See Appendix. □

We can now establish global convergence of our algorithm.

Theorem 3.2. *Algorithm SDP:Reduced is globally convergent, with the SDP optimality conditions (3)–(6) satisfied in the limit as $k \to \infty$.*

Proof. Since δ_x is bounded by Lemma 3.8, so is δ_ϵ by (45). Lemma 3.8 bounds δ. Therefore, $\widehat{\theta}$ defined by (43) is bounded away from 0. Thus the step length $\overline{\theta} \in [\widehat{\theta}, \breve{\theta}]$ is bounded away from 0. The result follows from Theorem 3.1. □

Next we prove that Algorithm SDP:Reduced converges in $O(n \ln(\epsilon_0/\epsilon))$ iterations, the same as the (unreduced) algorithm of [34],
where $\epsilon_0 = \max(\boldsymbol{X}_0 \bullet \boldsymbol{Z}_0, \|\boldsymbol{r}_p^0\|, \|\boldsymbol{r}_d^0\|)$ and ϵ is the required tolerance on the resulting residuals.

Lemma 3.9. *Suppose that $\mathbf{X}^0 = \mathbf{Z}^0 = \rho\mathbf{I}$, $\|\mathbf{X}^*\| \le \rho$ and $\|\mathbf{Z}^*\| \le \rho$ for $(\mathbf{X}^*, \mathbf{y}^*, \mathbf{Z}^*) \in \mathcal{F}^*$ and $\rho > 0$. Then the predictor step length $\overline{\theta}_k \in [\widehat{\theta}_k, \breve{\theta}_k]$ satisfies*

$$\overline{\theta}_k \ge \frac{1}{wn},$$

where

$$w = 1 + \frac{hq}{(\beta-\alpha)} + \sqrt{\frac{h(h+3.5)}{\beta-\alpha}},$$

and $h = 13/(0.5-q)$.

Proof. See Appendix. □

Note that if $q = 0$, then constraint reduction is not performed. In that case,

$$w = 1 + \sqrt{(26 \times 29.5)/(\beta-\alpha)} \le 1 + (29/\sqrt{\beta-\alpha}),$$

and $\overline{\theta}$ has the same lower bound as the unreduced algorithm by [34, Theorem 3.8].

Theorem 3.3. *For a given tolerance ϵ on $\max(\mathbf{X}^k \bullet \mathbf{Z}^k, \|\mathbf{r}_p^k\|, \|\mathbf{r}_d^k\|)$, Algorithm SDP:Reduced converges in $O(n\ln(\epsilon_0/\epsilon))$ iterations where $\epsilon_0 = \max(n\tau_0, \|\mathbf{r}_p^0\|, \|\mathbf{r}_d^0\|)$.*

Proof. By Theorem 3.1, we know

$$\begin{aligned} \epsilon_k &\le \max((1+\alpha/\sqrt{n})n\tau_k, \|\boldsymbol{r}_p^k\|, \|\boldsymbol{r}_d^k\|) \le \psi_k \max((1+\alpha/\sqrt{n})n\tau_0, \|\boldsymbol{r}_p^0\|, \|\boldsymbol{r}_d^0\|) \\ &\le \psi_k(1+\alpha/\sqrt{n})\epsilon_0. \end{aligned}$$

On the other hand, by the definition of ψ_k and Lemma 3.9,

$$\psi_k = \prod_{i=1}^{k}(1-\overline{\theta}_i) \le (1-\frac{1}{wn})^k.$$

Thus, if

$$\left(1-\frac{1}{wn}\right)^K (1+\alpha/\sqrt{n})\epsilon_0 \le \epsilon$$

after K iterations, then $\epsilon_K \le \epsilon$. By taking ln on both sides,

$$K \ln\left(1-\frac{1}{wn}\right) + \ln\left[(1+\alpha/\sqrt{n})\epsilon_0\right] \le \ln\epsilon$$

if and only if

$$K \ln\left(1-\frac{1}{wn}\right) \le \ln\epsilon - \ln\left[(1+\alpha/\sqrt{n})\epsilon_0\right] = \ln(\epsilon/\epsilon_0) - \ln\left[(1+\alpha/\sqrt{n})\right] \le \ln(\epsilon/\epsilon_0)$$

Hence, $\epsilon_K \le \epsilon$ if

$$K \ge \frac{\ln(\epsilon_0/\epsilon)}{-\ln\left(1-\frac{1}{wn}\right)}.$$

By the fact

$$\frac{-1}{\ln\left(1-\frac{1}{wn}\right)} \to wn, \quad \text{as } n \to \infty,$$

$K = O(n \ln(\epsilon_0/\epsilon))$. □

4 Conclusions

We proposed an infeasible *primal-dual predictor-corrector* interior point method with adaptive constraint reduction for block diagonal SDP problems. By adaptively selecting appropriate *inactive* constraint blocks, we retain global convergence and polynomial complexity, $O(n \ln(\epsilon_0/\epsilon))$. The polynomial complexity result is the first such result for *primal-dual* constraint reduced interior-point-methods for any problem class, and includes LP, QP, QCQP, and SOCP as special cases.

Our algorithm computes the Schur complement matrix twice for each iteration. Since most of the practical implementations reuse the Schur complement matrix in the corrector step, this is a disadvantage, but future research may remove this restriction.

Kojima, Shida and Shindoh [21] showed that an algorithm similar to that of Potra and Sheng [34] has superlinear local convergence if the generated sequence converges tangentially to the central path. They noted that tangential convergence can be achieved by repeating the corrector step of Potra and Sheng until $(\boldsymbol{X}^+, \boldsymbol{Z}^+)$ moves into $\mathcal{N}(g(\tau_k), \tau)$ for a given $g(\tau_k)$ such that $g(\tau_k) \to 0$ as $k \to \infty$. Similar algorithms using the AHO direction [22] or a carefully bounded scaled direction [32, 17] have been proven to have local superlinear convergence under simpler assumptions and without repeating the corrector. As future work, this approach might lead to a constraint-reduced algorithm with superlinear convergence.

A Appendix: Proofs

In the proofs, we frequently use properties of the Frobenius norm. For a matrix $\boldsymbol{E} \in \mathbb{R}^{n \times n}$,

$$\begin{aligned} \|\boldsymbol{E}\|_F^2 &= \sum_{i=1}^{n} \sigma_i^2(\boldsymbol{E}), \\ \|\boldsymbol{E}\| &= \sigma_{\max}(\boldsymbol{E}), \end{aligned}$$

where $\sigma_i(\boldsymbol{E})$ is the i-th singular value of $\boldsymbol{E}$, so

$$\|\boldsymbol{E}\|_F \leq \sqrt{n}\, \|\boldsymbol{E}\|,$$

If $\boldsymbol{E} \in \mathcal{S}^n$, then

$$|\lambda_i(\boldsymbol{E})| \leq \sigma_{\max}(\boldsymbol{E}) = \sqrt{\sigma_{\max}^2(\boldsymbol{E})} \leq \sqrt{\sum_{i=1}^{n} (\sigma_i^2(\boldsymbol{E}))} = \|\boldsymbol{E}\|_F,$$

so

$$-\|\boldsymbol{E}\|_F \leq \lambda_i(\boldsymbol{E}) \leq \|\boldsymbol{E}\|_F. \tag{71}$$

Lemma A.1. *If* $\mathbf{M} \in \mathbb{R}^{p \times p}$ *is nonsingular and* $\mathbf{E} \in \mathbb{R}^{p \times p}$ *has only real eigenvalues, then*

$$\lambda_{\max}(\mathbf{E}) \leq \lambda_{\max}\left(\operatorname{symm}\left(\mathbf{MEM}^{-1}\right)\right), \tag{72}$$

$$\lambda_{\min}(\mathbf{E}) \geq \lambda_{\min}\left(\operatorname{symm}\left(\mathbf{MEM}^{-1}\right)\right). \tag{73}$$

If $\mathbf{E} \in \mathcal{S}^p$, *then*

$$\|\mathbf{E}\|_F \leq \| \operatorname{symm}\left(\mathbf{MEM}^{-1}\right) \|_F. \tag{74}$$

Proof. See [24, Lemma 3.3] and [34, Lemma 2.2]. □

Next, we prove that the predictor step stays in a neighborhood of the central path.

Proof. of Lemma 3.2.
Let $\boldsymbol{X}(\theta) = \boldsymbol{X} + \theta\Delta\boldsymbol{X}$ and $\boldsymbol{Z}(\theta) = \boldsymbol{Z} + \theta\Delta\boldsymbol{Z}$, then

$$\begin{aligned}\boldsymbol{X}(\theta)\boldsymbol{Z}(\theta) - (1-\theta)\tau\boldsymbol{I} &= (\boldsymbol{X}+\theta\Delta\boldsymbol{X})(\boldsymbol{Z}+\theta\Delta\boldsymbol{Z}) - (1-\theta)\tau\boldsymbol{I} \\ &= (1-\theta)(\boldsymbol{X}\boldsymbol{Z}-\tau\boldsymbol{I}) + \theta(\boldsymbol{X}\boldsymbol{Z} + \boldsymbol{X}\Delta\boldsymbol{Z} + \Delta\boldsymbol{X}\boldsymbol{Z}) \\ &\quad +\theta^2\Delta\boldsymbol{X}\Delta\boldsymbol{Z}.\end{aligned}$$

Define

$$\begin{aligned}\boldsymbol{P}(\theta) &= \boldsymbol{Z}^{1/2}(\boldsymbol{X}(\theta)\boldsymbol{Z}(\theta) - (1-\theta)\tau\boldsymbol{I})\boldsymbol{Z}^{-1/2} \\ &= \boldsymbol{Z}^{1/2}((\boldsymbol{X}+\theta\Delta\boldsymbol{X})(\boldsymbol{Z}+\theta\Delta\boldsymbol{Z}) - (1-\theta)\tau\boldsymbol{I})\boldsymbol{Z}^{-1/2} \\ &= \boldsymbol{Z}^{1/2}(\boldsymbol{X}\boldsymbol{Z} + \theta(\Delta\boldsymbol{X}\boldsymbol{Z} + \boldsymbol{X}\Delta\boldsymbol{Z}) + \theta^2\Delta\boldsymbol{X}\Delta\boldsymbol{Z} - (1-\theta)\tau\boldsymbol{I})\boldsymbol{Z}^{-1/2} \\ &= (1-\theta)(\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2} - \tau\boldsymbol{I}) + \theta^2\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}\Delta\boldsymbol{Z}\boldsymbol{Z}^{-1/2} \\ &\quad +\theta\left[\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2} + \boldsymbol{Z}^{1/2}(\boldsymbol{X}\Delta\boldsymbol{Z} + \Delta\boldsymbol{X}\boldsymbol{Z})\boldsymbol{Z}^{-1/2}\right].\end{aligned}$$

Then, by (33P),

$$\begin{aligned}\mathrm{symm}\,(\boldsymbol{P}(\theta)) &= (1-\theta)(\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2} - \tau\boldsymbol{I}) + \theta^2\ \mathrm{symm}\left(\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}\Delta\boldsymbol{Z}\boldsymbol{Z}^{-1/2}\right) \\ &\quad + \theta\left[\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2} + \ \mathrm{symm}\left(\boldsymbol{Z}^{1/2}(\boldsymbol{X}\Delta\boldsymbol{Z} + \Delta\boldsymbol{X}\boldsymbol{Z})\boldsymbol{Z}^{-1/2}\right)\right] \\ &= (1-\theta)(\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2} - \tau\boldsymbol{I}) \\ &\quad + \theta^2\ \mathrm{symm}\left(\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}\Delta\boldsymbol{Z}\boldsymbol{Z}^{-1/2}\right) - \theta(\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}_\epsilon\ \boldsymbol{Z}^{1/2}).\end{aligned}$$

Thus, since $(\boldsymbol{X}, \boldsymbol{Z}) \in \mathcal{N}(\alpha, \tau)$, and using (36), (39), and (74), we have

$$\begin{aligned}\|\ \mathrm{symm}\,(\boldsymbol{P}(\theta))\,\|_F &\leq (1-\theta)\|\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2} - \tau\boldsymbol{I}\|_F + \theta^2\|\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}\Delta\boldsymbol{Z}\boldsymbol{Z}^{-1/2}\|_F \\ &\quad +\theta\|\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}_\epsilon\boldsymbol{Z}^{1/2}\|_F \qquad (75) \\ &= \alpha\tau(1-\theta) + \theta^2\delta\tau + \theta\delta_\epsilon\tau \\ &= \tau\left(\delta\theta^2 + (\delta_\epsilon - \alpha + \beta)\theta + (\alpha-\beta)\right) + \beta(1-\theta)\tau \\ &= \delta\tau(\theta-\theta_1)(\theta-\theta_2) + \beta(1-\theta)\tau, \qquad (76)\end{aligned}$$

where

$$\begin{aligned}\theta_1 &= \frac{(\alpha-\beta-\delta_\epsilon) + \sqrt{(\alpha-\beta-\delta_\epsilon)^2 + 4\delta(\beta-\alpha)}}{2\delta}, \\ \theta_2 &= \frac{(\alpha-\beta-\delta_\epsilon) - \sqrt{(\alpha-\beta-\delta_\epsilon)^2 + 4\delta(\beta-\alpha)}}{2\delta}.\end{aligned}$$

Since $\widehat{\theta} = \theta_1$ by definition (43) of $\widehat{\theta}$ and $\theta_2 < 0$, the first term in (76) becomes negative when $0 \leq \theta \leq \widehat{\theta}$, so

$$\|\ \mathrm{symm}\,(\boldsymbol{P}(\theta))\,\|_F \leq \beta(1-\theta)\tau, \quad \forall\theta \in [0, \widehat{\theta}].$$

By (74), with $\boldsymbol{M} = \boldsymbol{Z}^{1/2}$ and $\boldsymbol{E} = \boldsymbol{X}(\theta)\boldsymbol{Z}(\theta) - (1-\theta)\tau\boldsymbol{I}$,

$$\|\boldsymbol{X}(\theta)\boldsymbol{Z}(\theta) - (1-\theta)\tau\boldsymbol{I}\|_F \leq \beta(1-\theta)\tau, \ \forall\theta \in [0, \widehat{\theta}]. \tag{77}$$

Note that this implies that $\boldsymbol{X}(1)\boldsymbol{Z}(1) = \mathbf{0}$ when $\widehat{\theta} = 1$. From this result, if $(\boldsymbol{Z}(\theta))^{-1/2}$ exists for $\forall\theta \in [0, \widehat{\theta}]$, then, since $\|.\|_F$ is invariant under similarity transformation, (77) implies

$$\|\boldsymbol{Z}(\theta)^{1/2}\boldsymbol{X}(\theta)\boldsymbol{Z}(\theta)^{1/2} - (1-\theta)\tau\boldsymbol{I}\|_F \leq \beta(1-\theta)\tau, \ \forall\theta \in [0, \widehat{\theta}]. \tag{78}$$

To conclude, we show $\boldsymbol{X}(\theta) \succ 0$ and $\boldsymbol{Z}(\theta) \succ 0$ for $\forall\theta \in [0, \widehat{\theta}]$ when $\widehat{\theta} < 1$. (By continuity, (78) then holds for $\widehat{\theta} = 1$, too.) Otherwise, there must exist $\theta' \in [0, \widehat{\theta}]$ such that $\boldsymbol{X}(\theta')\boldsymbol{Z}(\theta')$ is singular, which implies that

$$\lambda_{\min}(\boldsymbol{X}(\theta')\boldsymbol{Z}(\theta') - (1-\theta')\tau\boldsymbol{I}) \leq -(1-\theta')\tau. \tag{79}$$

However, by (73) with $\boldsymbol{M} = \boldsymbol{Z}^{1/2}$ and $\boldsymbol{E} = \boldsymbol{X}(\theta')\boldsymbol{Z}(\theta') - (1-\theta')\tau\boldsymbol{I}$, and by (71),

$$\begin{aligned}\lambda_{\min}(\boldsymbol{X}(\theta')\boldsymbol{Z}(\theta') - (1-\theta')\tau\boldsymbol{I}) &\geq \lambda_{\min}\left(\,\mathrm{symm}\,(\boldsymbol{P}(\theta')\,)\,\right) \\ &\geq -\|\,\mathrm{symm}\,(\boldsymbol{P}(\theta'))\,\|_F \geq -\beta(1-\theta')\tau,\end{aligned}$$

which contradicts (79) since $\beta \in (0,1)$. Hence, $\boldsymbol{X}(\theta) \succ 0$ and $\boldsymbol{Z}(\theta) \succ 0$ for $\forall\theta \in [0, \widehat{\theta}]$. □

To prepare for the proof that the corrector step stays in a neighborhood of the central path, we need two technical lemmas.

Lemma A.2. *(Similar to* [24, Lemma 4.4 in p.671]*.) For* $(\mathbf{X}', \mathbf{Z}') \in \mathcal{N}(\gamma, \tau')$ *and* $(\Delta\mathbf{X}', \Delta\mathbf{y}', \Delta\mathbf{Z}')$ *such that*

$$\mathbf{A}_i \bullet \Delta\mathbf{X}' = 0 \quad \textit{for } i \;=\; 1, \ldots, m, \tag{80}$$

$$\sum_{i=1}^{m} \Delta y_i' \mathbf{A}_i + \Delta\mathbf{Z}' \;=\; \mathbf{0}, \tag{81}$$

define

$$\begin{aligned}\mathbf{H} &= \mathrm{symm}\left(\mathbf{Z}'^{1/2}(\mathbf{X}'\Delta\mathbf{Z}' + \Delta\mathbf{X}'\mathbf{Z}')\mathbf{Z}'^{-1/2}\right), \\ \delta_x' &= \|\mathbf{Z}'^{1/2}\Delta\mathbf{X}'\mathbf{Z}'^{1/2}\|_F, \\ \delta_z' &= \tau'\|\mathbf{Z}'^{-1/2}\Delta\mathbf{Z}'\mathbf{Z}'^{-1/2}\|_F.\end{aligned}$$

Then

$$\delta_x'\delta_z' \;\leq\; \frac{1}{2}(\delta_x'^2 + \delta_z'^2) \leq \frac{\|\mathbf{H}\|_F^2}{2(1-\gamma)^2}, \tag{82}$$

$$\delta_x' \;\leq\; \frac{\|\mathbf{H}\|_F}{1-\gamma}, \tag{83}$$

$$\delta_z' \;\leq\; \frac{\|\mathbf{H}\|_F}{1-\gamma}. \tag{84}$$

Proof. See Monteiro [24, Lemma 4.4], in which the roles of $\boldsymbol{X}$ and $\boldsymbol{Z}$ in $\boldsymbol{H}$ are switched. □

Lemma A.3. *Under Condition 2.2,* $\overline{\delta}_\epsilon < (1-\overline{\theta})(1-2\beta)$.

Proof. Recall that

$$s = \beta^2 - \beta + 1, \qquad t = 2\alpha(1-\beta)^2 - \beta^2,$$

by their definitions in Condition 2.2. By Condition 2.2 , it suffices to show

$$\sqrt{s^2+t} - s < 1 - 2\beta,$$

or equivalently, since $0 < \beta < 1/2$ and $s > 0$,

$$(s + (1-2\beta))^2 - (\sqrt{s^2+t})^2 > 0.$$

By (41), we have

$$\begin{aligned}
(s+(1-2\beta))^2 - (\sqrt{s^2+t}\,)^2 &= (1-2\beta)^2 + 2s(1-2\beta) - t \\
&= (1-2\beta)^2 + 2(\beta^2-\beta+1)(1-2\beta) - 2\alpha(1-\beta)^2 + \beta^2 \\
&> (1-2\beta)^2 + 2(\beta^2-\beta+1)(1-2\beta) - 2\beta(1-\beta)^2 + \beta^2 \\
&= -6\beta^3 + 15\beta^2 - 12\beta + 3 \\
&= 3(1-2\beta)(\beta-1)^2 > 0, \quad \forall \beta \in (0, 1/2).
\end{aligned}$$

So,

$$\overline{\delta}_\epsilon < (1-\overline{\theta})(1-2\beta).$$

□

We are now ready to establish Lemma 3.3.

Proof. of Lemma 3.3.

$$\begin{aligned}
\boldsymbol{X}^+\boldsymbol{Z}^+ - (1-\overline{\theta})\tau\boldsymbol{I} &= (\overline{\boldsymbol{X}} + \Delta\overline{\boldsymbol{X}})(\overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{Z}}) - (1-\overline{\theta})\tau\boldsymbol{I} \\
&= \overline{\boldsymbol{X}}\,\overline{\boldsymbol{Z}} - (1-\overline{\theta})\tau\boldsymbol{I} + \overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{X}}\,\overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}}.
\end{aligned}$$

Since $\overline{\theta} < 1$ due to step 3.(e) in Algorithm SDP:Reduced, we know that $\overline{\boldsymbol{X}} \succ 0$ and $\overline{\boldsymbol{Z}} \succ 0$ by Lemma 3.2. Thus, we can define

$$\begin{aligned}
\boldsymbol{P} &= \overline{\boldsymbol{Z}}^{1/2}(\boldsymbol{X}^+\boldsymbol{Z}^+ - (1-\overline{\theta})\tau\boldsymbol{I})\,\overline{\boldsymbol{Z}}^{(-1/2)} \\
&= \overline{\boldsymbol{Z}}^{1/2}((\overline{\boldsymbol{X}} + \Delta\overline{\boldsymbol{X}})(\overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{Z}}) - (1-\overline{\theta})\tau\boldsymbol{I})\,\overline{\boldsymbol{Z}}^{(-1/2)} \\
&= [\overline{\boldsymbol{Z}}^{1/2}\overline{\boldsymbol{X}}\,\overline{\boldsymbol{Z}}^{1/2} - (1-\overline{\theta})\tau\boldsymbol{I}] + \overline{\boldsymbol{Z}}^{1/2}(\overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{X}}\,\overline{\boldsymbol{Z}})\overline{\boldsymbol{Z}}^{(-1/2)} \\
&\quad + \overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}^{(-1/2)}.
\end{aligned}$$

By (33C), we have

$$\begin{aligned}\operatorname{symm}(\boldsymbol{P}) &= \left(\overline{\boldsymbol{Z}}^{1/2}\overline{\boldsymbol{X}}\,\overline{\boldsymbol{Z}}^{1/2} - (1-\overline{\theta})\tau\boldsymbol{I}\right)\\ &\quad + \operatorname{symm}\left(\overline{\boldsymbol{Z}}^{1/2}(\overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{X}}\,\overline{\boldsymbol{Z}})\overline{\boldsymbol{Z}}^{(-1/2)}\right)\\ &\quad + \operatorname{symm}\left(\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}^{(-1/2)}\right)\\ &= \operatorname{symm}\left(\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}^{(-1/2)}\right) - \overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}_\epsilon\overline{\boldsymbol{Z}}^{1/2}. \qquad (85)\end{aligned}$$

Since the corrector step satisfies (31C) - (32C) and $(\overline{\boldsymbol{X}}, \overline{\boldsymbol{Z}}) \in \mathcal{N}(\beta, (1-\theta)\tau)$, we can apply Lemma A.2 to $\overline{\boldsymbol{X}}$, $\overline{\boldsymbol{Z}}$, $\Delta\overline{\boldsymbol{X}}$, and $\Delta\overline{\boldsymbol{Z}}$. So, with $\gamma = \beta$ and replacing τ' with $(1-\theta)\tau$ and $(\boldsymbol{X}', \boldsymbol{Z}', \Delta\boldsymbol{X}', \Delta\boldsymbol{Z}')$ with $(\overline{\boldsymbol{X}}, \overline{\boldsymbol{Z}}, \Delta\overline{\boldsymbol{X}}, \Delta\overline{\boldsymbol{Z}})$, the inequality (82) divided by τ' becomes

$$\|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}\,\overline{\boldsymbol{Z}}^{1/2}\|_F\|\overline{\boldsymbol{Z}}^{(-1/2)}\Delta\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}^{(-1/2)}\|_F \le \frac{\|\boldsymbol{H}\|_F^2}{2(1-\beta)^2(1-\overline{\theta})\tau}, \qquad (86)$$

where

$$\boldsymbol{H} = \operatorname{symm}\left(\overline{\boldsymbol{Z}}^{1/2}(\overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{X}}\,\overline{\boldsymbol{Z}})\overline{\boldsymbol{Z}}^{(-1/2)}\right).$$

In addition, by (33C),

$$\begin{aligned}\|\boldsymbol{H}\|_F &= \|\operatorname{symm}\left(\overline{\boldsymbol{Z}}^{1/2}(\overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{X}}\,\overline{\boldsymbol{Z}})\overline{\boldsymbol{Z}}^{(-1/2)}\right)\|_F\\ &= \|\overline{\boldsymbol{Z}}^{1/2}(\overline{\boldsymbol{X}} + \Delta\overline{\boldsymbol{X}}_\epsilon)\overline{\boldsymbol{Z}}^{1/2} - (1-\overline{\theta})\tau\boldsymbol{I}\|_F\\ &\le \|\overline{\boldsymbol{Z}}^{1/2}\overline{\boldsymbol{X}}\,\overline{\boldsymbol{Z}}^{1/2} - (1-\overline{\theta})\tau\boldsymbol{I}\|_F + \|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}_\epsilon\overline{\boldsymbol{Z}}^{1/2}\|_F\\ &\le \beta(1-\overline{\theta})\tau + \overline{\delta}_\epsilon\tau \quad (\text{ since } (\overline{\boldsymbol{X}}, \overline{\boldsymbol{Z}}) \in \mathcal{N}(\beta, (1-\theta)\tau)). \qquad (87)\end{aligned}$$

By (86) and (87),

$$\begin{aligned}\|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}\,\overline{\boldsymbol{Z}}^{1/2}\|_F\|\overline{\boldsymbol{Z}}^{(-1/2)}\Delta\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}^{(-1/2)}\|_F &\le \frac{1}{2(1-\beta)^2(1-\overline{\theta})\tau}\left(\beta(1-\overline{\theta})\tau + \overline{\delta}_\epsilon\tau\right)^2\\ = \frac{\beta^2}{2(1-\beta)^2}(1-\overline{\theta})\tau + \frac{\beta}{(1-\beta)^2}\overline{\delta}_\epsilon\tau + \frac{\overline{\delta}_\epsilon^2\tau}{2(1-\beta)^2(1-\overline{\theta})}. \qquad (88)\end{aligned}$$

By Lemma A.2 again, using (84) divided by τ',

$$\begin{aligned}\frac{\delta_z'}{\tau'} = \|\overline{\boldsymbol{Z}}^{(-1/2)}\Delta\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}^{(-1/2)}\|_F \le \frac{\|\boldsymbol{H}\|_F}{(1-\beta)(1-\overline{\theta})\tau} &\le \frac{\beta(1-\overline{\theta})\tau + \overline{\delta}_\epsilon\tau}{(1-\beta)(1-\overline{\theta})\tau} \quad (\text{by } (87)),\\ < \frac{\beta}{1-\beta} + \frac{(1-\overline{\theta})(1-2\beta)}{(1-\beta)(1-\overline{\theta})} = 1 \quad (\text{by Lemma A.3}).\end{aligned}$$

So, by (71),

$$\lambda_{\min}(\overline{\boldsymbol{Z}}^{(-1/2)}\Delta\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}^{(-1/2)}) > -1.$$

This implies that $(\boldsymbol{I}+\overline{\boldsymbol{Z}}^{(-1/2)}\Delta\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}^{(-1/2)}) \succ 0$, so

$$\boldsymbol{Z}^{+} = \overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{Z}} = \overline{\boldsymbol{Z}}^{1/2}(\boldsymbol{I}+\overline{\boldsymbol{Z}}^{(-1/2)}\Delta\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}^{(-1/2)})\overline{\boldsymbol{Z}}^{1/2} \succ 0.$$

Therefore, $(\boldsymbol{Z}^{+})^{-1/2}$ exists. By defining

$$\boldsymbol{E} = (\boldsymbol{Z}^{+})^{1/2}\boldsymbol{X}^{+}(\boldsymbol{Z}^{+})^{1/2} - (1-\overline{\theta})\tau\boldsymbol{I}, \qquad \boldsymbol{M} = \overline{\boldsymbol{Z}}^{1/2}(\boldsymbol{Z}^{+})^{-1/2},$$

we can see that $\boldsymbol{P} = \boldsymbol{M}\boldsymbol{E}\boldsymbol{M}^{-1}$. Applying (74) with these $\boldsymbol{E} \in \mathcal{S}^n$ and $\boldsymbol{M}$ and using Condition 2.2 and (48), we have

$$\begin{aligned}
&\|(\boldsymbol{Z}^{+})^{1/2}\boldsymbol{X}^{+}(\boldsymbol{Z}^{+})^{1/2} - (1-\overline{\theta})\tau\boldsymbol{I}\|_F \le \|\,\mathrm{symm}\,(\boldsymbol{P})\,\|_F \\
&= \|\,\mathrm{symm}\left(\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}^{(-1/2)}\right) - \overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}_\epsilon\overline{\boldsymbol{Z}}^{1/2}\|_F \quad \text{(by (85))} \\
&\le \|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}^{(-1/2)}\|_F + \|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}_\epsilon\overline{\boldsymbol{Z}}^{1/2}\|_F \\
&\le \|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}\,\overline{\boldsymbol{Z}}^{1/2}\|_F\|\overline{\boldsymbol{Z}}^{(-1/2)}\Delta\overline{\boldsymbol{Z}}\,\overline{\boldsymbol{Z}}^{(-1/2)}\|_F + \|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}_\epsilon\overline{\boldsymbol{Z}}^{1/2}\|_F \\
&\le \frac{\beta^2}{2(1-\beta)^2}(1-\overline{\theta})\tau + \left(\frac{\beta}{(1-\beta)^2}+1\right)\overline{\delta}_\epsilon\tau + \frac{\overline{\delta}_\epsilon^2\tau}{2(1-\beta)^2(1-\overline{\theta})} \quad \text{(by (88) and (40))} \\
&= \frac{\tau}{2(1-\beta)^2(1-\overline{\theta})}\left[\beta^2(1-\overline{\theta})^2 + (1-\overline{\theta})(2\beta+2(1-\beta)^2)\overline{\delta}_\epsilon + \overline{\delta}_\epsilon^2\right] \\
&< \frac{\tau}{2(1-\beta)^2(1-\overline{\theta})}\left[\beta^2(1-\overline{\theta})^2 + 2(1-\overline{\theta})^2(\beta^2-\beta+1)(\sqrt{s^2+t}-s) + (1-\overline{\theta})^2(\sqrt{s^2+t}-s)^2\right] \\
&= \frac{\tau}{2(1-\beta)^2(1-\overline{\theta})}\left[\beta^2(1-\overline{\theta})^2 + 2(1-\overline{\theta})^2 s(\sqrt{s^2+t}-s) + (1-\overline{\theta})^2(s^2+t+s^2-2s\sqrt{s^2+t})\right] \\
&= \frac{(1-\overline{\theta})\tau}{2(1-\beta)^2}\left[\beta^2 + 2s(\sqrt{s^2+t}-s) - 2s(\sqrt{s^2+t}-s) + t\right] \\
&= \frac{(1-\overline{\theta})\tau}{2(1-\beta)^2}(\beta^2+t) = \frac{(1-\overline{\theta})\tau}{2(1-\beta)^2}(2(1-\beta)^2\alpha) = \alpha(1-\overline{\theta})\tau.
\end{aligned}$$

In addition, this implies that

$$\lambda_{\min}((\boldsymbol{Z}^{+})^{1/2}\boldsymbol{X}^{+}(\boldsymbol{Z}^{+})^{1/2} - (1-\overline{\theta})\tau\boldsymbol{I}) \ge -\alpha(1-\overline{\theta})\tau,$$

by (71), so

$$\lambda_{\min}((\boldsymbol{Z}^{+})^{1/2}\boldsymbol{X}^{+}(\boldsymbol{Z}^{+})^{1/2}) \ge -\alpha(1-\overline{\theta})\tau + (1-\overline{\theta})\tau = (1-\alpha)(1-\overline{\theta})\tau > 0.$$

Therefore, $(\boldsymbol{Z}^{+})^{1/2}\boldsymbol{X}^{+}(\boldsymbol{Z}^{+})^{1/2} \succ 0$, and $\boldsymbol{X}^{+} \succ 0$ as well. □

In the following proofs, we frequently use the inequality

$$\|\boldsymbol{M}_1\boldsymbol{M}_2\|_F \le \min(\|\boldsymbol{M}_1\|\|\boldsymbol{M}_2\|_F, \|\boldsymbol{M}_1\|_F\|\boldsymbol{M}_2\|), \quad \forall \boldsymbol{M}_1, \boldsymbol{M}_2 \in \mathbb{R}^{n\times n}. \tag{89}$$

(See Horn and Johnson [15, Exercise 20 in Section 5.6].) In addition, note that the Frobenius norm $\|\boldsymbol{E}\|_F$ for $\boldsymbol{E} \in \mathbb{R}^{n\times n}$ can be alternatively defined as

$$\|\boldsymbol{E}\|_F = \sqrt{\mathrm{tr}\left(\boldsymbol{E}^T\boldsymbol{E}\right)}. \tag{90}$$

Proof. of Lemma 3.6
First, we prove (61). By (90),

$$\begin{aligned}
\|\boldsymbol{X}^{1/2}(\boldsymbol{Z}^0)^{1/2}\|_F &= \sqrt{\mathrm{tr}\left((\boldsymbol{Z}^0)^{1/2}\boldsymbol{X}(\boldsymbol{Z}^0)^{1/2}\right)} = \sqrt{\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{Z}^0\right)} \\
&\leq \sqrt{\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{Z}^0\right) + \mathrm{tr}\left(\boldsymbol{X}^0\boldsymbol{Z}\right)} \quad (\text{since } \boldsymbol{X}^0 \in \mathcal{S}^n_+, \boldsymbol{Z} \in \mathcal{S}^n_+\ ) \\
&= \sqrt{\boldsymbol{X} \bullet \boldsymbol{Z}^0 + \boldsymbol{X}^0 \bullet \boldsymbol{Z}} \\
&\leq (n\tau_0)^{1/2}\left(2 + \zeta + \frac{\alpha}{\sqrt{n}}\right)^{1/2}. \quad (\text{by Lemma 3.5})
\end{aligned}$$

In a similar way, (62) can be proved.
Next, we prove (63).

$$\begin{aligned}
\|\boldsymbol{X}^{1/2}\|_F &= \|\boldsymbol{X}^{1/2}(\boldsymbol{Z}^0)^{1/2}(\boldsymbol{Z}^0)^{(-1/2)}\|_F \\
&\leq \|\boldsymbol{X}^{1/2}(\boldsymbol{Z}^0)^{1/2}\|_F\|(\boldsymbol{Z}^0)^{(-1/2)}\| \quad (\text{by (89)}) \\
&\leq \|(\boldsymbol{Z}^0)^{-1/2}\|\,(n\tau_0)^{1/2}\ \left(2 + \zeta + \frac{\alpha}{\sqrt{n}}\right)^{1/2} \quad (\text{by (61) proven above}).
\end{aligned}$$

In a similar way, we can also prove (64).
Next, we prove (65). The equality is satisfied since $\sigma^2_{\max}(\boldsymbol{E}) = \sigma_{\max}(\boldsymbol{E}^T\boldsymbol{E})$ for any matrix $\boldsymbol{E}$. Because $(\boldsymbol{X}, \boldsymbol{Z}) \in \mathcal{N}(\alpha, \tau)$,

$$\begin{aligned}
\|\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2} - \tau\boldsymbol{I}\| &\leq \|\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2} - \tau\boldsymbol{I}\|_F \leq \alpha\tau, \\
\|\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2}\| - \tau &\leq \alpha\tau, \\
\|\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2}\| &\leq \tau + \alpha\tau = (1+\alpha)\tau.
\end{aligned}$$

In a similar way, (66) can be proved. □

Proof. of Lemma 3.7
We will use Lemma A.2 with $(\boldsymbol{X}', \boldsymbol{y}', \boldsymbol{Z}') = (\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z})$, $(\Delta\boldsymbol{X}', \Delta\boldsymbol{y}', \Delta\boldsymbol{Z}') = (\Delta\boldsymbol{X} + \psi(\boldsymbol{X}^0 - \breve{\boldsymbol{X}}), \Delta\boldsymbol{y} + \psi(\boldsymbol{y}^0 - \breve{\boldsymbol{y}}), \Delta\boldsymbol{Z} + \psi(\boldsymbol{Z}^0 - \breve{\boldsymbol{Z}}))$, $\gamma = \alpha$, and $\tau' = \tau$. For a predictor direction $(\Delta\boldsymbol{X}, \Delta\boldsymbol{y}, \Delta\boldsymbol{Z})$, by (7) and (11),

$$\begin{aligned}
\boldsymbol{A}_i \bullet \Delta\boldsymbol{X} &= r_{pi}, \\
\psi(\boldsymbol{A}_i \bullet \boldsymbol{X}^0) &= \psi(b_i - r^0_{pi}) = \psi b_i - r_{pi},
\end{aligned}$$

and since $\breve{\boldsymbol{X}}$ is feasible,

$$\psi(\boldsymbol{A}_i \bullet \breve{\boldsymbol{X}}) = \psi b_i,$$

for $i = 1, \ldots, m$. Hence, $\boldsymbol{A}_i \bullet (\Delta\boldsymbol{X} + \psi(\boldsymbol{X}^0 - \breve{\boldsymbol{X}})) = 0$, thus satisfying (80).

Also, by (8) and (12)

$$\begin{aligned}\left(\sum_{i=1}^{m}\Delta y_i \boldsymbol{A}_i\right)+\Delta\boldsymbol{Z} &= \boldsymbol{R}_d,\\ \psi\left[\left(\sum_{i=1}^{m} y_i^0 \boldsymbol{A}_i\right)+\boldsymbol{Z}^0\right] &= \psi(\boldsymbol{C}-\boldsymbol{R}_d^0)=\psi\boldsymbol{C}-\boldsymbol{R}_d,\\ \psi\left[\left(\sum_{i=1}^{m}\breve{y}_i\boldsymbol{A}_i\right)+\breve{\boldsymbol{Z}}\right] &= \psi\boldsymbol{C}.\end{aligned}$$

Thus, $(\Delta\boldsymbol{y}+\psi(\boldsymbol{y}^0-\breve{\boldsymbol{y}}),\Delta\boldsymbol{Z}+\psi(\boldsymbol{Z}^0-\breve{\boldsymbol{Z}}))$ satisfies (81).
In addition, since $(\boldsymbol{X},\boldsymbol{Z})\in\mathcal{N}(\alpha,\tau)$, the prerequisite of Lemma A.2 is now verified. Then, using (33P), $\boldsymbol{H}$ in Lemma A.2 becomes $\boldsymbol{T}$.
Therefore, from Lemma A.2, using (83) and (84), we have the following inequalities,

$$\begin{aligned}\|\boldsymbol{Z}^{1/2}(\Delta\boldsymbol{X}+\psi(\boldsymbol{X}^0-\breve{\boldsymbol{X}}))\boldsymbol{Z}^{1/2}\|_F &\le \frac{\|\boldsymbol{T}\|_F}{1-\alpha},\\ \tau\|\boldsymbol{Z}^{-1/2}(\Delta\boldsymbol{Z}+\psi(\boldsymbol{Z}^0-\breve{\boldsymbol{Z}}))\boldsymbol{Z}^{-1/2}\|_F &\le \frac{\|\boldsymbol{T}\|_F}{1-\alpha}.\end{aligned}$$

Hence,

$$\begin{aligned}\|\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}\boldsymbol{Z}^{1/2}\|_F &\le \frac{\|\boldsymbol{T}\|_F}{1-\alpha}+\psi\|\boldsymbol{Z}^{1/2}(\boldsymbol{X}^0-\breve{\boldsymbol{X}})\boldsymbol{Z}^{1/2}\|_F,\\ &= \frac{\|\boldsymbol{T}\|_F}{1-\alpha}+\|\boldsymbol{T}_x\|_F\\ \tau\|\boldsymbol{Z}^{-1/2}\Delta\boldsymbol{Z}\boldsymbol{Z}^{-1/2}\|_F &\le \frac{\|\boldsymbol{T}\|_F}{1-\alpha}+\tau\psi\|\boldsymbol{Z}^{-1/2}(\boldsymbol{Z}^0-\breve{\boldsymbol{Z}})\boldsymbol{Z}^{-1/2}\|_F\\ &= \frac{\|\boldsymbol{T}\|_F}{1-\alpha}+\tau\|\boldsymbol{T}_z\|_F.\end{aligned}$$

□

Proof. of Lemma 3.8
First, we calculate bounds on $\|\boldsymbol{T}_x\|$, $\|\boldsymbol{T}_z\|$, and $\|\boldsymbol{T}\|$ in Lemma 3.7.
By Lemma 3.6, we have

$$\begin{aligned}\|\boldsymbol{T}_x\|_F &= \psi\|\boldsymbol{Z}^{1/2}(\boldsymbol{X}^0-\breve{\boldsymbol{X}})\boldsymbol{Z}^{1/2}\|_F\\ &= \psi\|\boldsymbol{Z}^{1/2}(\boldsymbol{X}^0)^{1/2}(\boldsymbol{X}^0)^{-1/2}(\boldsymbol{X}^0-\breve{\boldsymbol{X}})(\boldsymbol{X}^0)^{-1/2}(\boldsymbol{X}^0)^{1/2}\boldsymbol{Z}^{1/2}\|_F\\ &\le \psi\|\boldsymbol{Z}^{1/2}(\boldsymbol{X}^0)^{1/2}\|_F^2\ \ \|(\boldsymbol{X}^0)^{-1/2}(\boldsymbol{X}^0-\breve{\boldsymbol{X}})(\boldsymbol{X}^0)^{-1/2}\|_F\\ &\le \psi n\tau_0(2+\zeta+\alpha/\sqrt{n})d_0 = nd_0(2+\zeta+\alpha/\sqrt{n})\tau,\\ \|\boldsymbol{T}_z\|_F &= \psi\|\boldsymbol{Z}^{-1/2}(\boldsymbol{Z}^0-\breve{\boldsymbol{Z}})\boldsymbol{Z}^{-1/2}\|_F\\ &\le \psi\|\boldsymbol{Z}^{-1/2}\boldsymbol{X}^{-1/2}\|^2\ \ \|\boldsymbol{X}^{1/2}(\boldsymbol{Z}^0)^{1/2}\|_F^2\ \ \|(\boldsymbol{Z}^0)^{-1/2}(\boldsymbol{Z}^0-\breve{\boldsymbol{Z}})(\boldsymbol{Z}^0)^{-1/2}\|_F\\ &\le \frac{\psi n\tau_0\,(2+\zeta+\alpha/\sqrt{n})}{(1-\alpha)\tau}d_0 = \frac{nd_0\,(2+\zeta+\alpha/\sqrt{n})}{(1-\alpha)}.\end{aligned}$$

Similarly,

$$\begin{aligned}
\|\boldsymbol{T}\|_F &\le \psi\|\boldsymbol{Z}^{1/2}(\boldsymbol{X}^0-\breve{\boldsymbol{X}})\boldsymbol{Z}^{1/2}\|_F \\
&\quad+\psi\|\boldsymbol{Z}^{1/2}\boldsymbol{X}(\boldsymbol{Z}^0-\breve{\boldsymbol{Z}})\boldsymbol{Z}^{-1/2}\|_F+\|\boldsymbol{Z}^{1/2}(\boldsymbol{X}+\Delta\boldsymbol{X}_\epsilon)\boldsymbol{Z}^{1/2}\|_F \\
&= \psi\|\boldsymbol{Z}^{1/2}(\boldsymbol{X}^0)^{1/2}(\boldsymbol{X}^0)^{-1/2}(\boldsymbol{X}^0-\breve{\boldsymbol{X}})(\boldsymbol{X}^0)^{-1/2}(\boldsymbol{X}^0)^{1/2}\boldsymbol{Z}^{1/2}\|_F \\
&\quad+\psi\|\boldsymbol{Z}^{1/2}\boldsymbol{X}^{1/2}\boldsymbol{X}^{1/2}(\boldsymbol{Z}^0)^{1/2}(\boldsymbol{Z}^0)^{-1/2}(\boldsymbol{Z}^0-\breve{\boldsymbol{Z}})(\boldsymbol{Z}^0)^{-1/2}(\boldsymbol{Z}^0)^{1/2}\boldsymbol{X}^{1/2}\boldsymbol{X}^{-1/2}\boldsymbol{Z}^{-1/2}\|_F \\
&\quad+\|\boldsymbol{Z}^{1/2}(\boldsymbol{X}+\Delta\boldsymbol{X}_\epsilon)\boldsymbol{Z}^{1/2}\|_F \\
&\le \psi\|\boldsymbol{Z}^{1/2}(\boldsymbol{X}^0)^{1/2}\|_F^2\|(\boldsymbol{X}^0)^{-1/2}(\boldsymbol{X}^0-\breve{\boldsymbol{X}})(\boldsymbol{X}^0)^{-1/2}\|_F \\
&\quad+\psi\|\boldsymbol{Z}^{1/2}\boldsymbol{X}^{1/2}\|\|\boldsymbol{X}^{1/2}(\boldsymbol{Z}^0)^{1/2}\|_F^2\|(\boldsymbol{Z}^0)^{-1/2}(\boldsymbol{Z}^0-\breve{\boldsymbol{Z}})(\boldsymbol{Z}^0)^{-1/2}\|_F\|\boldsymbol{X}^{-1/2}\boldsymbol{Z}^{-1/2}\| \\
&\quad+\sqrt{n}\|\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2}\|+\|\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}_\epsilon\boldsymbol{Z}^{1/2}\|_F \\
&\le \psi n\tau_0(2+\zeta+\alpha/\sqrt{n})d_0+\psi n\tau_0(2+\zeta+\alpha/\sqrt{n})d_0\sqrt{\frac{1+\alpha}{1-\alpha}} \\
&\quad+\sqrt{n}(1+\alpha)\tau+\delta_\epsilon\tau \quad (\text{by definition of } \delta_\epsilon \text{ in (39)}) \\
&\le n\tau d_0(2+\zeta+\alpha/\sqrt{n})+n\tau d_0(2+\zeta+\alpha/\sqrt{n})\left(\frac{1+\alpha}{1-\alpha}\right)+\sqrt{n}(1+\alpha)\tau+\delta_\epsilon\tau \\
&\le \tau\left[\frac{2nd_0(2+\zeta+\alpha/\sqrt{n})}{1-\alpha}+\sqrt{n}(1+\alpha)\right]+\delta_\epsilon\tau \\
&\le \tau\left[\frac{2nd_0(2+\zeta+\alpha/\sqrt{n})}{1-\alpha}+\sqrt{n}(1+\alpha)\right]+\delta_x q \quad (\text{by Condition 2.1}).
\end{aligned}$$

Using (68)-(70), we can rewrite the bounds on $\|\boldsymbol{T}_x\|_F$, $\|\boldsymbol{T}_z\|$, and $\|\boldsymbol{T}\|$ as

$$\|\boldsymbol{T}_x\|_F\le C_x\tau,\quad \|\boldsymbol{T}_z\|_F\le C_z,\quad \|\boldsymbol{T}\|_F\le C_0\tau+\delta_x q.$$

By Lemma 3.7 with the bounds on $\|\boldsymbol{T}_x\|_F$ and $\|\boldsymbol{T}\|_F$ above, we have

$$\begin{aligned}
\delta_x &\le \|\boldsymbol{T}_x\|_F+\frac{\|\boldsymbol{T}\|_F}{1-\alpha}\le C_x\tau+\frac{C_0\tau+\delta_x q}{1-\alpha}, \\
\delta_x &\le \left(\frac{1-\alpha}{1-\alpha-q}\right)\left(C_x+\frac{C_0}{1-\alpha}\right)\tau \quad (\text{since } 1-\alpha-q>0 \text{ by (42) and (46)}).
\end{aligned}$$

In a similar way, by Lemma 3.7 with the bounds on $\|\boldsymbol{T}_z\|_F$ and $\|\boldsymbol{T}\|_F$ above, we have

$$\begin{aligned}
\delta_z &\le \tau\|\boldsymbol{T}_z\|_F+\frac{\|\boldsymbol{T}\|_F}{1-\alpha}\le C_z\tau+\frac{C_0\tau+\delta_x q}{1-\alpha}\le\left(C_z+\frac{C_0}{1-\alpha}\right)\tau+\frac{q}{1-\alpha}\delta_x \\
&\le \left(C_z+\frac{C_0}{1-\alpha}\right)\tau+\frac{q}{1-\alpha-q}\left(C_x+\frac{C_0}{1-\alpha}\right)\tau \quad (\text{by the bound of } \delta_x \text{ above}).
\end{aligned}$$

By definitions of C_x, C_z, and C_0, since $0<\alpha<1/2$,

$$\left(C_x+\frac{C_0}{1-\alpha}\right)<\left(C_z+\frac{C_0}{1-\alpha}\right), \tag{91}$$

so we have

$$\begin{aligned} \delta_z &\leq \left(C_z + \frac{C_0}{1-\alpha}\right)\tau + \frac{q}{1-\alpha-q}\left(C_x + \frac{C_0}{1-\alpha}\right)\tau \\ &\leq \left(\frac{1-\alpha}{1-\alpha-q}\right)\left(C_z + \frac{C_0}{1-\alpha}\right)\tau. \end{aligned}$$

Finally, by (67) and (91),

$$\begin{aligned} \delta &\leq \left(\frac{1-\alpha}{1-\alpha-q}\right)\left(C_x + \frac{C_0}{1-\alpha}\right)\left(C_z + \frac{C_0}{1-\alpha} + \frac{q}{1-\alpha-q}\left(C_x + \frac{C_0}{1-\alpha}\right)\right) \\ &= \left(\frac{1-\alpha}{1-\alpha-q}\right)\left(C_x + \frac{C_0}{1-\alpha}\right)\left(C_z + \frac{C_0}{1-\alpha}\right) \\ &\quad + \frac{q(1-\alpha)}{(1-\alpha-q)^2}\left(C_x + \frac{C_0}{1-\alpha}\right)^2 \\ &\leq \left(\frac{1-\alpha}{1-\alpha-q} + \frac{q(1-\alpha)}{(1-\alpha-q)^2}\right)\left(C_z + \frac{C_0}{1-\alpha}\right)^2 \\ &= \left(\frac{1-\alpha}{1-\alpha-q}\right)^2\left(C_z + \frac{C_0}{1-\alpha}\right)^2, \end{aligned}$$

and we obtain the final inequality. □

Proof. of Lemma 3.9
By Lemma 3.5, we have $\rho(\mathrm{tr}\,(\boldsymbol{X}) + \mathrm{tr}\,(\boldsymbol{Z})) \leq (2+\zeta+\alpha/\sqrt{n})n\tau_0 = (2+\zeta+\alpha/\sqrt{n})n\rho^2$, so

$$\sum_{i=1}^{n}(\lambda_i(\boldsymbol{X}) + \lambda_i(\boldsymbol{Z})) \leq (2+\zeta+\alpha/\sqrt{n})n\rho.$$

From (41), we have $\alpha/\sqrt{n} \leq \alpha \leq 1/2$, and since $\boldsymbol{X}^* \bullet \boldsymbol{Z}^* = 0$,

$$\zeta = (\boldsymbol{Z}^* \bullet \boldsymbol{X}^0 + \boldsymbol{X}^* \bullet \boldsymbol{Z}^0)/(\boldsymbol{X}^0 \bullet \boldsymbol{Z}^0) = (\mathrm{tr}\,(\boldsymbol{X}^*) + \mathrm{tr}\,(\boldsymbol{Z}^*))/(n\rho) \leq 1.$$

This implies

$$\|\boldsymbol{X}^{1/2}\|_F^2 + \|\boldsymbol{Z}^{1/2}\|_F^2 = \sum_{i=1}^{n}(\lambda_i(\boldsymbol{X}) + \lambda_i(\boldsymbol{Z})) \leq (3+\alpha/\sqrt{n})\rho n \leq 3.5\rho n. \tag{92}$$

In addition, we can see that $\|\boldsymbol{X}^0 - \boldsymbol{X}^*\| \leq \rho$ and $\|\boldsymbol{Z}^0 - \boldsymbol{Z}^*\| \leq \rho$. By (92) and Lemma 3.6,

$$\|\boldsymbol{Z}^{1/2}(\boldsymbol{X}^0 - \boldsymbol{X}^*)\boldsymbol{Z}^{1/2}\| \leq \|\boldsymbol{Z}^{1/2}\|_F^2\|\boldsymbol{X}^0 - \boldsymbol{X}^*\| \leq 3.5\rho^2 n, \tag{93}$$

$$\begin{aligned} \|\boldsymbol{Z}^{1/2}\boldsymbol{X}(\boldsymbol{Z}^0 - \boldsymbol{Z}^*)\boldsymbol{Z}^{-1/2}\| &\leq \|(\boldsymbol{Z}^{1/2}\boldsymbol{X}^{1/2})\boldsymbol{X}^{1/2}(\boldsymbol{Z}^0 - \boldsymbol{Z}^*)\boldsymbol{X}^{1/2}(\boldsymbol{X}^{-1/2}\boldsymbol{Z}^{-1/2})\| \\ &\leq \|(\boldsymbol{Z}^{1/2}\boldsymbol{X}^{1/2})\|\|(\boldsymbol{X}^{-1/2}\boldsymbol{Z}^{-1/2})\|\|\boldsymbol{X}^{1/2}\|_F^2\|(\boldsymbol{Z}^0 - \boldsymbol{Z}^*)\| \\ &\leq \left(\sqrt{\frac{1+\alpha}{1-\alpha}}\right)3.5\rho^2 n \leq 6.1\rho^2 n. \end{aligned} \tag{94}$$

By (93), Lemma 3.6, and Lemma 3.7 with $(\breve{\boldsymbol{X}}, \breve{\boldsymbol{y}}, \breve{\boldsymbol{Z}}) = (\boldsymbol{X}^*, \boldsymbol{y}^*, \boldsymbol{Z}^*)$,

$$\begin{aligned} \|\boldsymbol{T}_x\|_F &\leq \psi\|\boldsymbol{Z}^{1/2}(\boldsymbol{X}^0 - \boldsymbol{X}^*)\boldsymbol{Z}^{1/2}\|_F \leq 3.5\psi\rho^2 n = 3.5n\tau, \qquad (95) \\ \tau\|\boldsymbol{T}_z\|_F &\leq \tau\psi\|(\boldsymbol{Z}^{-1/2}\boldsymbol{X}^{-1/2})\boldsymbol{X}^{1/2}(\boldsymbol{Z}^0 - \boldsymbol{Z}^*)\boldsymbol{X}^{1/2}(\boldsymbol{X}^{-1/2}\boldsymbol{Z}^{-1/2})\|_F \\ &\leq \tau\psi\|(\boldsymbol{Z}^{-1/2}\boldsymbol{X}^{-1/2})\|^2\|\boldsymbol{X}^{1/2}\|_F^2\|(\boldsymbol{Z}^0 - \boldsymbol{Z}^*)\|_F \\ &\leq 3.5\tau\psi\rho^2 n/(0.5\tau) = 7n\tau. \qquad (96) \end{aligned}$$

Similarly, by (93), (94), (65), and (39)

$$\begin{aligned} \|\boldsymbol{T}\|_F &\leq \psi\|\boldsymbol{Z}^{1/2}(\boldsymbol{X}^0 - \boldsymbol{X}^*)\boldsymbol{Z}^{1/2}\|_F + \psi\|\boldsymbol{Z}^{1/2}\boldsymbol{X}(\boldsymbol{Z}^0 - \boldsymbol{Z}^*)\boldsymbol{Z}^{-1/2}\|_F \\ &\quad + \|\boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2}\|_F + \|\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}_\epsilon\boldsymbol{Z}^{1/2}\|_F \\ &\leq (3.5\psi\rho^2 n + 6.1\psi\rho^2 n + 1.5n\tau) + \delta_\epsilon\tau \leq 11.1n\tau + \delta_x q \quad \text{(by Condition 2.1).} \end{aligned}$$

By the bound of δ_x in Lemma 3.7,

$$\begin{aligned} \|\boldsymbol{T}\|_F &\leq 11.1n\tau + \left(\|\boldsymbol{T}_x\|_F + \frac{\|\boldsymbol{T}\|_F}{1-\alpha}\right)q, \\ \|\boldsymbol{T}\|_F &\leq \left(\frac{1-\alpha}{1-\alpha-q}\right)(11.1n\tau + q\|\boldsymbol{T}_x\|_F). \end{aligned}$$

Furthermore, by the bound of $\|\boldsymbol{T}_x\|_F$ above, we have

$$\|\boldsymbol{T}\|_F \leq \left(\frac{1-\alpha}{1-\alpha-q}\right)(11.1 + 3.5q)n\tau. \qquad (97)$$

By Lemma 3.7 with (95) and (97),

$$\begin{aligned} \delta_x &\leq \|\boldsymbol{T}_x\|_F + \frac{\|\boldsymbol{T}\|_F}{1-\alpha} \\ &\leq 3.5n\tau + \frac{(11.1 + 3.5q)n\tau}{1-\alpha-q} \\ &\leq 3.5n\tau + \frac{(11.1 + 3.5q)n\tau}{0.5-q} \quad \text{(since } \alpha < 0.5) \\ &\leq \left(3.5 + \frac{11.1 + 3.5q}{0.5-q}\right)n\tau = \left(\frac{12.85}{0.5-q}\right)n\tau, \end{aligned}$$

so, by the definition of h,

$$\delta_x \leq hn\tau. \qquad (98)$$

Similarly, by Lemma 3.7 with (96) and (97),

$$\begin{aligned} \delta_z &\leq \tau\|\boldsymbol{T}_z\|_F + \frac{\|\boldsymbol{T}\|_F}{1-\alpha} \\ &\leq 7n\tau + \frac{(11.1 + 3.5q)n\tau}{1-\alpha-q} \\ &\leq 7n\tau + \frac{(11.1 + 3.5q)n\tau}{0.5-q} \quad \text{(since } \alpha < 0.5) \\ &\leq \left(7 + \frac{11.1 + 3.5q}{0.5-q}\right)n\tau = \left(3.5 + \frac{12.85}{0.5-q}\right)n\tau, \end{aligned}$$

so, by the definition of h, $\delta_z \leq (h+3.5)n\tau$. Therefore, by (67) and (98),

$$\delta \leq \frac{1}{\tau^2}\delta_x\delta_z \leq \frac{1}{\tau^2}(hn\tau)\left((h+3.5)n\tau\right) \leq h(h+3.5)n^2. \tag{99}$$

By Condition 2.1 and (98),

$$\delta_\epsilon \leq \frac{q}{\tau}\delta_x \leq \frac{q}{\tau}(hn\tau) = qnh. \tag{100}$$

By the definition of $\widehat{\theta}$ in (43) and the fact that $\sqrt{x}+\sqrt{y} \geq \sqrt{x+y}$,

$$\begin{aligned}
\widehat{\theta} &= \frac{2}{\sqrt{\left(1+\frac{\delta_\epsilon}{\beta-\alpha}\right)^2+\frac{4\delta}{\beta-\alpha}}+\left(1+\frac{\delta_\epsilon}{\beta-\alpha}\right)} \\
&\geq \frac{2}{\left(1+\frac{\delta_\epsilon}{\beta-\alpha}\right)+\sqrt{\frac{4\delta}{\beta-\alpha}}+\left(1+\frac{\delta_\epsilon}{\beta-\alpha}\right)} \\
&= \frac{1}{\left(1+\frac{\delta_\epsilon}{\beta-\alpha}\right)+\sqrt{\frac{\delta}{\beta-\alpha}}} \geq \frac{1}{\left(n+\frac{\delta_\epsilon}{\beta-\alpha}\right)+\sqrt{\frac{\delta}{\beta-\alpha}}}.
\end{aligned}$$

Finally, by the bounds on δ and δ_ϵ in (99) and (100), we have

$$\begin{aligned}
\widehat{\theta} &\geq \frac{1}{\left(n+\frac{qnh}{\beta-\alpha}\right)+\sqrt{\frac{h(h+3.5)n^2}{\beta-\alpha}}} \\
&\geq \frac{1}{n\left(1+\frac{hq}{\beta-\alpha}+\sqrt{\frac{h(h+3.5)}{\beta-\alpha}}\right)} = \frac{1}{wn}.
\end{aligned}$$

□

References

[1] Farid Alizadeh, Jean-Pierre A. Haeberly, and Michael L. Overton. Primal-dual interior-point methods for semidefinite programming. Technical report, manuscript presented at the Math. Programming Symposium, Ann Arbor, MI, 1994.

[2] Farid Alizadeh, Jean-Pierre A. Haeberly, and Michael L. Overton. Primal-dual interior-point methods for semidefinite programming: Convergence rates, stability and numerical results. *SIAM J. Opt.*, 8(3):746–768, 1998.

[3] Christine Bachoc and Frank Vallentin. New upper bounds for kissing number from semidefinite programming. *J. Amer. Math. Soc.*, 21(3):909–924, 2007.

[4] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge Univ. Press, New York, NY, 2004.

[5] George B. Dantzig and Yinyu Ye. A build-up interior-point method for linear programming: Affine scaling form. Technical report, Stanford Univ., 1991.

[6] Etienne de Klerk. *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications*. Kluwer Academic Publishers, Norwell, MA, 2002.

[7] Etienne de Klerk, Dmitrii V. Pasechnik, and Renata Sotirov. On semidefinite programming relaxations of the traveling salesman problem. *SIAM J. Opt.*, 19(4):1559–1573, 2008.

[8] Jack J. Dongarra, Cleve B. Moler, James R. Bunch, and G. W. Stewart. *LINPACK Users' Guide*. SIAM, Philadelphia, PA, 1979.

[9] Katsuki Fujisawa, Masakazu Kojima, and Kazuhide Nakata. Exploiting sparsity in primal-dual interior-point methods for semidefinite programming. *Math. Prog.*, 79(1):235–253, 1997.

[10] Philip E. Gill, Gene H. Golub, Walter Murray, and Michael A. Saunders. Methods for modifying matrix factorizations. *Math. Comp.*, 28(126):505–535, 1974.

[11] William Hager. Condition estimates. *SIAM J. Sci. Stat. Comput.*, 5(2):311–316, 1984.

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

[13] Dick Den Hertog, Cornelis Roos, and Tamás Terlaky. Adding and deleting constraints in the path-following method for LP. In *Advances in Optimization and Approximation*, volume 1, pages 166–185. Springer, New York, 1994.

[14] Nicholas Higham. A survey of condition number estimation for triangular matrices. *SIAM Review*, 9(4):575–596, 1987.

[15] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge Univ. Press, New York, NY, 1985.

[16] Benjamin Jansen. *Interior Point Techniques in Optimization*. Kluwer Academic Publishers, Norwell, MA, 1997.

[17] Jun Ji, Florian A. Potra, and Rongqin Sheng. On the local convergence of a predictor-corrector method for semidefinite programming. *SIAM J. Opt.*, 10(1):195–210, 1999.

[18] Jin Hyuk Jung, Dianne P. O'Leary, and André L. Tits. Adaptive constraint reduction for training support vector machines. *Elec. Trans. Numer. Anal.*, 31:156–177, 2008.

[19] Jin Hyuk Jung, Dianne P. O'Leary, and André L. Tits. Adaptive constraint reduction for convex quadratic programming. *Comput. Optim. Appl.*, 51(1):125–157, 2012.

[20] John A. Kaliski and Yinyu Ye. A decomposition variant of the potential reduction algorithm for linear programming. *Manage. Sci.*, 39:757–776, 1993.

[21] Masakazu Kojima, Masayuki Shida, and Susumu Shindoh. Local convergence of predictor-corrector infeasible-interior-point algorithms for SDPs and SDLCPs. *Math. Prog.*, 80(2):129–160, 1998.

[22] Masakazu Kojima, Masayuki Shida, and Susumu Shindoh. A predictor-corrector interior-point algorithm for the semidefinite linear complementarity problem using the Alizadeh–Haeberly–Overton search direction. *SIAM J. Opt.*, 9(2):444–465, 1999.

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

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

[25] Renato D. C. Monteiro. Polynomial convergence of primal-dual algorithms for semidefinite programming based on Monteiro and Zhang family of directions. *SIAM J. Opt.*, 8(3):797–812, 1998.

[26] Renato D. C. Monteiro and Yin Zhang. A unified analysis for a class of long-step primal-dual path-following interior point algorithms for semidefinite programming. *Math. Prog.*, 81(3):281–299, 1998.

[27] Yurii E. Nesterov and Michael J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Math. Op. Res.*, 22(1):1–42, 1997.

[28] Yurii E. Nesterov and Michael J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM J. Opt.*, 8(2):324–364, 1998.

[29] Dianne P. O'Leary. Estimating matrix condition numbers. *SIAM J. Sci. Stat. Comput.*, 1(2):205–209, 1980.

[30] Christopher C. Paige and Michael A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal.*, 12(4):617–629, 1975.

[31] Sungwoo Park. *Matrix Reduction in Numerical Optimization*. PhD thesis, Computer Science Department, Univ. of Maryland, College Park, MD, 2011.

[32] Florian A. Potra and Rongqin Sheng. Superlinear convergence of a predictor-corrector method for semidefinite programming without shrinking central path neighborhood. Technical Report 91, Reports on Computational Mathematics, Department of Mathematics, Univ. of Iowa, 1996.

[33] Florian A. Potra and Rongqin Sheng. Superlinear convergence of interior-point algorithms for semidefinite programming. *J. Opt. Theory Appl.*, 99(1):103–119, 1998.

[34] Florian A. Potra and Rongqin Sheng. A superlinearly convergent primal-dual infeasible-interior-point algorithm for semidefinite programming. *SIAM J. Opt.*, 8(4):1007–1028, 1998.

[35] Alexander Schrijver. A comparison of the Delsarte and Loviász bounds. *IEEE Trans. Inform. Theory*, IT-25(4):425–429, 1979.

[36] Pete G. W. Stewart. Efficient generation of random orthogonal matrices with an application to condition estimators. *SIAM J. Numer. Anal.*, 17(3):403–409, 1980.

[37] André L. Tits, Pierre-Antoine Absil, and William P. Woessner. Constraint reduction for linear programs with many inequality constraints. *SIAM J. Opt.*, 17(1):119–146, 2006.

[38] Kim-Chuan Toh, Michael J. Todd, and Reha H. Tütüncü. On the implementation and usage of SDPT3 - a MATLAB software package for semidefinite-quadratic-linear programming, version 4.0. In Miguel F. Anjos and Jean B. Lasserre, editors, *Handbook on Semidefinite, Conic and Polynomial Optimization*, volume 166, pages 715–754. Springer, New York, 2012.

[39] Kaoru Tone. An active-set strategy in an interior point method for linear programming. *Math. Prog.*, 59(3):345–360, 1993.

[40] Jhacova Ashira Williams. The use of preconditioning for training support vector machines. Master's thesis, Applied Mathematics and Scientific Computing Program, Univ. of Maryland, College Park, MD, 2008.

[41] Luke B. Winternitz, Stacey O. Nicholls, André L. Tits, and Dianne P. O'Leary. A constraint-reduced variant of Mehrotra's predictor-corrector algorithm. *Comput. Optim. Appl.*, 51(3):1001–1036, 2012.

[42] Yin Zhang. On extending some primal-dual interior-point algorithms from linear programming to semidefinite programming. *SIAM J. Opt.*, 8(2):365–386, 1998.

[43] Qing Zhao, Stefan E. Karisch, Franz Rendl, and Henry Wolkowicz. Semidefinite programming relaxations for the quadratic assignment problem. *J. Combin. Opt.*, 2(1):71–109, 1998.