SIAM J. OPTIM.
Vol. 8, No. 3, pp. 746–768, August 1998

007

PRIMAL-DUAL INTERIOR-POINT METHODS FOR SEMIDEFINITE PROGRAMMING: CONVERGENCE RATES, STABILITY AND NUMERICAL RESULTS*

FARID ALIZADEH†, JEAN-PIERRE A. HAEBERLY‡, AND MICHAEL L. OVERTON§

Abstract. Primal-dual interior-point path-following methods for semidefinite programming are considered. Several variants are discussed, based on Newton's method applied to three equations: primal feasibility, dual feasibility, and some form of centering condition. The focus is on three such algorithms, called the XZ, XZ+ZX, and Q methods. For the XZ+ZX and Q algorithms, the Newton system is well defined and its Jacobian is nonsingular at the solution, under nondegeneracy assumptions. The associated Schur complement matrix has an unbounded condition number on the central path under the nondegeneracy assumptions and an additional rank assumption. Practical aspects are discussed, including Mehrotra predictor-corrector variants and issues of numerical stability. Compared to the other methods considered, the XZ+ZX method is more robust with respect to its ability to step close to the boundary, converges more rapidly, and achieves higher accuracy.

Key words. semidefinite programming, eigenvalue optimization, interior-point method, convex programming

AMS subject classifications. 65K10, 90C25

PII. S1052623496304700

1. Introduction.

Let $\mathcal{S}^n$ denote the vector space of real symmetric $n \times n$ matrices. Denote the dimension of this space by

$$n^{\overline{2}} = \frac{n(n+1)}{2}. \tag{1.1}$$

The standard inner product on $\mathcal{S}^n$ is

$$A \bullet B = \operatorname{tr} AB = \sum_{i,j} A_{ij} B_{ij}.$$

By $X \succeq 0$ ($X \succ 0$), where $X \in \mathcal{S}^n$, we mean that X is positive semidefinite (positive definite).

Consider the semidefinite program (SDP)

$$\begin{array}{ll} \min\limits_{X \in \mathcal{S}^n} & C \bullet X \\ \text{s.t.} & A_k \bullet X = b_k, \ k = 1, \ldots, m, \\ & X \succeq 0, \end{array} \tag{1.2}$$

*Received by the editors June 4, 1996; accepted for publication (in revised form) May 27, 1997; published electronically July 20, 1998.
http://www.siam.org/journals/siopt/8-3/30470.html

†RUTCOR, Rutgers University, New Brunswick, NJ 08854 (alizadeh@rutcor.rutgers.edu). The research of this author was supported in part by National Science Foundation grant CCR-9501941 and Office of Naval Research contract N00014-96-1-0704.

‡Mathematics Department, Fordham University, Bronx, NY 10458 (haeberly@murray.fordham.edu). The research of this author was supported in part by National Science Foundation grant CCR-9401119.

§Computer Science Department, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012 (overton@cs.nyu.edu). The research of this author was supported in part by National Science Foundation grant CCR-9625955.

where $b \in \mathbb{R}^m$, $C \in \mathcal{S}^n$, and $A_k \in \mathcal{S}^n$, $k = 1, \ldots, m$. The dual SDP is

$$\begin{array}{ll} \max\limits_{y \in \mathbb{R}^m, Z \in \mathcal{S}^n} & b^T y \\ \text{s.t.} & \sum_{k=1}^m y_k A_k + Z = C, \\ & Z \succeq 0. \end{array} \tag{1.3}$$

The following assumptions hold throughout the paper.

Assumption 1. There exists a primal feasible point $X \succ 0$, and a dual feasible point (y, Z) with $Z \succ 0$.

Assumption 2. The matrices A_k, $k = 1, \ldots, m$, are linearly independent; i.e., they span an m-dimensional linear space in $\mathcal{S}^n$.

The *central path* consists of points $(X^\mu, y^\mu, Z^\mu) \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n$ satisfying the primal and dual feasibility constraints as well as the centering condition

$$X^\mu Z^\mu = \mu I \tag{1.4}$$

for some $\mu \in \mathbb{R}$, $\mu > 0$. It is well known [11, 7] that under Assumptions 1 and 2 (X^μ, y^μ, Z^μ) exists and is unique for all $\mu > 0$, and that

$$(X, y, Z) = \lim_{\mu \to 0} (X^\mu, y^\mu, Z^\mu) \tag{1.5}$$

exists and solves the primal and dual SDPs. Furthermore, because X^μ and Z^μ commute, there exists an orthogonal matrix Q^μ such that

$$X^\mu = Q^\mu \ \mathrm{Diag}(\lambda_1^\mu, \ldots, \lambda_n^\mu) \ (Q^\mu)^T, \quad Z^\mu = Q^\mu \ \mathrm{Diag}(\omega_1^\mu, \ldots, \omega_n^\mu) \ (Q^\mu)^T, \tag{1.6}$$

where the λ_i^μ and ω_i^μ, respectively the eigenvalues of X^μ and Z^μ, satisfy

$$\lambda_i^\mu \omega_i^\mu = \mu, \quad i = 1, \ldots, n. \tag{1.7}$$

Without loss of generality, assume that

$$\lambda_1^\mu \geq \cdots \geq \lambda_n^\mu \quad \text{and} \quad \omega_1^\mu \leq \cdots \leq \omega_n^\mu. \tag{1.8}$$

As $\mu \to 0$, the centering condition (1.4) reduces to the complementarity condition $XZ = 0$, implying that

$$X = Q \ \mathrm{Diag}(\lambda_1, \ldots, \lambda_n) \ Q^T, \quad Z = Q \ \mathrm{Diag}(\omega_1, \ldots, \omega_n) \ Q^T \tag{1.9}$$

for some orthogonal matrix Q, with the eigenvalue complementarity condition $\lambda_i \omega_i = 0, i = 1, \ldots, n$. Observe that λ_i and ω_i are the limits of λ_i^μ and ω_i^μ as $\mu \to 0$, and Q may be taken to be a limit point (not necessarily unique) of the set $\{Q^\mu : \mu > 0\}$. We have

$$\lambda_1 \geq \cdots \geq \lambda_n \quad \text{and} \quad \omega_1 \leq \cdots \leq \omega_n. \tag{1.10}$$

Interior-point methods for semidefinite programming were originally introduced by [11, 4]. Early papers on primal-dual methods include [17] and [6]. A preliminary version of the present work appeared as [2]. Convergence analysis of primal-dual path-following methods for SDP appeared first in [7, 13, 12]. We are primarily concerned with four methods, which we call the XZ, XZ+ZX, Nesterov–Todd (NT), and Q methods. The XZ method first appeared in [6, 7]. The XZ+ZX method was introduced in

[2] and was recently analyzed in [8, 9]. The NT method was given in [13, 12] and its implementation was recently discussed in [16]. The Q method originally appeared in [1]. Many other papers on semidefinite programming have recently been announced.

The paper is organized as follows. In section 2 we introduce several algorithms in a common framework based on Newton's method, focusing on the XZ and XZ+ZX variants. In section 3 we study the Jacobian of the Newton system for the various methods under nondegeneracy assumptions, and discuss implications for local convergence rates. In section 4 we consider the conditioning of the Schur complement matrix on the central path, again under nondegeneracy assumptions. This leads to the issue of numerical stability, discussed in section 5. We introduce the Q method in section 6. In section 7, we present computational results.

Our main focus is on the nondegenerate case; this assumption (defined in section 3) implies unique primal and dual solutions. We take the view that it is important to understand how methods behave on nondegenerate problems. This does not discount the significance of degenerate problems that may arise in applications, as is common in linear programming (LP).

In practice, many semidefinite programs are block diagonal. Everything in this paper extends easily to the block-diagonal case. Note that LP is the special case where all block sizes are one.

A word about notation: we use the symbols X, y, and Z to mean several things. Depending on the context, they may refer to the variables of the SDP, the iterates generated by a method, or a solution of the SDP.

2. The methods in a general framework. We consider only primal-dual interior-point path-following methods, generating a sequence of iterates approximating the central path and converging to the primal and dual solutions. See [18] for a detailed discussion of such methods for LP. In LP, the basic iterative step can be readily derived using Newton's method. For SDP, points on the central path satisfy the nonlinear equation

$$\begin{bmatrix} \sum_{k=1}^{m} y_k A_k + Z - C \\ A_1 \bullet X - b_1 \\ \vdots \\ A_m \bullet X - b_m \\ XZ - \mu I \end{bmatrix} = 0. \tag{2.1}$$

However, the matrix XZ is not symmetric in general. Consequently, the domain and range of the function defined by the left-hand side of (2.1) are not the same spaces, and Newton's method is not directly applicable. For LP, on the other hand, the standard primal-dual interior-point method is obtained by applying Newton's method to (2.1). In this case, X and Z are diagonal, and XZ is also diagonal, so the domain and range of (2.1) reduce to $\mathbb{R}^{2n+m}$.

A key question in formulating primal-dual interior-point methods for SDP is therefore, how should one appropriately formulate Newton's method? We consider here two possibilities. Other choices are discussed at the end of this section.

The XZ method. Use the centering condition (1.4) directly and view the left-hand side of (2.1) as a function whose domain and range are both $\mathcal{U} = \mathbb{R}^{n\times n} \times \mathbb{R}^m \times \mathbb{R}^{n\times n}$. Then Newton's method is well defined, though the iterates are not symmetric matrices. (Actually, only the X iterates are not symmetric, since the dual feasibility equation forces Z to be symmetric.) The

X iterates can then be explicitly symmetrized before continuing with the next iteration. Consequently, this method is not strictly a Newton method. A different iteration is obtained by using $ZX = \mu I$ instead of (1.4).

The $XZ + ZX$ method. Rewrite (1.4) in the symmetric form

$$XZ + ZX = 2\mu I. \tag{2.2}$$

Substituting (2.2) for (1.4) in (2.1) gives a mapping with domain and range both given by $\mathcal{V} = \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n$. Application of Newton's method to (2.2) leads to symmetric matrix iterates X and Z.

We observe that (1.4) and (2.2) are equivalent when $X \succeq 0$ (or $Z \succeq 0$). That (1.4) implies (2.2) is immediate. That the converse holds for $X \succeq 0$ is seen by using $X = Q\Lambda Q^T$ to reduce (2.2) to $\Lambda(Q^T Z Q) + (Q^T Z Q)\Lambda = 2\mu I$, with Λ diagonal and nonnegative and $Q^T Q = I$. The entries on the left-hand side are $(\lambda_i + \lambda_j)(QZQ^T)_{ij}$, and so, since the off-diagonal entries must be zero, either $\lambda_i = \lambda_j = 0$ or $(QZQ^T)_{ij} = 0$ when $i \neq j$. Thus, $\Lambda(Q^T Z Q)$ is diagonal, and (1.4) holds.

We now examine the steps defined by these methods in more detail. The Newton step for the XZ method satisfies the linear equation

$$X\,\Delta Z + \Delta X\, Z = \mu I - XZ. \tag{2.3}$$

Let nvec map $\mathbb{R}^{n\times n}$ to $\mathbb{R}^{n^2}$, stacking the columns of a matrix in a vector. Then we may rewrite (2.3) in the form

$$(I \otimes X)\,\mathrm{nvec}\,(\Delta Z) + (Z \otimes I)\,\mathrm{nvec}\,(\Delta X) = \mathrm{nvec}\,(\mu I - XZ), \tag{2.4}$$

where $\otimes$ denotes the standard Kronecker product (see the Appendix, equation (7.3)).

To discuss the XZ+ZX method, we introduce a *symmetric* version of the Kronecker product. The Newton correction for (2.2) satisfies the linear equation

$$X\,\Delta Z + \Delta Z\, X + \Delta X\, Z + Z\,\Delta X = 2\mu I - XZ - ZX, \tag{2.5}$$

where ΔX and ΔZ are symmetric. Let svec be an isometry identifying $\mathcal{S}^n$ with $\mathbb{R}^{\overline{n^2}}$, so that $K \bullet L = \mathrm{svec}\,(K)^T\,\mathrm{svec}\,(L)$ for all $K, L \in \mathcal{S}^n$ (see Appendix). Then (2.5) can be written as

$$(Z \circledast I)\,\mathrm{svec}\,(\Delta X) + (X \circledast I)\,\mathrm{svec}\,(\Delta Z) = \mathrm{svec}\left(\mu I - \frac{1}{2}(XZ + ZX)\right), \tag{2.6}$$

where $\circledast$ denotes the symmetric Kronecker product defined in the Appendix (see (7.6)).

We shall now describe both methods in a common framework. Let vec denote either nvec or svec, depending on the context. Specifically, vec will mean nvec in the case of the XZ method and svec otherwise. The inverse of vec is denoted by mat. We shall use lower case letters x and z to denote $\mathrm{vec}\,X$ and $\mathrm{vec}\,Z$ respectively, and we shall use Δx and Δz interchangeably with $\mathrm{vec}\,\Delta X$ and $\mathrm{vec}\,\Delta Z$, to be defined shortly.

Let

$$\mathbf{A} = \begin{bmatrix} (\,\mathrm{vec}\,A_1)^T \\ \vdots \\ (\,\mathrm{vec}\,A_m)^T \end{bmatrix} \tag{2.7}$$

and define

$$r_p = b - \mathbf{A}x, \qquad R_d = C - Z - \operatorname{mat} \mathbf{A}^T y,$$

and

$$R_c = \left\{ \begin{array}{ll} \mu I - XZ & XZ \text{ method} \\ \mu I - \frac{1}{2}(XZ + ZX) & XZ + ZX \text{ method} \end{array} \right\} \tag{2.8}$$

with

$$r_d = \operatorname{vec} R_d, \qquad r_c = \operatorname{vec} R_c.$$

Let

$$G(x, y, z) = \begin{bmatrix} -r_d \\ -r_p \\ -r_c \end{bmatrix}. \tag{2.9}$$

Note that G maps $\mathcal{U}$ to $\mathcal{U}$ in the case of the XZ method and $\mathcal{V}$ to $\mathcal{V}$ otherwise. Application of one step of Newton's method to $G(x, y, z) = 0$ gives the linear system

$$\begin{bmatrix} 0 & \mathbf{A}^T & \mathbf{I} \\ \mathbf{A} & 0 & 0 \\ \mathbf{E} & 0 & \mathbf{F} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} r_d \\ r_p \\ r_c \end{bmatrix}. \tag{2.10}$$

Here

$$\mathbf{E} = \left\{ \begin{array}{ll} Z \otimes I & XZ \text{ method} \\ Z \circledast I & XZ + ZX \text{ method} \end{array} \right\}$$

and

$$\mathbf{F} = \left\{ \begin{array}{ll} I \otimes X & XZ \text{ method} \\ X \circledast I & XZ + ZX \text{ method} \end{array} \right\}$$

and $\mathbf{I}$ is the identity matrix of appropriate dimension ($I \otimes I$ for the XZ method and $I \circledast I$ for the XZ+ZX method). We denote the Jacobian matrix on the left-hand side of (2.10) by $\mathbf{J}$.

Applying block Gauss elimination, (2.10) reduces to the system

$$\begin{bmatrix} -\mathbf{F}^{-1}\mathbf{E} & \mathbf{A}^T \\ \mathbf{A} & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} r_d - \mathbf{F}^{-1} r_c \\ r_p \end{bmatrix}. \tag{2.11}$$

A second step of block Gauss elimination gives

$$\mathbf{M}\Delta y = r_p + \mathbf{A}\mathbf{E}^{-1}(\mathbf{F} r_d - r_c), \tag{2.12}$$

$$\Delta x = -\mathbf{E}^{-1}(\mathbf{F}(r_d - \mathbf{A}^T \Delta y) - r_c), \tag{2.13}$$

and, from the dual feasibility equation,

$$\Delta z = r_d - \mathbf{A}^T \Delta y, \tag{2.14}$$

where

$$\mathbf{M} = \mathbf{A}\mathbf{E}^{-1}\mathbf{F}\mathbf{A}^T. \tag{2.15}$$

We call $\mathbf{M}$ the *Schur complement* matrix. The main computational work is the formation and factorization of $\mathbf{M}$. The kth column of the matrix $\mathbf{E}^{-1}\mathbf{F}\mathbf{A}^T$ is

$$\left\{ \begin{array}{ll} \text{nvec}\,(XA_kZ^{-1}) & XZ \text{ method} \\ \text{svec}\,(G_k) & XZ + ZX \text{ method} \end{array} \right\},$$

where G_k is the solution of the Lyapunov equation (see Appendix)

$$ZG_k + G_kZ = XA_k + A_kX. \tag{2.16}$$

Formation of $\mathbf{M}$ thus requires $O(mn^3+m^2n^2)$ work, involving a Cholesky factorization of Z, in the case of the XZ method, and an eigenvalue factorization of Z, in the case of the XZ+ZX method (see Appendix). Neglecting sparsity considerations, the additional cost of the eigenvalue factorization is negligible in comparison to the other operations required to form $\mathbf{M}$.

It is clear that, as long as $X \succ 0$ and $Z \succ 0$, nonsingularity of the Jacobian matrix $\mathbf{J}$ is equivalent to nonsingularity of the Schur complement $\mathbf{M}$. In the case of the XZ method, $\mathbf{M}$ is symmetric and positive definite. In the case of the XZ+ZX method, $\mathbf{M}$ is not symmetric, but can be shown to be nonsingular if $XZ + ZX \succ 0$ [15]. An alternative condition guaranteeing the nonsingularity of $\mathbf{M}$ is given in [10]. Equation (2.12) is solved by using a Cholesky factorization of $\mathbf{M}$ in the case of the XZ method and an LU factorization of $\mathbf{M}$ in the case of the XZ+ZX method.

For the XZ+ZX method, the multiplications by $\mathbf{E}^{-1}$ in (2.12) and (2.13) require the solution of Lyapunov equations, using the eigenvalues of Z already computed to form $\mathbf{M}$.

Both methods are then described by the following iteration.

BASIC ITERATION.

1. Choose $0 \le \sigma < 1$ and define

$$\mu = \sigma \frac{X \bullet Z}{n}. \tag{2.17}$$

2. Determine ΔX, Δy, ΔZ from (2.10), equivalently (2.12)–(2.14).
3. In the case of the XZ method, replace ΔX by $\frac{1}{2}(\Delta X + \Delta X^T)$.
4. Choose steplengths α, β and update the iterates by

$$\begin{array}{rcl} X & \leftarrow & X + \alpha\, \Delta X, \\ y & \leftarrow & y + \beta\, \Delta y, \\ Z & \leftarrow & Z + \beta\, \Delta Z. \end{array}$$

Rules for defining σ will be discussed later. A simple steplength rule is given by choosing a parameter τ, $0 < \tau < 1$, and defining

$$\alpha = \min(1, \tau\,\hat{\alpha}), \qquad \hat{\alpha} = \sup\{\bar{\alpha} : X + \bar{\alpha}\,\Delta X \succeq 0\} \tag{2.18}$$

and

$$\beta = \min(1, \tau\,\hat{\beta}), \qquad \hat{\beta} = \sup\{\bar{\beta} : Z + \bar{\beta}\,\Delta Z \succeq 0\}. \tag{2.19}$$

Note that, except in the case $\Delta X \succeq 0$, we have $0 < \hat{\alpha} < \infty$ with

$$\hat{\alpha}^{-1} = \lambda_{\max}(-L^{-1}\Delta X L^{-T}),$$

where $\lambda_{\max}$ means the largest eigenvalue and L is the Cholesky factor of X, i.e., $X = LL^T$.

Other methods can also be defined in the same framework; two of these are discussed below. See [19] for a class of methods that includes the XZ+ZX method, and [7, 15] for another class that includes all those discussed here except the XZ+ZX method.

The X^{-1} method. Replace R_c in (2.9) by $R_c = \mu X^{-1} - Z$, so $\mathbf{E} = \mu X^{-1} \circledast X^{-1}$, $\mathbf{F} = I \circledast I$. A similar method can be defined with $R_c = \mu Z^{-1} - X$. In fact, the method given by [17] is based on a combination of these two steps.

The NT method. Use $R_c = \mu X^{-1} - Z$, $\mathbf{E} = W^{-1} \circledast W^{-1}$, $F = I \circledast I$, where $W = X^{1/2}(X^{1/2}ZX^{1/2})^{-1/2}X^{1/2}$. This form does not actually arise from applying Newton's method to (2.9). However, see [16] for a Newton interpretation of this method.

As long as $X \succ 0$ and $Z \succ 0$, $\mathbf{E}^{-1}\mathbf{F}$ is symmetric and positive definite for both these methods. However, in both cases, the function to which Newton's method is applied fails to exist at a solution. We call an algorithm a Newton method if $(\Delta X, \Delta y, \Delta Z)$ is derived by applying Newton's method to a function that is well defined for all $X \succeq 0$, $Z \succeq 0$. Under this definition, of the four variants defined so far, only XZ+ZX is a Newton method for SDP.

In the special case of LP (i.e., a block-diagonal SDP with block sizes all one), the XZ, XZ+ZX, and NT methods coincide, giving the XZ method for LP, which is a Newton method.

In order to understand the asymptotic behavior of Newton's method, it is important to analyze the Jacobian at the solution itself. This is done in the next section.

3. The Jacobian at the solution. In this section we study the Jacobian of the function G, appearing on the left-hand side of (2.10), under nondegeneracy assumptions. To do this, we use the notions of nondegeneracy that were introduced by the authors in [3].

DEFINITION 1. *Let (X, y, Z) solve* (1.2), (1.3) *with an orthogonal matrix Q satisfying* (1.9). *Let X have rank r, with positive eigenvalues $\lambda_1, \ldots, \lambda_r$, and partition $Q = [Q_1\ Q_2]$, where the columns of Q_1 are eigenvectors corresponding to $\lambda_1, \ldots, \lambda_r$. We say that (X, y, Z) satisfies the* strict complementarity *and* primal and dual nondegeneracy *conditions if the following hold:*

1. $\operatorname{rank}(Z) = n - r$,
2. *the matrices*

$$\begin{bmatrix} Q_1^T A_k Q_1 & Q_1^T A_k Q_2 \\ Q_2^T A_k Q_1 & 0 \end{bmatrix} \quad \text{for } k = 1, 2, \ldots, m \tag{3.1}$$

are linearly independent in $\mathcal{S}^n$, and

3. *the matrices*

$$Q_1^T A_k Q_1 \quad \text{for } k = 1, 2, \ldots, m \tag{3.2}$$

span the space $\mathcal{S}^r$.

These conditions are well defined even if Q is not unique. The first requirement is the strict complementarity condition. Conditions (3.1), (3.2) are respectively primal

and dual nondegeneracy conditions under the assumption of strict complementarity. They immediately imply the inequalities

$$r^{\overline{2}} \leq m \leq r^{\overline{2}} + r(n-r) \tag{3.3}$$

(recalling the notation (1.1)). They also imply uniqueness of the primal and dual solutions. Furthermore, the conditions are generic properties of SDPs, meaning roughly that they hold with probability one for an optimal solution triple, given random data with feasible solutions. For motivation of these conditions and further details, see [3]. The definitions are easily extended to the block-diagonal case, giving the usual LP nondegeneracy conditions when all blocks have size one.

The strict complementarity condition $\operatorname{rank}(X) = r$, $\operatorname{rank}(Z) = n - r$ implies, using (1.10), that

$$\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0 \tag{3.4}$$

and

$$0 = \omega_1 = \cdots = \omega_r < \omega_{r+1} \leq \cdots \leq \omega_n. \tag{3.5}$$

Let $B_k = Q^T A_k Q$. From (7.6), we have

$$\operatorname{svec} B_k = (Q^T \circledast Q^T) \operatorname{svec} A_k.$$

Recall the definition (2.7), and define

$$\mathbf{B} = \begin{bmatrix} (\operatorname{svec} B_1)^T \\ \vdots \\ (\operatorname{svec} B_m)^T \end{bmatrix}, \tag{3.6}$$

so that

$$\mathbf{A}(Q \circledast Q) = \mathbf{B}.$$

Each column of $\mathbf{B}$ corresponds to an index pair (i, j), identifying two columns of Q, with $1 \leq i \leq j \leq n$. By choosing the ordering used by the svec operator appropriately, we may write

$$\mathbf{B} = [\mathbf{C}_1 \ \mathbf{C}_2 \ \mathbf{C}_3], \tag{3.7}$$

where $\mathbf{C}_1$ contains $r^{\overline{2}}$ columns corresponding to $1 \leq i \leq j \leq r$, $\mathbf{C}_2$ contains $r(n-r)$ columns corresponding to $1 \leq i \leq r, \quad r+1 \leq j \leq n$, and $\mathbf{C}_3$ consists of $(n-r)^{\overline{2}}$ columns corresponding to $r+1 \leq i \leq j \leq n$. The primal nondegeneracy condition (3.1) holds exactly when the rows of $[\mathbf{C}_1 \ \mathbf{C}_2]$ are linearly independent, i.e., $[\mathbf{C}_1 \ \mathbf{C}_2]$ has rank m. The dual nondegeneracy condition (3.2) holds exactly when $\mathbf{C}_1$ has rank $r^{\overline{2}}$, i.e., the columns of $\mathbf{C}_1$ are linearly independent. Thus, the conditions (3.1) and (3.2) together imply that it is possible to choose $m - r^{\overline{2}}$ columns from $\mathbf{C}_2$ so that, together with *all* the columns of $\mathbf{C}_1$, they form a nonsingular $m \times m$ matrix. In other words, we can choose an ordering for the columns of $\mathbf{C}_2$, and therefore of $\mathbf{B}$, so that

$$\mathbf{B} = [\mathbf{B}_1 \ \mathbf{B}_2], \tag{3.8}$$

where $\mathbf{B}_1 \in \mathbb{R}^{m \times m}$ is nonsingular.

THEOREM 3.1. *Consider an SDP whose solution (X, y, Z) satisfies the strict complementarity and primal and dual nondegeneracy conditions. Let* $\mathbf{J}$ *be the Jacobian of the function G defining the XZ+ZX method, evaluated at (X, y, Z). Then* $\mathbf{J}$ *is nonsingular.*

Proof. We have

$$\mathbf{J} = \begin{bmatrix} 0 & \mathbf{A}^T & \mathbf{I} \\ \mathbf{A} & 0 & 0 \\ \mathbf{E} & 0 & \mathbf{F} \end{bmatrix},$$

where $\mathbf{E} = Z \circledast I$ and $\mathbf{F} = X \circledast I$. Let $\mathbf{P} = Q \circledast Q$, and let $\mathbf{S} = \mathrm{Diag}(\mathbf{P}, \mathbf{I}, \mathbf{P})$, so that

$$\mathbf{S}^T\mathbf{J}\mathbf{S} = \begin{bmatrix} 0 & \mathbf{B}^T & \mathbf{I} \\ \mathbf{B} & 0 & 0 \\ \boldsymbol{\Upsilon} & 0 & \boldsymbol{\Phi} \end{bmatrix},$$

with $\boldsymbol{\Phi} = \mathbf{P}^T\mathbf{F}\mathbf{P}$ and $\boldsymbol{\Upsilon} = \mathbf{P}^T\mathbf{E}\mathbf{P}$. Using Lemma 7.2 (see Appendix) and (1.9), we see that $\mathbf{P}^T\mathbf{P} = \mathbf{I}$, and $\boldsymbol{\Phi}$ and $\boldsymbol{\Upsilon}$ are diagonal with entries $\frac{1}{2}(\lambda_i + \lambda_j)$ and $\frac{1}{2}(\omega_i + \omega_j)$, $1 \le i \le j \le n$, respectively. Notice that the diagonal entry of $\boldsymbol{\Phi}$ corresponding to the index pair (i, j) is zero if and only if $r + 1 \le i \le j \le n$ (because of (3.4)), while the diagonal entry of $\boldsymbol{\Upsilon}$ corresponding to the pair (i, j) is zero if and only if $1 \le i \le j \le r$ (see (3.5)).

Using the partitioning of $\mathbf{B}$ in (3.8), we have

$$\mathbf{S}^T\mathbf{J}\mathbf{S} = \begin{bmatrix} 0 & 0 & \mathbf{B}_1^T & \mathbf{I} & 0 \\ 0 & 0 & \mathbf{B}_2^T & 0 & \mathbf{I} \\ \mathbf{B}_1 & \mathbf{B}_2 & 0 & 0 & 0 \\ \boldsymbol{\Upsilon}_1 & 0 & 0 & \boldsymbol{\Phi}_1 & 0 \\ 0 & \boldsymbol{\Upsilon}_2 & 0 & 0 & \boldsymbol{\Phi}_2 \end{bmatrix}, \tag{3.9}$$

where $\boldsymbol{\Upsilon} = \mathrm{Diag}(\boldsymbol{\Upsilon}_1, \boldsymbol{\Upsilon}_2)$ and $\boldsymbol{\Phi} = \mathrm{Diag}(\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2)$. We have $\boldsymbol{\Phi}_1 \succ 0$, since none of the columns of $\mathbf{C}_3$ are included in $\mathbf{B}_1$, and $\boldsymbol{\Upsilon}_2 \succ 0$, since all of the columns of $\mathbf{C}_1$ are included in $\mathbf{B}_1$.

Interchanging the first and third rows and the second and last columns of (3.9), we obtain

$$\begin{bmatrix} \mathbf{B}_1 & 0 & 0 & 0 & \mathbf{B}_2 \\ 0 & \mathbf{I} & \mathbf{B}_2^T & 0 & 0 \\ 0 & 0 & \mathbf{B}_1^T & \mathbf{I} & 0 \\ \boldsymbol{\Upsilon}_1 & 0 & 0 & \boldsymbol{\Phi}_1 & 0 \\ 0 & \boldsymbol{\Phi}_2 & 0 & 0 & \boldsymbol{\Upsilon}_2 \end{bmatrix}.$$

We shall demonstrate the nonsingularity of this matrix using block Gauss elimination. First, subtract $\boldsymbol{\Upsilon}_1\mathbf{B}_1^{-1}$ times the first block row from the fourth block row to eliminate $\boldsymbol{\Upsilon}_1$ from the (4,1) position. This does not otherwise change the lower triangle or the diagonal blocks, but only introduces $-\boldsymbol{\Upsilon}_1\mathbf{B}_1^{-1}\mathbf{B}_2$ into the (4,5) position. Second, subtract $\boldsymbol{\Phi}_2$ times the second block row from the fifth row, eliminating $\boldsymbol{\Phi}_2$ from the (5,2) position; this introduces $-\boldsymbol{\Phi}_2\mathbf{B}_2^T$ into the (5,3) position. This (5,3) block is then eliminated by adding $\boldsymbol{\Phi}_2\mathbf{B}_2^T\mathbf{B}_1^{-T}$ times the third row to the fifth row, introducing

$\mathbf{\Phi}_2\mathbf{B}_2^T\mathbf{B}_1^{-T}$ into the (5,4) position, giving

$$\begin{bmatrix} \mathbf{B}_1 & 0 & 0 & 0 & \mathbf{B}_2 \\ 0 & \mathbf{I} & \mathbf{B}_2^T & 0 & 0 \\ 0 & 0 & \mathbf{B}_1^T & \mathbf{I} & 0 \\ 0 & 0 & 0 & \mathbf{\Phi}_1 & -\mathbf{\Upsilon}_1\mathbf{H} \\ 0 & 0 & 0 & \mathbf{\Phi}_2\mathbf{H}^T & \mathbf{\Upsilon}_2 \end{bmatrix},$$

where $\mathbf{H} = \mathbf{B}_1^{-1}\mathbf{B}_2$. In order to show that this matrix is nonsingular we need only show that the trailing 2×2 block is nonsingular, or equivalently that its positive row scaling

$$\begin{bmatrix} \mathbf{I} & -\mathbf{\Phi}_1^{-1}\mathbf{\Upsilon}_1\mathbf{H} \\ \mathbf{\Upsilon}_2^{-1}\mathbf{\Phi}_2\mathbf{H}^T & \mathbf{I} \end{bmatrix}$$

is nonsingular. A final step of block Gauss elimination yields a block upper triangular matrix with last diagonal block given by

$$\mathbf{I} + \mathbf{\Upsilon}_2^{-1}\mathbf{\Phi}_2\mathbf{H}^T\mathbf{\Phi}_1^{-1}\mathbf{\Upsilon}_1\mathbf{H}.$$

This matrix is nonsingular, since it is of the form $I + N_1N_2$ with $N_1 \succeq 0$ and $N_2 \succeq 0$. (The product of two symmetric and positive semidefinite matrices, though nonsymmetric, has real nonnegative eigenvalues.) □

COROLLARY 3.2. *Consider an SDP whose solution (X,y,Z) satisfies the strict complementarity and primal and dual nondegeneracy conditions. Suppose that the XZ+ZX method uses* $\sigma = 0$ *and* $\alpha = \beta = 1$ *in the Basic Iteration. Then, there exists* $\epsilon > 0$ *such that, if the iteration is started at* (X_0, y_0, Z_0)*, with* $||(X_0, y_0, Z_0) - (X, y, Z)|| < \epsilon$*, the iterates converge Q-quadratically to* (X, y, Z).

The proof of Corollary 3.2 is immediate from the standard convergence theory for Newton's method. It is clear that Corollary 3.2 holds also for less restrictive assumptions on σ, α, and β. See [20] for relevant results for LP. There is no requirement that (X_0, y_0, Z_0) lie in a horn-shaped neighborhood of the central path, or even in the feasible region. Note that the assumptions of Corollary 3.2 do not guarantee positive definite iterates. These are not required to make (2.10) well defined, though the equivalence of (2.10) with (2.12) to (2.14) does not hold if X or Z is singular. In practice, conditions (2.18) and (2.19) ensure positive definite iterates.

A result like Theorem 3.1 does not hold for any of the other methods discussed so far. As already noted, the function to which Newton's method is applied is, in the case of the X^{-1} and NT methods, not defined at an optimal point. For the XZ method, the function G is defined at the solution, but it can be shown that the Jacobian $\mathbf{J}$ is always singular there. More importantly, bearing in mind the symmetrization step, an example can be constructed where $\mathbf{J}$ has a null vector $(\Delta X, \Delta y, \Delta Z)$ with $\Delta X + \Delta X^T \neq 0$.

It is well known that a result like Theorem 3.1 holds for the XZ method for LP, using LP nondegeneracy assumptions.

Nondegeneracy assumptions are not required to obtain superlinear convergence results. This has been known for some years for LP [18] and is the subject of active current research for SDPs. However, such results require that the iterates of a method stay close to the central path. Our point here is that classical Newton theory applies to the XZ+ZX method, under nondegeneracy assumptions, in SDPs just as in LP.

4. Conditioning of the Schur complement matrix. In this section, we study the conditioning of the Schur complement matrix $\mathbf{M}$, introduced in section 2, on the central path. It is important to note that, when started on the central path, all the methods discussed so far generate the same first iterate. On the central path, X and Z commute. Therefore,

$$\mathbf{E}^{-1}\mathbf{F} = \frac{1}{\mu} X \circledast X$$

in all cases except the XZ method for which we have $\mathbf{E}^{-1}\mathbf{F} = \frac{1}{\mu} X \otimes X$. In both cases the Schur complement matrix $\mathbf{M} = \mathbf{A}\mathbf{E}^{-1}\mathbf{F}\mathbf{A}^T$ is the same.

We now analyze the condition number of $\mathbf{M}$ on the central path, as $\mu \to 0$. We begin by considering its rank in the limit.

THEOREM 4.1. *Assume that (X^μ, y^μ, Z^μ) lies on the central path of an SDP whose solution $(X, y, Z) = \lim_{\mu\to 0}(X^\mu, y^\mu, Z^\mu)$ satisfies the dual nondegeneracy condition* (3.2)*, with $r = \operatorname{rank}(X)$. Let $\mathbf{M}^\mu$ be the Schur complement matrix defined at (X^μ, y^μ, Z^μ). Then*

$$\lim_{\mu\to 0}(\mu\,\mathbf{M}^\mu)$$

exists and has rank $r^{\overline{2}}$.

Proof. Clearly,

$$\mu\,\mathbf{M}^\mu \to \mathbf{N} = \mathbf{A}(X \circledast X)\mathbf{A}^T,$$

the matrix whose (l, k) element is tr (XA_lXA_k). Let Q and λ_i satisfy (1.9), and write $\Lambda_1 = \operatorname{Diag}(\lambda_1, \ldots, \lambda_r) \succ 0$, with corresponding eigenvectors collected in Q_1, so that $X = Q_1\Lambda_1 Q_1^T$. Let $\mathbf{C}_1$ be the $m \times r^{\overline{2}}$ matrix introduced in (3.7), and let $\mathbf{D}_1$ be the $r^{\overline{2}} \times r^{\overline{2}}$ diagonal matrix

$$\mathbf{D}_1 = \operatorname{Diag}(\lambda_i\lambda_j), \quad 1 \le i \le j \le r,$$

using consistent orderings for $\mathbf{C}_1$ and $\mathbf{D}_1$. Then

$$\mathbf{N} = \mathbf{C}_1\mathbf{D}_1\mathbf{C}_1^T,$$

since the (l, k) element of the right-hand side is

$$\text{tr } (\Lambda_1 Q_1^T A_l Q_1 \Lambda_1 Q_1^T A_k Q_1) = \text{tr } (XA_lXA_k).$$

Since, by the dual nondegeneracy assumption, $\mathbf{C}_1$ has linearly independent columns, and since $\mathbf{D}_1 \succ 0$, this completes the proof of the theorem. ∎

Recall that the condition number of a symmetric positive definite matrix is $\kappa_{\max}/\kappa_{\min}$, where $\kappa_{\max}$ and $\kappa_{\min}$ are respectively its largest and smallest eigenvalues.

THEOREM 4.2. *Suppose that the assumptions of Theorem* 4.1 *hold. Then, if $m > r^{\overline{2}} > 0$, the condition number of $\mathbf{M}^\mu$ (equivalently of $\mu\mathbf{M}^\mu$) is bounded below by a positive constant times $1/\mu$.*

Proof. Let Q^μ, λ_i^μ satisfy (1.6), and let $\mathbf{B}^\mu, \mathbf{C}^\mu$ be the matrices introduced in section 3, evaluated at (X^μ, y^μ, Z^μ). Using Lemma 7.2 (see Appendix), we have

$$\mu\,\mathbf{M}^\mu = \mathbf{A}(X^\mu \circledast X^\mu)\mathbf{A}^T = \mathbf{B}^\mu\mathbf{D}^\mu(\mathbf{B}^\mu)^T, \tag{4.1}$$

where $\mathbf{D}^\mu$ is the diagonal $n^{\overline{2}} \times n^{\overline{2}}$ matrix

$$\mathbf{D}^\mu = \mathrm{Diag}(\lambda_i^\mu \lambda_j^\mu), \qquad 1 \le i \le j \le n. \tag{4.2}$$

The primal solution rank r defines a splitting

$$\mathbf{D}^\mu = \mathrm{Diag}(\mathbf{D}_1^\mu, \mathbf{D}_2^\mu, \mathbf{D}_3^\mu)$$

consistent with (3.7), so that

$$\mu\,\mathbf{M}^\mu = \mathbf{C}_1^\mu \mathbf{D}_1^\mu (\mathbf{C}_1^\mu)^T + \mathbf{C}_2^\mu \mathbf{D}_2^\mu (\mathbf{C}_2^\mu)^T + \mathbf{C}_3^\mu \mathbf{D}_3^\mu (\mathbf{C}_3^\mu)^T. \tag{4.3}$$

Here the entries of the diagonal matrices $\mathbf{D}_1^\mu$, $\mathbf{D}_2^\mu$, and $\mathbf{D}_3^\mu$ are $\lambda_i^\mu \lambda_j^\mu$, with the indices $1 \le i \le j \le r$ for $\mathbf{D}_1^\mu$, $1 \le i \le r < j \le n$ for $\mathbf{D}_2^\mu$, and $r+1 \le i \le j \le n$ for $\mathbf{D}_3^\mu$. Although Q^μ and $\mathbf{C}^\mu$ do not generally converge as $\mu \to 0$, Theorem 4.1 shows that $\mu \mathbf{M}^\mu \to \mathbf{N} = \mathbf{C}_1 \mathbf{D}_1 \mathbf{C}_1^T$, with rank $r^{\overline{2}}$. By assumption, $m > r^{\overline{2}} > 0$, so the largest eigenvalue of $\mathbf{N}$ is positive and the smallest is zero. The norms of the second and third terms in (4.3) are $O(\mu)$, so the largest and smallest eigenvalues of $\mu\,\mathbf{M}^\mu$ are, respectively, bounded away from zero and $O(\mu)$. (Here we use the fact that eigenvalues of a symmetric matrix are Lipschitz continuous functions of the matrix entries.) □

Theorem 4.2 is easily extended to the block-diagonal case. When all block sizes are one, the condition on m in its hypothesis cannot hold under the nondegeneracy assumptions. Indeed, it is well known that for LP, under assumptions of nondegeneracy and strict complementarity, the condition number of the Schur complement matrix is bounded independent of μ.

5. Stability. We have seen in the previous section that, for nondegenerate SDPs, the condition number of the Schur complement matrix, evaluated on the central path, is bounded below by a positive constant times $1/\mu$ (ruling out the exceptional cases $r^{\overline{2}} = m$ and $r = 0$). Consequently, we expect that as $\mu \to 0$, the computation of Δy in (2.12) will become increasingly less accurate. Indeed, in our original implementations we observed numerical instability leading to significant loss of primal feasibility near a solution. Recently, however, Todd, Toh, and Tütüncü [16] found that high accuracy is achievable. The main issue is the choice of formulae for Δy and ΔX. Several mathematically equivalent choices are possible, but these have quite different stability properties.

Formulae for Δy and ΔX are given in (2.12) and (2.13). Both include the term $\mathbf{F} r_d - r_c$. For the XZ+ZX method, this term (in matrix form) is

$$\frac{1}{2}\left((X(C - Z - \mathrm{mat}\,\mathbf{A}^T y) + (C - Z - \mathrm{mat}\,\mathbf{A}^T y)X) - (2\mu I - XZ - ZX)\right),$$

which can be rewritten as

$$\frac{1}{2}\left((X(C - \mathrm{mat}\,\mathbf{A}^T y) + (C - \mathrm{mat}\,\mathbf{A}^T y)X) - 2\mu I\right).$$

However, using this simplification to modify (2.12) and (2.13) leads to instability and loss of primal feasibility. It is much better to implement (2.12) and (2.13) directly. This is done in the computational experiments reported in section 7.

The same issue applies to the XZ method. However, direct implementation of (2.12) and (2.13) does not give good results for the XZ method. Instead,[1] we use the fact that $\mathbf{E}^{-1}\mathbf{F}$ is symmetric positive definite to write

$$\mathbf{E}^{-1}\mathbf{F} = Z^{-1} \otimes X = \mathbf{G}^T\mathbf{G}, \qquad \mathbf{G} = M^{-1} \otimes L^T,$$

where L and M are respectively Cholesky factors of X and Z, i.e.,

$$X = LL^T, \qquad Z = MM^T.$$

Noting that the first block in the right-hand side of (2.11) is

$$u = \operatorname{vec} U = \operatorname{vec}(C - \mu X^{-1} - \operatorname{mat} \mathbf{A}^T y),$$

we see that (2.11) is equivalent to

$$\begin{bmatrix} -\mathbf{I} & \widetilde{\mathbf{A}}^T \\ \widetilde{\mathbf{A}} & 0 \end{bmatrix} \begin{bmatrix} \widetilde{\Delta x} \\ \Delta y \end{bmatrix} = \begin{bmatrix} \tilde{u} \\ r_p \end{bmatrix}, \tag{5.1}$$

$$\widetilde{\Delta x} = \mathbf{G}^{-T}\Delta x = \operatorname{vec}(L^{-1}\,\Delta X\, M) = \operatorname{vec} \widetilde{\Delta X},$$

$$\tilde{u} = \mathbf{G}u = \operatorname{vec}(L^T U M^{-T}) = \operatorname{vec} \widetilde{U},$$

and

$$\widetilde{\mathbf{A}} = \mathbf{A}\mathbf{G}^T = \begin{bmatrix} (\operatorname{vec} L^T A_1 M^{-T})^T \\ \vdots \\ (\operatorname{vec} L^T A_m M^{-T})^T \end{bmatrix}. \tag{5.2}$$

The solution is given by

$$(\widetilde{\mathbf{A}}\widetilde{\mathbf{A}}^T)\Delta y = r_p + \widetilde{\mathbf{A}}\tilde{u} \tag{5.3}$$

(which may be solved with a Cholesky factorization) and

$$\Delta X = L\widetilde{\Delta X}M^{-1} = L(\operatorname{mat} \widetilde{\mathbf{A}}^T\Delta y - \widetilde{U})M^{-1}. \tag{5.4}$$

This last equation can be written in many ways, three of which are

$$\Delta X = L\left(L^T(\operatorname{mat} \mathbf{A}^T\Delta y)M^{-T} - L^T U M^{-T}\right) M^{-1} \tag{5.5}$$

$$= LL^T(\operatorname{mat} \mathbf{A}^T\Delta y)M^{-T}M^{-1} - LL^T U M^{-T}M^{-1} \tag{5.6}$$

$$= LL^T\left((\operatorname{mat} \mathbf{A}^T\Delta y) - U\right) M^{-T}M^{-1}. \tag{5.7}$$

Of these four mathematically equivalent formulae, (5.4) and (5.5) give the highest accuracy, with smallest loss of primal feasibility. We used (5.4) in our computational experiments, with Δy defined by (5.3).

For the NT method, $\mathbf{E}^{-1}\mathbf{F}$ is also symmetric positive definite, so similar considerations apply; see [16].

[1] The discussion here is motivated by [16].

6. The Q method. In this section we change direction, deriving an alternative primal-dual interior-point method that generates iterates (X, y, Z) with the property that X and Z commute, i.e., $XZ = ZX$. This is motivated by the fact that this property holds for all points on the central path. Instead of treating the variables X and Z directly, we introduce as variables the eigenvalues of X and Z and their common set of eigenvectors. In other words, the variables consist of an orthogonal matrix Q, diagonal matrices Λ and Ω, and a vector $y \in \mathbb{R}^n$ that must satisfy

$$\begin{aligned} Q\Omega Q^T + \sum_{k=1}^{m} y_k A_k &= C, \\ A_k \bullet (Q\Lambda Q^T) &= b_k, \quad k = 1, \ldots, m, \\ \Lambda\Omega &= \mu I. \end{aligned} \tag{6.1}$$

This defines a map from $\mathcal{O}^n \times \mathbb{R}^{2n+m}$ to $\mathbb{R}^{n^{\overline{2}}+n+m}$, where $\mathcal{O}^n$ is the Lie group of orthogonal matrices with determinant one, whose dimension is $n(n-1)/2$. (Since the signs of eigenvectors are arbitrary, it is not a restriction to impose $\det Q = 1$.) The price paid for the diagonalization is the nonlinear appearance of the variable Q in the feasibility equations.

Let $\mathcal{K}^n$ denote the space of $n \times n$ skew-symmetric matrices, and consider the exponential map from $\mathcal{K}^n$ to $\mathcal{O}^n$ defined by

$$\exp(S) = I + S + \frac{1}{2}S^2 + \cdots.$$

This map is smooth, onto, and, in a neighborhood of 0, also one-to-one. Borrowing a technique used in [14], we derive a form of Newton's method based on parameterizing $\mathcal{O}^n$ near a given point Q by $Q\exp(S)$. Let kvec be an isometry from $\mathcal{K}^n$ to $\mathbb{R}^{n(n-1)/2}$, stacking the upper triangular entries of a skew-symmetric matrix in a vector, with a factor of $\sqrt{2}$ to preserve the inner product. Let us use the convention $s = \text{kvec}\,(S)$, $\Lambda = \text{Diag}(\lambda)$, and $\Omega = \text{Diag}(\omega)$. Define

$$G_Q(\lambda, y, \omega, s) = \begin{bmatrix} \text{vec}\,(C - Q\exp(S)\Omega\exp(-S)Q^T) - \mathbf{A}^T y \\ b - \mathbf{A}\,\text{vec}\,(Q\exp(S)\Lambda\exp(-S)Q^T) \\ \Lambda\Omega e - \mu e \end{bmatrix}. \tag{6.2}$$

The function G_Q maps $\mathbb{R}^{n^{\overline{2}}+n+m}$ to itself. Note that the third component of G_Q has the form familiar from LP.

Given an iterate $(X, y, Z) = (Q\Lambda Q^T, y, Q\Omega Q^T)$, we obtain a new iterate by applying Newton's method to the equation $G_Q = 0$ at the point $(\lambda, y, \omega, 0)$. The Newton step $(\Delta\lambda, \Delta y, \Delta\omega, s)$ is obtained by replacing $\exp(S)$ by $I + S$ and discarding second-order terms. The resulting $(n^{\overline{2}} + n + m) \times (n^{\overline{2}} + n + m)$ linear system is

$$\Delta\Omega + S\Omega - \Omega S + \sum_{k=1}^{m} \Delta y_k B_k = H - \Omega, \tag{6.3}$$

$$B_k \bullet (\Delta\Lambda + S\Lambda - \Lambda S) = b_k - B_k \bullet \Lambda, \qquad k = 1, ..., m, \tag{6.4}$$

$$\Lambda\,\Delta\Omega + \Omega\,\Delta\Lambda = \mu I - \Lambda\Omega, \tag{6.5}$$

where $B_k = Q^T A_k Q$ and $H = Q^T C Q - \sum_{k=1}^{m} y_k B_k$.

The basic iteration for the Q method is therefore

1. Choose $0 \le \sigma < 1$ and define
$$\mu = \sigma \frac{\lambda^T \omega}{n}.$$
2. Determine $(\Delta\lambda, \Delta y, \Delta\omega, s)$ from (6.3) to (6.5).
3. Choose steplengths α, β, γ and update the iterates by
$$\begin{array}{rcl} \Lambda & \leftarrow & \Lambda + \alpha\ \Delta\Lambda, \\ y & \leftarrow & y + \beta\ \Delta y, \\ \Omega & \leftarrow & \Omega + \beta\ \Delta\Omega, \\ Q & \leftarrow & Q(I + \frac{1}{2}\gamma S)(I - \frac{1}{2}\gamma S)^{-1}. \end{array}$$

A simple steplength rule is $\alpha = \min(1, \tau\,\hat{\alpha})$, $\beta = \min(1, \tau\,\hat{\beta})$, and $\gamma = \sqrt{\alpha\beta}$, where $\hat{\alpha}$ and $\hat{\beta}$ are steps to the boundary of the positive orthant. The multiplicative factor updating Q is the Cayley transform, an easily computed orthogonal matrix that approximates the matrix exponential to second order.

The equations defining the Q method can be rewritten as follows. First note that (6.4) can be rewritten as

$$B_k \bullet \Delta\Lambda + \mathrm{tr}\ ((\Lambda B_k - B_k \Lambda)S) = b_k - B_k \bullet \Lambda$$

and write

$$v = \begin{bmatrix} b_1 - B_1 \bullet \Lambda \\ \vdots \\ b_m - B_m \bullet \Lambda \end{bmatrix}.$$

Let $\mathrm{diag}(B_k)$ be the vector consisting of the n diagonal entries of B_k and $\mathrm{offdiag}(B_k)$ be the vector consisting of the $n(n-1)/2$ entries of the upper triangle of B_k, ordered consistently with the ordering chosen for the kvec operator. Define

$$\mathbf{L} = [\,\mathrm{diag}(B_1) \quad \cdots \quad \mathrm{diag}(B_m)\,]^T,$$

$$\mathbf{R} = [\,\mathrm{offdiag}(B_1) \quad \cdots \quad \mathrm{offdiag}(B_m)\,]^T.$$

Let

$$D = \mathrm{Diag}(\lambda_i - \lambda_j), \qquad E = \mathrm{Diag}(\omega_i - \omega_j)$$

be diagonal matrices of size $n(n-1)/2$ (corresponding to $1 \le i < j \le n$), whose orderings are also consistent with that of the kvec operator. Then, writing the diagonal and off-diagonal parts of (6.3) separately, we get the linear system

$$\begin{bmatrix} 0 & 0 & \mathbf{L}^T & I \\ 0 & E & \mathbf{R}^T & 0 \\ \mathbf{L} & \mathbf{R}D & 0 & 0 \\ \Omega & 0 & 0 & \Lambda \end{bmatrix} \begin{bmatrix} \Delta\lambda \\ s \\ \Delta y \\ \Delta\omega \end{bmatrix} = \begin{bmatrix} \mathrm{diag}(H - \Omega) \\ \mathrm{offdiag}(H) \\ v \\ \mu e - \Lambda\Omega e \end{bmatrix}. \tag{6.6}$$

We denote the matrix on the left-hand side of (6.6) by $\mathbf{J}_Q$.

Let (X, y, Z) be a solution of (1.2), (1.3) satisfying (3.4), (3.5). The matrix Q simultaneously diagonalizing X and Z is unique (up to signs of its columns) if and only if

$$\lambda_1 > \cdots > \lambda_r > 0 \quad \text{and} \quad 0 < \omega_{r+1} < \cdots < \omega_n. \tag{6.7}$$

THEOREM 6.1. *Let $(X, y, Z) = (Q\Lambda Q^T, y, Q\Omega Q^T)$ be a solution of* (1.2), (1.3) *satisfying the strict complementarity and primal and dual nondegeneracy conditions, and also condition* (6.7). *Then the matrix $\mathbf{J}_Q$, evaluated at the solution, is nonsingular.*

Proof. First note that the assumptions on the eigenvalues imply that the element of the diagonal matrix D corresponding to the index pair (i, j) is zero if and only if $r + 1 \le i < j \le n$, while the element of the diagonal matrix E corresponding to (i, j) is zero if and only if $1 \le i < j \le r$. Let us rewrite $\mathbf{J}_Q$ as

$$\begin{bmatrix} 0 & 0 & 0 & \mathbf{L}_1^T & I & 0 \\ 0 & 0 & 0 & \mathbf{L}_2^T & 0 & I \\ 0 & 0 & E & \mathbf{R}^T & 0 & 0 \\ \mathbf{L}_1 & \mathbf{L}_2 & \mathbf{R}D & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \Lambda_1 & 0 \\ 0 & \Omega_2 & 0 & 0 & 0 & 0 \end{bmatrix},$$

where $\Lambda_1 \succ 0$ and $\Omega_2 \succ 0$. As in the proof of Theorem 3.1, the nondegeneracy assumptions permit us to collect all r columns of $\mathbf{L}_1$ and $m-r$ columns of $\mathbf{R}$ together in a nonsingular $m \times m$ matrix $\mathbf{B}_1$. We collect the remaining $n(n-1)/2 - m + r$ columns of $\mathbf{R}$ in a matrix $\mathbf{R}_2$, and partition $D = \mathrm{Diag}(D_1, D_2)$ and $E = \mathrm{Diag}(E_1, E_2)$ accordingly. Observe that $D_1 \succ 0$ since the columns of $\mathbf{B}_1$ correspond to index pairs (i, j) with $\lambda_i > \lambda_j$. Likewise, $-E_2 \succ 0$ since all columns corresponding to index pairs (i, j) with $\omega_i = \omega_j = 0$ are contained in $\mathbf{B}_1$.

Let $\widetilde{D} = \mathrm{Diag}(I, D_1)$ and $\widetilde{E} = \mathrm{Diag}(0, E_1)$. Permuting the rows and columns, $\mathbf{J}_Q$ becomes

$$\begin{bmatrix} \widetilde{E} & 0 & 0 & \mathbf{B}_1^T & \widetilde{I} & 0 \\ 0 & 0 & 0 & \mathbf{L}_2^T & 0 & I \\ 0 & 0 & E_2 & \mathbf{R}_2^T & 0 & 0 \\ \mathbf{B}_1\widetilde{D} & \mathbf{L}_2 & \mathbf{R}_2 D_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \Lambda_1 & 0 \\ 0 & \Omega_2 & 0 & 0 & 0 & 0 \end{bmatrix},$$

where $\widetilde{I}$ is an m by r matrix containing r rows of the r by r identity matrix and $m-r$ zero rows. Interchanging the first and fourth rows and the second and last columns, this becomes

$$\begin{bmatrix} \mathbf{B}_1\widetilde{D} & 0 & \mathbf{R}_2 D_2 & 0 & 0 & \mathbf{L}_2 \\ 0 & I & 0 & \mathbf{L}_2^T & 0 & 0 \\ 0 & 0 & E_2 & \mathbf{R}_2^T & 0 & 0 \\ \widetilde{E} & 0 & 0 & \mathbf{B}_1^T & \widetilde{I} & 0 \\ 0 & 0 & 0 & 0 & \Lambda_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \Omega_2 \end{bmatrix}.$$

Performing Gauss block elimination on this matrix we see that its nonsingularity is equivalent to the nonsingularity of

$$\mathbf{B}_1^T + \widetilde{E}\widetilde{D}^{-1}\mathbf{B}_1^{-1}\mathbf{R}_2 D_2 E_2^{-1}\mathbf{R}_2^T.$$

Multiplying on the left by $\mathbf{B}_1^{-T}$ we obtain the matrix

$$I + (\mathbf{B}_1^{-T}\widetilde{E}\widetilde{D}^{-1}\mathbf{B}_1^{-1})(\mathbf{R}_2 D_2 E_2^{-1}\mathbf{R}_2^T).$$

This is nonsingular since it is of the form $I + N_1N_2$ with N_1, N_2 symmetric negative semidefinite. □

COROLLARY 6.2. *Consider an SDP whose solution satisfies the strict complementarity and primal and dual nondegeneracy conditions, and also condition* (6.7). *Suppose that the Q method uses* $\sigma = 0$ *and* $\alpha = \beta = \gamma = 1$. *Then, if the method is started with* λ, ω, y, *and* Q *initialized sufficiently close to their values at the solution, the iterates converge Q-quadratically to the solution.*

The proof of Corollary 6.2 is more technical than that of Corollary 3.2 and is omitted. It is necessary to establish that quadratic convergence is not impeded by either (a) the use of the Cayley transform to approximate the matrix exponential or (b) the dependence of the definition of G_Q on Q.

As with the other methods, we see how to efficiently implement the Q method by performing block Gauss elimination directly on $\mathbf{J}_Q$ without partitioning the blocks. The first step yields

$$\begin{bmatrix} -\Lambda^{-1}\Omega & 0 & \mathbf{L}^T \\ 0 & D^{-1}E & \mathbf{R}^T \\ \mathbf{L} & \mathbf{R} & 0 \end{bmatrix} \begin{bmatrix} \Delta\lambda \\ \tilde{s} \\ \Delta y \end{bmatrix} = \begin{bmatrix} \operatorname{diag}(H - \mu\Lambda^{-1}) \\ \operatorname{offdiag}(H) \\ v \end{bmatrix},$$

where $\widetilde{S} = \operatorname{kvec}\tilde{s}$ is the symmetric matrix defined by

$$\widetilde{S}_{ij} = (\lambda_i - \lambda_j)S_{ij}.$$

One more step of block elimination then gives the Schur complement

$$\mathbf{M}_Q = [\,\mathbf{L} \quad \mathbf{R}\,] \begin{bmatrix} \Lambda\Omega^{-1} & 0 \\ 0 & -DE^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{L}^T \\ \mathbf{R}^T \end{bmatrix}. \tag{6.8}$$

As in LP, the center factor of the Schur complement is diagonal, with entries

$$\frac{\lambda_i}{\omega_i},\ 1 \le i \le n, \quad \text{and} \quad \frac{\lambda_i - \lambda_j}{\omega_j - \omega_i},\ 1 \le i < j \le n.$$

Of course, the $\mathbf{L}$ and $\mathbf{R}$ blocks are not independent of the iteration count.

The Q method does not require computing eigenvalues. The variables Q, λ, and ω are all updated using rational operations. This is in contrast with the XZ+ZX method which requires the computation of eigenvalues in two places: the formation of the Schur complement matrix $\mathbf{M}$ (to solve the Lyapunov equations) and in the step-length computation (to find the step to the boundary). Finally, note that the Schur complement matrix is symmetric for the Q method, but not for the XZ+ZX method.

When evaluated on the central path, the Schur complement matrix $\mathbf{M}_Q^\mu$ for the Q method is equal to the Schur complement matrix $\mathbf{M}^\mu$ for the XZ and XZ+ZX methods, assuming that (6.7) holds. To see this, let $\mathbf{L}^\mu$, $\mathbf{R}^\mu$, D^μ, E^μ denote the matrices $\mathbf{L}$, $\mathbf{R}$, D, E evaluated on the central path. We have $\Lambda^\mu(\Omega^\mu)^{-1} = \operatorname{Diag}(\lambda_i^\mu/\omega_i^\mu) = \frac{1}{\mu}\operatorname{Diag}((\lambda_i^\mu)^2)$ and

$$-D^\mu(E^\mu)^{-1} = \operatorname{Diag}\left(\frac{\lambda_i^\mu - \lambda_j^\mu}{\omega_j^\mu - \omega_i^\mu}\right) = \frac{1}{\mu}\operatorname{Diag}(\lambda_i^\mu\lambda_j^\mu).$$

Thus,

$$\mathbf{M}_Q = \frac{1}{\mu}\mathbf{B}^\mu\mathbf{D}^\mu(\mathbf{B}^\mu)^T = \mathbf{M},$$

using (4.1) and (4.2).

Although the Q method has some attractive features, it is, at present, not a practical alternative to the other algorithms. When initialized far from the solution, convergence is generally not obtained. However, the quadratic local convergence established here is observed in practice.

7. Computational results. In this section we report on the results of some extensive numerical experiments. We start by discussing some important implementation details.

Mehrotra's predictor-corrector (PC) rule is a well-known technique in LP [18]. It can easily be extended to the XZ and XZ+ZX methods, as follows.

XZ and XZ+ZX methods with Mehrotra predictor-corrector rule.

1. Determine ΔX, Δy, ΔZ from (2.10) using $\mu = 0$ in (2.8), and symmetrize ΔX in the case of the XZ method.
2. Choose steplengths α, β using (2.18)–(2.19), and define

$$\sigma = \left(\frac{(X + \alpha\Delta X) \bullet (Z + \beta\Delta Z)}{X \bullet Z}\right)^3, \tag{7.1}$$

$$\mu = \sigma\frac{X \bullet Z}{n}.$$

3. Redetermine ΔX, Δy, ΔZ from (2.10), using

$$R_c = \left\{ \begin{array}{ll} \mu I - (XZ + \Delta X\,\Delta Z) & XZ \text{ method} \\ \mu I - \frac{1}{2}(XZ + ZX + \Delta X\,\Delta Z + \Delta Z\,\Delta X) & XZ + ZX \text{ method} \end{array} \right\},$$

symmetrize ΔX in the case of the XZ method, and update the iterates by

$$\begin{array}{rcl} X & \leftarrow & X + \alpha\ \Delta X, \\ y & \leftarrow & y + \beta\ \Delta y, \\ Z & \leftarrow & Z + \beta\ \Delta Z, \end{array}$$

with α, β given by (2.18) and (2.19).

See [16] for a definition of the PC version of the NT method. (Our experiments use (7.1) in the implementation of all the methods, although [16] uses the exponent 2 instead of 3 in (7.1).)

Computational results are presented in Tables 1 through 4. Tables 1, 2, and 3 report results for randomly generated problems, with $m = n$. The matrices $A_k, k = 1, \ldots, m$ were symmetric with entries uniformly distributed in the interval $[-1, 1]$. The vector b and the matrix C were chosen to ensure that Assumption 1 was satisfied. More precisely, random positive definite symmetric matrices $\tilde{X}$ and $\tilde{Z}$ and a random vector $\tilde{y}$ were generated, and b was defined by $b_k = A_k \bullet \tilde{X}$, $k = 1, \ldots, m$, while C was set to $\tilde{Z} + \sum_{k=1}^m \tilde{y}_k A_k$. All methods were initialized with the infeasible starting point $(X_0, y_0, Z_0) = (I, 0, I)$. Table 1 shows results for the XZ+ZX, XZ, and NT basic iteration, using $\sigma = 0.25$ in (2.17), with various choices for the steplength parameter τ in (2.18), (2.19). We also implemented the X^{-1} method but found it required many

TABLE 1

(a) *Number of iterations to reduce gap by* 10^{12} *averaged over 100 randomly generated problems. Basic iteration with* $\sigma = 0.25$*; starting infeasible;* $n = 20$*,* $m = 20$*; S: short step failure (not included in average); E: exceeded limit failure (not included in average).*

Method	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$
$XZ + ZX$	21.6	21.2	21.2
XZ	21.8	22.1	23.7 (S:11%, E:2%)
NT	21.6	22.0	29.8 (E:18%)

(b) *Log norm infeasibility averaged over same data.*

Method	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$
$XZ + ZX$	-12.6	-12.6	-12.6
XZ	-11.0	-10.9	-10.9
NT	-10.8	-10.7	-10.5

TABLE 2

(a) *Number of iterations to reduce gap by* 10^{12} *averaged over 100 randomly generated problems. Mehrotra predictor-corrector rule; starting infeasible;* $n = 20$*,* $m = 20$*; S: short step failure (not included in average); E: exceeded limit failure (not included in average).*

Method	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$
$XZ + ZX$	14.0	9.4	8.5
XZ	15.3	14.2	15.7 (S:63%, E:5%)
NT	14.5	22.8	(E:100%)

(b) *Log norm infeasibility averaged over same data.*

Method	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$
$XZ + ZX$	-10.7	-12.0	-12.2
XZ	-8.7	-8.8	-9.4
NT	-9.1	-8.3	

TABLE 3

(a) *Number of iterations to reduce gap by* 10^{12} *averaged over* 100 *randomly generated problems. Mehrotra predictor-corrector rule; starting infeasible;* S: *short step failure (not included in average).*

Method	$n = m = 20$	$n = m = 40$	$n = m = 80$
$XZ + ZX$, $\tau = 0.99$	9.4	9.9	10.0
$XZ + ZX$, $\tau = 0.999$	8.6	9.2 (S:3%)	9.5 (S:6%)

(b) *Log norm infeasibility averaged over same data.*

Method	$n = m = 20$	$n = m = 40$	$n = m = 80$
$XZ + ZX$, $\tau = 0.99$	-12.1	-11.2	-10.4
$XZ + ZX$, $\tau = 0.999$	-12.3	-11.4	-10.5

more iterations than the others with the same parameter choices. Table 2 shows results for the PC variants. Part (a) of both tables shows the number of iterations required to reduce the quantity $X \bullet Z$ by a factor of 10^{12}, averaged over 100 problems. Part (b) shows the final value of

$$\log_{10}\left(||r_p|| + ||R_d||\right),$$

averaged over the same data. A run was terminated reporting success when $X \bullet Z$ was reduced by the desired factor of 10^{12} and reporting failure if (i) the primal or dual steplength (α or β) dropped below 10^{-4} (indicated by the notation S in part (a) of the table), (ii) the number of iterations exceeded the maximum value 50 (indicated

by E in the table), or (iii) a Cholesky factorization failed (caused by rounding errors, impossible in exact arithmetic, and indicated by R in the table). Failures are *not* included in the average statistics. All experiments were conducted in Matlab, using IEEE double-precision arithmetic.

Let us first consider the results shown in Table 1 for the Basic Iteration without the PC rule. For $\tau = 0.9$, all three methods show essentially the same number of iterations. The XZ+ZX method achieves the highest accuracy (in terms of feasibility). More aggressive choices of the step-length parameter have little effect on the XZ+ZX method but cause difficulties for the XZ and NT methods. Choosing $\tau = 0.999$ causes the XZ and NT methods to fail in many cases. In the case of the XZ method, this was usually because the primal or dual steplength dropped below 10^{-4}, but for the NT method, failure generally occurred because the desired reduction in the duality gap was not achieved in 50 iterations.

Table 2 shows the same experiment using the PC rule. With $\tau = 0.9$, the PC rule greatly reduces the number of iterations, though with some loss of feasibility for the XZ and NT methods. More aggressive choices of τ gave a significantly reduced number of iterations (without loss of feasibility) for the XZ+ZX method, but led to many failures for the XZ and NT algorithms.

In Table 3, we show results for the XZ+ZX method when the problem size n is varied, using the PC rule and two choices of τ. We see an iteration count which is essentially constant as n increases, with occasional failures (with steps too short) for $\tau = 0.999$. In these cases, we found that success could generally be achieved by restarting with X_0 and Z_0 set to a larger multiple of the identity (alternatively, reducing τ). Note some loss of feasibility (due to rounding errors) for larger n. Primal feasibility can be regained by projecting onto the set $\{x : \mathbf{A}x = b\}$, but this generally fails to give a more accurate solution, as the duality gap usually increases.

For some classes of problem, the XZ and NT methods can be implemented very efficiently. This is the case, for example, for SDPs with only diagonal constraints on X (equivalently, off-diagonal entries in Z fixed). For such an SDP, we have $m = n$ and $A_k = e_k e_k^T$, $k = 1, ..., m$, where e_k is the kth column of the identity matrix. Consequently, for the XZ method we have $\mathbf{M}_{ij} = e_i^T X e_j e_j^T Z^{-1} e_i$, i.e. $\mathbf{M}$ is the Hadamard product of X and Z^{-1} [6], reducing the cost of forming $\mathbf{M}$ from $O(n^4)$ (the general case when $m = n$) to $O(n^3)$. It is not known how to implement the XZ+ZX method efficiently in this case. A similar observation applies to the SDP that computes the Lovász θ function for a graph [5], as long as the number of edges is not too large. In this case n is the number of vertices in the graph and $m - 1$ is the number of edges, with $b = e_1$, $-C$ the matrix of all ones, $A_1 = I$, and, for $k = 2, \ldots, m$, $A_k = e_i e_j^T + e_j e_i^T$, where the $(k-1)$th edge of the graph is from vertex i to vertex j.

Table 4 shows results comparing the XZ+ZX, XZ, and NT methods on the θ function for randomly generated graphs, with edge density 50%, using the general-purpose implementations. We set $n = 20$, so the expected value of m is $\frac{1}{4}n(n-1)+1 = 96$. For these runs, we used the initial feasible point $(X_0, y_0, Z_0) = ((1/n)I, -2ne_1, 2nI + C)$. Using an infeasible initial point did not significantly change the results. The XZ and NT methods often had difficulty reducing the duality gap by the desired factor, even with $\tau = 0.9$, because rounding errors caused a Cholesky factorization to fail. This was usually the Cholesky factorization of $\mathbf{M} = \widetilde{\mathbf{A}}\widetilde{\mathbf{A}}^T$ (see (2.15) and (5.3)), which is positive definite in exact arithmetic but may be numerically indefinite. Since the Schur complement for the XZ+ZX method is nonsymmetric, it is factored using

TABLE 4

(a) *Lovasz θ function: Number of iterations to reduce gap by 10^{12} averaged over 100 randomly generated problems. Mehrotra predictor-corrector rule; starting infeasible; $n = 20$, edge density 0.5 ($m \approx 96$);* S: *short step failure (not included in average);* E: *exceeded limit failure (not included in average);* R: *rounding failure (not included in average).*

Method	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$
$XZ + ZX$	15.2	11.0 (S:1%)	10.4 (S:1%)
XZ	15.5 (R:22%)	17.0 (E:15%, R:22%)	15.0 (S:68%, E:28%, R:3%)
NT	15.6 (R:25%)	21.3 (E:15%, R:22%)	(S:3%, E:97%)

(b) *Log norm infeasibility averaged over same data.*

Method	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$
$XZ + ZX$	-13.8	-13.7	-13.6
XZ	-12.6	-11.4	-12.4
NT	-12.5	-10.9	

an LU factorization, which fails only if the matrix is numerically singular, i.e., the factorization generates a zero pivot.

We also implemented the Q method and observed that it has essentially the same rapid local convergence and high accuracy properties as the XZ+ZX method, although when initialized far from the solution it generally fails to converge.

We conclude that the XZ+ZX PC method is the most efficient in terms of number of iterations, most accurate in terms of feasibility, and most robust with respect to its ability to step close to the boundary.

Appendix: Symmetric Kronecker products. Consider the linear operator on $\mathbb{R}^{n\times n}$ defined by the map

$$K \mapsto NKM^T, \tag{7.2}$$

where $M, N \in \mathbb{R}^{n\times n}$. It is standard to represent this linear operator by the Kronecker product

$$M \otimes N = \begin{bmatrix} M_{11}N & \cdots & M_{1n}N \\ \vdots & & \vdots \\ M_{n1}N & \cdots & M_{nn}N \end{bmatrix},$$

where nvec maps $\mathbb{R}^{n\times n}$ to $\mathbb{R}^{n^2}$, stacking the columns of a matrix in a vector, since then

$$(M \otimes N)\,\mathrm{nvec}\,(K) = \mathrm{nvec}\,(NKM^T). \tag{7.3}$$

Other Kronecker product identities include

$$(M \otimes N)^{-1} = M^{-1} \otimes N^{-1} \quad \text{and} \quad (M \otimes N)(K \otimes L) = MK \otimes NL. \tag{7.4}$$

Now consider the linear operator on $\mathcal{S}^n$ defined by the map

$$K \mapsto \frac{1}{2}(NKM^T + MKN^T), \tag{7.5}$$

where $M, N \in \mathbb{R}^{n\times n}$. To represent this map as a matrix, define $M \circledast N$ by the identity

$$(M \circledast N)\,\mathrm{svec}\,(K) = \mathrm{svec}\left(\frac{1}{2}(NKM^T + MKN^T)\right), \tag{7.6}$$

where svec maps $\mathcal{S}^n$ to $\mathbb{R}^{n^{\overline{2}}}$ by

$$\text{svec}\,(K) = \left[K_{11}, \sqrt{2}K_{12}, \ldots, \sqrt{2}K_{1n}, K_{22}, \ldots, \sqrt{2}K_{2n}, \ldots, K_{nn}\right]^T. \tag{7.7}$$

Note that

$$K \bullet L = \text{svec}\,(K)^T \text{svec}\,(L).$$

Of course, the ordering used in (7.7) is arbitrary: the important point is that each element of svec (K) is associated with an index pair (i, j), with $i \le j$. The ordering chosen for svec dictates a corresponding ordering for $\circledast$.

We call the matrix $M \circledast N$ a *symmetric Kronecker product*. Note the identity

$$M \circledast N = N \circledast M. \tag{7.8}$$

Furthermore $(M \circledast M)^{-1} = M^{-1} \circledast M^{-1}$, but $(M \circledast N)^{-1} \neq M^{-1} \circledast N^{-1}$, in general.

We need the following lemmas whose proofs are straightforward.

LEMMA 7.1. *Let $V \in \mathbb{R}^{n\times n}$ and let v_i, $1 \le i \le n$ denote the columns of V. The (i, j)th column of $V \circledast V$, $1 \le i \le j \le n$ is the vector*

$$\begin{cases} \text{svec}\,(v_i v_i^T) & \text{if } i = j, \\ \frac{1}{\sqrt{2}} \text{svec}\,\left(v_i v_j^T + v_j v_i^T\right) & \text{if } i < j. \end{cases}$$

LEMMA 7.2. *Let M, N be commuting symmetric matrices, and let $\alpha_1, \ldots, \alpha_n$, $\beta_1, \ldots, \beta_n$ denote their eigenvalues with $v_1, \ldots, v_n$ a common basis of orthonormal eigenvectors. The $n(n+1)/2$ eigenvalues of $M \circledast N$ are given by*

$$\frac{1}{2}(\alpha_i\beta_j + \beta_i\alpha_j), \quad 1 \le i \le j \le n,$$

with the corresponding set of orthonormal eigenvectors

$$\begin{cases} \text{svec}\,(v_i v_i^T) & \text{if } i = j, \\ \frac{1}{\sqrt{2}} \text{svec}\,(v_i v_j^T + v_j v_i^T) & \text{if } i < j. \end{cases}$$

In other words, if $V = [v_1 \cdots v_n]$, then $V \circledast V$ is an orthogonal matrix of size $n^{\overline{2}} \times n^{\overline{2}}$ which diagonalizes $M \circledast N$. The standard algorithm for solving the Lyapunov equation $MXN^T + NXM^T = B$ (when M and N commute) immediately follows: the solution is VCV^T, where C is found by computing $V^T BV$ and dividing its entries by the quantities $(\alpha_i\beta_j + \beta_i\alpha_j)$ componentwise.

Acknowledgments. The third author is grateful to Richard Tapia, Bob Vanderbei, and Steve Wright for introducing him to primal-dual interior-point methods for LP. The authors also thank Mike Todd for many helpful comments, especially a suggestion concerning the Q method that makes the development given here substantially simpler than the one given in [1].

REFERENCES

[1] F. ALIZADEH, J.-P. A. HAEBERLY, AND M. L. OVERTON, *A new primal-dual interior-point method for semidefinite programming*, in Proc. 5th SIAM Conference on Applied Linear Algebra, J.G. Lewis, ed., SIAM, Philadelphia, 1994, pp. 113–117.

[2] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton, *Primal-dual interior-point methods for semidefinite programming*, manuscript presented at the Math. Programming Symposium, Ann Arbor, MI, 1994.
[3] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton, *Complementarity and nondegeneracy in semidefinite programming*, Math. Programming, 77 (1997), pp. 111–128.
[4] F. Alizadeh, *Combinatorial Optimization with Interior Point Methods and Semidefinite Matrices*, Ph.D. thesis, University of Minnesota, Minneapolis, MN, 1991.
[5] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, Springer-Verlag, New York, 1988.
[6] C. Helmberg, F. Rendl, R. Vanderbei, and H. Wolkowicz, *An interior-point method for semidefinite programming*, SIAM J. Optim., 6 (1996), pp. 342–361.
[7] M. Kojima, S. Shindoh, and S. Hara, *Interior-point methods for the monotone linear complementarity problem in symmetric matrices*, SIAM J. Optim., 7 (1997), pp. 86–125.
[8] M. Kojima, M. Shida, and S. Shindoh, *A predictor-corrector interior-point algorithm for the semidefinite linear complementarity problem using the Alizadeh-Haeberly-Overton search direction*, SIAM J. Optim., to appear.
[9] R. Monteiro, *Polynomial convergence of primal-dual algorithms for semidefinite programming based on Monteiro and Zhang family of directions*, SIAM J. Optim., 8 (1998), pp. 797–812.
[10] R. Monteiro and P. Zanjacomo, *A note on the existence of the Alizadeh-Haeberly-Overton direction for semidefinite programming*, Math. Programming, 78 (1997), pp. 393–396.
[11] Y. Nesterov and A. Nemirovskii, *Interior Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, 1994.
[12] Y. Nesterov and M. Todd, *Primal-dual interior-point methods for self-scaled cones*, SIAM J. Optim., 8 (1998), pp. 324–364.
[13] Y. Nesterov and M. Todd, *Self-scaled barriers and interior-point methods for convex programming*, Math. Oper. Res., 22 (1997), pp. 1–42.
[14] M. L. Overton and R. S. Womersley, *Second derivatives for optimizing eigenvalues of symmetric matrices*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 697–718.
[15] M. Shida, S. Shindoh, and M. Kojima, *Existence and uniqeness of search directions in interior-point algorithms for the SDP and the monotone SDLCP*, SIAM J. Optim., 8 (1998), pp. 387–398.
[16] M. J. Todd, K. C. Toh, and R. H. Tütüncü, *On the Nesterov-Todd direction in semidefinite programming*, SIAM J. Optim., 8 (1998), pp. 769–796.
[17] L. Vandenberghe and S. Boyd, *Primal-dual potential reduction method for problems involving matrix inequalities*, Math. Programming B, 69 (1995), pp. 205–236.
[18] S. J. Wright, *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia, 1997.
[19] Y. Zhang, *On extending some primal-dual interior-point algorithms from linear programming to semidefinite programming*, SIAM J. Optim., 8 (1998), pp. 365–386.
[20] Y. Zhang, R. A. Tapia, and J. E. Dennis, *On the superlinear and quadratic convergence of primal-dual interior point linear programming algorithms*, SIAM J. Optim., 2 (1992), pp. 304–324.