A Constraint-Reduced Algorithm for Semidefinite Optimization Problems with Superlinear Convergence*

Sungwoo Park†

February 14, 2016

Abstract

Constraint reduction is an essential method because the computational cost of the interior point methods can be effectively saved. Park and O'Leary proposed a constraint-reduced predictor-corrector algorithm for semidefinite programming with polynomial global convergence, but they did not show its superlinear convergence. We first develop a constraint-reduced algorithm for semidefinite programming having both polynomial global and superlinear local convergences. The new algorithm repeats a corrector step to have an iterate tangentially approach a central path, by which superlinear convergence can be achieved. This study proves its convergence rate and shows its effective cost saving in numerical experiments.

Keywords: Semidefinite programming, Interior point methods, Constraint reduction, Primal dual infeasible, Local convergence.
AMS Classification: 90C22, 65K05, 90C51

1 Introduction

Constraint reduction methods originated from the question of whether we can save computational costs by ignoring a subset of constraints during the iterations of interior point methods (IPM). For semidefinite programming (SDP), construction of a Schur complement matrix is the most expensive part of the computation for each iteration of most IPM's (See [1]). A well-designed constraint reduction method can effectively reduce the computational cost by ignoring unimportant constraints without reducing convergence rate. This paper studies a constraint-reduced IPM for SDP, that has both polynomial global and superlinear local convergence.

*I would like to thank Professor Dianne P. O'Leary for careful review of the manuscript and insightful advices to enhance the convergence analysis.

†KCG holdings, 545 Washington BLVD, Jersey City, NJ 07310, USA. swpark81@gmail.com

For IPM of SDP problems, different search directions have been proposed: HKM direction (Helmberg-Rendl-Vanderbei-Wolkowicz/Kojima-Shindoh-Hara/Monteiro) [2, 3, 4], AHO direction (Alizadeh-Haeberly-Overton) [1], and NT direction (Nesterov-Todd direction) [5]. SDP algorithms [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15] adopt different directions and have different global and local convergence rates. For example, Potra et al. [14] developed a predictor-corrector infeasible IPM algorithm using the HKM direction with polynomial global convergence and superlinear local convergence. Kojima et al. [8] proposed a modified algorithm repeating a corrector step to achieve tangential convergence by which the fast-centering assumption in [14] can be removed. Later, Potra et al. [13] proved the local convergence of the algorithms without nondegeneracy assumption used in [8].

Many practical SDP packages [16, 17] exploit a block diagonal structure to avoid unnecessary computations for off-diagonal blocks. More recently, Fukuda et al. [18] developed two algorithms: one to convert sparse SDP into multiple small block variables using positive semidefinite matrix completion and the other to incorporate the method into primal-dual IPM. Our constraint-reduced algorithm adaptively excludes unnecessary block constraints to save computational costs. Therefore, from the algorithmic perspective, our method is potentially applicable to the previous methods exploiting the block structure.

There have been many efforts to apply constraint reduction to optimization problems, for example, linear programming (LP) [19, 20, 21, 22, 23, 24, 25], support vector machine (SVM) [26, 27], and quadratic programming (QP) [28] . Most recently, Park and O'Leary [29, 30] established a constraint-reduced predictor-corrector algorithm for block diagonal SDP by using a constraint-reduced HKM direction. They proved polynomial global convergence of their algorithm, but they did not show its superlinear convergence. This paper extends their study in that we develop a new constraint-reduced algorithm having not only polynomial global convergence but also superlinear local convergence. We utilize the idea of repeating the corrector step to achieve tangential convergence, so our algorithm is a constraint-reduced version of Kojima et al. and Potra et al. [8, 13]. To the author's best knowledge, this is the first constraint-reduced IPM for SDP that achieves superlinear local convergence.

This paper is organized as follows. In Sect. 2, we introduce our block-constrained SDP problem. In Sect. 3, we summarize the constraint-reduced predictor-corrector algorithm, established by Park and O'Leary [29, 30], and its polynomial convergence property, that will be used later for local convergence analysis. In Sect. 4, we propose a new algorithm adopting new constraint reduction criteria and show its superlinear convergence. In Sect. 5, through the numerical experiments, we demonstrate how effectively constraint reduction saves computational costs. Finally, in Sect. 6, we conclude the paper with summarizing this study and future work.

2 Block-constrained Semidefinite Programming

We frequently use the notation in Table 1 in this paper. The primal and dual SDP problems are defined as

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

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

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:

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

where $\boldsymbol{A}_{ij} \in \mathcal{S}^{n_j}$, $\boldsymbol{C}_j \in \mathcal{S}^{n_j}$ for $i = 1, \dots, m$ and $j = 1, \dots, p$. Many SDP problems have the block diagonal structures. See [31, Section 4.4.1] for examples. We let $\boldsymbol{X}_j$ and $\boldsymbol{Z}_j$ denote the corresponding block of $\boldsymbol{X}$ and $\boldsymbol{Z}$. By the block structure, we can decompose each constraint in (1) and (2) into p independent block constraints. The block diagonal SDP is of interest because the Schur complement can be broken down into matrices corresponding to each diagonal block. Constraint-reduced algorithms reduce computations by ignoring unnecessary matrices. We will discuss the relation of block structure and constraint reduction on Schur complement matrix in Section 3.

In the iteration of IPM, some block constraints make relatively insignificant contributions to determine a search direction. We call these *inactive* constraints. Constraint reduction has a goal to save computational cost by ignoring such block constraints while preserving the convergence property of the original IPM. We will introduce rigorous criteria to select the *inactive* block constraints to achieve superlinear convergence later in Sect. 4.

Without loss of generality, we assume that we can partition the matrices into the *active* block and the *inactive* block with appropriate reordering, so

$$\boldsymbol{A}_i = \begin{bmatrix} \widehat{\boldsymbol{A}}_i & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{A}}_i \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} \widehat{\boldsymbol{C}} & \boldsymbol{0} \\ \boldsymbol{0} & \widetilde{\boldsymbol{C}} \end{bmatrix},$$

where $\widehat{\boldsymbol{A}}_i$ and $\widehat{\boldsymbol{C}}$ contain the *active* blocks while $\widetilde{\boldsymbol{A}}_i$ and $\widetilde{\boldsymbol{C}}$ contain the *inactive* blocks. In a similar way, we can also partition the optimizing matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ into $\widehat{\boldsymbol{X}}$, $\widetilde{\boldsymbol{X}}$, $\widehat{\boldsymbol{Z}}$, and $\widetilde{\boldsymbol{Z}}$. In addition, let $\mathcal{A} \in \mathbb{R}^{m \times n^2}$ denote the matrix whose ith row is $\operatorname{vec}(\boldsymbol{A}_i)^T$. Then the matrix $\mathcal{A}$ can be partitioned into the *active* part $\widehat{\mathcal{A}}$ and the *inactive* part $\widetilde{\mathcal{A}}$, whose rows are $\operatorname{vec}\left(\widehat{\boldsymbol{A}}_i\right)^T$ and $\operatorname{vec}\left(\widetilde{\boldsymbol{A}}_i\right)^T$, so $\mathcal{A} = [\widehat{\mathcal{A}}, \widetilde{\mathcal{A}}]$. We also define a matrix $\mathcal{A}_j$ whose rows are $\operatorname{vec}(\boldsymbol{A}_{ij})^T$ for $j = 1, \dots, q$. We have computational benefits by using $\operatorname{svec}()$ instead of $\operatorname{vec}()$

Table 1: Notation for the SDP.

$\mathcal{S}^n$	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
$\boldsymbol{X} \succ \boldsymbol{0}$	A positive definite matrix
$\boldsymbol{X} \succeq \boldsymbol{0}$	A positive semidefinite matrix
$\boldsymbol{A} \bullet \boldsymbol{B} = \operatorname{tr}\left(\boldsymbol{A}\boldsymbol{B}^T\right)$	The dot-product of matrices
$\mu = (\boldsymbol{X} \bullet \boldsymbol{Z})/n$	The duality gap
$\boldsymbol{x} = \operatorname{vec}(\boldsymbol{X})$	The vectorization of a given matrix $\boldsymbol{X}$, a stack of columns of $\boldsymbol{X}^T$
$\boldsymbol{k} = \operatorname{svec}(\boldsymbol{K})$	The symmetric vectorization of a given symmetric matrix $\boldsymbol{K}$,
$\operatorname{mat}(\boldsymbol{x})$	The inverse of $\operatorname{vec}(\boldsymbol{X})$
$\operatorname{symm}(\boldsymbol{X}) = \frac{1}{2}(\boldsymbol{X} + \boldsymbol{X}^T)$	The symmetric part of $\boldsymbol{X}$
$\boldsymbol{G} \otimes \boldsymbol{H}$	Kronecker product of matrices $\boldsymbol{G}$ and $\boldsymbol{H}$
$\Vert\boldsymbol{A}\Vert$	The 2-norm of a matrix $\boldsymbol{A}$
$\Vert\boldsymbol{A}\Vert_F = (\sum_{ij} a_{ij}^2)^{1/2}$	The Frobenius norm of a matrix $\boldsymbol{A}$
$\boldsymbol{G}^k = O(1)$	$\exists\ \Gamma > 0$ such that $\left\Vert\boldsymbol{G}^k\right\Vert \le \Gamma$
$\boldsymbol{G}^k = \Omega(1)$	$\exists\ \Gamma > 0$ such that $1/\Gamma \le \left\Vert\boldsymbol{G}^k\right\Vert \le \Gamma$
$\boldsymbol{G}^k = O(\eta_k)$	$\boldsymbol{G}^k/\eta_k = O(1)$
$\boldsymbol{G}^k = \Omega(\eta_k)$	$\boldsymbol{G}^k/\eta_k = \Omega(1)$

because we can avoid duplicate computations for off-diagonal elements of symmetric $\boldsymbol{A}_i$. For simplicity of notations in the following equations, we use $\operatorname{vec}()$ for the rest of this paper.

3 Constraint-reduced SDP Method

3.1 Constraint-reduced HKM direction

In this section, we introduce a constraint-reduced HKM direction. Throughout this paper, we assume the Slater condition.

Assumption 3.1 (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 the assumption, $(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z})$ is a solution of (1) and (2) if and only if

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

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

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

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

For a given current iterate $(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z})$, the primal residual, dual residual, and complementarity residuals are defined as

$$\begin{aligned} r_{pi} &= b_i - \boldsymbol{A}_i \bullet \boldsymbol{X} \quad \text{for } i = 1, \ldots, m, \\ \boldsymbol{R}_d &= \boldsymbol{C} - \boldsymbol{Z} - \sum_{i=1}^{m} y_i \boldsymbol{A}_i, \\ \boldsymbol{R}_c &= \overline{\mu} \boldsymbol{I} - \boldsymbol{X}\boldsymbol{Z}, \end{aligned}$$

where $\overline{\mu}$ is a target duality gap.

The HKM direction $(\Delta\boldsymbol{X}, \Delta\boldsymbol{y}, \Delta\boldsymbol{Z}) \in \mathcal{S}^n \times \mathbb{R}^n \times \mathcal{S}^n$ can be found by solving the following equations:

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

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

$$\text{symm} \left(\boldsymbol{Z}^{1/2} (\boldsymbol{X}\Delta\boldsymbol{Z} + \Delta\boldsymbol{X}\boldsymbol{Z}) \boldsymbol{Z}^{-1/2} \right) = \overline{\mu}\boldsymbol{I} - \boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2}. \tag{9}$$

Kojima et al. [3, Theorem 4.2] and Monteiro [4, Lemma 2.1 ff] showed that the equations above have a unique solution for $(\boldsymbol{X}, \boldsymbol{Z}) \in \mathcal{S}^n_{++} \times \mathcal{S}^n_{++}$.

On the other hand, we can obtain a solution of (7)–(9) by solving a reduced equation,

$$\boldsymbol{M}\Delta\boldsymbol{y} = \boldsymbol{g}, \tag{10}$$

where

$$\begin{aligned} \boldsymbol{M} &= \widehat{\boldsymbol{M}} + \widetilde{\boldsymbol{M}}, \\ \widehat{\boldsymbol{M}} &= \widehat{\mathcal{A}} (\widehat{\boldsymbol{X}} \otimes \widehat{\boldsymbol{Z}}^{-1}) \widehat{\mathcal{A}}^T, \\ \widetilde{\boldsymbol{M}} &= \widetilde{\mathcal{A}} (\widetilde{\boldsymbol{X}} \otimes \widetilde{\boldsymbol{Z}}^{-1}) \widetilde{\mathcal{A}}^T, \\ \boldsymbol{g} &= r_p + \mathcal{A}(\boldsymbol{X} \otimes \boldsymbol{Z}^{-1}) \boldsymbol{r}_d - \mathcal{A}(\boldsymbol{I} \otimes \boldsymbol{Z}^{-1}) \boldsymbol{r}_c. \end{aligned}$$

When solving (10), the most computationally expensive part is to compute the Schur complement matrix $\boldsymbol{M}$, $O(mn^3 + m^2n^2)$, which is even more expensive than its Cholesky decomposition $O(n^3)$, as Alizadeh et al. [1] explained. For a sophisticated flop count analysis of constructing the Schur complement matrix, see Fujisawa et al. [32].

To save the computational cost, Park and O'Leary [29, 30] proposed to solve the equations below,

$$\widehat{\boldsymbol{M}} \quad \Delta\boldsymbol{y} \quad = \boldsymbol{g}, \tag{11}$$

$$\Delta\boldsymbol{X} = \text{symm} \left(\text{mat} \left(\begin{bmatrix} (\boldsymbol{I} \otimes \widehat{\boldsymbol{Z}}^{-1}) \boldsymbol{r}_c - (\widehat{\boldsymbol{X}} \otimes \widehat{\boldsymbol{Z}}^{-1}) (\boldsymbol{r}_d - \widehat{\mathcal{A}}^T \Delta\boldsymbol{y}) \\ (\boldsymbol{I} \otimes \widetilde{\boldsymbol{Z}}^{-1}) \boldsymbol{r}_c - (\widetilde{\boldsymbol{X}} \otimes \widetilde{\boldsymbol{Z}}^{-1}) \boldsymbol{r}_d \end{bmatrix} \right) \right) \tag{12}$$

$$\Delta\boldsymbol{Z} = \boldsymbol{R}_d - \sum_{i=1}^{m} \boldsymbol{A}_i \Delta y_i. \tag{13}$$

Compared to (10), (11) replaces $\boldsymbol{M}$ with $\widehat{\boldsymbol{M}}$ to avoid the computation of $\widetilde{\boldsymbol{M}}$, by which the constraints associated with $\widetilde{\boldsymbol{M}}$ are implicitly reduced, so we call it a *constraint-reduced* HKM direction. Further discussion about efficient ways of updating $\widehat{\boldsymbol{M}}$ and solving linear system (11) can be found in Park and O'Leary [29, Section 2.3]. The following lemma explains how constraint reduction affects the original HKM direction.

Lemma 3.1 (Constraint-reduced HKM direction)**.** *A solution* $(\Delta\mathbf{X}, \Delta\mathbf{y}, \Delta\mathbf{Z}) \in \mathcal{S}^n \times \mathbb{R}^n \times \mathcal{S}^n$ *of (11)–(13) satisfies (7), (8), and a perturbed complementarity equation,*

$$\operatorname{symm}\left(\mathbf{Z}^{1/2}(\mathbf{X}\Delta\mathbf{Z} + \Delta\mathbf{X}\mathbf{Z})\mathbf{Z}^{-1/2}\right) = \overline{\mu}\mathbf{I} - \mathbf{Z}^{1/2}(\mathbf{X} + \Delta\mathbf{X}_\epsilon)\mathbf{Z}^{1/2}, \qquad (9^*)$$

where

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

Proof. See Park and O'Leary [30, Lemma 2.1 ff]. □

As a result of constraint reduction, a perturbation term $\Delta\boldsymbol{X}_\epsilon$ is added in the complementarity equation (9*).

3.2 Constraint-reduced Algorithm

We summarize Algorithm SDP:Reduced, the constraint-reduced predictor-corrector IPM, in which Park and O'Leary [30] used the constraint-reduced HKM direction. The algorithm defines a constant ρ based on unknown optimal $\boldsymbol{X}^*$ and $\boldsymbol{Z}^*$. Our convergence result holds for any $\rho > 0$, but in practical implementation, we can choose ρ based on the given input matrices as discussed in Toh et al. [16, Section 3.4]. We will develop a modified algorithm in Sect. 4. First, we establish a few essential notations to explain the algorithm. We define a set $\mathcal{F}$ of feasible solutions, a set $\mathcal{F}^*$ of optimal solutions, and a neighborhood $\mathcal{N}(\gamma, \tau)$ of the central path as

$$\begin{aligned} \mathcal{F} &= \{(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z}) \in \mathcal{S}^n_+ \times \mathbb{R}^m \times \mathcal{S}^n_+ : \ (\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z}) \text{ satisfies (3) and (4). } \}, \\ \mathcal{F}^* &= \{(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z}) \in \mathcal{F} : \boldsymbol{X} \bullet \boldsymbol{Z} = 0\}, \\ \mathcal{N}(\gamma, \tau) &= \{(\boldsymbol{X}, \boldsymbol{Z}) \in \mathcal{S}^n_{++} \times \mathcal{S}^n_{++} : \left\| \boldsymbol{Z}^{1/2}\boldsymbol{X}\boldsymbol{Z}^{1/2} - \tau\boldsymbol{I} \right\|_F \le \gamma\tau\}. \end{aligned}$$

The algorithm uses two fixed positive parameters α and β satisfying

$$\frac{\beta^2}{2(1-\beta)^2} < \alpha < \beta \le \frac{\beta}{1-\beta} < 1. \qquad (14)$$

For example, we can choose $(\alpha, \beta) = (0.17, 0.3)$. For a starting point, we can use a standard choice of $(\boldsymbol{X}^0, \boldsymbol{y}^0, \boldsymbol{Z}^0) = (\rho_p\boldsymbol{I}, \boldsymbol{0}, \rho_d\boldsymbol{I})$ for $\rho_p > 0$ and $\rho_d > 0$, which satisfies $(\boldsymbol{X}^0, \boldsymbol{Z}^0) \in \mathcal{N}(\alpha, \tau_0)$ that is required in Step 2 of the algorithm. In

Algorithm 1 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 (14); convergence tolerance τ^*; $\omega \in (0, 0.5)$ for the perturbation bound of the primal direction in the predictor step.
2. Choose $(\boldsymbol{X}^0, \boldsymbol{y}^0, \boldsymbol{Z}^0)$ such that $(\boldsymbol{X}^0, \boldsymbol{Z}^0) \in \mathcal{N}(\alpha, \tau_0)$, and set $\tau = \tau_0$.
3. Repeat until $\tau < \tau^*$: For $k = 0, 1, \ldots,$

   (a) **(Predictor step):** Set $(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z}) = (\boldsymbol{X}^k, \boldsymbol{y}^k, \boldsymbol{Z}^k)$, $\boldsymbol{r}_p = \boldsymbol{r}_p^k$, $\boldsymbol{r}_d = \boldsymbol{r}_d^k$, and $\tau = \tau_k$.

      i. Find $\widehat{\boldsymbol{M}}_{\widehat{p}} = \sum_{j=1}^{\widehat{p}} \mathcal{A}_j(\boldsymbol{X}_j \otimes \boldsymbol{Z}_j^{-1})\mathcal{A}_j^T$ such that $\widehat{p} \le p$ where p is the number of blocks, $\widehat{\boldsymbol{M}}_{\widehat{p}}$ is full-rank, and Condition 3.1 is satisfied.

      ii. Solve (11) with $\widehat{\boldsymbol{M}} = \widehat{\boldsymbol{M}}_{\widehat{p}}$ and $\overline{\mu} = 0$ in $\boldsymbol{r}_c$ to find $(\Delta\boldsymbol{X}, \Delta\boldsymbol{y}, \Delta\boldsymbol{Z})$ satisfying (7), (8), and (9*). Choose a step length $\overline{\theta} \in [\widehat{\theta}, \breve{\theta}]$ defined by (19) and (20).
      Set $\overline{\boldsymbol{X}} = \boldsymbol{X} + \overline{\theta}\Delta\boldsymbol{X}$, $\overline{\boldsymbol{y}} = \boldsymbol{y} + \overline{\theta}\Delta\boldsymbol{y}$, $\overline{\boldsymbol{Z}} = \boldsymbol{Z} + \overline{\theta}\Delta\boldsymbol{Z}$.

      iii. If $\overline{\theta} = 1$, terminate the iteration with optimal solution $(\overline{\boldsymbol{X}}, \overline{\boldsymbol{y}}, \overline{\boldsymbol{Z}})$.

   (b) **(Corrector step): Set** $\overline{\tau} = (1 - \overline{\theta})\tau$

      i. Find $\widehat{\boldsymbol{M}}_{\widehat{p}} = \sum_{j=1}^{\widehat{p}} \mathcal{A}_j(\overline{\boldsymbol{X}}_j \otimes \overline{\boldsymbol{Z}}_j^{-1})\mathcal{A}_j^T$ such that $\widehat{p} \le p$ where p is the number of blocks, $\widehat{\boldsymbol{M}}_{\widehat{p}}$ is full-rank, and Condition 3.2 is satisfied.

      ii. Solve (11) with $\widehat{\boldsymbol{M}} = \widehat{\boldsymbol{M}}_{\widehat{p}}$, $\boldsymbol{r}_p = \boldsymbol{0}$, $\boldsymbol{r}_d = \boldsymbol{0}$, and $\overline{\mu} = \overline{\tau}$ in $\boldsymbol{r}_{\widehat{p}}$ to find $(\Delta\overline{\boldsymbol{X}}, \Delta\overline{\boldsymbol{y}}, \Delta\overline{\boldsymbol{Z}})$ satisfying (7), (8), and (9*). Take a full step as
      $\overline{\boldsymbol{X}}^+ = \overline{\boldsymbol{X}} + \Delta\overline{\boldsymbol{X}}$, $\overline{\boldsymbol{y}}^+ = \overline{\boldsymbol{y}} + \Delta\overline{\boldsymbol{y}}$, $\overline{\boldsymbol{Z}}^+ = \overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{Z}}$.

   (c) Update $(\boldsymbol{X}^{k+1}, \boldsymbol{y}^{k+1}, \boldsymbol{Z}^{k+1}) = (\overline{\boldsymbol{X}}^+, \overline{\boldsymbol{y}}^+, \overline{\boldsymbol{Z}}^+)$,
   $\boldsymbol{r}_p^{k+1} = \boldsymbol{b} - \mathcal{A}\boldsymbol{x}^{k+1}$, $\boldsymbol{r}_d^{k+1} = \boldsymbol{c} - \boldsymbol{z}^{k+1} - \mathcal{A}^T\boldsymbol{y}^{k+1}$, and $\tau_{k+1} = \overline{\tau}$,

---

the algorithm, both predictor and corrector steps solve the constraint-reduced equations (11)–(13), but with different settings of $\boldsymbol{r}_p$, $\boldsymbol{r}_d$, and $\boldsymbol{r}_c$. First, for a given $(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z}) \in \mathcal{N}(\alpha, \tau)$, the predictor step finds a solution $(\Delta\boldsymbol{X}, \Delta\boldsymbol{y}, \Delta\boldsymbol{Z})$ by setting $\overline{\mu} = 0$ in $\boldsymbol{r}_c$ and updates the iterate as

$$(\overline{\boldsymbol{X}}, \overline{\boldsymbol{y}}, \overline{\boldsymbol{z}}) = (\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z}) + \overline{\theta}(\Delta\boldsymbol{X}, \Delta\boldsymbol{y}, \Delta\boldsymbol{Z}),$$

where $\overline{\theta}$ is the predictor's step size. Second, by using $(\overline{\boldsymbol{X}}, \overline{\boldsymbol{y}}, \overline{\boldsymbol{z}})$ for $(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z})$ in the equations, the corrector step finds $(\Delta\overline{\boldsymbol{X}}, \Delta\overline{\boldsymbol{y}}, \Delta\overline{\boldsymbol{Z}})$ by setting

$(\boldsymbol{r}_p, \boldsymbol{r}_d, \overline{\mu}) = (\boldsymbol{0}, \boldsymbol{0}, (1-\overline{\theta})\tau)$ and updates the iterate as

$$(\boldsymbol{X}^+, \boldsymbol{y}^+, \boldsymbol{z}^+) = (\overline{\boldsymbol{X}}^+, \overline{\boldsymbol{y}}^+, \overline{\boldsymbol{z}}^+) = (\overline{\boldsymbol{X}}, \overline{\boldsymbol{y}}, \overline{\boldsymbol{Z}}) + (\Delta\overline{\boldsymbol{X}}, \Delta\overline{\boldsymbol{y}}, \Delta\overline{\boldsymbol{Z}}).$$

For the predictor step's direction $(\Delta\boldsymbol{X}, \Delta\boldsymbol{y}, \Delta\boldsymbol{Z})$, we define

$$\delta := \frac{1}{\tau}\left\|\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}\Delta\boldsymbol{Z}\boldsymbol{Z}^{-1/2}\right\|_F. \tag{15}$$

Additionally , let $\Delta\boldsymbol{X}_\epsilon$ and $\Delta\overline{\boldsymbol{X}}_\epsilon$ be the perturbation matrices in (9*), the former for the predictor step and the latter for the corrector step. The perturbations for constraint reduction are quantified as

$$\begin{aligned} \delta_\epsilon &:= \frac{1}{\tau}\left\|\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}_\epsilon\boldsymbol{Z}^{1/2}\right\|_F, && (16)\\ \overline{\delta}_\epsilon &:= \frac{1}{\tau}\left\|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}_\epsilon\overline{\boldsymbol{Z}}^{1/2}\right\|_F. && (17)\end{aligned}$$

Based on the perturbations, conditions below will guide us to find how many block constraints should be included in constructing Schur complement matrix. We will say that the included block are *active* and the others are *inactive*.

Condition 3.1 (Requirement on predictor step's perturbation)**.** *For a given constant ω,*

$$\delta_\epsilon \le \frac{\omega}{\tau}\delta_x,$$

where $0 < \omega < 0.5$ and $\delta_x := \left\|\mathbf{Z}^{1/2}\Delta\mathbf{X}\mathbf{Z}^{1/2}\right\|_F$.

Condition 3.2 (Requirement on corrector step's perturbation)**.**

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

where $s = \beta^2 - \beta + 1$ and $t = 2\alpha(1-\beta)^2 - \beta^2$.

For the construction of the Schur complement matrix in 3-(a)-i and 3-(b)-i of Algorithm SDP:Reduced, we may incrementally build the matrix $\widehat{\boldsymbol{M}}_{\widehat{p}}$ as

$$\widehat{\boldsymbol{M}}_{\widehat{p}} \leftarrow \widehat{\boldsymbol{M}}_{\widehat{p}-1} + \mathcal{A}_{\widehat{p}}(\boldsymbol{X}_{\widehat{p}} \otimes \boldsymbol{Z}_{\widehat{p}}^{-1})\mathcal{A}_{\widehat{p}}^T,$$

until Condition 3.1 or Condition 3.2 is satisfied. In Algorithm ExploratoryConstruction, we propose two adaptations to improve efficiency. First, Lemma 3.1 suggests adding the blocks in order of decreasing $\left\|\boldsymbol{X}_j \otimes \boldsymbol{Z}_j^{-1}\right\|_F$, but for efficiency, we add in order of decreasing $\|\boldsymbol{X}_j\|_F \left\|\boldsymbol{Z}_j^{-1}\right\|_F$. Second, checking the condition for each increment is too expensive, so we make at most four tries. We first form $\widehat{\boldsymbol{M}}_{\widehat{p}}$ for a $\widehat{p}$ somewhat smaller than that used at the previous iteration. If that matrix fails to satisfy the condition, we add in blocks to try the previous $\widehat{p}$ and then a somewhat larger $\widehat{p}$. If all of those fail, we use all blocks, no reduction. This exploratory construction uses the fact that *inactive*

Algorithm 2 ExploratoryConstruction

Exploratory Construction of Reduced Schur Complement Matrix

1. Input: $r_s < 1$, $r_e > 1$, $\mathcal{A}_j$, $\boldsymbol{X}_j$, and $\boldsymbol{Z}_j$ for $j = 1, \dots, p$.
   The number of *active* blocks $\widehat{p}_0$ at the previous iteration (initially $\widehat{p}_0 = p$).
   Constraint reduction condition C: Condition 3.1 or Condition 3.2.
2. Sort the block constraints in decreasing order of $\|\boldsymbol{X}_j\|_F \left\|\boldsymbol{Z}_j^{-1}\right\|_F$.
3. **(Shrink step)** Build $\widehat{\boldsymbol{M}}_{\widehat{p}}$ for $\widehat{p} = \max(1, r_s\widehat{p}_0)$. If $\widehat{\boldsymbol{M}}_{\widehat{p}}$ satisfies C, return $\widehat{\boldsymbol{M}}_{\widehat{p}}$.
4. **(Rollback step)** Incrementally build $\widehat{\boldsymbol{M}}_{\widehat{p}}$ for $\widehat{p} = \widehat{p}_0$ from $\widehat{\boldsymbol{M}}_{\widehat{p}}$ of step 3. If $\widehat{\boldsymbol{M}}_{\widehat{p}}$ satisfies C, return $\widehat{\boldsymbol{M}}_{\widehat{p}}$.
5. **(Expand step)** Incrementally build $\widehat{\boldsymbol{M}}_{\widehat{p}}$ for $\widehat{p} = \min(\widehat{p}, r_e\widehat{p}_0)$ from $\widehat{\boldsymbol{M}}_{\widehat{p}}$ of step 4. If $\widehat{\boldsymbol{M}}_{\widehat{p}}$ satisfies C, return $\widehat{\boldsymbol{M}}_{\widehat{p}}$.
6. **(Full Schur complement)** Incrementally build $\widehat{\boldsymbol{M}}_{\widehat{p}}$ for $\widehat{p} = p$ from $\widehat{\boldsymbol{M}}_{\widehat{p}}$ of step 5, and return $\widehat{\boldsymbol{M}}_{\widehat{p}}$.

blocks at the previous iteration tend to be *inactive* again. Later in Lemma 4.4, asymptotic behavior of $\boldsymbol{X}$ and $\boldsymbol{Z}$ will disclose this property.

Park and O'Leary [29, 30] presented a range for the predictor's step size $\overline{\theta}$ as

$$\widehat{\theta} \leq \overline{\theta} \leq \breve{\theta}, \tag{18}$$

where

$$\widehat{\theta} = \frac{(\alpha - \beta - \delta_\epsilon) + \sqrt{(\alpha - \beta - \delta_\epsilon)^2 + 4\delta(\beta - \alpha)}}{2\delta}, \tag{19}$$

$$\breve{\theta} = \max\{\tilde{\theta} \in [0,1] : (\boldsymbol{X} + \theta\Delta\boldsymbol{X}, \boldsymbol{y} + \theta\Delta\boldsymbol{y}, \boldsymbol{Z} + \theta\Delta\boldsymbol{Z}) \in \mathcal{N}(\beta, (1-\theta)\tau), \forall\theta \in [0, \tilde{\theta}]\}. \tag{20}$$

They proved polynomial convergence using the following results.

Lemma 3.2 (After Predictor Step)**.** *For* $(\mathbf{X}, \mathbf{Z}) \in \mathcal{N}(\alpha, \tau)$*, after the predictor step,*

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

In addition, if $\overline{\theta} < 1$*,*

$$(\overline{\mathbf{X}}, \overline{\mathbf{Z}}) \in \mathcal{N}(\beta, \tau^+),$$

where $\tau^+ = \overline{\tau} = (1 - \overline{\theta})\tau$*. Otherwise,* $(\overline{\mathbf{X}}, \overline{\mathbf{y}}, \overline{\mathbf{Z}})$ *is a solution of the SDP.*

Proof. See Park and O'Leary [30, Lemma 3.1–3.2]. □

Algorithm 3 SDP:ReducedLocal

Modified Primal-Dual Infeasible Constraint-Reduced Predictor-Corrector Algorithm for Block Diagonal SDP with repeating corrector step.
Step 1 and 2 are same as SDP:Reduced but using α, β, and ζ satisfying (14) and (23).

3. Repeat until $\tau < \tau^*$: For $k = 0, 1, \ldots,$

   (a) **(Predictor step):** Same as SDP:Reduced but replacing Condition 3.1 with Condition 4.1.

   (b) **(Corrector step):** Initially, set $(\overline{\boldsymbol{X}}^0, \overline{\boldsymbol{y}}^0, \overline{\boldsymbol{Z}}^0) = (\overline{\boldsymbol{X}}, \overline{\boldsymbol{y}}, \overline{\boldsymbol{Z}})$.
   Repeat 3(b) of SDP:Reduced until $(\overline{\boldsymbol{X}}^q, \overline{\boldsymbol{Z}}^q) \in \mathcal{N}(\min(\overline{\tau}^\sigma, \alpha), \overline{\tau})$ for $\sigma > 0$:
   For $q = 0, 1, \ldots,$ update $(\overline{\boldsymbol{X}}^q, \overline{\boldsymbol{y}}^q, \overline{\boldsymbol{Z}}^q) = (\overline{\boldsymbol{X}}^+, \overline{\boldsymbol{y}}^+, \overline{\boldsymbol{Z}}^+)$.

   (c) Update $(\boldsymbol{X}^{k+1}, \boldsymbol{y}^{k+1}, \boldsymbol{Z}^{k+1}) = (\overline{\boldsymbol{X}}^+, \overline{\boldsymbol{y}}^+, \overline{\boldsymbol{Z}}^+)$,
   $\boldsymbol{r}_p^{k+1} = \boldsymbol{b} - \mathcal{A}\boldsymbol{x}^{k+1}$, $\boldsymbol{r}_d^{k+1} = \boldsymbol{c} - \boldsymbol{z}^{k+1} - \mathcal{A}^T\boldsymbol{y}^{k+1}$, and $\tau_{k+1} = \overline{\tau}$.

Lemma 3.3 (After Corrector Step)**.** *For* $(\overline{\mathbf{X}}, \overline{\mathbf{Z}}) \in \mathcal{N}(\beta, \tau^+)$*, after the corrector steps,*

$$(\mathbf{X}^+, \mathbf{Z}^+) \in \mathcal{N}(\alpha, \tau^+).$$

Proof. See Park and O'Leary [30, Lemma 3.3]. □

4 Algorithm with Polynomial and Superlinear Convergence

In this section, we propose a new constraint-reduced algorithm, as presented in Algorithm SDP:ReducedLocal, and show its superlinear local convergence. The new algorithm replaces the previous constraint reduction criteria, Condition 3.1 and 3.2, with Condition 4.1 and 4.2. Plus, the corrector step has its own iteration so as to obtain a tangential convergence to a central path in the sense that

$$\frac{\left\| (\boldsymbol{Z}^k)^{1/2} \boldsymbol{X}^k (\boldsymbol{Z}^k)^{1/2} - \tau_k \boldsymbol{I} \right\|_F}{\tau_k} \to 0.$$

We define a few parameters to explain the new algorithm. We use indices k and q to denote the parameters at the k-th outer iteration and at the corrector step's q-th inner iteration. However, when the meaning of notation is obvious by the context, we omit the indices to simplify notation. First, at the k-th iteration, we define ϕ_k as

$$\phi_k := \max\left(\frac{\left\| (\boldsymbol{Z}^k)^{1/2} \boldsymbol{X}^k (\boldsymbol{Z}^k)^{1/2} - \tau_k \boldsymbol{I} \right\|_F}{\tau_k}, \sqrt{\tau_k} \right). \tag{21}$$

Second, we let $(\overline{\boldsymbol{X}}^q, \overline{\boldsymbol{y}}^q, \overline{\boldsymbol{Z}}^q)$ denote the q-th iterate of the corrector step's iteration in SDP:ReducedLocal, and we define its relative distance $\overline{\gamma}_q$ to the central path as

$$\overline{\gamma}_q := \frac{\left\| (\overline{\boldsymbol{Z}}^q)^{1/2} \overline{\boldsymbol{X}}^q (\overline{\boldsymbol{Z}}^q)^{1/2} - \overline{\tau} \boldsymbol{I} \right\|_F}{\overline{\tau}}, \tag{22}$$

where $\overline{\tau} = (1 - \overline{\theta})\tau$. By the definition, $(\overline{\boldsymbol{X}}^q, \overline{\boldsymbol{Z}}^q) \in \mathcal{N}(\overline{\gamma}_q, \overline{\tau})$ for $(\overline{\boldsymbol{X}}^q, \overline{\boldsymbol{Z}}^q) \in \mathcal{S}^n_{++} \times \mathcal{S}^n_{++}$ Third, we define a constant ζ such that

$$0 < \zeta < \frac{1}{\beta} - 2. \tag{23}$$

Because $0 < \beta < 0.5$ by (14), such a ζ exists. Using the parameters, we introduce the new constraint reduction criteria.

Condition 4.1 (New requirement on predictor step's perturbation)**.** *For a given positive constant* C_ϵ*,*

$$\delta_\epsilon \le \min(\frac{\omega}{\tau_k}\delta_x, C_\epsilon \phi_k),$$

where ω *and* δ_x *are as defined in Condition 3.1.*

Condition 4.2 (New requirement on corrector step's perturbation)**.**

$$\overline{\delta}_\epsilon < \zeta \overline{\gamma}_q^2 (1 - \overline{\theta}).$$

As Lemma 3.1 reveals, we can reduce δ_ϵ and $\overline{\delta}_\epsilon$ as we include more constraints in the *active* set. When taking all constraints *active*, δ_ϵ and $\overline{\delta}_\epsilon$ become zero and no constraints are reduced. Therefore, we can always satisfy Condition 4.1 and 4.2 by taking enough block constraints.

4.1 Polynomial Convergence

In this section, we show SDP:ReducedLocal has polynomial convergence, since Lemma 3.2 and 3.3 hold for SDP:Reduced. First, it is obvious that Condition 4.1 is stricter than Condition 3.1, so the result of Lemma 3.2 also holds for SDP:ReducedLocal. Second, the stopping condition of the corrector step guarantees that

$$(\boldsymbol{X}^+, \boldsymbol{Z}^+) \in \mathcal{N}(\min(\overline{\tau}^\sigma, \alpha), \overline{\tau}), \tag{24}$$

which is even stricter than the result of Lemma 3.3. To meet the stopping condition, the corrector step's iteration must converge and its convergence rate needs to be fast enough to have the algorithm practical. The following lemmas show that the corrector step's iteration converges with a quadratic rate. Note that initially $\overline{\gamma} \le \beta$ because $(\overline{\mathbf{X}}, \overline{\mathbf{Z}}) \in (\beta, \overline{\tau})$ by Lemma 3.2.

Lemma 4.1 (Quadratic convergence of $\overline{\gamma}$)**.** *Let* $\overline{\gamma}$ *be the relative distance to the central path as defined in* (22)*. For* $\overline{\gamma} \le \beta$*, after each iteration of the corrector step 3(b),*

$$(\overline{\mathbf{X}}^+, \overline{\mathbf{Z}}^+) \in \mathcal{N}(\overline{\gamma}^+, \overline{\tau}),$$

and

$$\overline{\gamma}^+ < \frac{\overline{\gamma}^2}{\beta} \le \overline{\gamma},$$

so $\overline{\gamma}$ *quadratically converges to* 0.

Proof. See Appendix. □

Lemma 4.2. *The corrector step 3(b) in SDP:ReducedLocal requires* $O\left(\log\left(\log\left(1/\overline{\tau}^{\sigma}\right)\right)\right)$ *iterations to satisfy the stopping condition (24),* $(\overline{\mathbf{X}}^q, \overline{\mathbf{Z}}^q) \in \mathcal{N}(\min(\overline{\tau}^{\sigma}, \alpha), \overline{\tau})$.

Proof. See Appendix. □

By Lemma 4.2, the stopping condition (24) of the corrector step can be satisfied by a finite number of iterations, so the iterate tangentially approaches the central path by the repeated corrector step. From Lemma 3.2 and (24), SDP:ReducedLocal has polynomial global convergence like SDP:Reduced.

Theorem 4.1 (Polynomial Global Convergence)**.** *After k-th iteration of Algorithm SDP:ReducedLocal, the iterate* $(\mathbf{X}^k, \mathbf{y}^k, \mathbf{Z}^k)$ *satisfies*

$$\mathbf{r}_p^k = \psi_k \mathbf{r}_p^0, \quad \mathbf{r}_d^k = \psi_k \mathbf{r}_d^0, \quad \tau_k = \psi_k \tau_0, \tag{25}$$

$$(\mathbf{X}^k, \mathbf{Z}^k) \in \mathcal{N}(\alpha, \tau_k), \quad \text{and} \tag{26}$$

$$(1 - \frac{\alpha}{\sqrt{n}})\tau_k \le \mu_k = \frac{1}{n}(\mathbf{X}^k \bullet \mathbf{Z}^k) \le (1 + \frac{\alpha}{\sqrt{n}})\tau_k, \tag{27}$$

where $\psi_k = \prod_{i=0}^{k-1}(1 - \overline{\theta}_i)$.
In addition, the step length $\overline{\theta}_k$ *is bounded away from zero, so Algorithm SDP:ReducedLocal is globally convergent. Defining* $\epsilon_k := \max(\mathbf{X}^k \bullet \mathbf{Z}^k, \left\|\mathbf{r}_p^k\right\|, \left\|\mathbf{r}_d^k\right\|)$*, Algorithm SDP:ReducedLocal converges in* $O(n \ln(\epsilon_0/\epsilon))$ *iterations for a given tolerance* ϵ.

Proof. (25)–(26) are direct consequences of Lemma 3.2 and (24). We can derive (27) from (26). See Park and O'Leary [30, Theorem 3.1]. The rest of result can be shown from (25)–(27) as in Park and O'Leary [30, Sect. 3.2]. □

4.2 Superlinear Convergence

We discuss the asymptotic bounds of $\boldsymbol{X}^k$ and $\boldsymbol{Z}^k$ under a strict complementarity assumption. The notation for the asymptotic bounds, like $O(\cdot)$ and $\Omega(\cdot)$, are defined in Table 1. In the asymptotic bounds for a matrix, we assume that the matrix size n is constant because iteration-by-iteration change of matrix norm is more of our interest. By using the asymptotic bounds, we prove that SDP:ReducedLocal has superlinear convergence.

Assumption 4.1 (Strict Complementarity)**.** *The SDP problem has a solution* $(\mathbf{X}^*, \mathbf{y}^*, \mathbf{Z}^*)$*, so* $\mathbf{X}^* + \mathbf{Z}^* \succ 0$.

The following proofs are organized as follows. Lemma 4.3 will show that the optimal matrices $\boldsymbol{X}^*$ and $\boldsymbol{Z}^*$ share eigenvectors. Based on the property, the asymptotic behavior of $\boldsymbol{X}^k$ and $\boldsymbol{Z}^k$ will be discovered in Lemma 4.4. Lemma 4.5 introduces important inequalities that will be frequently used in the following proofs. By using Lemma 4.4 and 4.5, we show an asymptotic bound of $(1-\overline{\theta}_k)$ in Lemma 4.6. Finally, Theorem 4.2 proves the superlinear convergence from Theorem 4.1, Lemma 4.4 and Lemma 4.6.

Lemma 4.3. *Let* $(\mathbf{X}^*, \mathbf{y}^*, \mathbf{Z}^*) \in \mathcal{F}^*$. *Then, there exists an orthogonal matrix* $\mathbf{Q} = [\mathbf{q}_1, \cdots, \mathbf{q}_n]$ *such that* $\mathbf{Q}^T\mathbf{X}^*\mathbf{Q}$ *and* $\mathbf{Q}^T\mathbf{Z}^*\mathbf{Q}$ *are diagonal matrices.*

Proof. See Alizadeh, Haeberly, and Overton [6, Lemma 1 ff]. □

Let $(\boldsymbol{X}^*, \boldsymbol{y}^*, \boldsymbol{Z}^*)$ be the strict complementary solution, and let $\boldsymbol{Q}$ be the orthogonal matrix whose columns are the eigenvectors of $\boldsymbol{X}^*$ and $\boldsymbol{Z}^*$. By the strict complementarity, we define two sets $\boldsymbol{B}$ and $\boldsymbol{N}$ as

$$\boldsymbol{B} := \{i : \boldsymbol{q}_i^T\boldsymbol{X}^*\boldsymbol{q}_i > 0\}, \qquad \boldsymbol{N} := \{i : \boldsymbol{q}_i^T\boldsymbol{Z}^*\boldsymbol{q}_i > 0\},$$

where $\boldsymbol{B} \cup \boldsymbol{N} = \{1, 2, \ldots, n\}$ and $\boldsymbol{q}_i$ is the i-th column of $\boldsymbol{Q}$. We also define a set $\mathcal{M}$ of $(\boldsymbol{X}', \boldsymbol{y}', \boldsymbol{Z}')$ as

$$\mathcal{M} := \{(\boldsymbol{X}', \boldsymbol{y}', \boldsymbol{Z}') \in \mathcal{F}_0 : \boldsymbol{q}_i^T\boldsymbol{X}'\boldsymbol{q}_j = 0 \text{ if } i \text{ or } j \in \boldsymbol{N}, \quad \boldsymbol{q}_i^T\boldsymbol{Z}'\boldsymbol{q}_j = 0 \text{ if } i \text{ or } j \in \boldsymbol{B}.\}$$

where $\mathcal{F}_0 := \{(\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z}) \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n : \ (\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{Z}) \text{ satisfies (3) and (4)}\}$. Then, we consider the following minimization problem:

$$\min_{\boldsymbol{X}', \boldsymbol{y}', \boldsymbol{Z}'} \left\| (\boldsymbol{Z}^k)^{1/2}(\boldsymbol{X}^k - \boldsymbol{X}')(\boldsymbol{Z}^k - \boldsymbol{Z}')(\boldsymbol{Z}^k)^{-1/2} \right\|_F \tag{28}$$

such that $(\boldsymbol{X}', \boldsymbol{y}', \boldsymbol{Z}') \in \mathcal{M}$ and $\left\| [\boldsymbol{X}', \boldsymbol{Z}'] \right\|_F \leq \Gamma$, where Γ is a constant such that $\left\| [\boldsymbol{X}^k, \boldsymbol{Z}^k] \right\|_F \leq \Gamma$ for $\forall k$. Let $(\breve{\boldsymbol{X}}^k, \breve{\boldsymbol{y}}^k, \breve{\boldsymbol{Z}}^k)$ denote the solution[1] of (28), and we define η_k as

$$\eta_k := \frac{1}{\tau_k} \left\| (\boldsymbol{Z}^k)^{1/2}(\boldsymbol{X}^k - \breve{\boldsymbol{X}}^k)(\boldsymbol{Z}^k - \breve{\boldsymbol{Z}}^k)(\boldsymbol{Z}^k)^{-1/2} \right\|_F. \tag{29}$$

By definition of $\mathcal{M}$,

$$\breve{\boldsymbol{X}}^k \breve{\boldsymbol{Z}}^k = \mathbf{0}. \tag{30}$$

Now, the lemma below from Potra et al. [12, 14] reveals asymptotic bounds for $\boldsymbol{X}^k$, $\boldsymbol{Z}^k$, and η_k.

Lemma 4.4 (Asymptotic Bounds)**.** *Let* $(\mathbf{X}^k, \mathbf{y}^k, \mathbf{Z}^k)$ *denote the iterate satisfying the properties (25)–(27) in Theorem 4.1, and let* η_k *be as defined in (29). Then,*

$$\left\| (\mathbf{X}^k)^{1/2}(\mathbf{Z}^k)^{1/2} \right\|^2 = \left\| (\mathbf{Z}^k)^{1/2}\mathbf{X}^k(\mathbf{Z}^k)^{1/2} \right\| \leq (1+\alpha)\tau_k. \tag{31}$$

[1] See Potra et al. [14, pp18–19] for the existence of the solution.

Under Assumption 4.1, we have

$$\mathbf{Q}^T(\mathbf{X}^k)^{1/2}\mathbf{Q} = \begin{bmatrix} O(1) & O(\sqrt{\tau_k}) \\ O(\sqrt{\tau_k}) & O(\sqrt{\tau_k}) \end{bmatrix},\ \mathbf{Q}^T(\mathbf{X}^k)^{-1/2}\mathbf{Q} = \begin{bmatrix} O(1) & O(1) \\ O(1) & O(1/\sqrt{\tau_k}) \end{bmatrix}, \tag{32}$$

$$\mathbf{Q}^T(\mathbf{Z}^k)^{1/2}\mathbf{Q} = \begin{bmatrix} O(\sqrt{\tau_k}) & O(\sqrt{\tau_k}) \\ O(\sqrt{\tau_k}) & O(1) \end{bmatrix},\ \mathbf{Q}^T(\mathbf{Z}^k)^{-1/2}\mathbf{Q} = \begin{bmatrix} O(1/\sqrt{\tau_k}) & O(1) \\ O(1) & O(1) \end{bmatrix}, \tag{33}$$

$$\mathbf{Q}^T\mathbf{X}^k\mathbf{Q} = \begin{bmatrix} O(1) & O(\sqrt{\tau_k}) \\ O(\sqrt{\tau_k}) & O(\tau_k) \end{bmatrix},\ \mathbf{Q}^T\mathbf{Z}^k\mathbf{Q} = \begin{bmatrix} O(\tau_k) & O(\sqrt{\tau_k}) \\ O(\sqrt{\tau_k}) & O(1) \end{bmatrix}, \tag{34}$$

and

$$\eta_k = O(\phi_k). \tag{35}$$

Proof. See Potra et al. [14, Corollary 3.3 and Lemma 4.4] and [12, Theorem 4.2]. Note that, in the course of their proofs, the roles of $\boldsymbol{X}^k$ and $\boldsymbol{Z}^k$ are switched. □

As Theorem 4.1 suggests, the superlinear convergence can be established by showing that

$$\lim_{k\to\infty}(1-\overline{\theta}_k) \to 0.$$

Potra et al. [14, Theorem 4.7] proved this by showing that $(1-\overline{\theta}_k) = O(\eta_k)$. However, their result is not directly applicable to our algorithm due to the perturbation δ_ϵ by constraint reduction. Instead, Lemma 4.6 shows that

$$(1-\overline{\theta}_k) = O(\eta_k + \delta_\epsilon).$$

Then, Theorem 4.2 will establish the superlinear convergence by showing $\eta_k \to 0$ and $\delta_\epsilon \to 0$. In the proofs, we utilize the following preliminary lemma.

Lemma 4.5. *For $(\mathbf{X}', \mathbf{Z}') \in \mathcal{N}(\gamma, \tau')$ and $\mathbf{H} \in \mathbb{R}^{n\times n}$, let $(\Delta\mathbf{X}', \Delta\mathbf{y}', \Delta\mathbf{Z}')$ be a solution of*

$$\mathbf{A}_i \bullet \Delta\mathbf{X}' = 0 \quad \textit{for } i = 1,\ldots,m, \tag{36}$$

$$\sum_{i=1}^m \Delta y_i'\mathbf{A}_i + \Delta\mathbf{Z}' = \mathbf{0}, \tag{37}$$

$$\operatorname{symm}\left(\mathbf{Z}'^{1/2}(\mathbf{X}'\Delta\mathbf{Z}' + \Delta\mathbf{X}'\mathbf{Z}')\mathbf{Z}'^{-1/2}\right) = \mathbf{H}. \tag{38}$$

Then

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

$$\delta_x' \le \frac{\|\mathbf{H}\|_F}{1-\gamma}, \quad \delta_z' \le \frac{\|\mathbf{H}\|_F}{1-\gamma},$$

where

$$\delta_x' = \left\|\mathbf{Z}'^{1/2}\Delta\mathbf{X}'\mathbf{Z}'^{1/2}\right\|_F \quad \textit{and} \quad \delta_z' = \tau'\left\|\mathbf{Z}'^{-1/2}\Delta\mathbf{Z}'\mathbf{Z}'^{-1/2}\right\|_F.$$

Proof. See Monteiro [4, Lemma 4.4 in p.671], in which the roles of $\boldsymbol{X}$ and $\boldsymbol{Z}$ in $\boldsymbol{H}$ are switched. $\square$

Lemma 4.6 (Similar to [14, Theorem 4.7])**.** *Under Assumption 4.1,*

$$(1-\overline{\theta}_k) = O(\eta_k + \delta_\epsilon).$$

Proof. For simplicity, we omit the index k in the following equations. By Theorem 4.1, $(\boldsymbol{X}, \boldsymbol{Z}) \in \mathcal{N}(\alpha, \tau)$, and it is easy to show that (36) and (37) in Lemma 4.5 are satisfied by substituting

$$(\Delta \boldsymbol{X}', \Delta \boldsymbol{y}', \Delta \boldsymbol{Z}') = (\Delta \boldsymbol{X} + \boldsymbol{X} - \breve{\boldsymbol{X}}, \Delta \boldsymbol{y} + \boldsymbol{y} - \breve{\boldsymbol{y}}, \Delta \boldsymbol{Z} + \boldsymbol{Z} - \breve{\boldsymbol{Z}}),$$

where $(\breve{\boldsymbol{X}}, \breve{\boldsymbol{y}}, \breve{\boldsymbol{Z}})$ is a solution of (28). Thus, we can use Lemma 4.5 by substituting $(\boldsymbol{X}', \boldsymbol{Z}')$ with $(\boldsymbol{X}, \boldsymbol{Z})$. By using (30) and (9*) in Lemma 3.1 with $\overline{\mu} = 0$, we can rewrite $\boldsymbol{H}$ in Lemma 4.5 as

$$\begin{aligned}
\boldsymbol{H} &= \operatorname{symm}\left(\boldsymbol{Z}^{1/2}\left(\boldsymbol{X}(\Delta \boldsymbol{Z} + \boldsymbol{Z} - \breve{\boldsymbol{Z}}) + (\Delta \boldsymbol{X} + \boldsymbol{X} - \breve{\boldsymbol{X}})\boldsymbol{Z}\right)\boldsymbol{Z}^{-1/2}\right) \\
&= \operatorname{symm}\left(\boldsymbol{Z}^{1/2}\left(\boldsymbol{X}\Delta \boldsymbol{Z} + \boldsymbol{X}\boldsymbol{Z} - \boldsymbol{X}\breve{\boldsymbol{Z}} + \Delta \boldsymbol{X}\boldsymbol{Z} + \boldsymbol{X}\boldsymbol{Z} - \breve{\boldsymbol{X}}\boldsymbol{Z}\right)\boldsymbol{Z}^{-1/2}\right) \\
&= \operatorname{symm}\left(\boldsymbol{Z}^{1/2}\left(\boldsymbol{X}\boldsymbol{Z} - \boldsymbol{X}\breve{\boldsymbol{Z}} - \breve{\boldsymbol{X}}\boldsymbol{Z} + \breve{\boldsymbol{X}}\breve{\boldsymbol{Z}}\right)\boldsymbol{Z}^{-1/2}\right) \\
&\quad + \operatorname{symm}\left(\boldsymbol{Z}^{1/2}\left(\boldsymbol{X}\Delta \boldsymbol{Z} + \Delta \boldsymbol{X}\boldsymbol{Z} + \boldsymbol{X}\boldsymbol{Z}\right)\boldsymbol{Z}^{-1/2}\right) \qquad (\because \breve{\boldsymbol{X}}\breve{\boldsymbol{Z}} = \boldsymbol{0} \text{ by (30).}) \\
&= \operatorname{symm}\left(\boldsymbol{Z}^{1/2}(\boldsymbol{X} - \breve{\boldsymbol{X}})(\boldsymbol{Z} - \breve{\boldsymbol{Z}})\boldsymbol{Z}^{-1/2}\right) - \boldsymbol{Z}^{1/2}\Delta \boldsymbol{X}_\epsilon \boldsymbol{Z}^{1/2} \qquad (\because \text{(9*) with } \overline{\mu} = 0) \\
&= \operatorname{symm}(\Delta) - \Delta_\epsilon
\end{aligned}$$

where Δ and Δ_ϵ are defined as

$$\Delta := \boldsymbol{Z}^{1/2}(\boldsymbol{X} - \breve{\boldsymbol{X}})(\boldsymbol{Z} - \breve{\boldsymbol{Z}})\boldsymbol{Z}^{-1/2} \quad \text{and} \quad \Delta_\epsilon := \boldsymbol{Z}^{1/2}\Delta \boldsymbol{X}_\epsilon \boldsymbol{Z}^{1/2},$$

and, by (16) and (29),

$$\|\Delta_\epsilon\| = \delta_\epsilon \tau, \quad \|\Delta\| = \eta_k \tau. \tag{39}$$

Thus, we have

$$\|\boldsymbol{H}\|_F = \|\operatorname{symm}(\Delta) - \Delta_\epsilon\|_F \le \|\Delta\|_F + \|\Delta_\epsilon\|_F = \eta_k \tau + \delta_\epsilon \tau = (\eta_k + \delta_\epsilon)\tau.$$

We also define Δ_x and Δ_z as

$$\Delta_x := \boldsymbol{Z}^{1/2}(\Delta \boldsymbol{X} + \boldsymbol{X} - \breve{\boldsymbol{X}})\boldsymbol{Z}^{1/2} \quad \text{and} \quad \Delta_z := \boldsymbol{Z}^{-1/2}(\Delta \boldsymbol{Z} + \boldsymbol{Z} - \breve{\boldsymbol{Z}})\boldsymbol{Z}^{-1/2}.$$

By Lemma 4.5, because $(\boldsymbol{X}, \boldsymbol{Z}) \in \mathcal{N}(\alpha, \tau)$, we have

$$\begin{aligned}
\delta_z' = \tau \|\Delta_z\|_F &\le \frac{\|\boldsymbol{H}\|_F}{1-\alpha} \le \frac{(\eta_k + \delta_\epsilon)\tau}{1-\alpha} \\
\|\Delta_z\|_F &\le \frac{\eta_k + \delta_\epsilon}{1-\alpha}.
\end{aligned} \tag{40}$$

Similarly,

$$\delta_x' = \|\Delta_x\|_F \le \frac{\|\boldsymbol{H}\|_F}{1-\alpha} \le \frac{(\eta_k + \delta_\epsilon)\tau}{1-\alpha}. \tag{41}$$

Let $\boldsymbol{v}_i$ denote columns of of $\boldsymbol{Q}^T(\boldsymbol{X})^{-1/2}\boldsymbol{Q}$, so

$$\boldsymbol{Q}^T(\boldsymbol{X})^{-1/2}\boldsymbol{Q} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n].$$

Then, by using (32) in Lemma 4.4 together with $(\breve{\boldsymbol{X}}, \breve{\boldsymbol{y}}, \breve{\boldsymbol{Z}}) \in \mathcal{M}$ and $\left\|[\breve{\boldsymbol{X}}, \breve{\boldsymbol{Z}}]\right\| \le \Gamma$,

$$\begin{aligned}
\left\|(\boldsymbol{X})^{-1/2}(\boldsymbol{X} - \breve{\boldsymbol{X}})(\boldsymbol{X})^{-1/2}\right\|_F &= \left\|\boldsymbol{I} - (\boldsymbol{X})^{-1/2}\breve{\boldsymbol{X}}(\boldsymbol{X})^{-1/2}\right\|_F \\
&\le \|\boldsymbol{I}\|_F + \left\|(\boldsymbol{X})^{-1/2}\breve{\boldsymbol{X}}(\boldsymbol{X})^{-1/2}\right\|_F \\
&= \sqrt{n} + \left\|\boldsymbol{Q}^T(\boldsymbol{X})^{-1/2}\boldsymbol{Q}\boldsymbol{Q}^T\breve{\boldsymbol{X}}\boldsymbol{Q}\boldsymbol{Q}^T(\boldsymbol{X})^{-1/2}\boldsymbol{Q}\right\|_F \\
&= \sqrt{n} + \left\|[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n]\boldsymbol{Q}^T\breve{\boldsymbol{X}}\boldsymbol{Q}[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n]^T\right\|_F \\
&= \sqrt{n} + \left\|\sum_{i,j\in \boldsymbol{B}}(\boldsymbol{q}_i^T\breve{\boldsymbol{X}}\boldsymbol{q}_j)\boldsymbol{v}_i\boldsymbol{v}_j^T\right\|_F \quad (\because (\breve{\boldsymbol{X}}, \breve{\boldsymbol{y}}, \breve{\boldsymbol{Z}}) \in \mathcal{M}) \\
&= \sqrt{n} + \Gamma\left\|\sum_{i,j\in \boldsymbol{B}}\boldsymbol{v}_i\boldsymbol{v}_j^T\right\|_F \quad (\because \left\|[\breve{\boldsymbol{X}}, \breve{\boldsymbol{Z}}]\right\| \le \Gamma) \\
&= O(1). \quad (\because (32))
\end{aligned} \tag{42}$$

Similarly, we can show that

$$\left\|\boldsymbol{Z}^{-1/2}(\boldsymbol{Z} - \breve{\boldsymbol{Z}})\boldsymbol{Z}^{-1/2}\right\|_F = O(1). \tag{43}$$

Next,

$$\begin{aligned}
& \boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}\Delta\boldsymbol{Z}\boldsymbol{Z}^{-1/2} \\
= & \left(\boldsymbol{Z}^{1/2}\Delta\boldsymbol{X}\boldsymbol{Z}^{1/2}\right)\left(\boldsymbol{Z}^{-1/2}\Delta\boldsymbol{Z}\boldsymbol{Z}^{-1/2}\right) \\
= & \left(\Delta_x - \boldsymbol{Z}^{1/2}(\boldsymbol{X} - \breve{\boldsymbol{X}})\boldsymbol{Z}^{1/2}\right)\left(\Delta_z - \boldsymbol{Z}^{-1/2}(\boldsymbol{Z} - \breve{\boldsymbol{Z}})\boldsymbol{Z}^{-1/2}\right) \\
= & \Delta_x\Delta_z - \Delta_x\boldsymbol{Z}^{-1/2}(\boldsymbol{Z} - \breve{\boldsymbol{Z}})\boldsymbol{Z}^{-1/2} - \boldsymbol{Z}^{1/2}(\boldsymbol{X} - \breve{\boldsymbol{X}})\boldsymbol{Z}^{1/2}\Delta_z + \Delta. \\
= & \Delta_x\Delta_z - \Delta_x\boldsymbol{Z}^{-1/2}(\boldsymbol{Z} - \breve{\boldsymbol{Z}})\boldsymbol{Z}^{-1/2} \\
& -(\boldsymbol{Z}^{1/2}\boldsymbol{X}^{1/2})\left(\boldsymbol{X}^{-1/2}(\boldsymbol{X} - \breve{\boldsymbol{X}})\boldsymbol{X}^{-1/2}\right)(\boldsymbol{X}^{1/2}\boldsymbol{Z}^{1/2})\Delta_z + \Delta.
\end{aligned}$$

Thus, by (31) in Lemma 4.4, and (39)–(43), we can calculate the upper bound of δ in (15) as

$$
\begin{aligned}
\delta &= \frac{1}{\tau}\left\| \boldsymbol{Z}^{1/2}\Delta \boldsymbol{X}\Delta \boldsymbol{Z}\boldsymbol{Z}^{-1/2}\right\|_F \\
&\leq \frac{1}{\tau}\|\Delta_x\Delta_z\|_F + \frac{1}{\tau}\|\Delta_x\|_F \left\|\boldsymbol{Z}^{-1/2}(\boldsymbol{Z}-\breve{\boldsymbol{Z}})\boldsymbol{Z}^{-1/2}\right\|_F \\
&\quad + \frac{1}{\tau}\|\Delta_z\|_F \left\|\boldsymbol{X}^{1/2}\boldsymbol{Z}^{1/2}\right\|_2^2 \left\|\boldsymbol{X}^{-1/2}(\boldsymbol{X}-\breve{\boldsymbol{X}})\boldsymbol{X}^{-1/2}\right\|_F + \frac{1}{\tau}\|\Delta\|_F \\
&= O(\eta_k+\delta_\epsilon). \qquad (44)
\end{aligned}
$$

By (18) and the definition of $\widehat{\theta}$ in (19),

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

Thus, by (44),

$$
1-\overline{\theta} \leq \frac{\delta+\delta_\epsilon}{\beta-\alpha} = O(\eta_k+\delta_\epsilon)+O(\delta_\epsilon) = O(\eta_k+\delta_\epsilon).
$$

□

Finally, we prove the superlinear local convergence by using that $\eta_k = O(\phi_k)$ and $\delta_\epsilon = O(\phi_k)$.

Theorem 4.2 (Superlinear Convergence)**.** *Under Assumption 4.1, Algorithm SDP:ReducedLocal converges superlinearly with Q-order of at least* $1+\min(\sigma, 0.5)$.

Proof. Since $\overline{\tau} = \tau^+$ in (24), we can rewrite it as $(\boldsymbol{X}^k, \boldsymbol{Z}^k) \in \mathcal{N}(\min(\tau_k^\sigma, \alpha), \tau_k)$, so

$$
\begin{aligned}
\left\|(\boldsymbol{Z}^k)^{1/2}\boldsymbol{X}^k(\boldsymbol{Z}^k)^{1/2}-\tau_k\boldsymbol{I}\right\|_F &\leq \min(\tau_k^\sigma,\alpha)\tau_k \leq (\tau_k)^{1+\sigma}, \\
\frac{\left\|(\boldsymbol{Z}^k)^{1/2}\boldsymbol{X}^k(\boldsymbol{Z}^k)^{1/2}-\tau_k\boldsymbol{I}\right\|_F}{\tau_k} &= O(\tau_k^\sigma).
\end{aligned}
$$

Thus, by the definition of ϕ_k, $\phi_k = O\left(\max(\tau_k^{\sigma}, \tau_k^{0.5})\right) = O(\tau_k^{\min(\sigma,0.5)})$. Then, by Lemma 4.6, Condition 4.1, and (35) in Lemma 4.4,

$$(1-\overline{\theta}_k) = O(\eta_k + \delta_\epsilon) = O(\phi_k) = O(\tau_k^{\min(\sigma,0.5)}),$$

which implies superlinear convergence by Theorem 4.1. Therefore,

$$\tau_{k+1} = (1-\overline{\theta}_k)\tau_k = O(\tau_k^{1+\min(\sigma,0.5)}).$$

□

5 Experiments

This section summarizes numerical experiments to evaluate the effectiveness of constraint-reduced algorithms. MATLAB codes for the experiments are available at `http://www.mathworks.com/matlabcentral/fileexchange/54117`. To evaluate the constraint-reduced algorithms, we compare their computational costs with an algorithm with no constraint reduction. We can disable constraint reduction by constructing a full Schur complement matrix with no condition checks in SDP:ReducedLocal, which is equivalent to the algorithm by Kojima et al. [8]. We use this unreduced algorithm as our benchmark.

Reducing zero blocks of $\boldsymbol{X}$ makes the perturbations δ_ϵ and $\overline{\delta}_\epsilon$ zero by Lemma 3.1, so the constraint reduction conditions are immediately satisfied. From the perspective, the number of $\boldsymbol{X}_j^* = 0$ blocks at the solution $(\boldsymbol{X}^*, \boldsymbol{y}^*, \boldsymbol{Z}^*) \in \mathcal{F}^*$ may determine an applicability of constraint reduction. In this experiment, we consider three groups of SDP problems, having different proportions of zero $\boldsymbol{X}_j^*$ blocks: 0%, 25%, and 50%. For each group, we randomly generate SDP problems for different m, n, and block size n_j. For simplicity, we only consider the case of identical block sizes n_j.

First, we evaluate constraint reduction in the predictor step, disabling the constraint reduction in the corrector step. Fig. 1 shows the average number of iterations and the average number of reduced blocks varying the parameter ω in Condition 3.1 and Condition 4.1. When $\omega = 0$, no constraint reduction is performed, so the iteration counts for the two algorithms are the same as that for the unreduced algorithm. When $\omega \geq 0.1$, the algorithms reduce blocks but require additional iterations, which generally increase the total amount of work in the experiment. We can understand this from the relation of $\widehat{\theta}$ and δ_ϵ in (19) together with its effect on the converging rate in Lemma 3.2. Results for different dimensions and different ratios of zero $\boldsymbol{X}_j^*$ blocks are similar.

Next, we evaluate the constraint reduction in the corrector step with disabling the reduction in the predictor step. First, we use $m = 256$, $n_j = 16$, and vary $n = 128, 256, 512, 1024$, and 2048. Second, we fix $m = 256$, $n = 512$, and vary $n_j = 4, 8, 16, 32$, and 64. Varying m gives similar results, not presented. For each set of dimensions, we generate 5 random SDP problems and take an average of results. To evaluate the computational cost saving, we count the total number of flops to construct all of the Schur complement matrices during

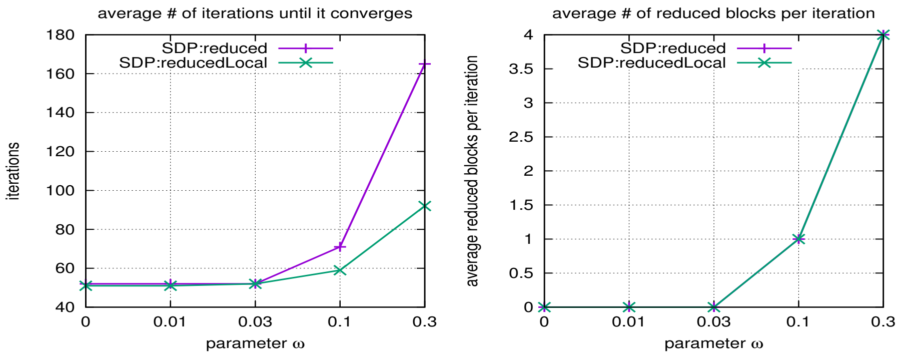

Figure 1: When $m = 64$, $n = 256$, $n_j = 16$, and 0% zero $\boldsymbol{X}_j^*$ blocks, the number of iteration until convergence and the average reduced blocks per iteration by SDP:Reduced and SDP:ReducedLocalwith varying $\omega = 0.0, 0.01, 0.03, 0.1$, and 0.3.

the corrector steps until each algorithm converges. In case of the constraint-reduced algorithms, all the additional computations for constraint reduction are also included in flop counts.

Fig. 2 shows the % of flop savings in the corrector steps by SDP:Reduced and SDP:ReducedLocal against the unreduced algorithm with varying matrix size n and block size n_j. For the same settings of n and n_j, Fig. 3 demonstrates an average number of reduced blocks per iteration. As matrix size n grows, constraint reduction saves more flops because the algorithm has more block candidates to be reduced. On the other hand, constraint reduction tends to save the more flops for larger block sizes because Schur complement matrix construction costs $O(mn_j^3 + m^2n_j^2)$ for each block. Particularly at block size $n_j = 64$, the % of saving drops because we have only a few blocks, so fewer candidates for constraint reduction.

The constraint reduction effectively reduces flop counts even for SDP problems whose zero $\boldsymbol{X}_j^*$ block ratio is 0%, which indicates that constraint reduction is not limited to SDP problems having many zero $\boldsymbol{X}_j^*$ blocks. However, compared to the case of 25% of zero $\boldsymbol{X}_j^*$ blocks, the flops saving is not improved very much for the case of 50%. Thus, the effectiveness of constraint reduction is more related to the contributions of blocks along the iterations rather than the number of zero blocks at the final optimum.

When we enable constraint reduction only for the corrector step, both SDP:Reduced and SDP:ReducedLocal converge as fast as the unreduced algorithm in terms of the number of iterations. However, SDP:ReducedLocal saves more flops than SDP:Reduced. Fig. 3 shows that SDP:ReducedLocal reduced more blocks than SDP:Reduced, so the constraint reduction conditions for SDP:ReducedLocal are more effective than those of SDP:Reduced while achieving a theoretically faster convergence rate.

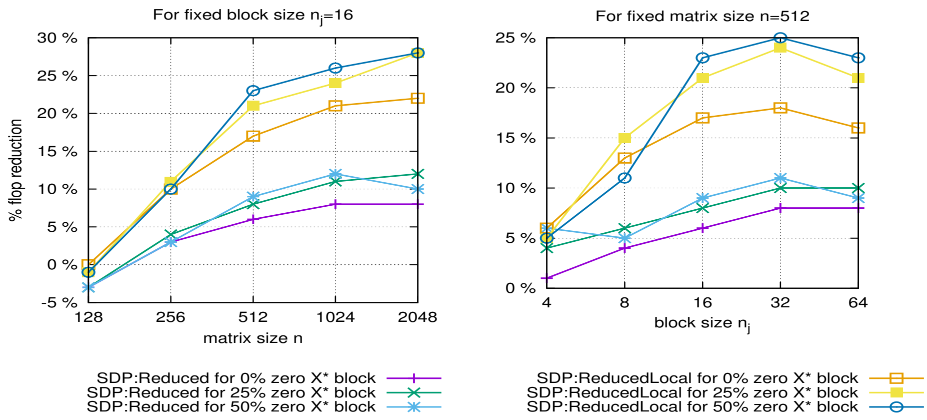

Figure 2: When $m = 256$, the % of flops saving in the corrector steps with varying matrix size n and block size n_j by SDP:Reduced and SDP:ReducedLocal compared with unreduced algorithm.

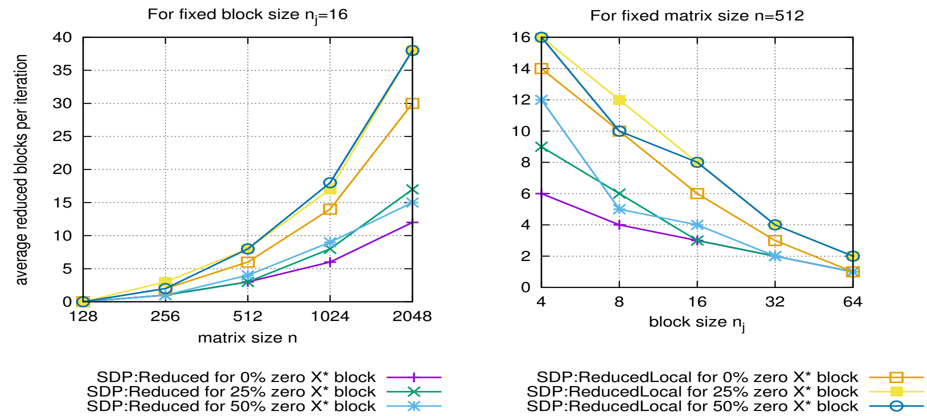

Figure 3: When $m = 256$, the average number of reduced blocks in the corrector steps with varying matrix size n and block size n_j by SDP:Reduced and SDP:ReducedLocal.

6 Conclusions

In this paper, we developed a constraint-reduced predictor-corrector algorithm for block-constrained SDP and showed its polynomial global and superlinear local convergence under Slater and strict complementarity assumptions. To accomplish the superlinear convergence, Algorithm SDP:ReducedLocal adopts new constraint reduction criteria and repeats the corrector step so that the iterate tangentially approaches the central path. By the numerical experiments, we demonstrated its computational cost saving especially for the corrector step.

We applied a constraint reduction method to the predictor-corrector methods using the HKM direction. It will be also interesting to apply constraint reduction to the algorithm using other directions. For example, Kojima et al.

[9] developed an algorithm using AHO direction, that has a quadratic local convergence. Inspired by the fast centering effect of AHO direction, Potra et al. [12] developed an algorithm using HKM direction in the predictor step and AHO direction in the corrector step, that has the superlinear local convergence. Later, Ji, Potra, and Sheng [7] generalized the idea by revealing the relation between the convergence rate and the condition numbers of scaling matrix for MZ-family directions. The advantage of the algorithms using the AHO direction is that we need neither a fast centering assumption nor repeated corrector steps for local convergence. The study of constraint reduction method for these algorithms will extend the scope of application for constraint reduction.

A Appendix: Proofs

Lemma A.1. *Suppose that* $\mathbf{W} \in \mathcal{R}^{n\times n}$ *is a nonsingular matrix. Then, for any* $\mathbf{E} \in \mathcal{S}^{n\times n}$*, we have*

$$-\|\mathbf{E}\|_F \le \lambda_i(\mathbf{E}) \le \|\mathbf{E}\|_F\,,$$
$$\|\mathbf{E}\|_F \le \left\| \operatorname{symm}\left(\mathbf{W}\mathbf{E}\mathbf{W}^{-1}\right)\right\|_F .$$

Proof. The first inequality comes from

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

For the second inequality, see Monteiro [4, Lemma 3.3]. □

Proof. of Lemma 4.1.
By definition of $\overline{\gamma}$,

$$\overline{\gamma} = \frac{\left\|(\overline{\boldsymbol{Z}})^{1/2}\overline{\boldsymbol{X}}(\overline{\boldsymbol{Z}})^{1/2} - \overline{\tau}\boldsymbol{I}\right\|_F}{\overline{\tau}} \quad \text{and} \quad \overline{\gamma}^+ = \frac{\left\|(\overline{\boldsymbol{Z}}^+)^{1/2}\overline{\boldsymbol{X}}^+(\overline{\boldsymbol{Z}}^+)^{1/2} - \overline{\tau}\boldsymbol{I}\right\|_F}{\overline{\tau}}.$$

where

$$(\overline{\boldsymbol{X}}^+, \overline{\boldsymbol{y}}^+, \overline{\boldsymbol{Z}}^+) = (\overline{\boldsymbol{X}} + \Delta\overline{\boldsymbol{X}}, \overline{\boldsymbol{y}} + \Delta\overline{\boldsymbol{y}}, \overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{Z}}).$$

Because $(\overline{\boldsymbol{X}}^0, \overline{\boldsymbol{Z}}^0) \in \mathcal{N}(\beta, (1-\overline{\theta})\tau) = \mathcal{N}(\beta, \overline{\tau})$ by Lemma 3.2, the initial relative distance $\overline{\gamma}_0$ satisfies

$$\overline{\gamma}_0 \le \beta.$$

Thus, it suffices to show that, for $0 < \overline{\gamma} \le \beta$,

$$\overline{\gamma}^+ < \overline{\gamma}, \quad \text{and} \tag{45}$$
$$\overline{\gamma}^+ = O(\overline{\gamma}^2). \tag{46}$$

By Lemma 4.5, we have

$$
\begin{aligned}
\left\|\overline{\boldsymbol{Z}}^{\,-1/2}\Delta\overline{\boldsymbol{Z}}\overline{\boldsymbol{Z}}^{\,-1/2}\right\|_F &\le \frac{\|\boldsymbol{H}\|_F}{(1-\overline{\gamma})\overline{\tau}} \le \frac{\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}{(1-\overline{\gamma})\overline{\tau}} \\
&= \frac{\left\|\overline{\tau}\boldsymbol{I}-\overline{\boldsymbol{Z}}^{1/2}\overline{\boldsymbol{X}}\overline{\boldsymbol{Z}}^{1/2}+\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}_\epsilon\overline{\boldsymbol{Z}}^{1/2}\right\|_F}{(1-\overline{\gamma})\overline{\tau}} \quad (\because (9*)) \\
&\le \frac{\left\|\overline{\tau}\boldsymbol{I}-\overline{\boldsymbol{Z}}^{1/2}\overline{\boldsymbol{X}}\overline{\boldsymbol{Z}}^{1/2}\right\|_F+\left\|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}_\epsilon\overline{\boldsymbol{Z}}^{1/2}\right\|_F}{(1-\overline{\gamma})\overline{\tau}} \\
&= \frac{\overline{\gamma}\overline{\tau}+\tau\overline{\delta}_\epsilon}{(1-\overline{\gamma})\overline{\tau}} \\
&= \frac{\overline{\gamma}}{1-\overline{\gamma}}+\frac{\overline{\delta}_\epsilon}{(1-\overline{\gamma})(1-\overline{\theta})} \quad (\because (17) \text{ and } \overline{\tau}=(1-\overline{\theta})\tau) \\
&\le \frac{\overline{\gamma}}{1-\overline{\gamma}}+\frac{\zeta\overline{\gamma}^2(1-\overline{\theta})}{(1-\overline{\gamma})(1-\overline{\theta})} \quad (\because \text{Condition 4.2}) \\
&= \frac{\overline{\gamma}}{1-\overline{\gamma}}(1+\zeta\overline{\gamma}) \;<\; \frac{\overline{\gamma}}{1-\overline{\gamma}}\left(1+\left(\frac{1}{\beta}-2\right)\overline{\gamma}\right) \quad (\because (23)) \\
&= \overline{\gamma}\left[\frac{1}{1-\overline{\gamma}}+\frac{-\overline{\gamma}}{1-\overline{\gamma}}+\left(\frac{\overline{\gamma}}{1-\overline{\gamma}}\right)\left(\frac{1}{\beta}-1\right)\right] \\
&= \overline{\gamma}\left[1+\left(\frac{1-\beta}{\beta}\right)\left(\frac{\overline{\gamma}}{1-\overline{\gamma}}\right)\right] \\
&\le \overline{\gamma}\left[1+\left(\frac{1-\overline{\gamma}}{\overline{\gamma}}\right)\left(\frac{\overline{\gamma}}{1-\overline{\gamma}}\right)\right] = 2\overline{\gamma} < 1 \quad (\because \overline{\gamma}\le\beta<0.5)
\end{aligned}
$$

Thus, by Lemma A.1,

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

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

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

so $\overline{\boldsymbol{Z}}^{+}\succ\boldsymbol{0}$. Therefore, $(\overline{\boldsymbol{Z}}^{+})^{1/2}$ exists and is invertible. Define $\boldsymbol{Q}$, $\boldsymbol{E}$, and $\boldsymbol{W}$ as

$$
\begin{aligned}
\boldsymbol{Q} &:= \overline{\boldsymbol{Z}}^{1/2}(\overline{\boldsymbol{X}}^{+}\overline{\boldsymbol{Z}}^{+}-\overline{\tau}\boldsymbol{I})\overline{\boldsymbol{Z}}^{\,-1/2}, \\
\boldsymbol{E} &:= (\overline{\boldsymbol{Z}}^{+})^{1/2}\overline{\boldsymbol{X}}^{+}(\overline{\boldsymbol{Z}}^{+})^{1/2}-\overline{\tau}\boldsymbol{I}, \\
\boldsymbol{W} &:= \overline{\boldsymbol{Z}}^{1/2}(\overline{\boldsymbol{Z}}^{+})^{-1/2}.
\end{aligned}
$$

Then,

$$
\begin{aligned}
\boldsymbol{W}\boldsymbol{E}\boldsymbol{W}^{-1} &= (\overline{\boldsymbol{Z}}^{1/2}(\overline{\boldsymbol{Z}}^{+})^{-1/2})\left[(\overline{\boldsymbol{Z}}^{+})^{1/2}\overline{\boldsymbol{X}}^{+}(\overline{\boldsymbol{Z}}^{+})^{1/2}-\overline{\tau}\boldsymbol{I}\right](\overline{\boldsymbol{Z}}^{1/2}(\overline{\boldsymbol{Z}}^{+})^{-1/2})^{-1} \\
&= \overline{\boldsymbol{Z}}^{1/2}(\overline{\boldsymbol{X}}^{+}\overline{\boldsymbol{Z}}^{+}-\overline{\tau}\boldsymbol{I})\overline{\boldsymbol{Z}}^{\,-1/2}=\boldsymbol{Q}.
\end{aligned}
$$

Thus, by Lemma A.1 and the equation above,

$$\overline{\tau}\,\overline{\gamma}^{+} = \left\| (\overline{\boldsymbol{Z}}^{+})^{1/2}\overline{\boldsymbol{X}}^{+}(\overline{\boldsymbol{Z}}^{+})^{1/2} - \overline{\tau}\boldsymbol{I} \right\|_F = \|\boldsymbol{E}\|_F \leq \left\| \operatorname{symm}\left(\boldsymbol{W}\boldsymbol{E}\boldsymbol{W}^{-1}\right)\right\|_F = \|\operatorname{symm}(\boldsymbol{Q})\|_F\,. \tag{47}$$

On the other hand, by (9*) in Lemma 3.1 with $\overline{\mu} = \overline{\tau}$, $(\boldsymbol{X}, \boldsymbol{Z}) = (\overline{\boldsymbol{X}}, \overline{\boldsymbol{Z}})$, and $(\Delta\boldsymbol{X}, \Delta\boldsymbol{Z}, \Delta\boldsymbol{X}_\epsilon) = (\Delta\overline{\boldsymbol{X}}, \Delta\overline{\boldsymbol{Z}}, \Delta\overline{\boldsymbol{X}}_\epsilon)$, we have

$$\operatorname{symm}\left(\overline{\boldsymbol{Z}}^{1/2}(\overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{X}}\overline{\boldsymbol{Z}})\overline{\boldsymbol{Z}}^{\,-1/2}\right) = \overline{\mu}\boldsymbol{I} - \overline{\boldsymbol{Z}}^{1/2}(\overline{\boldsymbol{X}} + \Delta\overline{\boldsymbol{X}}_\epsilon)\overline{\boldsymbol{Z}}^{1/2}, \tag{48}$$

Thus, by using the equation above,

$$\begin{aligned}
\operatorname{symm}(\boldsymbol{Q}) &= \operatorname{symm}\left(\overline{\boldsymbol{Z}}^{1/2}(\overline{\boldsymbol{X}}^{+}\overline{\boldsymbol{Z}}^{+} - \overline{\tau}\boldsymbol{I})\overline{\boldsymbol{Z}}^{\,-1/2}\right) \\
&= \operatorname{symm}\left(\overline{\boldsymbol{Z}}^{1/2}((\overline{\boldsymbol{X}} + \Delta\overline{\boldsymbol{X}})(\overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{Z}}) - \overline{\tau}\boldsymbol{I})\overline{\boldsymbol{Z}}^{\,-1/2}\right) \\
&= \operatorname{symm}\left(\overline{\boldsymbol{Z}}^{1/2}(\overline{\boldsymbol{X}}\overline{\boldsymbol{Z}} + \overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{X}}\overline{\boldsymbol{Z}} + \Delta\overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}} - \overline{\tau}\boldsymbol{I})\overline{\boldsymbol{Z}}^{\,-1/2}\right) \\
&= \operatorname{symm}\left(\overline{\boldsymbol{Z}}^{1/2}\left[(\overline{\boldsymbol{X}}\overline{\boldsymbol{Z}} - \overline{\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}})\right]\overline{\boldsymbol{Z}}^{\,-1/2}\right) \\
&= \left[(\overline{\boldsymbol{Z}}^{1/2}\overline{\boldsymbol{X}}\overline{\boldsymbol{Z}}^{1/2} - \overline{\tau}\boldsymbol{I}) + \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)\right] \\
&\quad + \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} + \operatorname{symm}\left(\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}}\overline{\boldsymbol{Z}}^{\,-1/2}\right) \quad (\because (48)).
\end{aligned}$$

By the definition of $\overline{\gamma}$, $(\overline{\boldsymbol{X}}, \overline{\boldsymbol{Z}}) \in \mathcal{N}(\overline{\gamma}, \overline{\tau})$. By using (17), (48), and Lemma 4.5

together with the fact $(\overline{\boldsymbol{X}}, \overline{\boldsymbol{Z}}) \in \mathcal{N}(\overline{\gamma}, \overline{\tau})$,

$$
\begin{aligned}
\| \operatorname{symm}(\boldsymbol{Q})\|_F &= \left\| \frac{1}{2}(\boldsymbol{Q}+\boldsymbol{Q}^T)\right\|_F \le \|\boldsymbol{Q}\|_F \le \left\|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}_\epsilon\overline{\boldsymbol{Z}}^{1/2}\right\|_F + \left\|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}\Delta\overline{\boldsymbol{Z}}\overline{\boldsymbol{Z}}^{\,-1/2}\right\|_F \\
&\le \left\|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}_\epsilon\overline{\boldsymbol{Z}}^{1/2}\right\|_F + \left\|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}\ (\overline{\boldsymbol{Z}}^{1/2}\overline{\boldsymbol{Z}}^{\,-1/2})\ \Delta\overline{\boldsymbol{Z}}\overline{\boldsymbol{Z}}^{\,-1/2}\right\|_F \\
&\le \left\|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}_\epsilon\overline{\boldsymbol{Z}}^{1/2}\right\|_F + \left\|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}\overline{\boldsymbol{Z}}^{1/2}\right\|_F \left\|\overline{\boldsymbol{Z}}^{\,-1/2}\Delta\overline{\boldsymbol{Z}}\overline{\boldsymbol{Z}}^{\,-1/2}\right\|_F \\
&\le \tau\overline{\delta}_\epsilon + \frac{\|\boldsymbol{H}\|_F^2}{2\overline{\tau}(1-\overline{\gamma})^2} \quad (\because (17) \text{ and Lemma 4.5}) \\
&\le \tau\overline{\delta}_\epsilon + \frac{\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^2}{2\overline{\tau}(1-\overline{\gamma})^2} \\
&\le \tau\overline{\delta}_\epsilon + \frac{\left\|\overline{\tau}\boldsymbol{I}-\overline{\boldsymbol{Z}}^{1/2}\overline{\boldsymbol{X}}\overline{\boldsymbol{Z}}^{1/2}-\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}_\epsilon\overline{\boldsymbol{Z}}^{1/2}\right\|_F^2}{2\overline{\tau}(1-\overline{\gamma})^2} \quad (\because (48)) \\
&\le \tau\overline{\delta}_\epsilon + \frac{1}{2\overline{\tau}(1-\overline{\gamma})^2}\left[\left\|\overline{\tau}\boldsymbol{I}-\overline{\boldsymbol{Z}}^{1/2}\overline{\boldsymbol{X}}\overline{\boldsymbol{Z}}^{1/2}\right\|_F^2 + \left\|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}_\epsilon\overline{\boldsymbol{Z}}^{1/2}\right\|_F^2\right] \\
&\quad + \frac{2}{2\overline{\tau}(1-\overline{\gamma})^2}\left\|\overline{\tau}\boldsymbol{I}-\overline{\boldsymbol{Z}}^{1/2}\overline{\boldsymbol{X}}\overline{\boldsymbol{Z}}^{1/2}\right\|_F \left\|\overline{\boldsymbol{Z}}^{1/2}\Delta\overline{\boldsymbol{X}}_\epsilon\overline{\boldsymbol{Z}}^{1/2}\right\|_F \\
&\quad (\because \|\boldsymbol{A}+\boldsymbol{B}\|_F^2 \le \|\boldsymbol{A}\|_F^2 + \|\boldsymbol{B}\|_F^2 + 2\,\|\boldsymbol{A}\|_F\,\|\boldsymbol{B}\|_F) \\
&\le \tau\overline{\delta}_\epsilon + \frac{\left(\overline{\tau}^2\overline{\gamma}^2 + \tau^2\overline{\delta}_\epsilon^2 + 2\overline{\tau}\tau\overline{\gamma}\overline{\delta}_\epsilon\right)}{2\overline{\tau}(1-\overline{\gamma})^2} \quad (\because (\overline{\boldsymbol{X}}, \overline{\boldsymbol{Z}}) \in \mathcal{N}(\overline{\gamma}, \overline{\tau}) \text{ and } (17)).
\end{aligned}
$$

In addition, by Condition 4.2,

$$
\begin{aligned}
\| \operatorname{symm}(\boldsymbol{Q})\|_F &\le \tau\zeta(1-\overline{\theta})\overline{\gamma}^2 + \frac{1}{2\overline{\tau}(1-\overline{\gamma})^2}\left(\overline{\tau}^2\overline{\gamma}^2 + \tau^2\zeta^2(1-\overline{\theta})^2\overline{\gamma}^4 + 2\overline{\tau}\tau\zeta(1-\overline{\theta})\overline{\gamma}^3\right) \\
&= \overline{\tau}\zeta\overline{\gamma}^2 + \frac{1}{2\overline{\tau}(1-\overline{\gamma})^2}\left(\overline{\tau}^2\overline{\gamma}^2 + \zeta^2\overline{\tau}^2\overline{\gamma}^4 + 2\zeta\overline{\tau}^2\overline{\gamma}^3\right) \quad (\because \overline{\tau} = (1-\overline{\theta})\tau) \\
&= \overline{\tau}\overline{\gamma}^2\left[\zeta + \frac{1}{2}\left(\frac{1+\zeta\overline{\gamma}}{1-\overline{\gamma}}\right)^2\right].
\end{aligned}
$$

Therefore, together with (47), we have

$$
\begin{aligned}
\overline{\tau}\ \overline{\gamma}^+ &\le \| \operatorname{symm}(\boldsymbol{Q})\|_F \le \overline{\tau}\overline{\gamma}^2\left[\zeta + \frac{1}{2}\left(\frac{1+\zeta\overline{\gamma}}{1-\overline{\gamma}}\right)^2\right], \\
\overline{\gamma}^+ &\le \overline{\gamma}^2\left[\zeta + \frac{1}{2}\left(\frac{1+\zeta\overline{\gamma}}{1-\overline{\gamma}}\right)^2\right]. \qquad (49)
\end{aligned}
$$

By (23), for any $\overline{\gamma} \le \beta$,

$$
\begin{aligned}
\zeta &< \frac{1}{\beta} - 2 \le \frac{1}{\overline{\gamma}} - 2, \qquad (50) \\
\zeta\overline{\gamma} &< 1 - 2\overline{\gamma}. \qquad (51)
\end{aligned}
$$

Thus, by (49) – (51) together with $\overline{\gamma} \leq \beta$, we have

$$\overline{\gamma}^{+} \leq \overline{\tau}\overline{\gamma}^2 \left[\zeta + \frac{1}{2}\left(\frac{1+\zeta\overline{\gamma}}{1-\overline{\gamma}}\right)^2\right] < \overline{\gamma}^2 \left[\zeta + \frac{1}{2}\left(\frac{1+(1-2\overline{\gamma})}{1-\overline{\gamma}}\right)^2\right] = \overline{\gamma}^2(\zeta+2) \leq \frac{\overline{\gamma}^2}{\beta} \leq \overline{\gamma},$$

Now, we finish the proof by showing $\overline{\boldsymbol{X}}^{+} \succ \boldsymbol{0}$. By the inequality above,

$$\left\|(\overline{\boldsymbol{Z}}^{+})^{1/2}\overline{\boldsymbol{X}}(\overline{\boldsymbol{Z}}^{+})^{1/2} - \overline{\tau}\boldsymbol{I}\right\|_F < \overline{\gamma\tau}. \tag{52}$$

By Lemma A.1 and (52),

$$\begin{aligned} \lambda_{\min}\left((\overline{\boldsymbol{Z}}^{+})^{1/2}\overline{\boldsymbol{X}}(\overline{\boldsymbol{Z}}^{+})^{1/2} - \overline{\tau}\boldsymbol{I}\right) &\geq -\left\|(\overline{\boldsymbol{Z}}^{+})^{1/2}\overline{\boldsymbol{X}}(\overline{\boldsymbol{Z}}^{+})^{1/2} - \overline{\tau}\boldsymbol{I}\right\|_F > -\overline{\gamma\tau}, \\ \lambda_{\min}\left((\overline{\boldsymbol{Z}}^{+})^{1/2}\overline{\boldsymbol{X}}(\overline{\boldsymbol{Z}}^{+})^{1/2}\right) &> (1-\overline{\gamma})\overline{\tau} > 0, \end{aligned}$$

so $(\overline{\boldsymbol{Z}}^{+})^{1/2}\overline{\boldsymbol{X}}^{+}(\overline{\boldsymbol{Z}}^{+})^{1/2} \succ \boldsymbol{0}$. Therefore, $\overline{\boldsymbol{X}}^{+} \succ \boldsymbol{0}$ because $\overline{\boldsymbol{Z}}^{+} \succ \boldsymbol{0}$ as we showed above.

□

Proof. of Lemma 4.2.
To simplify notation, let ϵ denote the target distance, $\min(\overline{\tau}^{\sigma}, \alpha)$. By mathematical induction, for $q \geq 2$, we can rewrite the inequality in Lemma 4.1 as

$$\overline{\gamma}_q < \beta\left(\frac{\overline{\gamma}_1}{\beta}\right)^{2^{q-1}}.$$

Thus, at worst, we can reach the target distance by

$$\begin{aligned} \overline{\gamma}_q &< \beta\left(\frac{\overline{\gamma}_1}{\beta}\right)^{2^{q-1}} < \epsilon \\ 2^{q-1} &> \frac{\log(\epsilon/\beta)}{\log(\overline{\gamma}_1/\beta)} = \frac{\log(\beta/\epsilon)}{\log(\beta/\overline{\gamma}_1)} \quad (\because \overline{\gamma}_1 < \overline{\gamma}_0 \leq \beta) \\ q &> 1 + \log_2\left(\frac{\log(\beta/\epsilon)}{\log(\beta/\overline{\gamma}_1)}\right). \end{aligned}$$

Therefore, the required iteration Q is bounded by

$$Q \leq 2 + \log_2\left(\frac{\log(\beta/\epsilon)}{\log(\beta/\overline{\gamma}_1)}\right) = 2 + \log_2\left(\frac{\log(\beta/\min(\overline{\tau}^{\sigma}, \alpha))}{\log(\beta/\overline{\gamma}_1)}\right),$$

so, $Q = O\left(\log\left(\log(1/\overline{\tau}^{\sigma})\right)\right)$ for given α, β, and $\overline{\gamma}_1$.

□

References

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

[2] Helmberg, C., Rendl, F., Vanderbei, R.J., Wolkowicz, H.: An interior-point method for semidefinite programming. SIAM J. Opt. **6**(2), 342–361 (1996)

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

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

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

[6] Alizadeh, F., Haeberly, J.P.A., Overton, M.L.: Complementarity and nondegeneracy in semidefinite programming. Math. Prog. **77**(1), 111–128 (1997)

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

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

[9] Kojima, M., Shida, M., Shindoh, S.: 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)

[10] Monteiro, R.D.C.: 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)

[11] Monteiro, R.D.C., Zhang, Y.: 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)

[12] Potra, F.A., Sheng, R.: Superlinear convergence of a predictor-corrector method for semidefinite programming without shrinking central path neighborhood. Tech. Rep. 91, Reports on Computational Mathematics, Department of Mathematics, Univ. of Iowa (1996)

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

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

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

[16] Toh, K.C., Todd, M.J., Tütüncü, R.H.: On the implementation and usage of SDPT3 - a MATLAB software package for semidefinite-quadratic-linear programming, version 4.0. In: M.F. Anjos, J.B. Lasserre (eds.) Handbook on Semidefinite, Conic and Polynomial Optimization, vol. 166, pp. 715–754. Springer, New York (2012)

[17] Fujisawa, K., Kojima, M., Nakata, K.: Sdpa (semidefinite programming algorithm) Users Manual B-308. Tech. rep., Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Japan (1996). URL `ftp://ftp.is.titech.ac.jp/pub/OpRes/software/SDPA`

[18] Fukuda, M., Kojima, M., Murota, K., Nakata, K.: Exploiting sparsity in semidefinite programming via matrix completion I: implementation and numerical results. SIAM J. Opt. **11**(3), 647–674 (2000)

[19] Dantzig, G.B., Ye, Y.: A build-up interior-point method for linear programming: Affine scaling form. Tech. rep., Stanford Univ. (1991)

[20] Hertog, D.D., Roos, C., Terlaky, T.: Adding and deleting constraints in the path-following method for LP. In: Advances in Optimization and Approximation, vol. 1, pp. 166–185. Springer, New York (1994)

[21] Kaliski, J.A., Ye, Y.: A decomposition variant of the potential reduction algorithm for linear programming. Manage. Sci. **39**, 757–776 (1993)

[22] Tits, A.L., Absil, P.A., Woessner, W.P.: Constraint reduction for linear programs with many inequality constraints. SIAM J. Opt. **17**(1), 119–146 (2006)

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

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

[25] Ye, Y.: An $O(n^3L)$ potential reduction algorithm for linear programming. Math. Prog. **50**(2), 239–258 (1991)

[26] Jung, J.H., O'Leary, D.P., Tits, A.L.: Adaptive constraint reduction for training support vector machines. Elec. Trans. Numer. Anal. **31**, 156–177 (2008)

[27] Williams, J.A.: 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)

[28] Jung, J.H., O'Leary, D.P., Tits, A.L.: Adaptive constraint reduction for convex quadratic programming. Comput. Optim. Appl. **51**(1), 125–157 (2012)

[29] Park, S., O'Leary, D.P.: A polynomial time constraint-reduced algorithm for semidefinite optimization problems. J. Opt. Theory Appl. (2015)

[30] Park, S., O'Leary, D.P.: A polynomial time constraint reduced algorithm for semidefinite optimization problems, with convergence proofs. Tech. rep., Univ. of Maryland, College Park, MD (2015). URL http://www.optimization-online.org/DB_HTML/2013/08/4011.html

[31] Park, S.: Matrix reduction in numerical optimization. Ph.D. thesis, Computer Science Department, Univ. of Maryland, College Park, MD (2011). URL http://drum.lib.umd.edu/handle/1903/11751

[32] Fujisawa, K., Kojima, M., Nakata, K.: Exploiting sparsity in primal-dual interior-point methods for semidefinite programming. Math. Prog. **79**(1), 235–253 (1997)