Primal-Dual Interior-Point Methods for Semidefinite Programming: Convergence Rates, Stability and Numerical Results *

Farid Alizadeh[†] Jean-Pierre A. Haeberly[‡] Michael L. Overton[§]

May 16, 1996

Abstract

Primal-dual interior-point path-following methods for semidefinite programming (SDP) 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 respectively the XZ, $XZ + ZX$ and Q methods. For the $XZ + ZX$ and Q algorithms, the Newton system is well-defined and its Jabobian 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. We observe that, 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.

*Updated versions will be made available by anonymous ftp to `cs.nyu.edu`, in the file `pub/local/overton/pdsdp.ps.gz`.

†RUTCOR, Rutgers University, New Brunswick, NJ. Supported in part by National Science Foundation grant CCR-9501941 and Office of Naval Research contract N00014-96-1-0704. E-mail: `alizadeh@rutcor.rutgers.edu`

‡Mathematics Department, Fordham University, Bronx, NY. Supported in part by National Science Foundation grant CCR-9401119. E-mail:`haeberly@murray.fordham.edu`

§Computer Science Department, Courant Institute of Mathematical Sciences, New York University, New York, NY. Supported in part by National Science Foundation grant CCR-9401136. E-mail: `overton@cs.nyu.edu`

1 Introduction

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

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

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

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

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)

$$\begin{array}{ll} \min_{X \in \mathcal{S}^n} & C \bullet X \\ \text{s.t.} & A_k \bullet X = b_k, \ \ k = 1, \ldots, m \\ & X \succeq 0, \end{array} \tag{2}$$

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

$$\begin{array}{ll} \max_{y \in \mathbb{R}^m, Z \in \mathcal{S}^n} & b^T y \\ \text{s.t.} & \sum_{k=1}^m y_k A_k + Z = C, \\ & Z \succeq 0. \end{array} \tag{3}$$

The following are assumed to 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

$$X^\mu Z^\mu = \mu I \tag{4}$$

for some $\mu \in \mathbb{R}$, $\mu > 0$. It is well known [NN94] that, under Assumptions 1 and 2, (X^μ, y^μ, Z^μ) exists and is unique for all $\mu > 0$, and that

$$(X, y, Z) = \lim_{\mu \to 0} (X^\mu, y^\mu, Z^\mu) \tag{5}$$

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

$$X^\mu = Q^\mu \ \mathbf{Diag}(\lambda_1^\mu, \ldots, \lambda_n^\mu)\ (Q^\mu)^T, \quad Z^\mu = Q^\mu \ \mathbf{Diag}(\omega_1^\mu, \ldots, \omega_n^\mu)\ (Q^\mu)^T, \tag{6}$$

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

$$\lambda_i^\mu \omega_i^\mu = \mu, \quad i = 1, \ldots, n. \tag{7}$$

Without loss of generality, assume that

$$\lambda_1^\mu \geq \cdots \geq \lambda_n^\mu \quad \text{and} \quad \omega_1^\mu \leq \cdots \leq \omega_n^\mu. \tag{8}$$

As $\mu \to 0$, the centering condition (4) reduces to the complementarity condition $XZ = 0$, implying that

$$X = Q \ \mathbf{Diag}(\lambda_1, \ldots, \lambda_n)\ Q^T, \quad Z = Q \ \mathbf{Diag}(\omega_1, \ldots, \omega_n)\ Q^T, \tag{9}$$

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

$$\lambda_1 \geq \cdots \geq \lambda_n \quad \text{and} \quad \omega_1 \leq \cdots \leq \omega_n. \tag{10}$$

Interior point methods for semidefinite programming were originally introduced by [NN94, Ali91]. Early papers on primal-dual methods include [VB95] and [HRVW96]. A preliminary version of the present work appeared as [AHO94b]. Convergence analysis of primal-dual path-following methods for SDP appeared first in [KSH94, NT94, NT95]. We are primarily concerned with four methods, which we call respectively the XZ, $XZ + ZX$, Nesterov-Todd (NT) and Q methods. The XZ method first appeared in [HRVW96, KSH94]. The $XZ + ZX$ method was introduced in [AHO94b] and was recently analyzed by [KSS96]. The NT method was given by [NT94, NT95] and its implementation was recently discussed in [TTT96]. The Q method originally appeared in [AHO94a]. 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).

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 [Wri96] 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

$$\begin{bmatrix} \sum_{k=1}^m y_k A_k + Z - C \\ A_1 \bullet X - b_1 \\ \vdots \\ A_m \bullet X - b_m \\ XZ - \mu I \end{bmatrix} = 0. \tag{11}$$

However, the matrix XZ is not symmetric in general. Consequently, the domain and range of the function defined by the left-hand side of (11) 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 (11). In this case, X and Z are diagonal, and XZ is also diagonal, so the domain and range of (11) 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 (4) directly and view the left-hand side of (11) 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 (4).

The $XZ + ZX$ Method. Rewrite (4) in the symmetric form

$$XZ + ZX = 2\mu I. \tag{12}$$

Substituting (12) for (4) in (11) 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 (12) leads to symmetric matrix iterates X and Z.

We observe that (4) and (12) are equivalent when $X \succeq 0$ (or $Z \succeq 0$). That (4) implies (12) is immediate. That the converse holds for $X \succeq 0$ is seen by using $X = Q\Lambda Q^T$ to reduce (12) to $\Lambda(Q^TZQ) + (Q^TZQ)\Lambda = 2\mu I$, with Λ diagonal and nonnegative and $Q^TQ = I$. 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, $\Lambda(Q^TZQ)$ is diagonal, and (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

$$X\,\Delta Z + \Delta X\,Z = \mu I - XZ. \tag{13}$$

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 (13) in the form

$$(I \otimes X)\,\mathbf{nvec}\,(\Delta Z) + (Z \otimes I)\,\mathbf{nvec}\,(\Delta X) = \mathbf{nvec}\,(\mu I - XZ). \tag{14}$$

where $\otimes$ denotes the standard Kronecker product (see Appendix, equation (59)).

To discuss the $XZ + ZX$ method, we introduce a *symmetric* version of the Kronecker product. The Newton correction for (12) satisfies the linear equation

$$X\,\Delta Z + \Delta Z\,X + \Delta X\,Z + Z\,\Delta X = 2\mu I - XZ - ZX, \tag{15}$$

where ΔX and ΔZ are symmetric. Let **svec** be an isometry identifying $\mathcal{S}^n$ with $\mathbb{R}^{\overline{n^2}}$, so that $K \bullet L = \mathbf{svec}\,(K)^T\,\mathbf{svec}\,(L)$ for all $K, L \in \mathcal{S}^n$ (see Appendix). Then (15) can be written as:

$$(Z \circledast I)\,\mathbf{svec}\,(\Delta X) + (X \circledast I)\,\mathbf{svec}\,(\Delta Z) = \mathbf{svec}\,(\mu I - \frac{1}{2}(XZ + ZX)) \tag{16}$$

where $\circledast$ denotes the symmetric Kronecker product defined in the Appendix (see (62)).

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 $\mathbf{vec}\, X$ and $\mathbf{vec}\, Z$ respectively, and we shall use Δx and Δz interchangeably with $\mathbf{vec}\, \Delta X$ and $\mathbf{vec}\, \Delta Z$, to be defined shortly.

Let

$$\mathbf{A} = \begin{bmatrix} (\mathbf{vec}\, A_1)^T \\ \vdots \\ (\mathbf{vec}\, A_m)^T \end{bmatrix} \tag{17}$$

and define

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

and

$$R_c = \left\{ \begin{array}{ll} \mu I - XZ & XZ \text{ method} \\ \mu I - \frac{1}{2}(XZ + ZX) & XZ + ZX \text{ method} \end{array} \right\} \tag{18}$$

with

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

Let

$$G(x, y, z) = \begin{bmatrix} -r_d \\ -r_p \\ -r_c \end{bmatrix}. \tag{19}$$

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

$$\begin{bmatrix} 0 & \mathbf{A}^T & \mathbf{I} \\ \mathbf{A} & 0 & 0 \\ \mathbf{E} & 0 & \mathbf{F} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = \begin{bmatrix} r_d \\ r_p \\ r_c \end{bmatrix}. \tag{20}$$

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 (20) by $\mathbf{J}$.

Applying block Gauss elimination, (20) reduces to the system

$$\begin{bmatrix} -\mathbf{F}^{-1}\mathbf{E} & \mathbf{A}^T \\ \mathbf{A} & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} r_d - \mathbf{F}^{-1} r_c \\ r_p \end{bmatrix} \tag{21}$$

A second step of block Gauss elimination gives

$$\mathbf{M}\Delta y = r_p + \mathbf{A}\mathbf{E}^{-1}(\mathbf{F}r_d - r_c), \tag{22}$$

$$\Delta x = -\mathbf{E}^{-1}(\mathbf{F}(r_d - \mathbf{A}^T\Delta y) - r_c), \tag{23}$$

and, from the dual feasibility equation,

$$\Delta z = r_d - \mathbf{A}^T \Delta y, \tag{24}$$

where

$$\mathbf{M} = \mathbf{A}\mathbf{E}^{-1}\mathbf{F}\mathbf{A}^T. \tag{25}$$

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

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

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

$$ZL_k + L_kZ = XA_k + A_kX. \tag{26}$$

Formation of $\mathbf{M}$ thus requires 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. If $m \gg n$ and we neglect sparsity considerations, the additional cost of the eigenvalue factorization is negligible in comparison to the other operations required to form and factor $\mathbf{M}$.

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 $XZ + ZX \succ 0$ [SSK96]. Equation (22) 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 (22) and (23) 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:

Basic Iteration.

1. Choose $0 \le \sigma < 1$ and define
$$\mu = \sigma \frac{X \bullet Z}{n}. \tag{27}$$
2. Determine ΔX, Δy, ΔZ from (20), equivalently (22)–(24).
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
$$\alpha = \min(1, \tau\, \hat{\alpha}) \qquad \hat{\alpha} = \sup\{\bar{\alpha} : X + \bar{\alpha}\, \Delta X \succeq 0\} \tag{28}$$
and
$$\beta = \min(1, \tau\, \hat{\beta}) \qquad \hat{\beta} = \sup\{\bar{\beta} : Z + \bar{\beta}\, \Delta Z \succeq 0\}. \tag{29}$$
Note that, except in the case $\Delta X \succeq 0$, we have $0 < \hat{\alpha} < \infty$ with
$$\hat{\alpha}^{-1} = \lambda_{\max}(-\Delta X, X),$$
where $\lambda_{\max}(A, B)$ means the largest eigenvalue of the pencil (A, B), i.e. the largest root of $\det(\lambda B - A) = 0$.

Other methods can also be defined in the same framework; two of these are discussed below. See [Zha95] for a class of methods that includes the $XZ + ZX$ method, and [KSH94, SSK96] for another class that includes all those discussed here except the $XZ + ZX$ method.

The X^{-1} Method. Replace R_c in (19) 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 [VB95] is based on a combination of these two steps.

The Nesterov-Todd Method. Use $R_c = \mu X^{-1} - Z$, $\mathbf{E} = W^{-1} \circledast W^{-1}$, $F = I \circledast I$, where $W = X^{1/2}(X^{1/2} Z X^{1/2})^{-1/2} X^{1/2}$. This form does not actually arise from applying Newton's method to (19). However, see [TTT96] 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 that SDP is an LP, the XZ, $XZ + ZX$ and Nesterov-Todd 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 (20), under nondegeneracy assumptions. To do this, we use the notions of nondegeneracy that were introduced by the authors in [AHO95].

Definition 1. Let (X, y, Z) solve SDP, with an orthogonal matrix Q satisfying (9). Let X have rank r, with positive eigenvalues $\lambda_1, \dots, \lambda_r$, and partition $Q = [Q_1\ Q_2]$, where the columns of Q_1 are eigenvectors corresponding to $\lambda_1, \dots, \lambda_r$. We say that (X, y, Z) satisfies the *strict complementarity and primal and dual nondegeneracy* conditions if the following hold:

1. $\mathbf{rank}(Z) = n - r$,

2. the matrices
$$\begin{bmatrix} Q_1^T A_k Q_1 & Q_1^T A_k Q_2 \\ Q_2^T A_k Q_1 & 0 \end{bmatrix}, \quad \text{for } k = 1, 2, \dots, m, \tag{30}$$
are linearly independent in $\mathcal{S}^n$, and

3. the matrices
$$Q_1^T A_k Q_1, \quad \text{for } k = 1, 2, \dots, m, \tag{31}$$
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 (30), (31) are respectively primal and dual nondegeneracy conditions under the assumption of strict complementarity. They immediately imply the inequalities
$$\overline{r^2} \le m \le \overline{r^2} + r(n - r) \tag{32}$$

(recalling the notation (1)). They also imply uniqueness of the primal and dual solutions. Furthermore, the conditions are generic properties of SDP, 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 [AHO95].

The strict complementarity condition $\mathbf{rank}(X) = r$, $\mathbf{rank}(Z) = n - r$ implies, using (10), that

$$\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \ldots = \lambda_n = 0, \tag{33}$$

and

$$0 = \omega_1 = \cdots = \omega_r < \omega_{r+1} \leq \ldots \leq \omega_n, \tag{34}$$

Let $B_k = Q^T A_k Q$. From (62), we have

$$\mathbf{svec}\, B_k = (Q \circledast Q)\, \mathbf{svec}\, A_k.$$

Recall the definition (17), and define

$$\mathbf{B} = \begin{bmatrix} (\mathbf{vec}\, B_1)^T \\ \vdots \\ (\mathbf{vec}\, B_m)^T \end{bmatrix}, \tag{35}$$

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

$$\mathbf{B} = [\mathbf{C}_1\ \mathbf{C}_2\ \mathbf{C}_3] \tag{36}$$

where $\mathbf{C}_1$ contains $\overline{r^2}$ 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 $\overline{(n-r)^2}$ columns corresponding to $r+1 \leq i \leq j \leq n$. The primal nondegeneracy condition (30) 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 (31) holds exactly when $\mathbf{C}_1$ has rank $\overline{r^2}$, i.e. the columns of $\mathbf{C}_1$ are linearly independent. Thus, the conditions (30) and (31) together imply that it is possible to choose $m - \overline{r^2}$ 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

$$\mathbf{B} = [\mathbf{B}_1\ \mathbf{B}_2] \tag{37}$$

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

Theorem 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 $\mathbf{S} = \mathbf{Diag}(\mathbf{P}, \mathbf{I}, \mathbf{P})$, so that

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

with $\mathbf{\Phi} = \mathbf{P}^T\mathbf{F}\mathbf{P}$ and $\mathbf{\Upsilon} = \mathbf{P}^T\mathbf{E}\mathbf{P}$. Using Lemma 2 (see Appendix) and (9), we see that $\mathbf{P}^T\mathbf{P} = \mathbf{I}$ and $\mathbf{\Phi}$ and $\mathbf{\Upsilon}$ are diagonal with entries $\lambda_i + \lambda_j$ and $\omega_i + \omega_j$, $1 \le i \le j \le n$, respectively. Notice that the diagonal entry of $\mathbf{\Phi}$ corresponding to the index pair (i, j) is zero if and only if $r + 1 \le i \le j \le n$ (because of (33)), while the diagonal entry of $\mathbf{\Upsilon}$ corresponding to the pair (i, j) is zero if and only if $1 \le i \le j \le r$ (see (34)).

Using the partitioning of $\mathbf{B}$ in (37), we have

$$\mathbf{S}^T\mathbf{J}\mathbf{S} = \begin{bmatrix} 0 & 0 & \mathbf{B}_1^T & \mathbf{I} & 0 \\ 0 & 0 & \mathbf{B}_2^T & 0 & \mathbf{I} \\ \mathbf{B}_1 & \mathbf{B}_2 & 0 & 0 & 0 \\ \mathbf{\Upsilon}_1 & 0 & 0 & \mathbf{\Phi}_1 & 0 \\ 0 & \mathbf{\Upsilon}_2 & 0 & 0 & \mathbf{\Phi}_2 \end{bmatrix} \tag{38}$$

where $\mathbf{\Upsilon} = \mathbf{Diag}(\mathbf{\Upsilon}_1, \mathbf{\Upsilon}_2)$ and $\mathbf{\Phi} = \mathbf{Diag}(\mathbf{\Phi}_1, \mathbf{\Phi}_2)$. We have $\mathbf{\Phi}_1 \succ 0$, since none of the columns of $\mathbf{C}_3$ are included in $\mathbf{B}_1$, and $\mathbf{\Upsilon}_2 \succ 0$, 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 (38), we obtain

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

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

$\square$

Corollary 1 *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 1 is immediate from the standard convergence theory for Newton's method. It is clear that Corollary 1 holds also for less restrictive assumptions on σ, α and β. See [ZTD92] 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 also that Corollary 1 does not guarantee positive definite iterates. These are not required to make the method well-defined, though this condition is imposed by the steplength rule in practice.

A result like Theorem 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 Nesterov-Todd 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 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 [Wri96] and is the subject of active current research for SDP. 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 SDP 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 2 *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 (31), with $r = \mathbf{rank}(X)$. 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 $\overline{r^2}$.

Proof: Clearly,

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

the matrix whose (l,k) element is $\mathbf{tr}\ (XA_lXA_k)$. Let Q and λ_i satisfy (9), and write $\Lambda_1 = \mathbf{Diag}(\lambda_1, \ldots, \lambda_r) \succ 0$, with corresponding eigenvectors collected in Q_1, so that $X = Q_1\Lambda_1Q_1^T$. Let $\mathbf{C}_1$ be the $m \times \overline{r^2}$ matrix introduced in (36), and let $\mathbf{D}_1$ be the $\overline{r^2} \times \overline{r^2}$ diagonal matrix

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

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

$$\mathbf{tr}\ (\Lambda_1Q_1^TA_lQ_1\Lambda_1Q_1^TA_kQ_1) = \mathbf{tr}\ (XA_lXA_k).$$

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 its smallest eigenvalues.

Theorem 3 *Suppose that the assumptions of Theorem 2 hold. Then, if $m > \overline{r^2} > 0$, 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, ω_i^μ satisfy (6), and let $\mathbf{B}^\mu, \mathbf{C}^\mu, \boldsymbol{\Upsilon}^\mu, \boldsymbol{\Phi}^\mu$ be the quantities introduced in Section 3, evaluated at (X^μ, y^μ, Z^μ). We have

$$\mu\,\mathbf{M}^\mu = \mathbf{B}^\mu\mathbf{D}^\mu(\mathbf{B}^\mu)^T \tag{39}$$

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

$$\mathbf{D}^\mu = \mu\,(\boldsymbol{\Upsilon}^\mu)^{-1}\boldsymbol{\Phi}^\mu = \mu\,\mathbf{Diag}\left(\frac{\lambda_i^\mu + \lambda_j^\mu}{\omega_i^\mu + \omega_j^\mu}\right) = \mathbf{Diag}\lambda_i^\mu\lambda_j^\mu, \tag{40}$$

using (7). The primal solution rank r defines a splitting

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

consistent with (36), so that

$$\mu\,\mathbf{M}^\mu = \mathbf{C}_1^\mu\mathbf{D}_1^\mu(\mathbf{C}_1^\mu)^T + \mathbf{C}_2^\mu\mathbf{D}_2^\mu(\mathbf{C}_2^\mu)^T + \mathbf{C}_3^\mu\mathbf{D}_3^\mu(\mathbf{C}_3^\mu)^T. \tag{41}$$

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 $1 \leq i \leq j \leq r$ for $\mathbf{D}_1^\mu$, $1 \leq i \leq r < j \leq n$ for $\mathbf{D}_2^\mu$, and $r+1 \leq i \leq j \leq n$ for $\mathbf{D}_3^\mu$. Although Q^μ and $\mathbf{C}^\mu$ do not generally converge as $\mu \to 0$, Theorem 2 shows that $\mu \mathbf{M}^\mu \to \mathbf{N} = \mathbf{C}_1 \mathbf{D}_1 \mathbf{C}_1^T$, with rank $\overline{r^2}$. By assumption, $m > \overline{r^2} > 0$, so the largest eigenvalue of $\mathbf{N}$ is positive and the smallest is zero. The norms of the second and third terms in (41) 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 [Kat82]). $\square$

For LP, a result like Theorem 3 does not hold. On the contrary, under assumptions of LP nondegeneracy and strict complementarity, the condition number of the LP Schur complement matrix is bounded independent of μ.

5 Stability

We have seen in the previous section that, for nondegenerate SDP's, 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 $\overline{r^2} = m$ and $r = 0$). Consequently, one expects that as $\mu \to 0$, the computation of Δy in (22) 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 [TTT96] found that much higher accuracy is achievable. The main issue is the choice of formulas for Δy and ΔX. Several mathematically equivalent choices are possible, but these have quite different stability properties.

Formulas for Δy and ΔX are given in (22) and (23). Both include the term $\mathbf{F} r_d - r_c$. For the $XZ + ZX$ method, this term is

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

which can be rewritten as

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

However, using this simplification to modify (22) and (23) leads to instability and loss of primal feasibility. It is much better to implement (22) and (23) directly. This is done in the computational experiments reported in Section 7.

The same issue applies to the XZ method. However, direct implementation of (22) and (23) 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 (21) is

$$U = C - \mu X^{-1} - \mathbf{mat}\,\mathbf{A}^T y,$$

we see that (21) is equivalent to

$$\begin{bmatrix} -\mathbf{I} & \widetilde{\mathbf{A}}^T \\ \widetilde{\mathbf{A}} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{vec}\,\widetilde{\Delta X} \\ \Delta y \end{bmatrix} = \begin{bmatrix} \tilde{u} \\ r_p \end{bmatrix} \tag{42}$$

where

$$\widetilde{\Delta X} = \mathbf{G}^{-T}\Delta X = L^{-1}\,\Delta X\, M, \quad \tilde{u} = \mathbf{vec}\,\widetilde{U}, \quad \widetilde{U} = \mathbf{G}U = L^T U M^{-T},$$

and

$$\widetilde{\mathbf{A}} = \mathbf{A}\mathbf{G}^T = \begin{bmatrix} (\mathbf{vec}\,L^T A_1 M^{-T})^T \\ \vdots \\ (\mathbf{vec}\,L^T A_m M^{-T})^T \end{bmatrix}. \tag{43}$$

The solution is given by

$$(\widetilde{\mathbf{A}}\widetilde{\mathbf{A}}^T)\Delta y = r_p + \widetilde{\mathbf{A}}\tilde{u} \tag{44}$$

(which may be solved with a Cholesky factorization) and

$$\Delta X = L\widetilde{\Delta X}M^{-1} = L(\,\mathbf{mat}\,\widetilde{\mathbf{A}}^T\Delta y - \widetilde{U})M^{-1}. \tag{45}$$

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

$$\begin{aligned} \Delta X &= L\left(L^T(\,\mathbf{mat}\,\mathbf{A}^T\Delta y)M^{-T} - L^T U M^{-T}\right)M^{-1} && (46)\\ &= LL^T(\,\mathbf{mat}\,\mathbf{A}^T\Delta y)M^{-T}M^{-1} - LL^T U M^{-T}M^{-1} && (47)\\ &= LL^T\left((\,\mathbf{mat}\,\mathbf{A}^T\Delta y) - U\right)M^{-T}M^{-1}. && (48) \end{aligned}$$

Of these four mathematically equivalent formulas, (45) and (46) give the highest accuracy, with smallest loss of primal feasibility. We used (45) in our computational experiments, with Δy defined by (44).

For the Nesterov-Todd method, $\mathbf{E}^{-1}\mathbf{F}$ is also symmetric positive definite, so similar considerations apply; see [TTT96].

[1]The discussion here is motivated by [TTT96].

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

$$\begin{aligned} Q\Omega Q^T + \sum_{k=1}^{m} y_k A_k &= C, \\ A_k \bullet (Q\Lambda Q^T) &= b_k, \quad k = 1, \ldots, m \\ \Lambda\Omega &= \mu I. \end{aligned} \tag{49}$$

This defines a map from $\mathcal{O}^n \times \mathbb{R}^{2n+m}$ to $\mathbb{R}^{n^2+n+m}$, 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 by [OW95], 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 $s = \mathbf{kvec}\,(S)$, $\Lambda = \mathbf{Diag}(\lambda)$ and $\Omega = \mathbf{Diag}(\omega)$. Define

$$G_Q(\lambda, y, \omega, s) = \begin{bmatrix} \mathbf{vec}\,(C - Q\exp(S)\Omega\exp(-S)Q^T) - \mathbf{A}^T y \\ b - \mathbf{A}\,\mathbf{vec}\,(Q\exp(S)\Lambda\exp(-S)Q^T) \\ \Lambda\Omega e - \mu e \end{bmatrix}, \tag{50}$$

The function G_Q maps $\mathbb{R}^{n^2+n+m}$ 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 $(n^2+n+m)\times(n^2+n+m)$ linear system is:

$$
\begin{aligned}
\Delta\Omega + S\Omega - \Omega S + \sum_{k=1}^{m} \Delta y_k B_k &= H - \Omega && (51)\\
B_k \bullet (\Delta\Lambda + S\Lambda - \Lambda S) &= b_k - B_k \bullet \Lambda, \qquad k = 1, ..., m && (52)\\
\Lambda\,\Delta\Omega + \Omega\,\Delta\Lambda &= \mu I - \Lambda\Omega && (53)
\end{aligned}
$$

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 (51)–(53).
3. Choose steplengths α, β, γ and update the iterates by
$$
\begin{aligned}
\Lambda &\leftarrow \Lambda + \alpha\ \Delta\Lambda\\
y &\leftarrow y + \beta\ \Delta y\\
\Omega &\leftarrow \Omega + \beta\ \Delta\Omega\\
Q &\leftarrow Q(I + \tfrac{1}{2}\gamma S)(I - \tfrac{1}{2}\gamma S)^{-1}.
\end{aligned}
$$

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 (52) can be rewritten as

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

and write

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

Let $\mathbf{diag}(B_k)$ be the vector consisting of the n diagonal entries of B_k and $\mathbf{offdiag}(B_k)$ 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

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

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

Let

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

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 (51) separately, we get the linear system

$$\begin{bmatrix} 0 & 0 & \mathbf{L}^T & I \\ 0 & E & \mathbf{R}^T & 0 \\ \mathbf{L} & \mathbf{R}D & 0 & 0 \\ \Omega & 0 & 0 & \Lambda \end{bmatrix} \begin{bmatrix} \Delta\lambda \\ s \\ \Delta y \\ \Delta\omega \end{bmatrix} = \begin{bmatrix} \mathbf{diag}(H-\Omega) \\ \mathbf{offdiag}(H) \\ v \\ \mu e - \Lambda\Omega e \end{bmatrix}. \tag{54}$$

We denote the matrix on the left-hand side of (54) by $\mathbf{J}_Q$.

Let (X, y, Z) be a solution of SDP satisfying (33), (34). The matrix Q simultaneously diagonalizing X and Z is unique (up to signs of its columns) if and only if

$$\lambda_1 > \cdots > \lambda_r > 0 \quad \text{and} \quad 0 < \omega_{r+1} < \cdots < \omega_n. \tag{55}$$

Theorem 4 *Let $(X, y, Z) = (Q\Lambda Q^T, y, Q\Omega Q^T)$ be a solution of SDP satisfying the strict complementarity and primal and dual nondegeneracy conditions, and also condition (55). 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 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 $D = \mathbf{Diag}(D_1, D_2)$ and $E = \mathbf{Diag}(E_1, E_2)$ 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 $\widetilde{D} = \mathbf{Diag}(I, D_1)$ and $\widetilde{E} = \mathbf{Diag}(0, E_1)$. Permuting the rows and columns accordingly, $\mathbf{J}_Q$ becomes

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

Interchanging the first and fourth rows and the second and last columns, this becomes

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

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 $I + N_1 N_2$ with N_1, N_2 symmetric negative semidefinite. □

Corollary 2 *Consider an SDP whose solution satisfies the strict complementarity and primal and dual nondegeneracy conditions, and also condition (55). 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 2 is more technical than that of Corollary 1, 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

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

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

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

As in LP, this system contains a leading diagonal block, actually two blocks of the form

$$\mathbf{Diag}\left(\frac{\lambda_i}{\omega_i}\right),\ 1 \le i \le n \quad \text{and} \quad \mathbf{Diag}\left(\frac{\lambda_i - \lambda_j}{\omega_j - \omega_i}\right),\ 1 \le i < j \le n.$$

Note the appearance of ratios of eigenvalue differences in contrast to the ratio of eigenvalue sums appearing in (40). Of course, the $\mathbf{L}$ and $\mathbf{R}$ blocks are not independent of the iteration count, as they are in LP. One more step of block elimination then gives the Schur complement

$$\mathbf{M}_Q = [\,\mathbf{L} \quad \mathbf{R}\,] \begin{bmatrix} \Lambda\Omega^{-1} & 0 \\ 0 & -DE^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{L}^T \\ \mathbf{R}^T \end{bmatrix}. \tag{56}$$

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 steplength computation (to find the step to the boundary). Finally, note that the Schur complement matrix is symmetric for the Q matrix, 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 (55) 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 $\Lambda^\mu(\Omega^\mu)^{-1} = \mathbf{Diag}(\lambda_i^\mu/\omega_i^\mu) = \frac{1}{\mu}\mathbf{Diag}((\lambda_i^\mu)^2)$ and

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

Thus,

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

using (39) and (40).

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 [Wri96]. It can easily be extended to the XZ and $XZ + ZX$ methods, as follows. In our implementation, we use the PC rule only if

$$||r_p|| + ||R_d|| < \rho,$$

where ρ is a given threshold value (10^{-4} in the runs reported here). This prevents the PC rule from being applied until an approximately feasible point has been reached. When the condition does not hold, we use the Basic Iteration instead, with $\sigma = 0.25$.

XZ and $XZ{+}ZX$ Methods with Mehrotra Predictor-Corrector Rule

1. Determine ΔX, Δy, ΔZ from (20), using $\mu = 0$ in (18), and symmetrize ΔX in the case of the XZ method.

2. Choose steplengths α, β using (28)–(29), and define

$$\sigma = \left(\frac{(X + \alpha\Delta X) \bullet (Z + \beta\Delta Z)}{X \bullet Z}\right)^3 \tag{57}$$

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

3. Redetermine ΔX, Δy, ΔZ from (20), 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 (28)–(29).

See [TTT96] for a PC version of the Nesterov-Todd method. (Our experiments use (57) in the implementation of all the methods, although [TTT96] use the exponent 2 instead of 3 in (57).)

Computational results are presented in Tables 1 through 4. Tables 1, 2 and 3 report results for randomly generated problems, for $m = n$, with the A_k generated first, and b and C taken to ensure existence of strictly feasible primal and dual points. Such problems are expected to be nondegenerate, and no feasible starting point is known. All methods were initialized with $X = Z = I$, $y = 0$. Table 1 shows results for the $XZ + ZX$, XZ and NT Basic Iteration, using $\sigma = 0.25$ in (27), with various choices for the steplength parameter τ in (28), (29). Table 2 shows results for the PC variants, with threshold $\rho = 10^{-4}$. Table 3 compares the PC algorithms (with $\tau = 0.99$) for different values of n. In all cases, part (a) of the table shows the number of iterations required to reduce the duality gap by a factor of 10^{12}, averaged over 20 to 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 considered to be a failure (indicated by F) if the duality gap and infeasibility norm were not reduced to at least 10^{-4}; these runs are not included in the average statistics. The $>$ symbol in some iteration counts indicates that, in at least one run, the duality gap was not reduced by the desired factor of 10^{12} before the maximum number of iterations was exceeded. 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 higher accuracy (in terms of feasibility). A more aggressive choice of the steplength parameter, $\tau = 0.99$, has little effect on the $XZ + ZX$ method but causes difficulties for the XZ and NT methods. Choosing $\tau = 0.999$ causes the XZ and NT methods to fail in many cases.

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 give a significantly reduced number of iterations (without loss of feasibility) for the $XZ+ZX$ method, but lead to many failures for the XZ and NT algorithms.

Turning to Table 3, we see an iteration count which is essentially constant as n increases, though with some loss of feasibility for larger n.

Table 4 shows results for the Lovász θ function [GLS88] for randomly generated graphs with varying edge density. Here n is the number of vertices in the graph and $m-1$ is the number of edges.

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.

The results show clearly that the $XZ+ZX$ PC method is the most efficient in terms of number of iterations. We make no claim regarding the best overall method in practice, when costs per iteration are taken into account.

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 [AHO94a].

Method	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$
$XZ + ZX$	21.5	21.1	21.1
XZ	21.7	22.1	> 23.3 (F: 11%)
NT	21.6	> 21.9	> 24.9 (F: 25%)

Number of iterations to reduce gap by 10^{12}
Averaged over 100 randomly generated problems
Basic iteration with $\sigma = 0.25$
Starting infeasible, $n = 20$, $m = 20$
F: Failed to reduce gap and feasibility norm to 10^{-4}
Table 1a

Method	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$
$XZ + ZX$	-12.6	-12.6	-12.6
XZ	-11.1	-11.0	-11.0 (F: 11%)
NT	-10.9	-10.8	-10.9 (F: 25%)

Log norm infeasibility
Averaged over same data
Table 1b

Method	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$
$XZ + ZX$	15.0	10.2	9.3
XZ	16.3	14.7	> 16.9 (F: 44%)
NT	15.6	> 22.8	> 30.0 (F: 98%)

Number of iterations to reduce gap by 10^{12}
Averaged over 100 randomly generated problems
Mehrotra predictor-corrector rule
Starting infeasible, $n = 20$, $m = 20$
F: Failed to reduce gap and feasibility norm to 10^{-4}
Table 2a

Method	$\tau = 0.9$	$\tau = 0.99$	$\tau = 0.999$
$XZ + ZX$	-12.1	-12.1	-12.2
XZ	-8.8	-8.8	-9.5 (F: 44%)
NT	-9.1	-8.5	-11.3 (F: 98%)

Log norm infeasibility
Averaged over same data
Table 2b

Method	$n = m = 20$	$n = m = 40$	$n = m = 80$
$XZ + ZX$	10.2	10.5	10.4
XZ	14.7	14.8	14.6
NT	> 22.8	> 23.1	> 22.0

Number of iterations to reduce gap by 10^{12}
Averaged over 20 randomly generated problems
Mehrotra predictor-corrector rule with $\tau = 0.99$
Starting infeasible
Table 3a

Method	$n = m = 20$	$n = m = 40$	$n = m = 80$
$XZ + ZX$	-12.1	-11.3	-10.5
XZ	-8.8	-8.4	-7.8
NT	-8.5	-7.9	-7.5

Log norm infeasibility
Averaged over same data
Table 3b

Method	Density$= 0.25$	Density$= 0.5$	Density$= 0.75$
$XZ + ZX$	9.1	9.2	8.7
XZ	> 13.6 (F: 7%)	> 17.4 (F: 9%)	12.4 (F: 1%)
NT	> 12.3	13.6	13.1

Lovasz θ function
Number of iterations to reduce gap by 10^{12}
Averaged over 100 randomly generated problems
Mehrotra predictor-corrector rule with $\tau = 0.99$
Starting infeasible, $n = 10$
F: Failed to reduce gap and feasibility norm to 10^{-4}
Table 4a

Method	Density$= 0.25$	Density$= 0.5$	Density$= 0.75$
$XZ + ZX$	-14.5	-14.5	-14.6
XZ	-12.2 (F: 7%)	-12.4 (F: 9%)	-13.2 (F: 1%)
NT	-12.0	-12.1	-12.5

Log norm infeasibility
Averaged over same data
Table 4b

Appendix. Symmetric Kronecker Products

Consider the linear operator on $\mathbb{R}^{n\times n}$ defined by the map

$$K \mapsto NKM^T \tag{58}$$

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

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

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

$$(M \otimes N)\,\mathbf{nvec}\,(K) = \mathbf{nvec}\,(NKM^T). \tag{59}$$

Other Kronecker product identities include

$$(M \otimes N)^{-1} = M^{-1} \otimes N^{-1} \quad \text{and} \quad (M \otimes N)(K \otimes L) = MK \otimes NL. \tag{60}$$

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

$$K \mapsto \frac{1}{2}(NKM^T + MKN^T) \tag{61}$$

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

$$(M \circledast N)\,\mathbf{svec}\,(K) = \mathbf{svec}\,(\frac{1}{2}(NKM^T + MKN^T)) \tag{62}$$

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

$$\mathbf{svec}\,(K) = \left[K_{11}, \sqrt{2}K_{12}, \ldots, \sqrt{2}K_{1n}, K_{22}, \ldots, \sqrt{2}K_{2n}, \ldots\ldots, K_{nn}\right]^T. \tag{63}$$

Note that

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

Of course, the ordering used in (63) is arbitrary: the important point is that each element of $\mathbf{svec}\,(K)$ is associated with an index pair (i, j), with $i \leq j$. 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

$$M \circledast N = N \circledast M. \tag{64}$$

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 proof are straightforward.

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

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

Lemma 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*

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

with the corresponding set of orthonormal eigenvectors

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

In other words, if $V = [v_1 \cdots v_n]$, then $V \circledast V$ is an orthogonal matrix of size $\overline{n^2} \times \overline{n^2}$ 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^TBV and dividing its entries by the quantities $(\alpha_i\beta_j + \beta_i\alpha_j)$ componentwise.

References

[AHO94a] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton. A new primal-dual interior-point method for semidefinite programming. In J.G. Lewis, editor, *Proceedings of the Fifth SIAM Conference on Applied Linear Algebra*, pages 113–117, Philadelphia, 1994. SIAM.

[AHO94b] 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, 1994.

[AHO95] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton. Complementarity and nondegeneracy in semidefinite programming. Technical Report 681 (Revised Dec 1995), NYU Computer Science Dept, 1995. To appear in Mathematical Programming (Series B).

[Ali91] F. Alizadeh. *Combinatorial Optimization with Interior Point Methods and Semidefinite Matrices.* PhD thesis, University of Minnesota, 1991.

[GLS88] M. Grötschel, L. Lovász, and A. Schriver. *Geometric Algorithms and Combinatorial Optimization.* Springer-Verlag, New York, 1988.

[HRVW96] C. Helmberg, F. Rendl, R. Vanderbei, and H. Wolkowicz. An interior-point method for semidefinite programming. *SIAM Journal on Optimization*, 1996.

[Kat82] T. Kato. *A Short Introduction to Perturbation Theory for Linear Operators.* Springer-Verlag, New York, 1982.

[KSH94] M. Kojima, S. Shindoh, and S. Hara. Interior-point methods for the monotone linear complementarity problem in symmetric matrices. Technical Report B-282(Revised Oct 95), Department of Information Sciences, Tokyo Inst. of Technology, 1994. To appear in SIAM J. Optimization.

[KSS96] 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. Technical Report B-311, Department of Information Sciences, Tokyo Inst. of Technology, 1996.

[NN94] Y. Nesterov and A. Nemirovskii. *Interior Point Polynomial Methods in Convex Programming.* SIAM, Philadelphia, 1994.

[NT94] Y. Nesterov and M. Todd. Self-scaled barriers and interior-point methods for convex programming. Technical Report TR=1091, School of Operations Research and Industrial Engineering, Cornell University, 1994. To appear in Mathematics of Operations Research.

[NT95] Y. Nesterov and M. Todd. Primal-dual interior-point methods for self-scaled cones. Technical Report TR=1125, School of Operations Research and Industrial Engineering, Cornell University, 1995. To appear in SIAM J. Optimization.

[OW95] M.L. Overton and R.S. Womersley. Second derivatives for optimizing eigenvalues of symmetric matrices. *SIAM Journal on Matrix Analysis and Applications*, 16:697–718, 1995.

[SSK96] M. Shida, S. Shindoh, and M. Kojima. Existence of search directions in interior-point algorithms for the SDP and the monotone SDLCP. Technical Report B-310, Department of Information Sciences, Tokyo Inst. of Technology, 1996.

[TTT96] M.J. Todd, K.C. Toh, and R.H. Tütüncü. On the Nesterov-Todd direction in semidefinite programming. Technical report, School of Operations Research and Industrial Engineering, Cornell University, 1996.

[VB95] L. Vandenberghe and S. Boyd. Primal-dual potential reduction method for problems involving matrix inequalities. *Mathematical Programming (Series B)*, 69:205–236, 1995.

[Wri96] S.J. Wright. Primal-Dual Interior-Point Methods. Manuscript, 1996.

[Zha95] Y. Zhang. On extending primal-dual interior-point algorithms from linear programming to semidefinite programming. Technical Report TR95-20, Dept of Math. and Statistics, University of Maryland, Baltimore County, 1995.

[ZTD92] Y. Zhang, R.A. Tapia, and J.E. Dennis. On the superlinear and quadratic convergence of primal-dual interior point linear programming algorithms. *SIAM Journal on Optimization*, 2:304–324, 1992.