Mathematical Programming manuscript No.
(will be inserted by the editor)

Kim-Chuan Toh

An inexact primal-dual path following algorithm for convex quadratic SDP

November 24, 2006

Abstract We propose primal-dual path-following Mehrotra-type predictor-corrector methods for solving convex quadratic semidefinite programming (QSDP) problems of the form: $\min_X\{\frac{1}{2}X\bullet\mathcal{Q}(X)+C\bullet X : \mathcal{A}(X)=b, X\succeq 0\}$, where $\mathcal{Q}$ is a self-adjoint positive semidefinite linear operator on $\mathcal{S}^n$, $b\in R^m$, and $\mathcal{A}$ is a linear map from $\mathcal{S}^n$ to R^m. At each interior-point iteration, the search direction is computed from a dense symmetric indefinite linear system (called the augmented equation) of dimension $m+n(n+1)/2$. Such linear systems are typically very large and can only be solved by iterative methods. We propose three classes of preconditioners for the augmented equation, and show that the corresponding preconditioned matrices have favorable asymptotic eigenvalue distributions for fast convergence under suitable nondegeneracy assumptions. Numerical experiments on a variety of QSDPs with n up to 1600 are performed and the computational results show that our methods are efficient and robust.

Keywords semidefinite programming – semidefinite least squares – interior point method – inexact search direction – Krylov iterative method

Mathematics Subject Classification (2000) 90C22, 90C25, 90C51, 65F10

1 Introduction

Let $\mathcal{S}^n$ be the space of $n\times n$ symmetric matrices endowed with the standard trace inner product denoted by "$\bullet$". We consider the following convex

Department of Mathematics, National University of Singapore, 2 Science Drive 2, Singapore 117543 (mattohkc@nus.edu.sg); and Singapore-MIT Alliance, 4 Engineering Drive 3, Singapore 117576. Research supported in part by Academic Research Grant R146-000-076-112.

quadratic semidefinite program (QSDP):

$$(QSDP)\ \ \min_X \Big\{ X \bullet \mathcal{Q}(X)/2 + C \bullet X \ :\ \mathcal{A}(X) = b, \quad X \succeq 0 \Big\}, \qquad (1)$$

where $\mathcal{Q} : \mathcal{S}^n \to \mathcal{S}^n$ is a given self-adjoint positive semidefinite linear operator, $\mathcal{A} : \mathcal{S}^n \to I\!R^m$ is a linear map, and $b \in I\!R^m$. The notation $X \succeq 0$ means that X is positive semidefinite. Let $\mathcal{A}^T$ be the adjoint of $\mathcal{A}$. The Lagrangian dual problem of (1) is given by:

$$\max_{X,y,S} \Big\{ -X \bullet \mathcal{Q}(X)/2 + b^T y \ :\ \mathcal{A}^T(y) - \mathcal{Q}(X) + S = C,\ S \succeq 0 \Big\}. \qquad (2)$$

We use the following notation and terminology. For an integer n, we let $\bar{n} = n(n+1)/2$. Given $U \in I\!R^{q\times l}, V \in I\!R^{q\times n}$, the symmetrized Kronecker product $U \circledast V$ is the linear map from $I\!R^{n\times l}$ to $\mathcal{S}^q$ defined by $U \circledast V(M) = (VMU^T + UM^TV^T)/2$. For $U \in I\!R^{p\times l}$ and $V \in I\!R^{q\times n}$, the Kronecker product $U \otimes V$ is the liner map from $I\!R^{n\times l}$ to $I\!R^{q\times p}$ defined by $U \otimes V(M) = VMU^T$; see [15, p. 254]. The Hadamard product, $U \circ V$, of two matrices with the same dimensions is defined by $(U \circ V)_{ij} = U_{ij}V_{ij}$. The set of symmetric positive semidefinite (definite) matrices is denoted by $\mathcal{S}^n_+$ ($\mathcal{S}^n_{++}$). We use $\|\cdot\|_2$ to denote the vector 2-norm or matrix 2-norm, and $\|\cdot\|_F$ to denote the Frobenius norm. We denote the identity matrix or operator of dimension d by I_d. The vector of ones is denoted by e. We let $I\!R^{n\times n}_+$ be the set of nonnegative matrices. The notation $x = \Theta(\nu)$ means that there exist constants $c_1, c_2 > 0$ independent of ν such that $c_1\nu \le x \le c_2\nu$.

Let p be the rank of $\mathcal{Q}$. By considering the Cholesky factorization of $\mathcal{Q}$, it is readily shown that (1) can be reformulated as a standard semidefinite-quadratic-linear programming (SQLP) by introducing an additional p linear constraints and $p+1$ variables. Unfortunately, the computational cost and memory space needed by a standard primal-dual interior-point method to solve the reformulated problem grow at least like $\Theta((m+p)^3)$ and $\Theta((m+p)^2)$, respectively. Thus unless $m+p$ is small, it not viable to solve (1) by reformulating it into a standard SQLP.

The problem (1) can also be reformulated as a semidefinite linear complementarity problem (SDLCP) [17]. However, the computational cost at each interior-point iteration for the SDLCP problem is the same as that for (1) because both need to solve linear systems with coefficient matrices having the same dimensions and numerical properties. In [17], polynomial iteration complexities of some theoretical path-following and potential reduction methods were established. But as far as we know, there is little research on the efficient numerical computation of the solution of (1) or the SDLCP problem derived from it. In [23], a theoretical primal-dual potential reduction algorithm was proposed for (1) and (2). At each iteration, the search direction is computed from a *dense* augmented system of dimension $\bar{n} + m$. The authors suggested using the conjugate gradient method to compute an approximate direction. However, as the focus of [23] was on establishing polynomial iteration complexity, the crucial issue of preconditioning for the conjugate gradient method was not discussed. There was also no numerical implementation to test the proposed method.

A prime example of QSDP is the nearest correlation matrix (NCM) problem, where given a data matrix $K \in \mathcal{S}^n$ and a self-adjoint linear operator $\mathcal{L}$ on $\mathcal{S}^n$, we want to solve

$$\min_X \big\{ \|\mathcal{L}(X-K)\|_F^2/2 \,:\, \mathrm{diag}(X) = e, X \succeq 0 \big\}. \tag{3}$$

The QSDP resulting from (3) has $\mathcal{Q} = \mathcal{L}^2$ and $C = -\mathcal{L}^2(K)$. Previous research on QSDP were mainly on the NCM problem with the special choice $\mathcal{L} = U^{1/2} \circledast U^{1/2}$ (hence $\mathcal{Q} = U \circledast U$) for a given $U \in \mathcal{S}^n_+$. One of the earliest work on such a special case for (3) was by Higham [14] who proposed a modified alternating projection solution method. The paper [14] also briefly considered the problem (3) with $\mathcal{L}(X) = \Sigma \circ X$ for a given $\Sigma \in \mathcal{S}^n \cap \mathbb{R}^{n\times n}_+$ (correspondingly, $\mathcal{Q}(X) = U \circ X$, with $U = \Sigma \circ \Sigma$). However, no practical solution method was proposed in [14].

Subsequent works on (3) for the special case where $\mathcal{L} = I$ (hence $\mathcal{Q} = I$), but extendable to the case $\mathcal{L} = U^{1/2} \circledast U^{1/2}$, include Anjos et al. [4], Malick [21], and Sun et al. [25]. The latter two papers used a Lagrangian dual approach that relied critically on the assumption that $\mathcal{Q} = U \circledast U$ to derive an analytical formula for the projection of $\mathbb{R}^{n\times n}$ onto $\mathcal{S}^n_+$ with respect to the norm $\|U^{1/2}(\cdot)U^{1/2}\|_F$. The recent paper [29] for (1) also focused on the special case $\mathcal{Q} = U \circledast U$. It proposed efficient methods for computing the search direction at each interior-point iteration for (1) by applying an iterative solver with two suitably designed preconditioners to the $m \times m$ Schur complement equation. All the previous techniques for solving (1) with $\mathcal{Q} = U \circledast U$, however, do not extend to the case of a general positive semidefinite self-adjoint linear operator $\mathcal{Q}$. The last statement holds true even if $\mathcal{Q}$ is a diagonal operator where $\mathcal{Q}(X) = U \circ X$ for some $U \in \mathcal{S}^n \cap \mathbb{R}^{n\times n}_+$.

The problem (1) also arises from the nearest Euclidean distance matrix (EDM) problem [3]:

$$\min \big\{ \|H \circ (B - \mathcal{L}_{\mathrm{E}}(X))\|_F^2/2 : (\mathcal{L}_{\mathrm{E}}(X))_{ij} = B_{ij} \,\forall\, (ij) \in \mathcal{E},\, X \succeq 0 \big\}, \tag{4}$$

where $B \in \mathcal{S}^n$ is a given dissimilarity matrix, H is a given weight matrix which typically has the same sparsity pattern as B, $\mathcal{E}$ is a given set of indices, and $\mathcal{L}_{\mathrm{E}}(X) = \mathrm{diag}(X_V)e^T + e\,\mathrm{diag}(X_V)^T - 2X_V$ with $X_V = VXV^T$. Here $V \in \mathbb{R}^{n\times(n-1)}$ is a given matrix with orthonormal columns and $V^Te = 0$. Note that the operator $\mathcal{Q}$ associated with (4) is typically not positive definite but positive semidefinite. In [3], the EDM problem (4) was solved by a primal-dual interior-point method for which the search direction at each iteration was computed from a linear system of dimension n^2 by a direct method. However, the direct approach in [3] is computationally viable only for small problems, say with n less than a hundred. We refer the reader to [3] wherein some other applications of QSDP are cited. More recently, QSDP has also been formulated in local compliance estimation in deformable object modelling [18].

While the search direction at each interior-point iteration for (1) can be computed from the $m \times m$ Schur complement equation when $\mathcal{Q} = U \circledast U$ [29], the corresponding direction for a general QSDP for which $\mathcal{Q}$ does not have this special form (even if $\mathcal{Q} = U \circledast V$ where $U \neq V$ or $\mathcal{Q}$ is a

diagonal operator) must be computed from a *dense* augmented equation of dimension $\bar{n}+m$. This is because reducing the augmented equation to the Schur complement equation is not viable due to the excessive memory ($\Theta(n^4)$ bytes) and computational cost ($\Theta(n^6)$ flops) required. Unfortunately, the augmented equation is generally very large, even for a moderate n, say 500. Thus it is impossible to solve the augmented equation by a direct method on a personal computer when n is say more than a hundred. The only other alternative is to use a preconditioned Krylov subspace iterative method. But because the augmented and Schur complement equations are quite different in structure, the preconditioning techniques described in [29] are not applicable to the augmented equation.

Our inexact primal-dual path-following method for (1) and (2) follows the same framework as the primal-dual path-following method with Mehrotra's predictor-corrector appeared in [29]. The key difference is in the computation of search directions. Here, the search direction at each iteration is computed from a *dense* augmented equation of dimension $m+\bar{n}$. We apply a preconditioned symmetric quasi-minimal residual (PSQMR) method [11] to solve the augmented equation.

In this paper, we are primarily interested in QSDPs for which following assumption holds:

Assumption A1. The linear map $\mathcal{A}$ is sparse or structured in the sense that if A_k denotes $\mathcal{A}^T$ acting on the kth unit vector of $\mathbb{R}^m$, then the matrices $A_1,\ldots,A_m$ are either sparse or low-rank. We assume that m is a moderate number, say less than 5000, so that a matrix of the form $\mathcal{A}U\circledast V\mathcal{A}^T$ and its Cholesky factorization can be computed at a moderate cost. The dimension of the primal matrix X is restricted to the range of say less than 2000 so that its full eigenvalue decomposition can be computed at a moderate cost.

Our contributions in this paper are as follows. We first analyze the asymptotic spectral property of the augmented matrix arising at each interior-point iteration for (1). Then we design three classes of preconditioners for the augmented matrix. Under suitable conditions including nondegeneracy of the optimal solution, the preconditioned matrices are shown to have favorable asymptotic spectral distributions to accelerate the convergence of the PSQMR method used to solve the augmented equation. But note that for one of the classes of preconditioners, no nondegeneracy assumption is needed. We also addressed numerous implementation issues to make our inexact interior-point method for (1) practical.

Preconditioning for the augmented equations arising from interior-point methods for *sparse* linearly constrained convex quadratic programming (LCCQP) problems of the form

$$\min\left\{x^TQx/2+c^Tx\,:\,Ax=b,x\in\mathbb{R}^n_+\right\} \tag{5}$$

has been studied in [9]. The preconditioners constructed in [9] are based on those proposed in [16], with the $(1,1)$ block $-X^{-1}S-Q$ of the augmented matrix being approximated by $-X^{-1}S-\mathrm{diag}(Q)$. Note that X and S are positive definite diagonal matrices for LCCQP. A reader who is familiar with SDP would realize that the augmented equations in LCCQP and QSDP are different in a fundamental way. For LCCQP, the augmented matrices

are sparse if A and Q are sparse, but those in QSDP are typically dense even if $\mathcal{A}$ and $\mathcal{Q}$ are sparse. Thus the construction of preconditioners for the augmented matrices in QSDP is considerably more difficult and the analysis and computation involved are more complex.

The problem (1) can be viewed as a generalization of (5) just as a linear SDP is a generalization of a linear program. We note that when the variable X in (1) is restricted to a diagonal matrix of the form $X = \mathrm{diag}(x)$, then (1) reduces to (5). Based on this observation, many of the results derived for (1) in this paper can be modified to suit (5). However, to keep the paper coherent, we will not separately state the corresponding results for (5).

The paper is organized as follows. In the next section, we derive the augmented equation from which the search direction at each interior-point iteration for (1) is computed. In Section 3, the asymptotic spectrum and conditioning of the augmented matrix are analyzed. This motivates the construction of preconditioners for the augmented matrix in Section 4. Three classes of preconditioners are constructed, and the asymptotic spectra of the associated preconditioned matrices are analyzed. In Section 5, we discuss the construction of symmetrized Kronecker product approximations for self-adjoint linear operators. Section 6 presents numerical experiments to test the performance of our inexact interior-point methods that employ iterative solvers with appropriate preconditioners to solve the augmented equation at each iteration. In the last section we give the conclusion.

1.1 Additional notation and terminology

For a linear map $\mathcal{T} : (\mathcal{X}, \bullet) \to (\mathcal{Y}, \bullet)$, where $\mathcal{X} = \mathbb{R}^{k\times l}$ or $\mathcal{S}^l$, and $\mathcal{Y} = \mathbb{R}^{p\times q}$ or $\mathcal{S}^q$, we define $\|\mathcal{T}\| = \max\{\|\mathcal{T}(M)\|_F \,:\, \|M\|_F \le 1\}$. The adjoint of $\mathcal{T}$ is denoted by $\mathcal{T}^T$. The null space is denoted by $\mathcal{N}(\mathcal{T})$. The matrix representation of $\mathcal{T}$ with respect to the canonical orthonormal bases of $\mathcal{X}$ and $\mathcal{Y}$ is denoted by $\mathrm{mat}(\mathcal{T})$. We will typically identify $\mathcal{T}$ with $\mathrm{mat}(\mathcal{T})$ and a phrase such as "the matrix $\mathcal{T}$" means the matrix representation of $\mathcal{T}$. Note that $\|\mathcal{T}\| = \|\mathrm{mat}(\mathcal{T})\|_2$. Given a self-adjoint linear operator $\mathcal{V}$ defined on a finite dimensional inner product space, we let $\lambda_j(\mathcal{V})$ be the jth eigenvalue (sorted in ascending order) and $\mathrm{eig}(\mathcal{V})$ be the set of eigenvalues. The largest and smallest eigenvalues in magnitudes are denoted by $\lambda_{\max}(\mathcal{V})$ and $\lambda_{\min}(\mathcal{V})$, respectively. The condition number is denoted by $\kappa(\mathcal{V}) := |\lambda_{\max}(\mathcal{V})|/|\lambda_{\min}(\mathcal{V})|$. Note that $\|\mathcal{V}\| = |\lambda_{\max}(\mathcal{V})|$, and if $\mathcal{V}$ is invertible, then $\|\mathcal{V}^{-1}\| = 1/|\lambda_{\min}(\mathcal{V})|$. For two matrices P and Q, $[P; Q]$ denotes the matrix obtained by appending Q to the last row of P. For self-adjoint linear operators $\mathcal{S}, \mathcal{T} : (\mathcal{X}, \bullet) \to (\mathcal{X}, \bullet)$, the notation $\mathcal{S} \preceq \mathcal{T}$ means that $M \bullet \mathcal{S}(M) \le M \bullet \mathcal{T}(M)$ for all $M \in \mathcal{X}$.

2 Computation of search direction

Our interior-point method for (1) is a primal-dual path-following method with Mehrotra's predictor-corrector. It is based on the perturbed KKT conditions associated with (1) and (2), which are given by

$$-\mathcal{Q}(X) + \mathcal{A}^T(y) + S = C, \quad \mathcal{A}(X) = b, \quad XS = \nu I, \quad X, S \succ 0, \qquad (6)$$

where $\nu > 0$ is a parameter that is to be driven to zero explicitly.

Let $\rho \geq 0$ be a given constant. We may observe that by adding $-\rho\mathcal{A}^T\mathcal{A}(X) = -\rho\mathcal{A}^T b$ to the first condition in (6), we get an equivalent condition:

$$-\mathcal{Q}_\rho(X) + \mathcal{A}^T(y) + S = C_\rho, \tag{7}$$

where $\mathcal{Q}_\rho := \mathcal{Q} + \rho\mathcal{A}^T\mathcal{A}$ and $C_\rho := C - \rho\mathcal{A}^T b$. Thus we can replace the first condition in (6) by (7). The motivation for using (7) is given in Remark 1.

The framework of our interior-point algorithm for (1) is described in Algorithm IP-QSDP in [29]. At a given iterate (X, y, S) with $X, S \succ 0$ (positive definite), the search direction $(\Delta X, \Delta y, \Delta S)$ at the current iteration is the solution of the following symmetrized Newton system:

$$\begin{array}{rrcl} -\mathcal{Q}_\rho(\Delta X) + \mathcal{A}^T(\Delta y) + \Delta S & = R_d & = & C_\rho - S - \mathcal{A}^T y + \mathcal{Q}_\rho(X) \\ \mathcal{A}(\Delta X) & = R_p & = & b - \mathcal{A}(X) \\ \mathcal{F}_S\Delta X + \mathcal{F}_X\Delta S & = R_c & = & \sigma\mu I - H_K(XS), \end{array} \tag{8}$$

where $\mathcal{F}_X$ and $\mathcal{F}_S$ are linear operators on $\mathcal{S}^n$ that depend on the symmetrization scheme $H_K(\cdot)$ chosen; for more details, see for example [27]. Here $\mu = X \bullet S/n$, and $\sigma \in (0, 1)$ is the centering parameter.

By eliminating ΔS, we get the following augmented equation of dimension $\bar{n} + m$:

$$\mathcal{B}_\rho \begin{bmatrix} \Delta X \\ \Delta y \end{bmatrix} = \begin{bmatrix} R_a \\ R_p \end{bmatrix}, \quad \text{where } \mathcal{B}_\rho = \begin{bmatrix} -\mathcal{K}_\rho & \mathcal{A}^T \\ \mathcal{A} & 0 \end{bmatrix}, \tag{9}$$

with $\mathcal{K}_\rho = \mathcal{F}_X^{-1}\mathcal{F}_S + \mathcal{Q}_\rho$ and $R_a = R_d - \mathcal{F}_X^{-1}R_c$. In this paper, we will consider only the Nesterov-Todd (NT) symmetrization scheme [27] for which $\mathcal{F}_X^{-1}\mathcal{F}_S = W^{-1} \circledast W^{-1}$, where $W \in \mathcal{S}^n_{++}$ is the unique matrix satisfying $WSW = X$. But note that the preconditioning strategies we are describing later are also applicable to the purely primal (purely dual) scheme for which $\mathcal{F}_X^{-1}\mathcal{F}_S = X^{-1} \circledast X^{-1}$ $(S \circledast S)$.

By further eliminating ΔX from (9), we get the Schur complement equation of dimension m below:

$$M_\rho\Delta y = R_p + \mathcal{A}\mathcal{K}_\rho^{-1}R_a, \quad \text{where } M_\rho = \mathcal{A}\mathcal{K}_\rho^{-1}\mathcal{A}^T. \tag{10}$$

When $\mathcal{Q}$ has the special form $U \circledast U$ and $\rho = 0$, the Schur complement matrix M_ρ can be computed at a cost of at most $4mn^3 + m^2n^2$ floating point operations (flops); see [29]. This is done by exploiting the fact that $\mathcal{K}_\rho^{-1}$ has an analytical expression and $\mathcal{K}_\rho^{-1}(V)$ can be computed with $\Theta(n^3)$ flops for any given $V \in \mathcal{S}^n$. However, for a general $\mathcal{Q}$, $\mathcal{K}_\rho^{-1}$ does not have an analytical expression and the computation of M_ρ requires the inversion of $\mathcal{K}_\rho$ (with dimension $\bar{n}$) that costs $\Theta(n^6)$ flops. Such a cost is prohibitively expensive when n is larger than a hundred. Thus for a general $\mathcal{Q}$, computing the search direction via (10) is not a viable approach. The practical alternative is to use the augmented equation (9).

However, the linear system (9), with dimension $\bar{n} + m$, is a large system even for a moderate n. In addition, it is typically dense since $\mathcal{F}_X^{-1}\mathcal{F}_S$ is

typically so. Thus unless $\bar{n} + m$ is of moderate size, it is impossible to solve (9) by a direct method, and iterative methods are the only alternatives. In this paper, we use the PSQMR method as the iterative solver for (9). The unpreconditioned PSQMR method is mathematically equivalent to the well known minimal residual (MINRES) method [11,26]. It has the same work and storage requirements as MINRES and usually converges in about the same number of iterations. But the PSQMR method has the advantage that it allows the use symmetric indefinite preconditioners whereas MINRES allows only symmetric positive definite preconditioners.

As iterative methods do not solve a linear system exactly (modulo rounding errors), we need the following result to gauge the quality of the computed direction.

Lemma 1 *Suppose the residual in computed direction $(\Delta X, \Delta y)$ for (9) is given by $(\eta_1\,,\,\eta_2)$, and that ΔS is computed exactly from the first equation in (8) based on the computed $(\Delta X, \Delta y)$. Then the residual of the direction $(\Delta X, \Delta y, \Delta S)$ with respect to (8) is given by $(0\,,\,\eta_2\,,\,-\mathcal{F}_X(\eta_1))$.*

Proof. We omit the proof since it is straightforward. □

In the numerical experiments in Section 6, we deem a direction $(\Delta X, \Delta y)$ computed by an iterative solver from (9) to be sufficiently accurate if

$$\max\{\|\eta_2\|_2, \|\mathcal{F}_X(\eta_1)\|_F\} \;\le\; 0.01\max\{\|R_d\|_F, \|R_p\|_2, \|R_c\|_F\}. \qquad (11)$$

Observe that the above criterion required the symmetrized Newton system (8) be solved with a relative accuracy of 0.01. While we do not investigate the question of whether such a criterion is sufficient for establishing a globally polynomial convergent interior-point algorithm, we note that it is motivated by a similar criterion used in the globally polynomial convergent inexact interior-point algorithms for SDP in [32].

We end this section by stating some basic facts concerning a 2×2 block matrix of the form $G = [-U,\, V^T;\, V,\, W]$. The inversion of such a matrix is at the heart of the computation of search directions. When the (1,1) block U and the associated Schur complement matrix $Y := W + VU^{-1}V^T$ are invertible, the inverse of G can be computed from the following analytical expression whose proof can be found for example in [26, p. 389]:

$$G^{-1} \;=\; \begin{bmatrix} -U^{-1} + U^{-1}V^TY^{-1}VU^{-1} & U^{-1}V^TY^{-1} \\ Y^{-1}VU^{-1} & Y^{-1} \end{bmatrix}. \qquad (12)$$

Lemma 2 *Suppose $U \in \mathcal{S}^p_{++}$, $W \in \mathcal{S}^q_+$, and $V \in \mathbb{R}^{q\times p}$ has full row rank. Let $G = [-U,\, V^T;\, V,\, W]$ and $Y = W + VU^{-1}V^T$. The following results hold:*

$$\max\{\|U\|, \|V\|, \|W\|\} \;\le\; \|G\| \;\le\; 2\max\{\|U\|, \|V\|, \|W\|\},$$

$$\|Y^{-1}\| \;\le\; \|G^{-1}\| \;\le\; 2\max\{\|U^{-1}\|,\, \|Y^{-1}\|\}.$$

Proof. We shall prove only the second inequality since the first follows from the inequality, $\|G\|^2 \le \|U\|^2 + 2\|V\|^2 + \|W\|^2$, proven in [6]. It is readily shown that the (1,1) block of G^{-1} can be expressed as $U^{-1/2}(\Pi - I)U^{-1/2}$, where $\Pi = U^{-1/2}V^TY^{-1}VU^{-1/2}$. By noting that $0 \preceq \Pi \preceq I$, the required result can be proven easily. □

3 Conditioning of the augmented matrix $\mathcal{B}_\rho^k$

We made the following assumptions on (1) and (2).
Assumption A2. The problems (1) and (2) are strictly feasible and that $\mathcal{A}$ is surjective. Note that the last condition implies that $m \le \bar{n}$.

Assumption A2 stated the necessary and sufficient conditions for the existence and uniqueness of solutions (X^ν, y^ν, S^ν) of the central path equations (6). Also, these solutions converge to some optimal solution (X^*, y^*, S^*) as $\nu \downarrow 0$; see [13] and [20]. We further assume that (X^*, y^*, S^*) satisfies the following assumption.
Assumption A3. Strict complementarity holds for (X^*, y^*, S^*) in the sense defined in [2]. Thus the ranks of X^* and S^* sum to n.

Suppose $\{\nu_k\}$ is a monotonically decreasing sequence with $\lim_{k\to\infty} \nu_k = 0$. Let the coefficient matrices in (9) and (10) corresponding to (X^ν, y^ν, S^ν) be $\mathcal{B}_\rho^\nu$ and M_ρ^ν, respectively. For simplicity of notation, we write $\mathcal{B}_\rho^k$, M_ρ^k, X^k, S^k, etc., for $\mathcal{B}_\rho^{\nu_k}$, $M_\rho^{\nu_k}$, X^{ν_k}, S^{ν_k}, and so on. Since X^k and S^k commute, there is an orthogonal matrix P^k that simultaneously diagonalizes X^k and S^k so that

$$X^k = P^k \Lambda^k (P^k)^T, \quad S^k = P^k \Sigma^k (P^k)^T,$$

where the eigenvalue matrices $\Lambda^k = \mathrm{diag}(\lambda_1^k, \ldots, \lambda_n^k)$, $\Sigma^k = \mathrm{diag}(\sigma_1^k, \ldots, \sigma_n^k)$ satisfy $\lambda_i^k \sigma_i^k = \nu_k$, and the eigenvalues are ordered such that

$$\lambda_1^k \ge \cdots \ge \lambda_n^k > 0, \qquad 0 < \sigma_1^k \le \cdots \le \sigma_n^k.$$

Let P^* be a limit point of the set $\{P^k\}$. We refine the sequence if necessary so that $\{P^k\}$ converges to P^*. Then P^* is an orthogonal matrix that simultaneously diagonalizes X^* and S^* with

$$X^* = P^* \Lambda^* (P^*)^T, \quad S^* = P^* \Sigma^* (P^*)^T, \tag{13}$$

where $\Lambda^* = \mathrm{diag}(\lambda_1^*, \ldots, \lambda_n^*)$, $\Sigma^* = \mathrm{diag}(\sigma_1^*, \ldots, \sigma_n^*)$ satisfy $\lambda_i^* \sigma_i^* = 0$, and

$$\lambda_1^* \ge \ldots \lambda_r^* > \lambda_{r+1}^* = \cdots = \lambda_n^* = 0, \;\; 0 = \sigma_1^* = \cdots = \sigma_{n-s}^* < \sigma_{n-s+1}^* \le \ldots \sigma_n^*.$$

Here r and s are the ranks of X^* and S^*, respectively.

We are assuming that (X^*, y^*, S^*) satisfies the strict complementarity condition, i.e., $r + s = n$. To analyze the spectrum of $\mathcal{B}_\rho^k$, we will identify the space $\mathcal{S}^n$ with the space $\mathcal{S}^r \times \mathbf{R}^{r\times s} \times \mathcal{S}^s$ as follows. For an element $X \in \mathcal{S}^n$, consider the partition $X = [X_1,\ X_2;\ X_2^T,\ X_3]$, where $X_1 \in \mathcal{S}^r$, $X_2 \in \mathbf{R}^{r\times s}$, and $X_3 \in \mathcal{S}^s$. Then X is identified with the element $[X_1;\ X_2;\ X_3]$ in $\mathcal{S}^r \times \mathbf{R}^{r\times s} \times \mathcal{S}^s$. The notation $[X_1;\ X_2;\ X_3]$ means that the objects X_1, X_2, X_3 are placed in a column format. The space $\mathcal{S}^r \times \mathbf{R}^{r\times s} \times \mathcal{S}^s$ is endowed with the inner product $[X_1;\ X_2;\ X_3] \bullet [Y_1;\ Y_2;\ Y_3] = X_1 \bullet Y_1 + 2X_2 \bullet Y_2 + X_3 \bullet Y_3$ so that the identification of $\mathcal{S}^n$ with $\mathcal{S}^r \times \mathbf{R}^{r\times s} \times \mathcal{S}^s$ is an isometry. Note the factor of 2 in front of the inner product $X_2 \bullet Y_2$. Thus the space $\mathbf{R}^{r\times s}$ is $\boldsymbol{I\!R}^{r\times s}$ but with the inner product for $X_2, Y_2 \in \mathbf{R}^{r\times s}$ given by $2X_2 \bullet Y_2$.

Let P_1^* and P_2^* be the submatrices denoting the first r and the last $n-r$ columns of P^*, respectively. Let $\mathcal{P}^* = P^* \circledast P^*$. Based on the identification

of $\mathcal{S}^n$ with $\mathcal{S}^r \times \mathbf{R}^{r\times s} \times \mathcal{S}^s$, $\mathcal{P}^*$ can be partitioned as follows. For $X = [X_1,\, X_2;\, X_2^T,\, X_3]$, we have

$$\mathcal{P}^*(X) = \mathcal{P}_1^*(X_1) + \mathcal{P}_2^*(X_2) + \mathcal{P}_3^*(X_3), \tag{14}$$

where $\mathcal{P}_1^* = P_1^* \circledast P_1^* : \mathcal{S}^r \to \mathcal{S}^n$, $\mathcal{P}_2^* = 2P_1^* \circledast P_2^* : \mathbf{R}^{r\times s} \to \mathcal{S}^n$, and $\mathcal{P}_3^* = P_2^* \circledast P_2^* : \mathcal{S}^s \to \mathcal{S}^n$. We may write $\mathcal{P}^* = [\mathcal{P}_1^*,\, \mathcal{P}_2^*,\, \mathcal{P}_3^*]$ so that $\mathcal{P}^*(X) = [\mathcal{P}_1^*,\, \mathcal{P}_2^*,\, \mathcal{P}_3^*][X_1; X_2; X_3]$.

Similarly, we let $\widetilde{\mathcal{A}}^* := \mathcal{A}\mathcal{P}^* = [\widetilde{\mathcal{A}}_1^*,\, \widetilde{\mathcal{A}}_2^*,\, \widetilde{\mathcal{A}}_3^*]$ with $\widetilde{\mathcal{A}}_j^* = \mathcal{A}\mathcal{P}_j^*$ for $j = 1, 2, 3$, and let $\widetilde{\mathcal{Q}}_\rho^* = (\mathcal{P}^*)^T \mathcal{Q}_\rho \mathcal{P}^*$ with $(\widetilde{\mathcal{Q}}_\rho^*)_{ij} = (\mathcal{P}_i^*)^T \mathcal{Q}_\rho \mathcal{P}_j^*$ for $i, j = 1, 2, 3$. We define $\widetilde{\mathcal{Q}}^*$ and $(\widetilde{\mathcal{Q}}^*)_{ij}$ similarly.

Based on the partitions of $\mathcal{P}^*$ and $\widetilde{\mathcal{A}}^*$, we can define the degeneracy of the optimal solution (X^*, y^*, S^*) as follows.

Definition 1 Suppose the optimal solution (X^*, y^*, S^*) satisfies the strict complementarity condition.
(a) The solution X^* is said to be primal nondegenerate [2] if the linear map $[\widetilde{\mathcal{A}}_1^*,\, \widetilde{\mathcal{A}}_2^*] : \mathcal{S}^r \times \mathbf{R}^{r\times s} \to I\!R^m$ defined by $[\widetilde{\mathcal{A}}_1^*,\, \widetilde{\mathcal{A}}_2^*][U; V] = \mathcal{A}(P_1^* U (P_1^*)^T) + \mathcal{A}(P_1^* V (P_2^*)^T + P_2^* V^T (P_1^*)^T)$ is surjective (this is an equivalent definition: see Theorem 6 in [2]).
(b) The solution S^* is said to be dual nondegenerate [2] if the linear map $\widetilde{\mathcal{A}}_1^* : \mathcal{S}^r \to I\!R^m$ defined by $\widetilde{\mathcal{A}}_1^*(U) = \mathcal{A}(P_1^* U (P_1^*)^T)$ is injective (this is again an equivalent definition: see Theorem 9 in [2]).

The importance of nondegeneracy of the optimal solution will become clear when we analyze the asymptotic spectra of $\{\mathcal{B}_\rho^k\}$ and those of the preconditioned matrices later.

Note that for the QSDP arising from (3), the optimal solution (X^*, y^*, S^*) is always primal nondegenerate; see [29].

On the central path $\{(X^\nu, y^\nu, S^\nu) : \nu > 0\}$, and under Assumptions A2 and A3, the eigenvalue decomposition of $(W^k)^{-1}$, where W^k is the NT scaling matrix corresponding to (X^k, y^k, S^k), must have the following form:

$$(W^k)^{-1} \;=\; P^k D^k (P^k)^T \;=\; P_1^k D_1^k (P_1^k)^T \;+\; P_2^k D_2^k (P_2^k)^T, \tag{15}$$

where $D_1^k = \mathrm{diag}(d_1^k) \in I\!R^{r\times r}$, $P_1^k \in I\!R^{n\times r}$ correspond to the small eigenvalues of order $\Theta(\sqrt{\nu_k})$, and $D_2^k = \mathrm{diag}(d_2^k) \in I\!R^{s\times s}$, $P_2^k \in I\!R^{n\times s}$ correspond to the large eigenvalues of the order $\Theta(1/\sqrt{\nu_k})$. Recall that the notation $\gamma = \Theta(\sqrt{\nu_k})$ means that there are constants $c_1, c_2 > 0$ such that $c_1\sqrt{\nu_k} \le \gamma \le c_2\sqrt{\nu_k}$ for all k.

It is easily shown that the following eigenvalue decomposition holds [27]:

$$(W^k)^{-1} \circledast (W^k)^{-1} = (P^k \circledast P^k)(D^k \circledast D^k)(P^k \circledast P^k)^T = \mathcal{P}^k \mathcal{D}^k (\mathcal{P}^k)^T, \tag{16}$$

where $\mathcal{P}^k = P^k \circledast P^k$ and $\mathcal{D}^k = D^k \circledast D^k$. Let $\mathcal{D}_1^k = D_1^k \circledast D_1^k$, $\mathcal{D}_2^k = D_2^k \otimes D_1^k$ and $\mathcal{D}_3^k = D_2^k \circledast D_2^k$. Then the diagonal entries of the matrix representations of $\mathcal{D}_1^k$, $\mathcal{D}_2^k$, and $\mathcal{D}_3^k$ consist of $\bar{r}$, rs, and $\bar{s}$ eigenvalues of $(W^k)^{-1} \circledast (W^k)^{-1}$ of the orders $\Theta(\nu_k)$, $\Theta(1)$, and $\Theta(1/\nu_k)$, respectively. We assume that there are constants $\underline{\tau}, \overline{\tau} > 0$ such that $\underline{\tau} I \preceq \mathcal{D}_2^k \preceq \overline{\tau} I$ for all k. Without loss

of generality, we may choose $\overline{\tau}$ large enough and $\underline{\tau}$ small enough so that $0 \preceq \mathcal{D}_1^k \preceq \overline{\tau} I$ and $\mathcal{D}_3^k \succeq \underline{\tau} I$ for all k. Note that we have $\mathcal{D} = \mathrm{diag}(\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3)$.

Consider the partition $\mathcal{P}^k = [\mathcal{P}_1^k,\, \mathcal{P}_2^k,\, \mathcal{P}_3^k]$ with $\mathcal{P}_1^k = P_1^k \circledast P_1^k$, $\mathcal{P}_2^k = 2P_1^k \circledast P_2^k$ and $\mathcal{P}_3^k = P_2^k \circledast P_2^k$ corresponding to $\mathcal{D}_1^k$, $\mathcal{D}_2^k$, and $\mathcal{D}_3^k$, respectively. Let $\widetilde{\mathcal{A}}^k = \mathcal{A}\mathcal{P}^k$, with the partition $\widetilde{\mathcal{A}}^k = [\widetilde{\mathcal{A}}_1^k\,,\, \widetilde{\mathcal{A}}_2^k\,,\, \widetilde{\mathcal{A}}_3^k] = [\mathcal{A}\mathcal{P}_1^k\,,\, \mathcal{A}\mathcal{P}_2^k\,,\, \mathcal{A}\mathcal{P}_3^k]$ conforming to that of $\mathrm{diag}(\mathcal{D}_1^k, \mathcal{D}_2^k, \mathcal{D}_3^k)$. Suppose $\widetilde{\mathcal{Q}}^k = (\mathcal{P}^k)^T \mathcal{Q} \mathcal{P}^k$. Then $\widetilde{\mathcal{Q}}_\rho^k := (\mathcal{P}^k)^T \mathcal{Q}_\rho \mathcal{P}^k = \widetilde{\mathcal{Q}}^k + \rho(\widetilde{\mathcal{A}}^k)^T \widetilde{\mathcal{A}}^k$. Let $(\widetilde{\mathcal{Q}}_\rho^k)_{ij} = (\mathcal{P}_i^k)^T \mathcal{Q}_\rho \mathcal{P}_j^k$ for $i, j = 1, 2, 3$. We define $\widetilde{\mathcal{Q}}_{ij}^k$ similarly. Note that since $\mathcal{P}^k \to \mathcal{P}^*$ as $k \uparrow \infty$, we have $\widetilde{\mathcal{A}}^k \to \widetilde{\mathcal{A}}^*$ and $\widetilde{\mathcal{Q}}_\rho^k \to \widetilde{\mathcal{Q}}_\rho^*$ as $k \uparrow \infty$.

Based on the eigenvalue decomposition (16), it is readily shown that the matrix $\mathcal{B}_\rho^k$ in (9) has the following decomposition:

$$\mathcal{B}_\rho^k = \begin{bmatrix} \mathcal{P}^k & 0 \\ 0 & I_m \end{bmatrix} \widetilde{\mathcal{B}}_\rho^k \begin{bmatrix} (\mathcal{P}^k)^T & 0 \\ 0 & I_m \end{bmatrix}, \text{ where } \widetilde{\mathcal{B}}_\rho^k = \begin{bmatrix} -D^k \circledast D^k - \widetilde{\mathcal{Q}}_\rho^k & (\widetilde{\mathcal{A}}^k)^T \\ \widetilde{\mathcal{A}}^k & 0 \end{bmatrix}.$$

For later usage, note that we have the following partition:

$$D^k \circledast D^k + \widetilde{\mathcal{Q}}_\rho^k = \begin{bmatrix} \mathcal{D}_1^k + (\widetilde{\mathcal{Q}}_\rho^k)_{11} & (\widetilde{\mathcal{Q}}_\rho^k)_{12} & (\widetilde{\mathcal{Q}}_\rho^k)_{13} \\ (\widetilde{\mathcal{Q}}_\rho^k)_{21} & \mathcal{D}_2^k + (\widetilde{\mathcal{Q}}_\rho^k)_{22} & (\widetilde{\mathcal{Q}}_\rho^k)_{23} \\ (\widetilde{\mathcal{Q}}_\rho^k)_{31} & (\widetilde{\mathcal{Q}}_\rho^k)_{32} & \mathcal{D}_3^k + (\widetilde{\mathcal{Q}}_\rho^k)_{33} \end{bmatrix}. \quad (17)$$

We note that it is possible to have the case where $r = n$ and $s = 0$. In that case, $\mathcal{D}_2^k$ is a null matrix for k sufficiently large, and correspondingly $\mathcal{D}_2^k, \mathcal{D}_3^k$ and $\mathcal{P}_2^k, \mathcal{P}_3^k$ are null maps. Similarly, $(\widetilde{\mathcal{Q}}_\rho^k)_{ij}$, $i, j = 1, 2, 3$, are null maps except for $(\widetilde{\mathcal{Q}}_\rho^k)_{11}$. As we shall see from the following lemma, the norm $\|\mathcal{B}_\rho^k\|$ behaves differently for $r = n$ and $r < n$.

Lemma 3 *Let r and s be the ranks of X^* and S^*, respectively.*
(a) Suppose $r = n$ and $s = 0$. Then $\limsup_{k\to\infty} \|\mathcal{B}_\rho^k\| \le 2\max\{\|\mathcal{A}\|, \|\mathcal{Q}_\rho\|\}$.
(b) Suppose $r < n$ and $s = n - r > 0$. Then $\|\mathcal{B}_\rho^k\| \ge \max\{\|\mathcal{Q}_\rho\|, \|\mathcal{D}_3^k\|\} = \Theta(1/\nu_k)$. Hence $\|\mathcal{B}_\rho^k\| \to \infty$ as $k \uparrow \infty$.

Proof. (a) In this case, $\mathcal{D}^k = \Theta(\nu_k)$, and the required result follows from Lemma 2. (b) By Lemma 2 and the fact that $\mathcal{D}_3^k = \Theta(1/\nu_k)$, the required result follows. □

We state the following lemma before analyzing the conditioning of $\mathcal{B}_\rho^k$.

Lemma 4 *Suppose that Assumptions A2 and A3 hold for the optimal solution (X^*, y^*, S^*), and $(\widetilde{\mathcal{Q}}_\rho^*)_{11} \succ 0$.*
(a) Then we have

$$\Upsilon_\rho^* := \begin{bmatrix} (\widetilde{\mathcal{Q}}_\rho^*)_{11} & (\widetilde{\mathcal{Q}}_\rho^*)_{12} \\ (\widetilde{\mathcal{Q}}_\rho^*)_{21} & \underline{\tau} I + (\widetilde{\mathcal{Q}}_\rho^*)_{22} \end{bmatrix} \succ 0. \quad (18)$$

(b) Let

$$M_\rho^* = [\widetilde{\mathcal{A}}_1^*,\, \widetilde{\mathcal{A}}_2^*](\Upsilon_\rho^*)^{-1}[\widetilde{\mathcal{A}}_1^*,\, \widetilde{\mathcal{A}}_2^*]^T. \quad (19)$$

Suppose further that (X^*, y^*, S^*) *is primal nondegenerate. Then* $M_\rho^* \succ 0$.

Proof. (a) Since the sub-matrix of maps, $[(\widetilde{\mathcal{Q}}_\rho^*)_{11},\ (\widetilde{\mathcal{Q}}_\rho^*)_{12};\ (\widetilde{\mathcal{Q}}_\rho^*)_{21},\ (\widetilde{\mathcal{Q}}_\rho^*)_{22}]$, of $\widetilde{\mathcal{Q}}_\rho$ is positive semidefinite, we have $\underline{\tau}I + (\widetilde{\mathcal{Q}}_\rho^*)_{22} - (\widetilde{\mathcal{Q}}_\rho^*)_{21}(\widetilde{\mathcal{Q}}_\rho^*)_{11}^{-1}(\widetilde{\mathcal{Q}}_\rho^*)_{12} \succeq \underline{\tau}I$. By Theorem 7.7.6 in [15], Υ_ρ^* in (18) is positive definite.

(b) Under the assumption that X^* is primal nondegenerate, the linear map $[\widetilde{\mathcal{A}}_1^*,\ \widetilde{\mathcal{A}}_2^*]$ is surjective. From here, it is easy to show that $M_\rho^* \succ 0$. □

Remark 1 (a) If $\widetilde{\mathcal{Q}}_{11}^*$ is positive definite on $\mathcal{N}(\sqrt{\rho}\widetilde{\mathcal{A}}_1^*) = \mathcal{N}(\rho(\widetilde{\mathcal{A}}_1^*)^T\widetilde{\mathcal{A}}_1^*)$, then $(\widetilde{\mathcal{Q}}_\rho^*)_{11} \succ 0$. This result follows from the fact that if $U, V \in \mathcal{S}_+^p$ are such that U is positive definite on $\mathcal{N}(V)$, then $U+V \succ 0$. Since it is easier for $(\widetilde{\mathcal{Q}}_\rho^*)_{11}$ to be positive definite compared to $\widetilde{\mathcal{Q}}_{11}^*$, this explains why we used (7) instead of the first condition in (6).
(b) If either $\rho > 0$ and (X^*, y^*, S^*) is dual nondegenerate, or $\widetilde{\mathcal{Q}}_{11}^*$ is positive definite, then $(\widetilde{\mathcal{Q}}_\rho^*)_{11} \succ 0$. Note that if $\mathcal{Q} \succ 0$, then $\widetilde{\mathcal{Q}}_{11}^* \succ 0$ since $\mathrm{eig}(\widetilde{\mathcal{Q}}_{11}^*) \subset \mathrm{eig}(\widetilde{\mathcal{Q}}^*) = \mathrm{eig}(\mathcal{Q})$.

Let

$$\Upsilon_\rho^k = \begin{bmatrix} \mathcal{D}_1^k + (\widetilde{\mathcal{Q}}_\rho^k)_{11} & (\widetilde{\mathcal{Q}}_\rho^k)_{12} \\ (\widetilde{\mathcal{Q}}_\rho^k)_{21} & \mathcal{D}_2^k + (\widetilde{\mathcal{Q}}_\rho^k)_{22} \end{bmatrix}, \quad \widehat{\Upsilon}_\rho^k = \begin{bmatrix} (\widetilde{\mathcal{Q}}_\rho^k)_{11} & (\widetilde{\mathcal{Q}}_\rho^k)_{12} \\ (\widetilde{\mathcal{Q}}_\rho^k)_{21} & \underline{\tau}I + (\widetilde{\mathcal{Q}}_\rho^k)_{22} \end{bmatrix}. \tag{20}$$

Note that Υ_ρ^k is positive definite, and

$$\widehat{\Upsilon}_\rho^k \ \preceq\ \Upsilon_\rho^k \ \preceq\ \overline{\tau}I + \widehat{\Upsilon}_\rho^k. \tag{21}$$

Recall that M_ρ^k denotes the Schur complement matrix in (10) corresponding to (X^k, y^k, S^k). We have the following three theorems concerning $\{\|M_\rho^k\|\}$, $\{\|(M_\rho^k)^{-1}\|\}$, and $\{\|(\mathcal{B}_\rho^k)^{-1}\|\}$.

Theorem 1 *Suppose that Assumptions A2 and A3 hold for the optimal solution* (X^*, y^*, S^*), *and* $(\widetilde{\mathcal{Q}}_\rho^*)_{11} \succ 0$. *Then the following results hold.*
(a) There exists $c > 0$ *such that* $\limsup_{k\to\infty} \|M_\rho^k\| \le c\|M_\rho^*\|$.
(b) Suppose further that (X^*, y^*, S^*) *is primal nondegenerate. Then there exists* $c > 0$ *such that* $\limsup_{k\to\infty} \|(M_\rho^k)^{-1}\| \le c\|(M_\rho^*)^{-1}\|$.

Proof. Let

$$J_\rho^k \ =\ \Upsilon_\rho^k - \big[\,(\widetilde{\mathcal{Q}}_\rho^k)_{13}\ ;\ (\widetilde{\mathcal{Q}}_\rho^k)_{23}\,\big]\big(\mathcal{D}_3^k + (\widetilde{\mathcal{Q}}_\rho^k)_{33}\big)^{-1}\big[\,(\widetilde{\mathcal{Q}}_\rho^k)_{31}\ (\widetilde{\mathcal{Q}}_\rho^k)_{32}\,\big]. \tag{22}$$

Note that the second term on the right-hand side converges to 0 as $k \uparrow \infty$, and Υ_ρ^k satisfies (21). Since $\widehat{\Upsilon}_\rho^k$ in (21) converges to Υ_ρ^* in (18) as $k \uparrow \infty$, and $\Upsilon_\rho^* \succ 0$ by Lemma 4(a), thus there exist $c_1, c_2 > 0$ such that $c_1^{-1}\Upsilon_\rho^* \preceq J_\rho^k \preceq c_2^{-1}\Upsilon_\rho^*$ for sufficiently large k. This implies that $c_1(\Upsilon_\rho^*)^{-1} \succeq (J_\rho^k)^{-1} \succeq c_2(\Upsilon_\rho^*)^{-1}$ for all k sufficiently large.
(a) Observe that $M_\rho^k = \widetilde{\mathcal{A}}^k(D^k \circledast D^k + \widetilde{\mathcal{Q}}_\rho^k)^{-1}(\widetilde{\mathcal{A}}^k)^T$. By applying the formula in (12) to the matrix in (17), we have

$$(D^k \circledast D^k + \widetilde{\mathcal{Q}}_\rho^k)^{-1} \ =\ \mathrm{diag}\big((J_\rho^k)^{-1}, 0\big) + O(\nu_k\|(J_\rho^k)^{-1}\|).$$

Thus, for k sufficiently large,

$$M_\rho^k = [\widetilde{\mathcal{A}}_1^k,\, \widetilde{\mathcal{A}}_2^k](J_\rho^k)^{-1}[\widetilde{\mathcal{A}}_1^k,\, \widetilde{\mathcal{A}}_2^k]^T + O(\nu_k\|(J_\rho^k)^{-1}\|\,\|\mathcal{A}\|^2) \tag{23}$$

$$\preceq c_1[\widetilde{\mathcal{A}}_1^k,\, \widetilde{\mathcal{A}}_2^k](\Upsilon_\rho^*)^{-1}[\widetilde{\mathcal{A}}_1^k,\, \widetilde{\mathcal{A}}_2^k]^T + O(\nu_k\|(\Upsilon_\rho^*)^{-1}\|\,\|\mathcal{A}\|^2).$$

Since the first term on the right-hand side converges to M_ρ^* in (19), the required result follows.
(b) From (23), we have that for k sufficiently large,

$$M_\rho^k \succeq c_2[\widetilde{\mathcal{A}}_1^k,\, \widetilde{\mathcal{A}}_2^k](\Upsilon_\rho^*)^{-1}[\widetilde{\mathcal{A}}_1^k,\, \widetilde{\mathcal{A}}_2^k]^T + O(\nu_k\|(\Upsilon_\rho^*)^{-1}\|\,\|\mathcal{A}\|^2).$$

Since the first term on the right-hand side converges to M_ρ^*, and $M_\rho^* \succ 0$ by Lemma 4(b), thus $\liminf_{k\to\infty} \lambda_{\min}(M_\rho^k) \geq c_2\lambda_{\min}(M_\rho^*) > 0$. From here, the required result follows. □

Theorem 2 *Under the assumptions stated in Theorem 1(b), there exists $c > 0$ such that*

$$\limsup_{k\to\infty} \|(\mathcal{B}_\rho^k)^{-1}\| \;\leq\; c\max\{\|(\Upsilon_\rho^*)^{-1}\|\,,\, \|(M_\rho^*)^{-1}\|\}. \tag{24}$$

Proof. By Lemma 2, we get $\|(\mathcal{B}_\rho^k)^{-1}\| \leq 2\max\{\|(\mathcal{K}_\rho^k)^{-1}\|,\, \|(M_\rho^k)^{-1}\|\}$. Now

$$(\mathcal{P}^k)^T(\mathcal{K}_\rho^k)^{-1}\mathcal{P}^k = (D^k \circledast D^k + \widetilde{\mathcal{Q}}_\rho^k)^{-1} = \mathrm{diag}\big((J_\rho^k)^{-1}, 0\big) + O(\nu_k\|(J_\rho^k)^{-1}\|),$$

where J_ρ^k is defined in (22), and there exist $c_1 > 0$ such that for k sufficiently large, $c_1(\Upsilon_\rho^*)^{-1} \succeq (J_\rho^k)^{-1}$. Thus $\|(\mathcal{K}_\rho^k)^{-1}\| = \|(J_\rho^k)^{-1}\|(1 + O(\nu_k)) \leq c_1\|(\Upsilon_\rho^*)^{-1}\|(1 + O(\nu_k))$, implying that $\limsup_{k\to\infty} \|(\mathcal{K}_\rho^k)^{-1}\| \leq c_1\|(\Upsilon_\rho^*)^{-1}\|$. By Theorem 1(b), we have $\limsup_{k\to\infty} \|(M_\rho^k)^{-1}\| \leq c_2\|(M_\rho^*)^{-1}\|$ for some constant $c_2 > 0$. From here, the required result in (24) follows. □

Theorem 3 *Suppose that Assumptions A2 and A3 hold for the optimal solution (X^*, y^*, S^*) and $\mathcal{Q}_\rho \succ 0$. If $S^* = 0$, then*

$$\limsup_{k\to\infty} \|M_\rho^k\| \leq \|\mathcal{A}\mathcal{Q}_\rho^{-1}\mathcal{A}^T\| \tag{25}$$

$$\limsup_{k\to\infty} \|(M_\rho^k)^{-1}\| \leq \|(\mathcal{A}\mathcal{Q}_\rho^{-1}\mathcal{A}^T)^{-1}\| \tag{26}$$

$$\limsup_{k\to\infty} \|(\mathcal{B}_\rho^k)^{-1}\| \leq 2\max\{\|\mathcal{Q}_\rho^{-1}\|\,,\, \|(\mathcal{A}\mathcal{Q}_\rho^{-1}\mathcal{A}^T)^{-1}\|\}. \tag{27}$$

Proof. Under the assumption of strict complementarity, $S^* = 0$ implies that X^* has full rank. In this case, $\mathcal{D}^k = \Theta(\nu_k)$. Now, $M_\rho^k = \widetilde{\mathcal{A}}^k(\mathcal{D}^k + \widetilde{\mathcal{Q}}_\rho^k)^{-1}(\widetilde{\mathcal{A}}^k)^T = \mathcal{A}\mathcal{Q}_\rho^{-1}\mathcal{A}^T + O(\nu_k)$. Thus the result in (25) follows. Since $\mathcal{A}\mathcal{Q}_\rho^{-1}\mathcal{A}^T$ is nonsingular, the inequality in (26) follows readily. The result in (27) can be proven similarly. □

Example 1. To illustrate the asymptotic result in Theorem 2 and later results on the spectra of preconditioned matrices, we consider an example of (1) arising from the NCM problem (3) with $n = 30$. The matrix K in (3) is

generated in a similar way as Higham did in [14] from the MATLAB segment: `x=10.^[-4:4/(n-1):0]; B=gallery('randcorr',n*x/sum(x))`. We take K = `B+1e-4*E`, where `E` is a random symmetric matrix with $\|E\|_F = 1$. The linear operator $\mathcal{Q}$ associated with (3) is chosen to be the positive definite operator $\mathcal{Q}(X) = U \circ X$, where U is randomly generated as follows: `tmp=rand(n); U=(tmp+tmp')/2`.

Note that for all the numerical examples in this paper, we take $\rho = 0$.

The spectra of $\mathcal{B}_\rho^k$ and M_ρ^k corresponding to the iterate (X^k, y^k, S^k) with complementarity gap $X^k \bullet S^k/n = 1.9 \times 10^{-10}$ are contained in the intervals $[-1.62, -0.01] \cup [0.62, 0.98]$ and $[1.00, 20.25]$, respectively. For this example, X^* has full rank and $S^* = 0$. It is clear that $\|\mathcal{B}_\rho^k\|$, $\|(\mathcal{B}_\rho^k)^{-1}\|$, $\|M_\rho^k\|$, and $\|(M_\rho^k)^{-1}\|$ remain bounded as $k \uparrow \infty$. This observation is consistent with the asymptotic results in Lemma 3 and Theorem 3.

Examples 2 and 3. The last example is not very interesting since $\mathcal{B}_\rho^k$ remains well conditioned as $k \uparrow \infty$. We now consider Examples 2 and 3, which are similar to Example 1 but with K taken to be `B + E` and `B + 1e2 * E`, respectively. Observe that we make the perturbation to the random correlation matrix `B` larger than `1e-4*E` considered in [14] so as to generate QSDPs with augmented matrices $\mathcal{B}_\rho^k$ that become increasing ill-conditioned as k increases.

Note that the optimal solution of a typical random instance of Example 2 is primal nondegenerate but dual degenerate, whereas that of Example 3 is both primal and dual nondegenerate.

Figure 1 shows the spectra of $\mathcal{B}_\rho^k$ corresponding to the final iterates for 10 random instances of Examples 2 and 3. From the spectra shown in plots, it is clear that $\|\mathcal{B}_\rho^k\|$ will tend to ∞ as $k \uparrow \infty$. On the other hand, $\|(\mathcal{B}_\rho^k)^{-1}\|$ remains bounded as $k \uparrow \infty$. This observation is consistent with Theorem 2. Note that for all the problem instances, the complementarity gaps of the final iterates are less than 10^{-8} and they are primal and dual feasible.

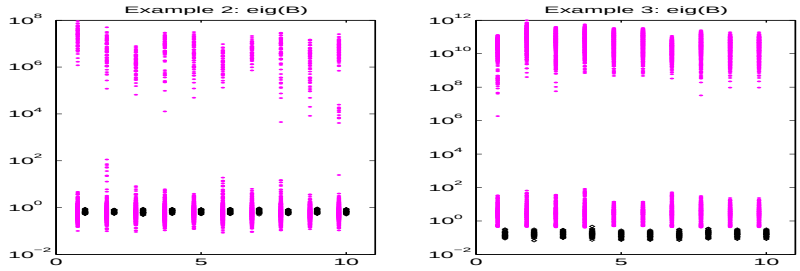

Fig. 1 The spectra of $\mathcal{B}_\rho^k$ corresponding to the final iterates for 10 random instances of Examples 2 and 3. The solid dots correspond to negative eigenvalues, and the diamonds correspond to positive eigenvalues. The x-coordinates are the indices of the problem instances, while the y-coordinates are the magnitudes of the eigenvalues in logarithmic scale.

4 Preconditioners for the augmented matrix $\mathcal{B}_\rho^k$

The convergence rate of a Krylov iterative method such as MINRES applied to the matrix $\mathcal{B}_\rho^k$ depends in a complicated manner on the eigenvalue distribution of the matrix; see [31]. Even though there is no accurate estimate of the convergence rate based on the condition number $\kappa(\mathcal{B}_\rho^k)$ like the case of a symmetric positive definite matrix, the condition number is generally still a reasonable measure of the convergence rate – the smaller the condition number is, the faster the convergence. Our purpose now is to design preconditioners for $\mathcal{B}_\rho^k$ so that the condition numbers of the preconditioned matrices are bounded independent of ν_k.

We have shown in the previous section that $\{\kappa(\mathcal{B}_\rho^k)\}_{k=1}^\infty$ is unbounded except when the optimal solution is primal nondegenerate, and $S^* = 0$ and $\mathrm{rank}(X^*) = n$. But note that even if $\{\kappa(\mathcal{B}_\rho^k)\}_{k=1}^\infty$ is bounded, $\kappa(\mathcal{B}_\rho^k)$ can still be large. Thus preconditioning for $\mathcal{B}_\rho^k$ is generally needed.

Let $\mathcal{H}_1^k : \mathcal{S}^r \to \mathcal{S}^r$, $\mathcal{H}_2^k : \mathbf{R}^{r\times s} \to \mathbf{R}^{r\times s}$, and $\mathcal{H}_3^k : \mathcal{S}^s \to \mathcal{S}^s$ be given self-adjoint linear operators such that $\mathcal{H}_j^k$ is a positive definite approximation of $\mathcal{D}_j^k + (\widetilde{\mathcal{Q}}_\rho^k)_{jj}$, $j = 1, 2, 3$. We will discuss specific choices of $\mathcal{H}_1^k$, $\mathcal{H}_2^k$, $\mathcal{H}_3^k$ later when needed. For later usage, we define the linear operator $\mathcal{E}^k : \mathcal{S}^n \to \mathcal{S}^n$:

$$\mathcal{E}^k(U) \;=\; \begin{bmatrix} (\mathcal{H}_1^k)^{-1}(U_1) & (\mathcal{H}_2^k)^{-1}(U_2) \\ ((\mathcal{H}_2^k)^{-1}(U_2))^T & (\mathcal{H}_3^k)^{-1}(U_3) \end{bmatrix}.$$

4.1 Part I

In this subsection, we assume that $\mathcal{H}_1^k$, $\mathcal{H}_2^k$, $\mathcal{H}_3^k$ satisfy the conditions below for all k:

$$\underline{\sigma}_1 I \;\preceq\; \mathcal{H}_1^k \;\preceq\; \overline{\sigma}_1 I,\;\; \underline{\sigma}_2 I \;\preceq\; \mathcal{H}_2^k \;\preceq\; \overline{\sigma}_2 I,\;\; \mathcal{D}_3^k \;\preceq\; \mathcal{H}_3^k \;\preceq\; \mathcal{D}_3^k + \overline{\sigma}_3 I, \quad (28)$$

for some positive constants $\underline{\sigma}_1, \overline{\sigma}_1$, $\underline{\sigma}_2, \overline{\sigma}_2$, and $\overline{\sigma}_3$. Let

$$N_\rho^k = \mathrm{diag}(\mathcal{H}_1^k, \mathcal{H}_2^k)^{-1/2}\, \Upsilon_\rho^k\, \mathrm{diag}(\mathcal{H}_1^k, \mathcal{H}_2^k)^{-1/2} \quad (29)$$

$$\underline{\theta}^k = \min\{1, \lambda_{\min}(N_\rho^k)\}, \quad \overline{\theta}^k = \max\{1, \lambda_{\max}(N_\rho^k)\}, \quad (30)$$

where Υ_ρ^k is defined in (20). We have the following results for $\{\underline{\theta}^k\}$ and $\{\overline{\theta}^k\}$.

Lemma 5 *Suppose that Assumptions A2 and A3 hold for the optimal solution (X^*, y^*, S^*), and $(\widetilde{\mathcal{Q}}_\rho^*)_{11} \succ 0$. Then*

$$\liminf_{k\to\infty} \underline{\theta}_k \geq \min\left\{1, \min(\overline{\sigma}_1^{-1}, \overline{\sigma}_2^{-1})\lambda_{\min}(\Upsilon_\rho^*)\right\} \;>\; 0, \quad (31)$$

$$\limsup_{k\to\infty} \overline{\theta}_k \leq \max\left\{1, \max(\underline{\sigma}_1^{-1}, \underline{\sigma}_2^{-1})\lambda_{\max}\left(\overline{\tau} I + \Upsilon_\rho^*\right)\right\} \;<\; \infty. \quad (32)$$

Consequently, there exist constants $c_1, c_2 > 0$ such that for k sufficiently large, $c_1 < \underline{\theta}^k \leq \overline{\theta}^k < c_2$.

Proof. By noting that $\underline{\sigma}_j I \preceq \mathcal{H}_j^k \preceq \overline{\sigma}_j I$, $j = 1, 2$, and $\underline{\tau} I \preceq \mathcal{D}_2^k \preceq \overline{\tau} I$ for all k, we have $N_\rho^k \succeq \min(\overline{\sigma}_1^{-1}, \overline{\sigma}_2^{-1})\widehat{\Upsilon}_\rho^k$, where $\widehat{\Upsilon}_\rho^k$ is defined in (20), and it converges to the matrix Υ_ρ^* in (18) as $k \uparrow \infty$. Since $\Upsilon_\rho^* \succ 0$ by Lemma 4(a), (31) follows. On the other hand, we have $N_\rho^k \preceq \max(\underline{\sigma}_1^{-1}, \underline{\sigma}_2^{-1})(\overline{\tau} I + \widehat{\Upsilon}_\rho^k)$. From here, the inequality in (32) follows easily. □

The first preconditioner we consider is the following block diagonal matrix whose partition conforms to that of $\mathcal{B}_\rho^k$:

$$\Phi_{\rho,\pm}^k \;=\; \begin{bmatrix} \pm\mathcal{P}^k \mathrm{diag}(\mathcal{H}_1^k, \mathcal{H}_2^k, \mathcal{H}_3^k)(\mathcal{P}^k)^T & 0 \\ 0 & \widehat{M}_\rho^k \end{bmatrix}, \tag{33}$$

where $\widehat{M}_\rho^k$ is a symmetric positive definite approximation of the matrix M_ρ^k in (10). An example would be $\widehat{M}_\rho^k = I_m$. We assume that $\widehat{M}_\rho^k$ satisfies the following condition: $\underline{\alpha} I \preceq \widehat{M}_\rho^k \preceq \overline{\alpha} I$ for all k for some constants $\underline{\alpha}, \overline{\alpha} > 0$. The motivation for assuming $\mathcal{H}_3^k \succeq \mathcal{D}_3^k$ in (28) for the (1,1) block of $\Phi_{\rho,\pm}^k$ is to obliterate the effect of large eigenvalues of $\mathcal{D}^k$ on the conditioning of $\mathcal{B}_\rho^k$. Note that for $\Phi_{\rho,\pm}^k$, the assumption that m is a moderate number in Assumption A1 is not needed. This is in contrast to the preconditioners we are considering in the next two parts. Given $[X; y] \in \mathcal{S}^n \times \mathbb{R}^m$, $(\Phi_{\rho,\pm}^k)^{-1}[X; y]$ can be evaluated efficiently through the steps given below:

Compute $\widehat{X} = (P^k)^T X P^k$

Compute $(\Phi_{\rho,\pm}^k)^{-1}[X; y] = [\pm P^k(\mathcal{E}^k(\widehat{X}))(P^k)^T;\, (\widehat{M}_\rho^k)^{-1} y]$.

The above computation requires four $n \times n$ matrix-matrix multiplications that cost $4n^3$ flops; see [22].

Block diagonal preconditioners for *sparse* 2×2 block symmetric indefinite systems have been studied extensively. We shall not give the literature review here but refer the reader to the recent survey paper [7]. As far as we know, this paper is the first attempt to apply a block diagonal preconditioner to a *dense* 2×2 block symmetric indefinite system.

It is well known that the convergence rate of a preconditioned Krylov iterative method such as PSQMR is determined primarily by the spectral distribution and condition number of the preconditioned matrix. Thus it is of great interest to estimate its spectrum. We will first establish an asymptotic result for the spectrum of $(\Phi_{\rho,+}^k)^{-1}\mathcal{B}_\rho^k$.

Theorem 4 *(a) Suppose that Assumptions A2 and A3 hold. Let $r = \mathrm{rank}(X^*)$ and $s = n - r$. Then the preconditioned matrix $(\Phi_{\rho,+}^k)^{-1}\mathcal{B}_\rho^k$ has $\bar{s}$ eigenvalues equal to $-1 + O(\sqrt{\nu_k}\max\{\|\mathcal{Q}_\rho\|, \|\mathcal{A}\|\})$, the remaining eigenvalues are $O(\sqrt{\nu_k}\max\{\|\mathcal{Q}_\rho\|, \|\mathcal{A}\|\})$ perturbations of those of the following matrix*

$$G_{\rho,+}^k \;:=\; \begin{bmatrix} -N_\rho^k & (L^k)^T \\ L^k & 0 \end{bmatrix},$$

where $L^k = (\widehat{M}_\rho^k)^{-1/2}[\widetilde{\mathcal{A}}_1^k,\, \widetilde{\mathcal{A}}_2^k]\mathrm{diag}(\mathcal{H}_1^k, \mathcal{H}_2^k)^{-1/2}$ and N_ρ^k is defined as in (29). Let $\widetilde{M}_\rho^k = [\widetilde{\mathcal{A}}_1^k,\, \widetilde{\mathcal{A}}_2^k](\Upsilon_\rho^k)^{-1}[\widetilde{\mathcal{A}}_1^k,\, \widetilde{\mathcal{A}}_2^k]^T$, and $\sigma_j^k = \big(1+4\lambda_j((\widehat{M}_\rho^k)^{-1}\widetilde{M}_\rho^k)\big)^{1/2}$,

$j = 1, \ldots, m$. *Let* $\sigma^k_{\max} = \max_j\{\sigma^k_j\}$ *and* $\sigma^k_{\min} = \min_j\{\sigma^k_j\}$. *The spectrum of* $G^k_{\rho,+}$ *is contained in the union of the following intervals:*

$$\Big[-\frac{\overline{\theta}^k}{2}(\sigma^k_{\max}+1), -\frac{\underline{\theta}^k}{2}(\sigma^k_{\min}+1)\Big], \Big[-\overline{\theta}^k, -\underline{\theta}^k\Big], \Big[\frac{\underline{\theta}^k}{2}(\sigma^k_{\min}-1), \frac{\overline{\theta}^k}{2}(\sigma^k_{\max}-1)\Big].$$

(b) Under the assumptions stated in Theorem 1(b), there exist constants c_1, c_2, c_3, c_4 *such that for* k *sufficiently large,* $0 < c_1 < \underline{\theta}^k \le \overline{\theta}^k < c_2$, *and* $1 < c_3 < \sigma^k_{\min} \le \sigma^k_{\max} < c_4$. *As a result,*

$$\mathrm{eig}(G^k_{\rho,+}) \subset [-c_2(c_4+1), -c_1(c_3+1)] \cup [-c_2, -c_1] \cup [c_1(c_3-1), c_2(c_4-1)].$$

Proof. For simplicity, we drop the superscript k in this proof. Note that the spectrum of $\Phi^{-1}_{\rho,+}\mathcal{B}_\rho$ is the same as that of $\Phi^{-1/2}_{\rho,+}\mathcal{B}_\rho\Phi^{-1/2}_{\rho,+}$.

(a) It is readily shown that up to a perturbation of $O(\sqrt{\nu}\max\{\|\mathcal{Q}_\rho\|, \|\mathcal{A}\|\})$, $\Phi^{-1/2}_{\rho,+}\mathcal{B}_\rho\Phi^{-1/2}_{\rho,+}$ is orthogonally similar to $\mathrm{diag}(G_{\rho,+}, -I_{\bar{s}})$. Thus the eigenvalues of $\Phi^{-1/2}_{\rho,+}\mathcal{B}_\rho\Phi^{-1/2}_{\rho,+}$ are $O(\sqrt{\nu}\max\{\|\mathcal{Q}_\rho\|, \|\mathcal{A}\|\})$ perturbations of those of $\mathrm{diag}(G_{\rho,+}, -I_{\bar{s}})$, whose eigenvalues are either -1 or those of $G_{\rho,+}$. Now, we can write $G_{\rho,+} = VUV^T$, where $V = \mathrm{diag}\big(\mathrm{diag}(\mathcal{H}_1, \mathcal{H}_2)^{-1/2}\Upsilon^{1/2}_\rho, I_m\big)$, and

$$U = \begin{bmatrix} -I_{\bar{r}+rs} & \Upsilon^{-1/2}_\rho[\widetilde{\mathcal{A}}_1, \widetilde{\mathcal{A}}_2]^T\widehat{M}^{-1/2}_\rho \\ \widehat{M}^{-1/2}_\rho[\widetilde{\mathcal{A}}_1, \widetilde{\mathcal{A}}_2]\Upsilon^{-1/2}_\rho & 0 \end{bmatrix}.$$

By Corollary 4.5.11 in [15], we have $\lambda_j(G_{\rho,+}) = \theta_j\,\lambda_j(U)$ for $j = 1, \ldots, \bar{r}+rs+m$, where θ_j satisfies the inequalities: $\underline{\theta} \le \theta_j \le \overline{\theta}$. It is easy to show that U has $\bar{r}+rs-m$ eigenvalues equal to -1, and the remaining $2m$ eigenvalues are given by $\big(-1 \pm \sigma_j\big)/2$, $j = 1, \ldots, m$. From here, we can show that $\mathrm{eig}(G_{\rho,+})$ is contained in the required intervals.

(b) The existence of $c_1, c_2 > 0$ is guaranteed by Lemma 5. The existence of positive constants c_3 and c_4 will follow if can prove the following results:

$$\liminf_{k\to\infty} \lambda_{\min}((\widehat{M}^k_\rho)^{-1}\widetilde{M}^k_\rho) \ge \underline{\alpha}\,\lambda_{\min}\left([\widetilde{\mathcal{A}}^*_1, \widetilde{\mathcal{A}}^*_2]\left(\overline{\tau}I + \Upsilon^*_\rho\right)^{-1}[\widetilde{\mathcal{A}}^*_1, \widetilde{\mathcal{A}}^*_2]^T\right) > 0 \quad (34)$$

$$\limsup_{k\to\infty} \lambda_{\max}((\widehat{M}^k_\rho)^{-1}\widetilde{M}^k_\rho) \le \overline{\alpha}\,\lambda_{\max}\left([\widetilde{\mathcal{A}}^*_1, \widetilde{\mathcal{A}}^*_2](\Upsilon^*_\rho)^{-1}[\widetilde{\mathcal{A}}^*_1, \widetilde{\mathcal{A}}^*_2]^T\right) < \infty. \quad (35)$$

Noting that $\underline{\alpha}I \preceq \widehat{M}^k_\rho \preceq \overline{\alpha}I$ and $\underline{\tau}I \preceq \mathcal{D}^k_2 \preceq \overline{\tau}I$ for all k, the first inequality in (34) is straightforward to prove. Under the assumptions of Theorem 1(b), $[\widetilde{\mathcal{A}}^*_1, \widetilde{\mathcal{A}}^*_2]$ is surjective and $\Upsilon^*_\rho \succ 0$, from here, the second inequality in (34) follows. We can prove (35) similarly. □

Next, we shall establish an asymptotic result for the spectrum of $(\Phi^k_{\rho,-})^{-1}\mathcal{B}^k_\rho$.

Theorem 5 *(a) Suppose that Assumptions A2 and A3 hold. Let* $r = rank(X^*)$ *and* $s = n - r$. *Then the preconditioned matrix* $(\Phi^k_{\rho,-})^{-1}\mathcal{B}^k_\rho$ *has* $\bar{s}$ *eigenvalues equal to* $1 + O(\sqrt{\nu_k}\max\{\|\mathcal{Q}_\rho\|, \|\mathcal{A}\|\})$, *the remaining eigenvalues are* $O(\sqrt{\nu_k}\max\{\|\mathcal{Q}_\rho\|, \|\mathcal{A}\|\})$ *perturbations of those of*

$$G^k_{\rho,-} = \begin{bmatrix} N^k_\rho & -(L^k)^T \\ L^k & 0 \end{bmatrix},$$

where N_ρ^k and L^k are defined as in Theorem 4(a). The real eigenvalues of $G_{\rho,-}^k$ are contained in the interval $[0,\, \overline{\theta}^k]$, whereas the complex eigenvalues (with non-zero imaginary parts) are contained in the region

$$\big\{x+iy : \underline{\theta}^k/2 \le x \le \overline{\theta}^k/2,\ |y| \le \|L^k\|\big\}.$$

(b) Under the assumptions stated in Theorem 1(b), there exist positive constants c_1, c_2, c_3 such that for k sufficiently large, the real eigenvalues of $G_{\rho,-}^k$ are contained in the interval $[0,\, 2c_2]$, whereas the complex eigenvalues are contained in the region: $\{x+iy : c_1 \le x \le c_2,\ |y| \le c_3\}$.

Proof. (a) We drop the superscript k in this proof. It is readily shown that up to a perturbation of $O(\sqrt{\nu}\max\{\|\mathcal{Q}_\rho\|, \|\mathcal{A}\|\})$, $\Phi_{\rho,-}^{-1}\mathcal{B}_\rho$ is similar to $\mathrm{diag}(G_{\rho,-}, I_{\bar{s}})$. Thus the eigenvalues of $\Phi_{\rho,-}^{-1}\mathcal{B}_\rho$ are $O(\sqrt{\nu}\max\{\|\mathcal{Q}_\rho\|, \|\mathcal{A}\|\})$ perturbations of those of $\mathrm{diag}(G_{\rho,-}, I_{\bar{s}})$, whose eigenvalues are either 1 or those of $G_{\rho,-}$. The remaining result follows from Proposition 2.11 in [8] and the definitions of $\underline{\theta}$ and $\overline{\theta}$ in (30).

(b) The existence of $c_1, c_2 > 0$ is guaranteed by Lemma 5. Since $\|L\| \le \underline{\alpha}^{-1/2}\max\{\underline{\sigma}_1^{-1/2}, \underline{\sigma}_2^{-1/2}\}\|\mathcal{A}\| =: c_3$, the required result is proved. □

Remark 2 (a) Unlike the matrix $G_{\rho,+}^k$ in Theorem 4, we are not able to show that the real eigenvalues of $G_{\rho,-}^k$ in Theorem 5 are bounded away from zero even with the assumptions stated in Theorem 1(b). However, even though the asymptotic result we are able to prove for $\mathrm{eig}((\Phi_{\rho,-}^k)^{-1}\mathcal{B}_\rho^k)$ is weaker than that for $\mathrm{eig}((\Phi_{\rho,+}^k)^{-1}\mathcal{B}_\rho^k)$, numerical results in [24] showed that a preconditioner that is analogous to $\Phi_{\rho,-}^k$ is typically more effective than the counterpart of $\Phi_{\rho,+}^k$. For the numerical experiments in Section 6, our empirical experience (which we do not report in this paper) confirmed that $\Phi_{\rho,-}^k$ is indeed more effective than $\Phi_{\rho,+}^k$.
(b) It is not difficult to show that $M_\rho^k = ((M_0^k)^{-1} + \rho I)^{-1}$. Thus when ρ is large, $\widehat{M}_\rho^k = \rho^{-1}I$ is a good choice to approximate M_ρ^k.

Finally, we illustrate the asymptotic result in Theorem 4 using the problems in Examples 2 and 3. In this case, we take

$$\mathcal{H}_1^k = \mathcal{D}_1^k + \|\mathcal{Q}_\rho\| I_{\bar{r}}, \quad \mathcal{H}_2^k = \mathcal{D}_2^k + \|\mathcal{Q}_\rho\| I_{rs}, \quad \mathcal{H}_3^k = \mathcal{D}_3^k, \quad \widehat{M}_\rho^k = I_m. \tag{36}$$

Observe that we have approximated $\mathcal{Q}_\rho$ by $\|\mathcal{Q}_\rho\| I_{\bar{n}}$, and hence $(\widetilde{\mathcal{Q}}_\rho^k)_{11} = (\mathcal{P}_1^k)^T\mathcal{Q}_\rho(\mathcal{P}_1^k) \approx \|\mathcal{Q}_\rho\| I_{\bar{r}}$, $(\widetilde{\mathcal{Q}}_\rho^k)_{22} = (\mathcal{P}_2^k)^T\mathcal{Q}_\rho(\mathcal{P}_2^k) \approx \|\mathcal{Q}_\rho\| I_{rs}$.

From the plots in Figure 2, it is clear that for k large, $\mathrm{eig}((\Phi_{\rho,+}^k)^{-1}\mathcal{B}_\rho^k)$ is contained in an interval of the form $[-c_1, -c_2] \cup [c_3, c_4]$ with $c_1, c_2, c_3, c_4 > 0$ independent of k. This observation is consistent with the result in Theorem 4(b) by noting that the optimal solutions for all the QSDPs in both examples are primal nondegenerate and $(\widetilde{\mathcal{Q}}_\rho^*)_{11} \succ 0$ (since $\mathcal{Q} \succ 0$).

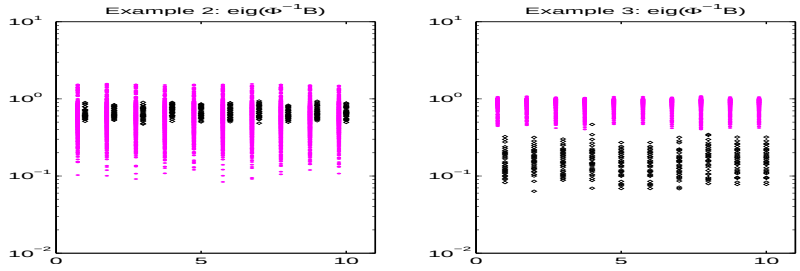

Fig. 2 Same as Fig. 1, but for the spectra of $(\Phi^k_{\rho,+})^{-1}\mathcal{B}^k_\rho$.

4.2 Part II

Our construction of the next preconditioner start with the observation that a matrix of the form given below can be inverted at a moderate cost under Assumption A1 when m is not too large (say, less than 5000):

$$\mathcal{V}^k_\rho \;=\; \begin{bmatrix} -V^k_\rho \circledast V^k_\rho & \mathcal{A}^T \\ \mathcal{A} & 0 \end{bmatrix}, \tag{37}$$

where $V^k_\rho \circledast V^k_\rho$ is a symmetric positive definite approximation of $(W^k)^{-1} \circledast (W^k)^{-1} + \mathcal{Q}_\rho$. Observe that the (1,1) block of $\mathcal{V}^k_\rho$ can be inverted easily and the corresponding Schur complement matrix $S^k_{\mathcal{V}} := \mathcal{A}(V^k_\rho)^{-1} \circledast (V^k_\rho)^{-1}\mathcal{A}^T$ can also be computed at a moderate cost under Assumption A1. Thus given $[X; y] \in \mathcal{S}^n \times \mathbb{R}^m$, $(\mathcal{V}^k_\rho)^{-1}[X; y]$ can be evaluated efficiently through the steps given below:

Compute $\widehat{X} = (V^k_\rho)^{-1}X(V^k_\rho)^{-1}$; $\quad u = \mathcal{A}(\widehat{X})$; $\quad v = (S^k_{\mathcal{V}})^{-1}(u + y)$;

Compute $(\mathcal{V}^k_\rho)^{-1}[X; y] = [(V^k_\rho)^{-1}(\mathcal{A}^T v - X)(V^k_\rho)^{-1}; v]$.

Assuming that $(V^k_\rho)^{-1}$ is pre-computed, the above computation requires four $n \times n$ matrix-matrix multiplications that cost $4n^3$ flops.

We construct the matrix V^k_ρ in (37) as follows. First we find a positive semidefinite symmetrized Kronecker product approximation, say $\Delta_\rho \circledast \Delta_\rho$, of $\mathcal{Q}_\rho$; see Section 5. Then consider the decomposition of the sum: $(\widehat{W}^k)^{-1} \circledast (\widehat{W}^k)^{-1} + \Delta_\rho \circledast \Delta_\rho = (P^k \circledast P^k)(\widetilde{D}^k \circledast \widetilde{D}^k + \widetilde{\Delta}^k_\rho \circledast \widetilde{\Delta}^k_\rho)(P^k \circledast P^k)^T$, with $\widetilde{\Delta}^k_\rho = (P^k)^T\Delta_\rho P^k$, and $\widetilde{D}^k = D^k + \gamma I$, where $\gamma \geq 0$ is a given constant. (Recall that D^k is the diagonal matrix of eigenvalues of $(W^k)^{-1}$. The motivation for adding γI to D^k is to make $\widetilde{D}^k$ more positive definite when ν_k tends to zero. For the linear operator $\widetilde{D}^k \circledast \widetilde{D}^k + \widetilde{\Delta}^k_\rho \circledast \widetilde{\Delta}^k_\rho$ in the decomposition, find a positive definite symmetrized Kronecker product approximation $\widetilde{V}^k_\rho \circledast \widetilde{V}^k_\rho$. Finally, take $V^k_\rho = P^k\widetilde{V}^k_\rho(P^k)^T$.

Observe that the construction of $\mathcal{V}^k_\rho$ does not depend on the clear separation of the eigenvalues of $(W^k)^{-1}$ into two distinct clusters as we had relied upon in the construction of $\Phi^k_{\rho,\pm}$. Thus when the complementarity gap ν_k is relatively large and the eigenvalues of $(W^k)^{-1}$ does not separate into two

distinct clusters, $\mathcal{V}_\rho^k$ is a suitable preconditioner to use. The last statement is especially applicable during the initial phase of the interior-point iteration.

We do not have asymptotic results for the spectrum of $(\mathcal{V}_\rho^k)^{-1}\mathcal{B}_\rho^k$ if the matrix V_ρ^k in (37) is an arbitrary positive definite matrix. However, for the special case where $\mathcal{V}_\rho^k = \Omega^k$, with

$$\Omega^k \;=\; \begin{bmatrix} -(W^k)^{-1} \circledast (W^k)^{-1} & \mathcal{A}^T \\ \mathcal{A} & 0 \end{bmatrix}, \tag{38}$$

we will prove an asymptotic result for the spectrum of $(\Omega^k)^{-1}\mathcal{B}_\rho^k$ in Theorem 6. The motivation for considering the special case Ω^k is as follows. Let $\widehat{\mathcal{Q}}_\rho^k = [\mathcal{Q}_\rho^k,\ 0;\ 0,\ 0]$. Since $\mathcal{B}_\rho^k = \Omega^k - \widehat{\mathcal{Q}}_\rho^k$, it is rather natural to consider using Ω^k as a preconditioner for $\mathcal{B}_\rho^k$. However, note that when the complementarity gap ν_k is small, Ω^k may have small eigenvalues of the order $\Theta(\nu_k)$ (which would happen if (X^*, y^*, S^*) is primal degenerate) and these small eigenvalues may blow up the norm of $(\Omega^k)^{-1}\widehat{\mathcal{Q}}_\rho^k$. Thus, from the equation $(\Omega^k)^{-1}\mathcal{B}_\rho^k = I - (\Omega^k)^{-1}\widehat{\mathcal{Q}}_\rho^k$, it is not clear whether Ω^k would be a good preconditioner. We will next show that when the optimal solution (X^*, y^*, S^*) is dual nondegenerate, Ω^k can in fact be an effective preconditioner for $\mathcal{B}_\rho^k$.

Before we proceed further, we state two lemmas that will be used in the analysis of the spectrum of the preconditioned matrix.

Lemma 6 *Suppose $U \in \mathcal{S}^p_{++}$, and $V \in \mathbb{R}^{m\times p}$ has full row rank. Let $G = [-U,\, V^T; V,\, 0]$. Suppose $\widehat{U}$ is a symmetric positive definite approximation of U, and we consider $\widehat{G} = [-\widehat{U},\, V^T; V,\, 0]$ as a preconditioner for G. Then $\widehat{G}^{-1}G$ has $2m$ eigenvalues equal to 1, and the remaining $p-m$ eigenvalues are those of the matrix $Z^T\widehat{U}^{-1/2}U\widehat{U}^{-1/2}Z$, where $Z \in \mathbb{R}^{p\times(p-m)}$ is a matrix whose columns form an orthonormal basis of $\mathcal{N}(V\widehat{U}^{-1/2})$.*

Proof. See Theorem 2 in [28]. □

Lemma 7 *Let $\left\{Z^l \;=\; [Z_1^l;\, Z_2^l;\, Z_3^l]\right\}_{l=1}^{\bar{n}-m}$ be an orthonormal set in $\mathcal{S}^r \times \mathbf{R}^{r\times s} \times \mathcal{S}^s$ that form a basis of $\mathcal{N}(\widetilde{\mathcal{A}}^k(\mathcal{D}^k)^{-1/2})$. Let $\mathcal{Z}_1^k : \mathbb{R}^{\bar{n}-m} \to \mathcal{S}^r$ be defined by $\mathcal{Z}_1^k y = \sum_{l=1}^{\bar{n}-m} y_l Z_1^l$. We define $\mathcal{Z}_2^k, \mathcal{Z}_3^k$ similarly. If (X^*, y^*, S^*) is dual nondegenerate, then for k sufficiently large,*

$$(\mathcal{Z}_1^k)^T(\mathcal{D}_1^k)^{-1}\mathcal{Z}_1^k \;\preceq\; \|(\widetilde{\mathcal{A}}_1^k)^\dagger\|^2\|\mathcal{A}\|^2\big((\mathcal{Z}_2^k)^T(\mathcal{D}_2^k)^{-1}\mathcal{Z}_2^k + (\mathcal{Z}_3^k)^T(\mathcal{D}_3^k)^{-1}\mathcal{Z}_3^k\big),$$

where $(\widetilde{\mathcal{A}}_1^k)^\dagger$ is the pseudo-inverse of $\widetilde{\mathcal{A}}_1^k$.

Proof. The solution (X^*, y^*, S^*) is dual nondegenerate implies that $\widetilde{\mathcal{A}}_1^*$ is injective. Since $\lim_{k\to\infty}\widetilde{\mathcal{A}}_1^k = \widetilde{\mathcal{A}}_1^*$, thus for k large, $\widetilde{\mathcal{A}}_1^k$ is injective, and $(\widetilde{\mathcal{A}}_1^k)^\dagger = \big((\widetilde{\mathcal{A}}_1^k)^T\widetilde{\mathcal{A}}_1^k\big)^{-1}(\widetilde{\mathcal{A}}_1^k)^T$. Now, $Z^l \in \mathcal{N}(\widetilde{\mathcal{A}}^k(\mathcal{D}^k)^{-1/2})$ for all l implies that $\widetilde{\mathcal{A}}_1^k(\mathcal{D}_1^k)^{-1/2}\mathcal{Z}_1^k = -[\widetilde{\mathcal{A}}_2^k,\, \widetilde{\mathcal{A}}_3^k]\,[(\mathcal{D}_2^k)^{-1/2}\mathcal{Z}_2^k;\, (\mathcal{D}_3^k)^{-1/2}\mathcal{Z}_3^k]$. Thus for k large,

$$(\mathcal{D}_1^k)^{-1/2}\mathcal{Z}_1^k \;=\; -(\widetilde{\mathcal{A}}_1^k)^\dagger[\widetilde{\mathcal{A}}_2^k,\, \widetilde{\mathcal{A}}_3^k]\,[(\mathcal{D}_2^k)^{-1/2}\mathcal{Z}_2^k;\, (\mathcal{D}_3^k)^{-1/2}\mathcal{Z}_3^k]. \tag{39}$$

By noting that $[\widetilde{\mathcal{A}}_2^k,\ \widetilde{\mathcal{A}}_3^k]^T((\widetilde{\mathcal{A}}_1^k)^\dagger)^T(\widetilde{\mathcal{A}}_1^k)^\dagger[\widetilde{\mathcal{A}}_2^k,\ \widetilde{\mathcal{A}}_3^k] \preceq \|(\widetilde{\mathcal{A}}_1^k)^\dagger\|^2\|\mathcal{A}\|^2 I$, the equation in (39) easily leads to the required result. □

Theorem 6 *Let $\mathcal{Z}_j^k$, $j = 1, 2, 3$, be defined as in Lemma 7 for $\mathcal{N}(\widetilde{\mathcal{A}}^k(\mathcal{D}^k)^{-1/2})$. Let $\mathcal{Z}^k = [\mathcal{Z}_1^k;\ \mathcal{Z}_2^k;\ \mathcal{Z}_3^k]$ and $\boldsymbol{G}_\rho^k = (\mathcal{Z}^k)^T(\mathcal{D}^k)^{-1/2}(\mathcal{D}^k + \widetilde{\mathcal{Q}}_\rho^k)(\mathcal{D}^k)^{-1/2}\mathcal{Z}^k$. We have the following results.*
(a) The preconditioned matrix $(\Omega^k)^{-1}\mathcal{B}_\rho^k$ has $2m$ eigenvalues equal to 1, and the remaining $\bar{n} - m$ eigenvalues are those of $\boldsymbol{G}_\rho^k$.
(b) Suppose the optimal solution (X^*, y^*, S^*) is dual nondegenerate. Then for k sufficiently large,

$$\mathrm{eig}(\boldsymbol{G}_\rho^k) \subset 1 + \Big[0,\ \|\mathcal{Q}_\rho\|\Big(\|(\widetilde{\mathcal{A}}_1^k)^\dagger\|^2\|\mathcal{A}\|^2 + 1\Big)\max(\underline{\tau}^{-1}, \Theta(\nu_k))\Big].$$

Proof. (a) The result follows from Lemma 6.

(b) For the rest of the proof, we assume that k is sufficiently large. We can write $\boldsymbol{G}_\rho^k = I + T_\rho^k$, where $T_\rho^k := (\mathcal{Z}^k)^T(\mathcal{D}^k)^{-1/2}\widetilde{\mathcal{Q}}_\rho^k(\mathcal{D}^k)^{-1/2}\mathcal{Z}^k$. Since $T_\rho^k \succeq 0$, it is clear that $\boldsymbol{G}_\rho^k \succeq I$. Now consider matrix T_ρ^k. Since $0 \preceq \widetilde{\mathcal{Q}}_\rho^k \preceq \|\mathcal{Q}_\rho\| I$, we have $T_\rho^k \preceq \|\mathcal{Q}_\rho\| \sum_{j=1}^3 (\mathcal{Z}_j^k)^T(\mathcal{D}_j^k)^{-1}\mathcal{Z}_j^k$. Hence

$$T_\rho^k \preceq \|\mathcal{Q}_\rho\|\,(\|(\widetilde{\mathcal{A}}_1^k)^\dagger\|^2\|\mathcal{A}\|^2 + 1)\big((\mathcal{Z}_2^k)^T(\mathcal{D}_2^k)^{-1}\mathcal{Z}_2^k + (\mathcal{Z}_3^k)^T(\mathcal{D}_3^k)^{-1}\mathcal{Z}_3^k\big).$$

Now $\underline{\tau} I \preceq \mathcal{D}_2^k, \Theta(1/\nu_k) I \preceq \mathcal{D}_3^k \Rightarrow (\mathcal{Z}_j^k)^T(\mathcal{D}_j^k)^{-1}\mathcal{Z}_j^k \preceq \max(\underline{\tau}^{-1}, \Theta(\nu_k))(\mathcal{Z}_j^k)^T\mathcal{Z}_j^k$ for $j = 2, 3$. Together with the fact that $\sum_{j=1}^3 (\mathcal{Z}_j^k)^T\mathcal{Z}_j^k = I$, we have $T_\rho^k \preceq \|\mathcal{Q}_\rho\|\,(\|(\widetilde{\mathcal{A}}_1^k)^\dagger\|^2\|\mathcal{A}\|^2 + 1)\max(\underline{\tau}^{-1}, \Theta(\nu_k))\, I$. From here, the required result follows readily. □

We consider again the problems in Examples 2 and 3 to illustrate the asymptotic result in Theorem 6(b). The spectra of $(\Omega^k)^{-1}\mathcal{B}_\rho^k$ for 10 random instances of Examples 2 and 3 are plotted in Figure 3. From the right-hand side plot, it is clear that for k large, $\mathrm{eig}((\Omega^k)^{-1}\mathcal{B}_\rho^k)$ for the instances in Example 3 are each contained in an interval of the form $[c_1, c_2]$ with $c_1, c_2 > 0$ independent of k. This observation is consistent with the result in Theorem 6(b) by noting that the optimal solutions for the instances in Example 3 are dual nondegenerate. On the other hand, the optimal solutions for the instances in Example 2 are dual degenerate, and the spectra of $(\Omega^k)^{-1}\mathcal{B}_\rho^k$ shown in the left-hand side plot of Figure 3 are not contained in a finite interval as $k \uparrow \infty$.

Remark 3 An alternative to $\mathcal{V}_\rho^k$ in (37) during the initial phase of the interior-point iteration is the following preconditioner:

$$\begin{bmatrix} -\mathrm{diag}((W^k)^{-1}) \circledast \mathrm{diag}((W^k)^{-1}) - \mathrm{diag}(\mathcal{Q}_\rho) & \mathcal{A}^T \\ \mathcal{A} & 0 \end{bmatrix}. \tag{40}$$

The above is analogous to the preconditioner proposed in [9] for the LCCQP problem (5). Unfortunately, it is not competitive at all compared to $\mathcal{V}_\rho^k$ for the test problems we are considering in Section 6. An important distinction

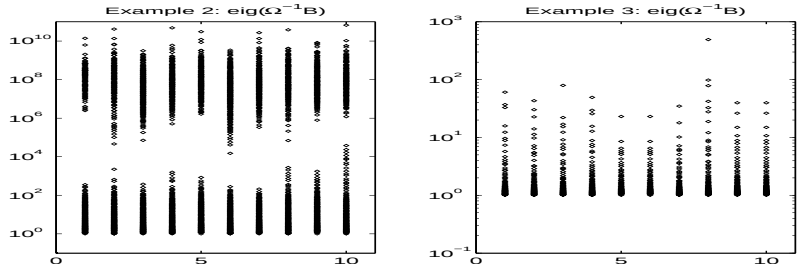

Fig. 3 Same as Fig. 1 but for the spectra of $(\Omega^k)^{-1}\mathcal{B}_\rho^k$.

between the preconditioner in (40) and its counterpart for LCCQP is that for the latter, the contribution by the interior-point iterate is fully incorporated into the (1,1) block of the preconditioner, but for the former, only a diagonal approximation of $(W^k)^{-1} \circledast (W^k)^{-1}$ is incorporated. Such a diagonal approximation appears to be too weak to make (40) an effective preconditioner for $\mathcal{B}_\rho^k$, even when the complementarity gap ν_k is relatively large. One may argue that the preconditioner (40) for (QSDP) is actually not the correct analog of the preconditioner proposed in [9] for LCCQP, and the correct version should have its (1,1) block equal to $-\mathcal{P}^k(\mathcal{D}^k + \mathrm{diag}(\widetilde{\mathcal{Q}}_\rho^k))(\mathcal{P}^k)^T$. However, it is expensive to compute $\mathrm{diag}(\widetilde{\mathcal{Q}}_\rho^k)$ since at least $\Theta(n^4)$ flops is needed in general. As such, the latter preconditioner is usually not a practical choice.

4.3 Part III

The next preconditioner we propose for $\mathcal{B}_\rho^k$ is constructed by suitably modifying $\mathcal{D}^k$ in the decomposition: $(W^k)^{-1} \circledast (W^k)^{-1} = \mathcal{P}^k \mathrm{diag}(\mathcal{D}_1^k, \mathcal{D}_2^k, \mathcal{D}_3^k)(\mathcal{P}^k)^T$. Specifically, the preconditioner is given by:

$$\Psi_\rho^k \;=\; \begin{bmatrix} -\mathcal{P}^k \, \Gamma_\rho^k \, (\mathcal{P}^k)^T & \mathcal{A}^T \\ \mathcal{A} & 0 \end{bmatrix}, \quad \text{where } \; \Gamma_\rho^k = \mathrm{diag}(\mathcal{H}_1^k, \mathcal{H}_2^k, \mathcal{H}_3^k). \tag{41}$$

For the above preconditioner, we typically want to choose the linear operators $\mathcal{H}_1^k$, $\mathcal{H}_2^k$, and $\mathcal{H}_3^k$ to have the following forms:

$$\mathcal{H}_1^k \;=\; H_1^k \circledast H_1^k, \quad \mathcal{H}_2^k \;=\; \widehat{H}_2^k \otimes H_2^k, \quad \mathcal{H}_3^k \;=\; H_3^k \circledast H_3^k, \tag{42}$$

where $H_1^k, H_2^k \in \mathcal{S}_{++}^r$, and $\widehat{H}_2^k, \widehat{H}_3^k \in \mathcal{S}_{++}^s$. The motivation for considering the operators in (42) will become apparent when we derive the Schur complement matrix associated with Ψ_ρ^k in the ensuing paragraph. Note that if (X^*, y^*, S^*) is primal degenerate (i.e., $[\widetilde{\mathcal{A}}_1^*, \widetilde{\mathcal{A}}_2^*]$ is not surjective), then Ψ_ρ^k may be nearly singular when k is large. Given $[X; y] \in \mathcal{S}^n \times \mathbb{R}^m$, $(\Psi_\rho^k)^{-1}[X; y]$ can be evaluated efficiently through the steps given below:

Compute $\widehat{X} = (P^k)^T X P^k$; $u = \mathcal{A}(P^k(\mathcal{E}^k(\widehat{X}))(P^k)^T)$; $v = (S_\Psi^k)^{-1}(u + y)$;

Compute $U = (P^k)^T(\mathcal{A}^T v - X)P^k$; $(\Psi_\rho^k)^{-1}[X; y] = [P^k(\mathcal{E}^k(U))(P^k)^T; v]$.

The above computation requires eight $n \times n$ matrix-matrix multiplications that cost $8n^3$ flops.

The Schur complement matrix associated with Ψ_ρ^k is given by $S_\Psi^k := \mathcal{A}\mathcal{P}^k(\Gamma_\rho^k)^{-1}(\mathcal{P}^k)^T\mathcal{A}^T$. For the choices of $\mathcal{H}_1^k$, $\mathcal{H}_2^k$, and $\mathcal{H}_3^k$ in (42), we have:

$$S_\Psi^k = \widetilde{\mathcal{A}}_1^k(\mathcal{H}_1^k)^{-1}(\widetilde{\mathcal{A}}_1^k)^T + \widetilde{\mathcal{A}}_2^k(\mathcal{H}_2^k)^{-1}(\widetilde{\mathcal{A}}_2^k)^T + \widetilde{\mathcal{A}}_3^k(\mathcal{H}_3^k)^{-1}(\widetilde{\mathcal{A}}_3^k)^T \quad (43)$$

$$= \mathcal{A}V_1^k \circledast V_1^k\mathcal{A}^T + 2\mathcal{A}V_2^k \circledast \widehat{V}_2^k\mathcal{A}^T + \mathcal{A}V_3^k \circledast V_3^k\mathcal{A}^T, \quad (44)$$

where $V_1^k = P_1^k(H_1^k)^{-1}(P_1^k)^T$, $V_2^k = P_1^k(H_2^k)^{-1}(P_1^k)^T$, $\widehat{V}_2^k = P_2^k(\widehat{H}_2^k)^{-1}(P_2^k)^T$, and $V_3^k = P_2^k(H_3^k)^{-1}(P_2^k)^T$. In this case, the computation of S_Ψ^k is similar to that of the Schur complement matrix in a linear SDP employing the NT direction. In particular, any sparsity structure in $\mathcal{A}$ can be exploited efficiently in the computation; see [12]. The formula in (44) for $\mathcal{S}_\Psi^k$ holds only for the special choices of $\mathcal{H}_1^k$, $\mathcal{H}_2^k$ and $\mathcal{H}_3^k$ in (42). Without using the special choices, $\mathcal{S}_\Psi^k$ must be computed from (43) for which the computation is akin to that of the Schur complement matrix of a linear SDP employing the AHO direction [1]. It is well known that the latter computation cannot exploit sparsity in $\mathcal{A}$ as effectively as the computation using (44). This explains the rationale for the special choices of $\mathcal{H}_1^k$, $\mathcal{H}_2^k$ and $\mathcal{H}_3^k$ in (42).

Theorem 7 *Suppose Assumptions A2, A3 hold. Let* $\left\{ Z^l = [Z_1^l;\, Z_2^l;\, Z_3^l] \right\}_{l=1}^{\bar{n}-m}$ *be an orthonormal set in* $\mathcal{S}^r \times \mathbf{R}^{r\times s} \times \mathcal{S}^s$ *that form a basis of* $\mathcal{N}(\widetilde{\mathcal{A}}^k(\Gamma_\rho^k)^{-1/2})$. *Let* $\mathcal{Z}_1^k : I\!R^{\bar{n}-m} \to \mathcal{S}^r$ *be defined by* $\mathcal{Z}_1^k y = \sum_{l=1}^{\bar{n}-m} y_l Z_1^l$. *We define* $\mathcal{Z}_2^k, \mathcal{Z}_3^k$ *similarly. Suppose* $\mathcal{Z}^k = [\mathcal{Z}_1^k; \mathcal{Z}_2^k; \mathcal{Z}_3^k]$. *Consider the matrix*

$$\boldsymbol{G}_\rho^k \;=\; (\mathcal{Z}^k)^T(\Gamma_\rho^k)^{-1/2}(\mathcal{D}^k + \widetilde{\mathcal{Q}}_\rho^k)(\Gamma_\rho^k)^{-1/2}\mathcal{Z}^k.$$

(a) Then the preconditioned matrix $(\Psi_\rho^k)^{-1}\mathcal{B}_\rho^k$ *has* $2m$ *eigenvalues equal to* 1. *The remaining* $\bar{n} - m$ *eigenvalues are those of the matrix* $\boldsymbol{G}_\rho^k$.
(b) Suppose $\mathcal{H}_1^k, \mathcal{H}_2^k, \mathcal{H}_3^k$ *in* Γ_ρ^k *satisfy the conditions in (28) and* $(\widetilde{\mathcal{Q}}_\rho^*)_{11} \succ 0$. *Then there exist* $c_1, c_2 > 0$ *such that* $\mathrm{eig}(\boldsymbol{G}_\rho^k) \;\subset\; [c_1,\, c_2]$ *for* k *large.*
(c) Suppose $\mathcal{H}_1^k, \mathcal{H}_2^k, \mathcal{H}_3^k$ *in* Γ_ρ^k *satisfy the conditions in (28) and* $\underline{\beta}\mathcal{D}^k \preceq \Gamma_\rho^k \preceq \overline{\beta}(\mathcal{D}^k + \widetilde{\mathcal{Q}}_\rho^k)$ *for all* k *for some constants* $\underline{\beta}, \overline{\beta} > 0$. *Then*

$$\mathrm{eig}(\boldsymbol{G}_\rho^k) \;\subset\; \big[\,\overline{\beta}^{-1},\, \underline{\beta}^{-1} + \|\mathcal{Q}_\rho\| \max\{\underline{\sigma}_1^{-1}, \underline{\sigma}_2^{-1}, \Theta(\nu_k)\}\big]. \quad (45)$$

(d) Suppose $\underline{\beta}(\mathcal{D}^k + \widetilde{\mathcal{Q}}_\rho^k) \preceq \Gamma_\rho^k \preceq \overline{\beta}(\mathcal{D}^k + \widetilde{\mathcal{Q}}_\rho^k)$ *for all* k *for some constants* $\underline{\beta}, \overline{\beta} > 0$. *Then* $\mathrm{eig}(\boldsymbol{G}_\rho^k) \subset \big[\,\overline{\beta}^{-1},\, \underline{\beta}^{-1}\big]$.

Proof. (a) The first result follows from Lemma 6.

(b) The matrix $\boldsymbol{G}_\rho^k$ can be written as $\boldsymbol{G}_\rho^k = (\mathcal{Z}^k)^T\mathrm{diag}(N_\rho^k, I_{\bar{s}})\mathcal{Z}^k + O(\sqrt{\nu_k}\|\mathcal{Q}_\rho\|)$, where N_ρ^k is defined in (29). Thus, using the definitions of $\underline{\theta}^k$ and $\overline{\theta}^k$ in (30), we have $\mathrm{eig}(\boldsymbol{G}_\rho^k) \subset [\underline{\theta}^k,\, \overline{\theta}^k] + O(\sqrt{\nu_k}\|\mathcal{Q}_\rho\|)$. By Lemma 5, the required result follows.

(c) It is easy to see that $\boldsymbol{G}_\rho^k \succeq \overline{\beta}^{-1}(\mathcal{Z}^k)^T\mathcal{Z}^k = \overline{\beta}^{-1}I$. On the other hand,

$$\boldsymbol{G}_\rho^k \preceq \underline{\beta}^{-1}I + (\mathcal{Z}^k)^T(\Gamma_\rho^k)^{-1/2}\widetilde{\mathcal{Q}}_\rho^k(\Gamma_\rho^k)^{-1/2}\mathcal{Z}^k \preceq \underline{\beta}^{-1}I + \|\mathcal{Q}_\rho\|(\mathcal{Y}^k)^T\mathcal{Y}^k,$$

where $\mathcal{Y}^k := [\mathcal{Y}_1^k; \mathcal{Y}_2^k; \mathcal{Y}_3^k] = (\Gamma_\rho^k)^{-1/2}\mathcal{Z}^k$. Now

$$\begin{aligned} I = (\mathcal{Z}^k)^T\mathcal{Z}^k = (\mathcal{Y}^k)^T\Gamma_\rho^k\mathcal{Y}^k &= (\mathcal{Y}_1^k)^T\mathcal{H}_1^k\mathcal{Y}_1^k + (\mathcal{Y}_2^k)^T\mathcal{H}_2^k\mathcal{Y}_2^k + (\mathcal{Y}_3^k)^T\mathcal{H}_3^k\mathcal{Y}_3^k \\ &\succeq \min\{\underline{\sigma}_1, \underline{\sigma}_2, \Theta(1/\nu_k)\}\,(\mathcal{Y}^k)^T\mathcal{Y}^k, \end{aligned}$$

hence $\boldsymbol{G}_\rho^k \preceq \underline{\beta}^{-1}I + \|\mathcal{Q}_\rho\|\max\{\underline{\sigma}_1^{-1}, \underline{\sigma}_2^{-1}, \Theta(\nu_k)\}I$. From here, (45) follows.

(d) Since $\overline{\beta}^{-1}I \preceq (\Gamma_\rho^k)^{-1/2}(\mathcal{D}^k + \widetilde{\mathcal{Q}}_\rho^k)(\Gamma_\rho^k)^{-1/2} \preceq \underline{\beta}^{-1}I$, and $(\mathcal{Z}^k)^T\mathcal{Z}^k = I$, thus $\overline{\beta}^{-1}I \preceq \boldsymbol{G}_\rho^k \preceq \underline{\beta}^{-1}I$, and the required result follows. □

Remark 4 (a) Observe that unlike Theorems 4(b) and 6(b), Theorem 7(b) does not assume primal or dual nondegeneracy for the optimal solution.
(b) As an example, we show how Theorem 7(b) can be modified to suit the LCCQP problem (5). Note that for a LCCQP problem, $\mathcal{D}_2^k$ does not exist, $\mathcal{D}^k = \mathrm{diag}(\mathcal{D}_1^k, \mathcal{D}_3^k)$ with $\mathrm{diag}(\mathcal{D}_1^k) = \Theta(\nu_k)$, $\mathrm{diag}(\mathcal{D}_3^k) = \Theta(1/\nu_k)$, and $\mathcal{P}^k$ is a permutation matrix. $\widetilde{\mathcal{A}}^k$ is actually A but with its columns permuted according to the partition in $\mathcal{D}^k$. Similarly, $\widetilde{\mathcal{Q}}_\rho^k$ is $Q_\rho := Q + \rho A^TA$ but with its rows and columns permuted. Assuming that $\mathcal{H}_1^k$ and $\mathcal{H}_3^k$ satisfy the conditions in (28) and $(\widetilde{\mathcal{Q}}_\rho)_{11} \succ 0$, then there exist $c_1, c_2 > 0$ such that $\mathrm{eig}(\Psi_\rho^k)^{-1}\mathcal{B}_\rho^k) \subset \{1\} \cup [c_1, c_2]$ for k large. We note that such a result is established under a weaker condition than the one required in Corollary 4.5 of [9], which assumed Q to be positive definite.

We consider again the problems in Examples 2 and 3 to illustrate the asymptotic result in Theorem 7(b). We take $\mathcal{H}_1^k$, $\mathcal{H}_2^k$, and $\mathcal{H}_3^k$ as in (36). The spectra of $(\Psi_\rho^k)^{-1}\mathcal{B}_\rho^k$ for 10 instances of Examples 2 and 3 are plotted in Figure 4. From the plots, it is clear that for each problem instances, $\mathrm{eig}((\Psi_\rho^k)^{-1}\mathcal{B}_\rho^k)$ is contained in a finite interval $[c_1, c_2]$ with $c_1, c_2 > 0$ independent of k.

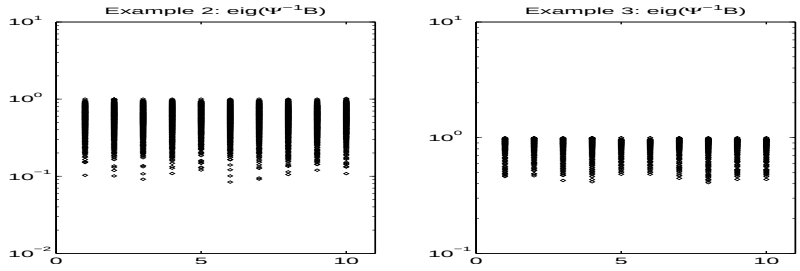

Fig. 4 Same as Fig. 1 but for the spectra of $(\Psi_\rho^k)^{-1}\mathcal{B}_\rho^k$.

5 A strategy for constructing symmetrized Kronecker product approximations

The construction of preconditioners in the last section relied to some extent on our ability to find a suitable symmetrized Kronecker product (SKP) approximation of $\mathcal{Q}_\rho$ or submatrices of $\widetilde{\mathcal{Q}}_\rho^k$. However, very little is known about such an approximation problem. Our objective in this section is to provide a strategy for constructing an SKP approximation.

Contrary to the SKP approximation problem, the problem of finding the nearest (in Frobenius norm) Kronecker product (NKP) approximation of a linear operator defined on $\mathbb{R}^{n\times n}$ has been studied in [30, Section 6]. (We refer the reader to the references therein for earlier work on the subject.) It is shown in [30] that the NKP approximation can be derived from the best rank-one approximation of a permuted version of the matrix representation of the linear operator with respect to the canonical basis of $\mathbb{R}^{n\times n}$. The special case of finding the NKP approximation of a sum of Kronecker products is studied in [19], wherein the solution is obtained by solving a small non-convex optimization problem. It is shown recently in [29] that the latter solution can actually be found analytically by solving a small eigenvalue problem.

By applying our knowledge on the NKP problem, a natural route for constructing a suitable SKP approximation of a linear operator $\mathcal{T}$ defined on $\mathcal{S}^n$ is as follows. We first extend the linear operator to $\widehat{\mathcal{T}} : R^{n\times n} \to R^{n\times n}$ defined by $\widehat{\mathcal{T}}(U) = \mathcal{T}((U+U^T)/2)$. Suppose $U\otimes V$ is the NKP approximation of $\widehat{\mathcal{T}}$. Then $U \circledast V$ can be used as an SKP approximation of $\mathcal{T}$. Moreover, the error in the SKP approximation is no larger than that for the NKP approximation in the following sense. Suppose the matrix representations of $\mathcal{T}$ and $\widehat{\mathcal{T}}$ in the canonical orthonormal bases of $\mathcal{S}^n$ and $\mathbb{R}^{n\times n}$ are denoted by $\mathrm{mat}(\mathcal{T})$ and $\mathrm{mat}(\widehat{\mathcal{T}})$, respectively. We define $\mathrm{mat}(U \otimes V)$ and $\mathrm{mat}(U \circledast V)$ similarly. Then $\mathrm{mat}(\mathcal{T} - U \circledast V) = \Pi^T \mathrm{mat}(\widehat{\mathcal{T}} - U \otimes V)\,\Pi$, where $\Pi \in \mathbb{R}^{n^2\times \bar{n}}$ has orthonormal columns; see [27]. As a result, $\|\mathrm{mat}(\mathcal{T}-U\circledast V)\|_F \le \|\mathrm{mat}(\widehat{\mathcal{T}} - U \otimes V)\|_F$.

Note that in the special case where $\mathcal{Q}$ is a diagonal operator defined by $\mathcal{Q}(X) = U \circ X$, where $U \in \mathcal{S}^n \cap \mathbb{R}^{n\times n}_+$ is given, a positive semidefinite SKP approximation for $\mathcal{Q}$ can be constructed readily as follows. Consider the best rank-one approximation uu^T of U such that $\|U - uu^T\|_F = \min$, then $\mathrm{diag}(u) \circledast \mathrm{diag}(u)$ is a positive semidefinite SKP approximation of $\mathcal{Q}$. The fact that the preceding SKP is positive semidefinite is a consequence of the Perron-Frobenius Theorem, which ensures that u can be chosen in $\mathbb{R}^n_+$ when the entries of U are non-negative; see [15, Theorem 8.3.1].

For the preconditioner Ψ_ρ^k in (41), a procedure for constructing an approximation of the form $\mathcal{H}_1^k = H_1^k \circledast H_1^k$ for $\mathcal{D}_1^k + (\widetilde{\mathcal{Q}}_\rho^k)_{11}$ in (42) is as follows.

Procedure SKPA:

(i) Construct a positive semidefinite SKP approximation $\Delta_\rho \circledast \Delta_\rho$ for $\mathcal{Q}_\rho$. Thus $(\widetilde{\mathcal{Q}}_\rho^k)_{11} \approx (P_1^k)^T \Delta_\rho P_1^k \circledast (P_1^k)^T \Delta_\rho P_1^k$.

(ii) From the sum of positive semidefinite Kronecker products, $D_1^k \otimes D_1^k + (P_1^k)^T \Delta_\rho P_1^k \otimes (P_1^k)^T \Delta_\rho P_1^k$, construct the NKP approximation $H_1^k \otimes H_1^k$

by using the method described in [19]. It is shown in [29] that H_1^k can be chosen to be positive semidefinite.

(iii) Take $\mathcal{H}_1^k = H_1^k \circledast H_1^k$.

The same procedure can similarly be applied to the construction of the approximations: $\mathcal{H}_2^k = \widehat{H}_2^k \otimes H_2^k$ for $\mathcal{D}_2^k + (\widetilde{\mathcal{Q}}_\rho^k)_{22}$, and $\mathcal{H}_3^k = H_3^k \circledast H_3^k$ for $\mathcal{D}_3^k + (\widetilde{\mathcal{Q}}_\rho^k)_{33}$.

6 Numerical experiments

Before we present the details of the numerical experiments, we should emphasize that our purpose here is not conduct extensive numerical testing but to demonstrate that the algorithms we have proposed for solving (1) are correct and they can potentially be efficient and robust. These algorithms may be adopted as prototypes from which more sophisticated and tailor-made algorithms can be designed for solving QSDPs coming from real-world applications.

To evaluate the performance of our interior-point algorithms for solving (1) and the effectiveness of the preconditioners constructed in Section 4 for solving (9), we consider the following classes of test problems:

E1. QSDPs arising from the NCM problem (3) with $\mathcal{Q}(X) = U \circ X$. We generate the matrix K in (3) as in Example 2. The matrix U is generated randomly as follows: `tmp = rand(n); U = (tmp+tmp')/2`. We generated 4 test problems with $n = 100, 400, 800, 1600$. For these problems, the condition numbers of $\mathcal{Q}$ range from 7.67×10^2 to 4.37×10^3. The norms $\|\mathcal{Q}\|$ are approximately 1 for all problems.

E2. Same as E1 but $K \in \mathcal{S}^n$ is generated as in Example 3.

E3. Same as E1 but U is given by `tmp = rand(n); U = 5*(tmp+tmp')`. In this case, the quadratic form $X \bullet \mathcal{Q}(X)$ has more weight in the objective function of (1). The norms $\|\mathcal{Q}\|$ are approximately 10 for all problems.

E4. Same as E2 but U is given by `tmp = rand(n); U = 5*(tmp+tmp')`.

E5. Same as E1 but with additional linear constraints added. Let J_n be a randomly chosen subset of $\{(i,j) : 1 \le i < j,\ j = 1, \ldots, n\}$. We set $X_{ij} = 0$ for all $(i,j) \in J_n$. The purpose here is to study the effect of having a more complicated linear map $\mathcal{A}(X)$ compared to the diagonal map $\mathrm{diag}(X)$ in (3). Again, we generated 4 test problems with $n = 100, 400, 800, 1600$. The number of additional linear constraints added are $120, 397, 772, 1566$, respectively.

E6–E8. Same as E2–E4 but with additional linear constraints similar to those in E5 added.

We use 5 variants of Algorithm IP-QSDP described in [29] to solve each test problem:

A0. Algorithm IP-QSDP that uses a direct solver to solve (9).

A1. Algorithm IP-QSDP that uses PSQMR to solve (9) with no preconditioning.

A2. Algorithm IP-QSDP that employs PSQMR to solve (9) with the preconditioner chosen as follows: it is taken to be $\mathcal{V}_\rho^k$ in (37) if $\kappa(W^k) \le 10^3$; otherwise, it is taken to be $\Phi_{\rho,-}^k$ in (33). As mentioned in Remark 2, $\Phi_{\rho,-}^k$ is empirically observed to be more effective than $\Phi_{\rho,+}^k$.

A3. Algorithm IP-QSDP that employs PSQMR to solve (9) using the preconditioner $\mathcal{V}_\rho^k$ in (37). Recall that $\mathcal{V}_\rho^k$ is a variant of the preconditioner Ω^k in (38). Even though we do not have any asymptotic result for $\mathrm{eig}((\mathcal{V}_\rho^k)^{-1}\mathcal{B}_\rho^k)$, we found that $\mathcal{V}_\rho^k$ is typically much more effective than Ω^k.

A4. Algorithm IP-QSDP that employs PSQMR to solve (9) with the preconditioner chosen as follows: it is taken to be $\mathcal{V}_\rho^k$ in (37) if $\kappa(W^k) \le 10^3$; otherwise, it is taken to be Ψ_ρ^k in (41).

When $\Phi_{\rho,-}^k$ is chosen as the preconditioner, the eigenvalues of $(W^k)^{-1}$ is partitioned according to the threshold value of 1. The linear operators $\mathcal{H}_1^k$, $\mathcal{H}_2^k$, $\mathcal{H}_3^k$, and $\widehat{M}_\rho^k$ in $\Phi_{\rho,-}^k$ are chosen as in (36).

When Ψ_ρ^k is chosen as the preconditioner, the eigenvalues of $(W^k)^{-1}$ is also partitioned according to the threshold value of 1. The linear operators $\mathcal{H}_1^k$, $\mathcal{H}_2^k$, and $\mathcal{H}_3^k$ in Ψ_ρ^k are chosen to have the forms in (42) and they are constructed from Procedure SKPA described in Section 5 with $\Delta_\rho = \sqrt{\|\mathcal{Q}_\rho\|}I_n$. Note that such a choice of Δ_ρ is particularly attractive because the resulting constituent matrices $H_1^k, H_2^k, \widehat{H}_2^k, H_3^k$ in $\mathcal{H}_1^k$, $\mathcal{H}_2^k$, $\mathcal{H}_3^k$ are diagonal matrices.

When $\mathcal{V}_\rho^k$ is chosen as the preconditioner, the operator $\mathcal{Q}(X) = U \circ X$ is approximated by $\Delta \circledast \Delta$, with $\Delta = \mathrm{diag}(u)$, where $u \in \mathbb{R}_+^n$ is the vector such that $\|U - uu^T\|_F = \min$.

For the test problems in E1–E8, the main computational cost at each PSQMR step lies in $n \times n$ matrix-matrix multiplications, and they are summarized as follows. The cost per PSQMR step for A1 is $2n^3$ flops; the corresponding cost for A2 or A3 is $6n^3$ flops; and for A4, it is $6n^3$ and $10n^3$ flops when the preconditioner is $\mathcal{V}_\rho^k$ and Ψ_ρ^k, respectively. For these problems, evaluating $\mathcal{A}(X)$, $\mathcal{A}^T y$, $\mathcal{Q}(X)$ cost at most $O(n^2)$ flops, and these terms are ignored.

We implemented the algorithms in MATLAB (version 7.0) and the experiments were conducted on a 3.0GHz Pentium 4 PC with 4GB of RAM. We stopped the algorithms when the accuracy measure ϕ defined by

$$\phi = \max\left\{ \frac{X \bullet S}{1 + |\mathrm{pobj}| + |\mathrm{dobj}|}\,,\ \frac{\|R_p\|_2}{1 + \|b\|_2}\,,\ \frac{\|R_d\|_F}{1 + \|C\|_F} \right\} \tag{46}$$

is less than 10^{-7}, or when the algorithms did not improve both the duality gap and infeasibilities. In (46), "pobj" and "dobj" denote the primal and dual objective values, respectively.

The stopping criterion used to solve (9) is described in (11). We also set the maximum number of PSQMR steps allowed to solve each augmented system to 1000. Moreover, we stopped the interior-point iteration when PSQMR hit the maximum number of steps allowed. The last condition indicated that the linear system was becoming very ill-conditioned and there were little to be gained in continuing the outer iteration unless the maximum number of

PSQMR steps allowed was increased. The initial iterate for all the algorithms was taken to be $X^0 = (n/\sqrt{2})I,\ y^0 = 0,\ S^0 = \sqrt{n}I$.

The performance results of our algorithms on E1–E8 are given in Table 1. The columns corresponding to "`it`" give the number of interior-point iterations taken, whereas the columns "`sq`" give the average number of PSQMR steps taken to solve each of the two linear systems (9) during the computation of predictor and corrector directions at each interior-point iteration. The time reported is in the format of "hour:minute:second". Note that we did not run Algorithm A0 or A1 for some of the larger problems. Those entries with an "$*$" mean that the algorithms were terminated because the PSQMR solver hit the maximum number of steps allowed. Table 1 contains a variety of information that we extract and summarize below.

1. The number of interior-point iterations required by our proposed interior-point method grows very modestly with the problem dimension n. In all the test problems, the number of iterations required is less than 20.
2. Algorithm A0 took more than 35 minutes to solve E1-100,...,E8-100 (not shown in Table 1). This indicates that solving (9) via a direct solver is extremely expensive. For the problems E1-100, ..., E8-100, it is at least 100 times more expensive than using an iterative solver with an appropriate preconditioner, say Ψ_ρ^k.
3. Based on the stopping criterion (11) for the PSQMR method used to solve (9), Algorithms A2, A3 and A4 took about the same number of interior-point iterations to converge compared to Algorithm A0 that uses a direct method. This indicates that the inexact search directions are computed to sufficient accuracy, and thus the residual errors do not degrade the outer iterations.
4. Algorithm A1, which employs no preconditioner, is not able to solve most of the test problems to the required accuracy in ϕ, defined in (46). All our test problems have ill-conditioned augmented matrices $\mathcal{B}_\rho^k$ when the complementarity gaps $X^k \bullet S^k/n$ are small.
5. Algorithm A3, which employs $\mathcal{V}_\rho^k$ in (37) as the preconditioner, performed fairly well on the problems in E1, E3, E5, and E7, but less so for those in E2, E4, E6 and E8. Note that the optimal solutions for all the problems are dual degenerate. This violates one of the condition in Theorem 6(b), and so there is no theoretical basis to expect $\mathcal{V}_\rho^k$ to be an effective preconditioner.
6. The preconditioners $\Phi_{\rho,-}^k$ and Ψ_ρ^k are very effective for $\mathcal{B}_\rho^k$, as attested by the good performance of Algorithms A2 and A4. We note that for the QSDPs arising from (3), the optimal solutions are always primal non-degenerate, and so the good performance of Algorithm A2 that involves $\Phi_{\rho,-}^k$ is consistent with Theorem 5(b).
7. Comparing Algorithms A2 and A4, we see that the preconditioner Ψ_ρ^k is more effective (in terms of the number of PSQMR steps) than $\Phi_{\rho,-}^k$ for the test problems in E1, E3, E5, and E7. For the problems E1, E3, E5, E7, with $n = 1600$, the average numbers of PSQMR steps required to solve (9) for Algorithm A4 is between 37% to 75% of the corresponding numbers for Algorithm A2. However, because evaluating $(\Psi_\rho^k)^{-1}[X; y]$ requires two times the number of $n \times n$ matrix-matrix multiplications for

$(\Phi^k_{\rho,-})^{-1}[X;y]$, the savings in computation time are not as impressive as the reductions in the number of PSQMR steps.
Notice that $\Phi^k_{\rho,-}$ is a more effective preconditioner than Ψ^k_ρ for the problems in E2, E4, E6, and E8. For these test problems, the choices $\mathcal{H}^k_1$, $\mathcal{H}^k_2$ and $\mathcal{H}^k_3$ in (36) used for $\Phi^k_{\rho,-}$ is more effective than the choices in (42) constructed from Procedure SKPA for Ψ^k_ρ.

8. We should emphasize that the dimension of the system in (9) is $m+n(n+1)/2$, and it is **dense**. Thus for $n = 1600$, the dimension is more than 1.28 millions. It is quite surprising that an iterative solver with appropriate preconditioning can solve such a large and ill-conditioned linear system in less than a hundred iterative steps. Because the preconditioning steps can be evaluated at moderate costs, the savings in the iterative steps enabled the test problems, E1-1600,...,E8-1600, to be solved in a few hours on a desktop computer.

To further evaluate the performance of our algorithms, we consider the following classes of QSDPs arising from the EDM problem (4):

EDM1. QSDPs arising from the EDM problem (4) with the dissimilarity matrix $B \in \mathcal{S}^n$ generated as follows. First we generate n random points, $x_1, \ldots, x_n$, in the unit cube centered at the origin in $\mathbb{R}^3$. Then we set $B_{ij} = \|x_i - x_j\|$ if the distance is less than a certain cut-off distance R; otherwise, set $B_{ij} = 0$. The non-negative weight matrix H is chosen to be the 0-1 matrix having the same sparsity pattern as B. The set of indices where the distances are fixed is given by $\mathcal{E} = \{(1,j) : B_{1j} \neq 0, j = 1, \ldots, n\}$. Note that the operators $\mathcal{Q}$ for these QSDPs are positive semidefinite, but not positive definite. We generated 4 test problems with $n = 100, 200, 400, 800$, and their corresponding dissimilarity matrices B have 16.6%, 8.6%, 4.5%, and 2.4% nonzero elements, respectively. Note that in the actual QSDP test problems, we added the term $-0.01I \bullet X$ to the objective function in (4) so as to induce a low-rank primal optimal solution.

EDM2. Same as EDM1 but the points are chosen to be coordinates of the atoms in the following protein molecules, `1PTQ`, `1HOE`, `1LFB`, `1PHT`, `1POA`, `1AX8`, taken from the Protein Data Bank [5]. These 6 test problems have dimension $n = 401, 557, 640, 813, 913, 1002$, with 17.4%, 13.1%, 11.0%, 8.3%, 10.9%, 7.4% nonzeros in B, respectively.

The performance results of our algorithms on EDM1 and EDM2 are given in Table 2. Basically, the observations we have in Items 1 to 4 for Table 1 are applicable also to Table 2. For the QSDPs in EDM1 and EDM2, Algorithm A4 is the most efficient, followed by A3, and then A2. It appears that for a QSDP where $\mathcal{Q}$ is not positive definite, the preconditioner Ψ^k_ρ is a better choice over $\Phi^k_{\rho,-}$.

7 Conclusion

We proposed an inexact primal-dual path-following Mehrotra-type predictor-corrector method for solving convex quadratic SDP problems. We suggested

computing the search direction at each iteration from the augmented equation by an iterative solver with preconditioners properly designed to overcome the ill-conditioning of the augmented matrix. The preconditioners are shown to have favorable asymptotic spectral distributions for the preconditioned systems to achieve fast convergence under suitable nondegeneracy assumptions. For one of the classes of preconditioners, no nondegeneracy assumption is needed.

Numerical experiments on a variety of convex QSDPs with matrices of dimension up to 1600 showed that our inexact interior-point method is quite efficient and robust. For the test problems considered in this paper, our inexact interior-point method can solve each QSDPs at a cost that is at most $\Theta(mn^3) + \Theta(m^2n^2) + \Theta(m^3)$. However, the computational complexity at each iteration of our interior-point method is inherently higher than first-order methods such as the nonlinear-programming based method of Burer, Monteiro, and Zhang [10], so it is of interest to ask whether such methods can be extended to QSDPs. It is also of interest to investigate whether variants of augmented Lagrangian dual methods can be applied to QSDPs and whether such methods are competitive to our inexact interior-point methods. The performance of our method here may serve as the benchmark for evaluating other (possibly more efficient) algorithms in the future.

While extensive testing on large QSDPs coming from practical problems remain to be done, we have provided the essential computational framework for such exploration to be possible.

References

1. Alizadeh,F., Haeberly,J.-P.A., Overton,M.L.: *Primal-dual interior-point methods for semidefinite programming: convergence results, stability and numerical results*. SIAM J. Optimization **8**, 746–768 (1998).
2. Alizadeh,F., Haeberly,J.-P.A., Overton,M.L.: *Complementarity and nondegeneracy in semidefinite programming*. Math. Program. **77**, 111–128 (1997).
3. Alfakih,A.Y., Khandani,A., Wolkowicz,H.: *Solving Euclidean distance matrix completion problems via semidefinite programming*. Comp. Optim. Appl. **12**, 13–30 (1999).
4. Anjos,M.F., Higham,N.J., Takouda,P.L., Wolkowicz,H.: *A semidefinite programming approach for the nearest correlation matrix problem*. Research Report, Department of Combinatorics and Optimization, U. of Waterloo, 2003.
5. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N., Bourne,P.E.: *The protein data bank*, Nucleic Acids Research **28**, 235–242 (2000).
6. Bhatia,R., Kittaneh,F.: *Norm inequalities for partitioned operators and an application*. Math. Ann. **287**, 719–726 (1990).
7. Benzi,M., Golub, G.H., Liesen,J.: *Numerical solution of saddle point problems*. Acta Numerica, 1–137 (2005).
8. Benzi,M., Simoncini,V.: *On the eigenvalues of a class of saddle point matrices*. Numerische Mathematik **103**, 173–196 (2006).
9. Bergamaschi,L., Gondzio,J., Zilli,G.: *Preconditioning indefinite systems in interior point methods for optimization*. Comp. Optim. Appl. **28**, 149-171 (2004).
10. Burer,S., Monteiro,R.D.C., Zhang,Y.: *Solving a class of semidefinite programs via nonlinear programming*. Math. Program. **93**, 97–122 (2002).
11. Freund,R.W., Nachtigal,N.M.: *A new Krylov-subspace method for symmetric indefinite linear systems*. Proceedings of the 14th IMACS World Congress on Computational and Applied Mathematics, Atlanta, USA, W.F. Ames ed., 1253–1256 (1994).

12. Fujisawa,K., Kojima,M., Nakata,K.: *Exploiting sparsity in primal-dual interior-point method for semidefinite programming*. Math. Program. **79**, 235–253 (1997).
13. Halicka,M., de Klerk,E., Roos,C.: *Limiting behaviour of the central path in semidefinite optimization*. Optimization Methods and Software **20**, 99–113 (2005).
14. Higham,N.J.: *Computing the nearest correlation matrix — a problem from finance*. IMA J. Numerical Analysis **22**, 329–343 (2002).
15. Horn,R., Johnson,C: *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1998.
16. Keller,C., Gould,N.I.M., Wathen,A.J.: *Constraint preconditioning for indefinite linear systems*. SIAM J. Matrix Anal. Appl. **21**, 1300–1317 (2000).
17. Kojima,M., Shindoh,S., Hara,S.: *Interior-point methods for the monotone linear complementarity problem in symmetric matrices*. SIAM J. Optimization **7**, 86–125 (1997).
18. Krislock,N., Lang,J., Varah,J., Pai,D.K., Seidel, H.-P.: *Local compliance estimation via positive semidefinite constrained least sqaures*. IEEE Trans. on Robotics **20**, 1007–1011 (2004).
19. Langville,A.N., Stewart,W.J.: *A Kronecker product approximate preconditioner for SANs*. Numer. Linear Algebra with Appl. **11**, 723–752 (2004).
20. Luo,Z.-Q., Sturm,J.F., Zhang,S.: *Superlinear convergence of a symmetric primal-dual path following algorithm for semidefinite programming*. SIAM J. Optimization **8**, 59–81 (1998).
21. Malick,J.: *A dual approach to semidefinite least-squares problems*. SIAM J. Matrix Anal. Appl. **26**, 272–284 (2004).
22. Monteiro,R.D.C., Zanjácomo,P.R.: *Implementation of primal-dual methods for semidefinite programming based on Monteiro and Tsuchiya Newton directions and their variants*. Optimization Methods and Software **11/12**, 91–140 (1999).
23. Nie,J.W., Yuan,Y.X.: *A predictor-corrector algorithm for QSDP combining Dikin-type and Newton centering steps*. Ann. Oper. Res. **103**, 115–133 (2001).
24. Phoon,K.K., Toh,K.C., Chan,S.H., Lee,F.H.: *An efficient diagonal preconditioner for finite element solution of Biot's consolidation equations*. Int. J. Numerical Methods in Engineering **55**, 377–400 (2002).
25. Qi,H., Sun,D.: *A quadratically convergent Newton method for computing the nearest correlation matrix*. SIAM J. Matrix Anal. Appl. **28**, 360–385 (2006).
26. Saad,Y.: *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company, Boston, 1996.
27. Todd,M.J., Toh,K.C., Tütüncü,R.H.: *On the Nesterov-Todd direction in semidefinite programming*. SIAM J. Optimization **8**, 769–796 (1998).
28. Toh,K.C., Phoon,K.K., Chan,S.H.: *Block preconditioners for symmetric indefinite linear systems*. Int. J. Numerical Methods in Engineering **60**, 1361–1381 (2004).
29. Toh,K.C., Tütüncü,R.H., Todd,M.J.: *Inexact primal-dual path-following algorithms for a special class of convex quadratic SDP and related problems*. Pacific J. Optimization, to appear.
30. Van Loan,C.F.: *The ubiquitous Kronecker product*. J. Computational and Applied Mathematics **123**, 85-100 (2000).
31. Wathen,A.J., Fischer,B., Silvester,D.J.: *The convergence of iterative solution methods for symmetric and indefinite linear systems*. In *Numerical Analysis 1997*, Eds. D.F. Griffiths and G.A. Watson, Addison Wesley Longman, Harlow, England, 230-243 (1997).
32. Zhou,G.L., Toh,K.C.: *Polynomiality of an inexact infeasible interior point algorithm for semidefinite programming*. Math. Program. **99**, 261–282 (2004).

Table 1: Performance of the Algorithms A1–A4 on the problem sets E1–E8. An entry with an "$*$" means that the algorithm was terminated before the accuracy of $\phi \leq 10^{-7}$ was achieved. For these problems, ϕ ranges from the level of 10^{-6} for E1 to the level of 10^{-3} for E8.

	A1 (I)			A2 ($\mathcal{V}_\rho$, $\Phi_{\rho,-}$)			A3 ($\mathcal{V}_\rho$)			A4 ($\mathcal{V}_\rho$, Ψ_ρ)		
n	it	time	sq	it	time	sq	it	time	sq	it	time	sq
E1 100	11*	0:15	167	11	0:05	23	12	0:08	28	12	0:06	15
400	13*	8:16	167	13	3:11	24	13	2:50	20	13	2:39	13
800	13*	36:40	122	13	18:48	22	13	14:23	15	13	13:29	10
1600				14	2:53:43	28	14	1:58:00	17	13	1:38:13	11
E2 100	7*	0:11	207	9	0:03	11	9	0:08	43	9	0:03	8
400	8*	6:39	221	11	1:44	14	11	7:58	78	11	3:24	22
800	8*	33:11	185	11	12:03	15	12	1:12:12	105	11	30:03	30
1600				12	1:43:36	18	12	7:05:16	90	12	4:42:59	38
E3 100	8*	0:13	206	9	0:05	27	9	0:05	24	10	0:06	19
400	9*	6:21	186	10	3:27	36	10	2:05	19	10	2:34	18
800	10*	40:19	180	11	28:55	44	11	14:03	18	11	21:38	21
1600				11	3:06:59	41	11	1:22:35	14	11	2:21:15	20
E4 100	7*	0:17	309	10	0:04	15	10	0:13	64	9	0:03	11
400	7*	7:21	283	10	2:01	19	10	7:28	80	10	2:47	19
800	7*	36:58	239	11	15:11	21	10	42:59	73	11	42:50	45
1600				11	1:56:01	23	11	7:18:52	102	11	4:09:31	37
E5 100	9*	0:14	195	11	0:07	30	11	0:12	49	11	0:06	18
400	11*	7:02	167	12	3:16	27	13	4:40	35	13	3:53	20
800	12*	53:52	201	13	24:52	30	13	22:58	26	13	19:11	15
1600				13	2:53:12	31	14	3:05:14	30	13	2:01:09	13
E6 100	7*	0:14	247	10	0:04	15	10	0:13	61	10	0:04	10
400	8*	7:35	253	11	1:54	16	11	8:23	81	11	4:26	28
800	8*	35:05	196	11	17:01	24	12	1:15:02	108	12	43:58	40
1600				12	1:54:35	20	12	7:11:38	91	12	4:59:57	40
E7 100	7*	0:18	326	9	0:05	28	9	0:10	53	9	0:05	17
400	8*	7:30	251	10	3:31	37	10	3:37	36	10	2:47	19
800	9*	50:48	256	11	29:49	46	11	25:43	37	11	22:37	22
1600				11	3:29:34	47	11	2:07:16	25	11	2:21:26	20
E8 100	6*	0:12	251	10	0:04	19	10	0:14	70	10	0:04	12
400	7*	8:00	308	10	2:08	20	10	7:07	75	10	4:01	29
800	7*	40:26	262	10	19:30	31	11	1:08:58	109	11	43:15	45
1600				11	2:23:34	30	10	3:53:29	57	11	5:10:25	46

Table 2: Performance of the Algorithms A1–A4 on the problem sets EDM1 and EDM2. An entry with an "$*$" means that the algorithm was terminated before the accuracy of $\phi \leq 10^{-7}$ was achieved. For these problems, ϕ is at the level of 10^{-1}.

	A1 (I)			A2 ($\mathcal{V}_\rho$, $\Phi_{\rho,-}$)			A3 ($\mathcal{V}_\rho$)			A4 ($\mathcal{V}_\rho$, Ψ_ρ)		
n	it	time	sq	it	time	sq	it	time	sq	it	time	sq
99	7*	0:23	331	16	0:22	63	16	0:17	44	16	0:15	32
199	8*	2:50	374	18	1:29	41	18	1:46	48	18	1:16	27
399	6*	8:40	278	19	9:46	46	19	9:45	44	19	6:56	23
799	7*	1:09:35	349	22	1:03:54	42	21	1:05:10	44	21	53:53	26
401	8*	17:37	421	21	41:14	188	21	11:13	46	21	10:50	34
557	8*	37:26	400	20	47:14	95	20	26:31	49	20	21:03	29
640	7*	32:00	281	22	1:11:21	93	22	47:34	58	22	32:44	30
813	8*	1:12:51	306	21	5:06:31	225	22	2:40:06	106	22	1:38:35	46
913				22	4:38:10	141	22	1:59:40	56	22	1:23:49	28
1002				23	3:41:58	82	23	3:07:46	67	23	1:53:04	28