Chapter 4

Search directions in interior-point methods for semidefinite programming

4.1 Introduction

In this chapter we continue our study of interior-point methods for convex programming problems by considering the semidefinite programming (SDP) problem. SDP is a generalization of the linear programming problem with matrix variables instead of vectors and is given in the following standard form:

$$
\begin{array}{lrcl}
(\mathcal{SDP}) \qquad \min_X & C \bullet X & & \\
& A_i \bullet X & = & b_i, \quad i = 1, \ldots, m \\
& X & \succeq & 0,
\end{array}
\tag{4.1}
$$

where $A_i \in S\mathbb{R}^{n\times n}$, $b \in \mathbb{R}^m$, $C \in S\mathbb{R}^{n\times n}$ are given, $X \in S\mathbb{R}^{n\times n}$ is the variable and the A_i's are assumed to be linearly independent. Here $S\mathbb{R}^{n\times n}$ denotes the space of $n \times n$ symmetric matrices, $X \succeq 0$ indicates that X is a positive semidefinite matrix and $G \bullet K := \mathbf{Tr}\, G^T K$ is the trace of the matrix $G^T K$.

$\mathcal{SDP}$ has a striking similarity to $\mathcal{LP}$. Indeed $C \bullet X = \sum_{i,j} c_{ij} x_{ij}$, the objective function of $\mathcal{SDP}$, is like the objective function of $\mathcal{LP}$: the standard inner product of two n^2-dimensional vectors that are obtained by stacking the columns of C and X in long column vectors (likewise for the equality constraints). Although the semidefiniteness constraint can not be replaced by the nonnegativity constraints on the components of the matrix X, the cone of positive semidefinite matrices and its intersection with any affine space are convex. Therefore, the SDP is a convex albeit often nonsmooth optimization problem.

As in the case of LP, there is a dual problem associated with $\mathcal{SDP}$:

$$
(\mathcal{SDD}) \qquad \begin{array}{rcl} \max_{y,S} \ b^T y & & \\ \sum_{i=1}^m y_i A_i + S & = & C \\ S \succeq 0, & & \end{array} \tag{4.2}
$$

where $y \in I\!\!R^m$ and $S \in SI\!\!R^{n \times n}$. Many interior-point algorithms for linear programming problems can be directly extended to solve SDPs, which have a wide range of applications in both continuous and combinatorial optimization as well as control theory, statistics, and eigenvalue problems. We refer the reader to [VB96] for an extensive list of applications.

An interesting application of the SDP arises within the context of the portfolio selection model of Markowitz [Mar52] we described in Chapter 3. There, we assumed that a positive semidefinite covariance matrix Q was available to set up the quadratic programming formulation of the problem. Often, the entries of the covariance matrix are estimated from historical data and one has lower and upper bounds for the entries of the matrix rather than point estimates. In this case the investor is faced with the problem of finding a positive semidefinite matrix Q that satisfies, say, $L \leq Q \leq U$, where the inequalities are componentwise and L and U are the lower and upper bound matrices. This problem can be cast as a semidefinite feasibility problem or as an SDP with some objective function (such as min (or max) $x_0^T Q x_0 = x_0 x_0^T \bullet Q$, for a particular portfolio x_0). This problem is also related to the matrix completion problems and interval problems described in [BGFB94].

We describe two more examples to give a better idea of the range of problems that can be expressed as SDPs. The first one is the *matrix norm minimization* problem. For this problem $p \times q$ real matrices A_k, $k = 0, \ldots, m$ are given and we look for a linear combination A of these matrices where $A := A_0 + \sum_{k=1}^m x_k A_k$ for some x_i, $i = 1, \ldots, m$ and the matrix 2-norm of A is minimized. Vandenberghe and Boyd [VB96] describe how this problem can be cast as an SDP with $m + 2$ square matrices of dimension $p + q$. This problem is closely related to the problem of minimizing the largest eigenvalue of the matrix A as defined above.

The last example of SDP formulations we consider comes from combinatorial optimization. Namely, we examine the semidefinite relaxation of the NP-hard *max-cut* problem which is to find a partition of the node set of an edge-weighted undirected graph into two parts so as to minimize the total weight of the edges that have their ends on different sides of the partition. To obtain a mathematical formulation of the problem we first define the matrix $L := \mathrm{diag}(Ae) - A$, where A is the weighted adjacency matrix of the graph and e is the vector of ones. The matrix L is often called the *Laplacian matrix* corresponding to a graph. Then, the max-cut problem can be formulated as:

$$(\mathcal{MC}) \qquad \max_x \qquad \frac{1}{4}x^T L x \tag{4.3}$$

$$\text{subject to} \quad x \in \{-1,1\}^n, \tag{4.4}$$

where n is the number of nodes and $x_i = 1$ or -1 depending on which side of the partition the node i is on. If we let $X = \frac{1}{4}xx^T$, (4.3) is equivalent to

$$\begin{array}{lllrcl} (\mathcal{MC}2) & \max_X & L \bullet X & & & \\ & & & \operatorname{diag}(X) & = & \frac{1}{4}e \\ & & & \operatorname{rank}(X) & = & 1 \\ & & & X & \succeq & 0, \end{array} \tag{4.5}$$

The semidefinite relaxation of the max-cut problem is obtained by simply dropping the rank condition from (4.5). Goemans and Williamson [GW95] proved that when $A \geq 0$, this relaxation yields a solution that is at most 14% suboptimal, which is much better than any of the previous bounds. Applications to other combinatorial optimization problems can be found in [Ali95].

Our purpose in this chapter is to develop a framework for an efficient implementation of Nesterov and Todd's interior-point method [NT94,NT95] for the solution of SDPs. More precisely, we show how the direction generated by the Nesterov-Todd method can be seen as a Newton direction by considering a reparametrization of the problem. Then, using this perspective we demonstrate how their direction can be used in a Mehrotra-type predictor-corrector approach [Meh92]. Mehrotra's approach has been very successful in the implementation of interior-point methods for LPs and this success gives us the motivation to use it in the case of SDPs. The development of the methods in Nesterov and Todd's papers has a different motivation and lacks a framework to compute the direction using a Mehrotra-type approach.

On the road to describing the Nesterov-Todd direction as a Newton direction we first focus on systems of equations that lead to symmetric search directions for the interior-point methods for the SDP. Systems that generate symmetric direction matrices are desirable because one would like to maintain the symmetry of the iterates. One way of doing this is to determine the direction matrices as a solution to a Newton-like system and then, if necessary, to symmetrize them explicitly. Alternatively, one can symmetrize (with respect to an invertible matrix P) the first order equation corresponding to the complementarity equation in the Newton system, as in Monteiro [Mon95] and Y. Zhang [Zha95a]. This symmetrization of the Newton system gives rise to different search directions for different P. We give a simple proof that the symmetrized Newton system has a unique solution under certain conditions related to those appearing in the work of Shida, Shindoh, and Kojima [SSK96]. Based on this uniqueness result, we show that the search directions described by Nesterov and Todd [NT94,NT95] can be expressed as the solution of a symmetrized Newton system. We note that very recently, independently of our work, Sturm and S. Zhang [SZ95] described how to generate the NT direction as a Newton search direction for primal-dual path-following algorithms for SDP using a symmetric primal-dual transformation, similar to Section 6 in [NT94]; this approach is very different from ours, and does not appear to lead directly to a Mehrotra-type predictor-corrector variant [Meh92]. We will demonstrate in the last section how our approach can be used to develop a Mehrotra-type predictor-corrector interior-point algorithm for SDPs.

Throughout this chapter, $I\!R^n$ denotes the n-dimensional Euclidean space and $\|\cdot\|$ denotes the Euclidean norm. The standard inner product between two real $n \times n$ matrices G and K is the number $G \bullet K := \mathbf{Tr}\, G^T K$, the trace of the matrix $G^T K$. A matrix $G \in I\!R^{n\times n}$ (not necessarily symmetric) is said to be positive definite (resp., positive semidefinite) if $v^T G v > 0$, ($v^T G v \geq 0$, resp.) for all $v \neq 0$. However, all positive (semi)definite matrices considered here are symmetric except in the statements and proofs of Lemma 4.3.1 and Theorem 4.4.1, so we shall assume that positive (semi)definiteness implies symmetry unless explicitly stated otherwise. We write $G \succ 0$ ($G \succeq 0$) if G is symmetric and positive definite (positive semidefinite). For $G \succeq 0$, $G^{1/2}$ denotes the unique symmetric positive semidefinite square root of G.

4.2 The central path and the Newton equation

Here we consider a parametric characterization of the central path for the SDP and use this characterization to derive the Newton equation for approximating points on the central path. We start by describing the optimality conditions for SDP.

If either $\mathcal{SDP}$ or $\mathcal{SDD}$ has a strictly feasible solution, i.e., with X or S positive definite, and both problems have finite optimal solutions, then the optimal values are equal and for any primal-dual optimal pair of solutions X and (y, S) we have that $X \bullet S = 0$ [VB96]. But then, since $X \bullet S = \mathbf{Tr}\, XS = \mathbf{Tr}\,(X^{1/2}S^{1/2})(S^{1/2}X^{1/2})$ it follows that the optimal pair of solutions X and (y, S) satisfy the complementary slackness condition $XS = SX = 0$. Therefore we can write the optimality conditions for $\mathcal{SDP}$ as follows:

$$
\begin{array}{rcrcl}
 & A_i \bullet X & & = & b_i, \quad \text{for } i = 1, \ldots, m, \\
\sum_{i=1}^m y_i A_i & & + \; S & = & C, \\
 & XS & & = & 0, \\
 & X, & S & \succeq & 0.
\end{array}
\tag{4.6}
$$

Similar to the case of LP, we obtain a system defining the central path by perturbing the complementarity equation in (4.6), i.e., for the SDP problem, the central path is defined as the set of solutions to the following equations:

$$
\begin{array}{rcrcl}
 & A_i \bullet X & & = & b_i, \quad \text{for } i = 1, \ldots, m, \\
\sum_{i=1}^m y_i A_i & & + \; S & = & C, \\
 & XS & & = & \mu I, \\
 & X, & S & \succeq & 0.
\end{array}
\tag{4.7}
$$

for all $\mu > 0$. In [Zha95a], Zhang defines the following linear transformation for a given invertible matrix P:

$$
H_P(M) \;\; := \;\; \frac{1}{2}\left[PMP^{-1} + P^{-T}M^TP^T\right], \tag{4.8}
$$

and makes the following observation:

Lemma 4.2.1 (Zhang [Zha95a]) *For $M \in \Re^{n\times n}$ with a real spectrum, nonsingular $P \in \Re^{n\times n}$ and a scalar μ,*

$$
H_P(M) = \mu I \Leftrightarrow M = \mu I.
$$

$\square$

This observation indicates that the last equation in (4.7) can be replaced by

$$
H_P(XS) \;\; = \;\; \mu I \tag{4.9}
$$

without affecting the definition of the central path, since XS is similar to the positive definite matrix $S^{1/2}XS^{1/2}$. (Note that matrices that are similar to a symmetric positive definite matrix have real spectra.)

Let (X, y, S) be the current iterate for the primal and dual semidefinite programming problems. We do not assume that (X, y, S) is feasible; hence the algorithms we consider here for solving SDPs are infeasible-interior-point methods. Most interior-point algorithms choose the next iterate simply as the first iterate of Newton's method applied to the nonlinear system of equations (4.7) (or a related system). Suppose one tries to approximate the point on the central path corresponding to a value $\sigma\mu$, where $\mu := X \bullet S/n$ and $\sigma \in [0, 1]$. If the solution is $(X + \Delta X, y + \Delta y, S + \Delta S)$, then we have

$$\begin{aligned}
A_i \bullet \Delta X \qquad\qquad &= b_i - A_i \bullet X, \quad \text{for } i = 1, \ldots, m, && \text{(4.10-a)} \\
\sum_{i=1}^{m} \Delta y_i A_i \qquad + \Delta S &= C - S - \sum_{i=1}^{m} y_i A_i, && \text{(4.10-b)} \\
\Delta X S \quad + X \Delta S &= \sigma\mu I - (XS + \Delta X \Delta S). && \text{(4.10-c)}
\end{aligned}$$

The linearization of (4.10-c) gives:

$$\Delta X\, S \;+\; X\, \Delta S \;=\; \sigma\mu I - XS. \qquad \text{(4.10-d)}$$

As noted in the introduction, to maintain the symmetry of the iterates X and S it is necessary to generate symmetric ΔX and ΔS matrices in interior-point methods for semidefinite programming. The equation (4.10-b) forces ΔS to be symmetric since the A_i's and C are symmetric. ΔX, on the other hand, is not necessarily symmetric. Indeed, it may happen that the unique solution to the system (4.10-a), (4.10-b), and (4.10-d) has an asymmetric ΔX (see [KSH94] for an example). Much of the recent debate on interior-point methods for SDPs has focused on the issue of the symmetrization of the equations (4.10-c) and (4.10-d). Zhang's characterization of the central path unifies these symmetrizations by viewing them as the transformation of the third equation for different values of the matrix P. In this case, (4.10-c) and (4.10-d) are replaced by

$$H_P(\Delta X\, S \;+\; X\, \Delta S) \;\;=\;\; \sigma\mu I - H_P(XS + \Delta X \Delta S) \qquad \text{(4.10-c}'\text{)}$$

and

$$H_P(\Delta X\, S \;+\; X\, \Delta S) \;\;=\;\; \sigma\mu I - H_P(XS). \qquad \text{(4.10-d}'\text{)}$$

Note that the resulting system (4.10-a), (4.10-b), and (4.10-d$'$) is a linear transformation from $I\!R^m \times SI\!R^{n\times n} \times SI\!R^{n\times n}$ to itself. When the transformation is onto, the solution $(\Delta y, \Delta X, \Delta S)$ to this system will have symmetric ΔX and ΔS. Therefore, we will restrict ΔX to be symmetric in the rest of our discussion. Therefore, when written in full detail the equation (4.10-d$'$) reads

$$\begin{aligned} &P(\Delta X\, S \;+\; X\, \Delta S)P^{-1} + \\ &P^{-T}(S\, \Delta X \;+\; \Delta S\, X)P^{T} \;\;=\;\; 2\sigma\mu I - PXSP^{-1} - P^{-T}SXP^{T}. \end{aligned} \qquad (4.11)$$

When $P = I$, (4.11) becomes

$$\Delta X\, S \;+\; X\, \Delta S \;+ S\, \Delta X \;+\; \Delta S\, X \;\;=\;\; 2\sigma\mu I - XS - SX, \qquad (4.12)$$

which is the symmetrization used by Alizadeh, Haeberly, and Overton [AHO96]. Monteiro used $P = X^{-1/2}$ and $P = S^{1/2}$ [Mon95]. The second of Monteiro's linearizations coincides with the direction introduced by Helmberg, Rendl, Vanderbei, and Wolkowicz [HRVW96]. However, the motivation of [HRVW96] is very different; their direction is obtained by symmetrizing the directions explicitly after solving the unsymmetrized Newton equation. Both $P = X^{-1/2}$ and $P = S^{1/2}$ give directions included in a class developed by Kojima, Shindoh, and Hara [KSH94] independently of Helmberg et al.; see also Lin and Saigal [LS95]. The motivation of [KSH94] is also different, so to view all these directions in a unified way, we follow Monteiro's approach. We denote these the AHO direction and the H..K..M direction. In Section 4.5, we shall demonstrate that the Nesterov-Todd direction also results from such a linearization for a suitable choice of the matrix P.

4.3 Kronecker product notation

To compute the Newton step $(\Delta X, \Delta y, \Delta S)$, it is easier to express the linear systems of equations (4.10-a), (4.10-b) and (4.10-d$'$) in the standard matrix-times-vector-equals-vector form. A useful tool to achieve this goal is the Kronecker product.

The Kronecker product of two matrices $G \in \Re^{m\times n}$ and $H \in \Re^{p\times q}$ denoted by $G \otimes H$ is defined as the partitioned matrix

$$G \otimes H \;\; := \;\; = \begin{bmatrix} g_{11}H & g_{12}H & \dots & g_{1n}H \\ g_{21}H & g_{22}H & \dots & g_{2n}H \\ \vdots & \vdots & & \vdots \\ g_{m1}H & g_{m2}H & \dots & g_{mn}H \end{bmatrix}, \qquad (4.13)$$

which is a matrix of order $mp \times nq$. A related operator is the vector valued function of a matrix $G \in \Re^{m \times n}$:

$$\mathbf{vec}(G) \;:=\; \begin{bmatrix} G_{\cdot 1} \\ G_{\cdot 2} \\ \vdots \\ G_{\cdot n} \end{bmatrix}, \tag{4.14}$$

i.e., the **vec** operator stacks the columns of a matrix in a long column vector. We denote the inverse map of **vec** by **mat** . Our interest in the Kronecker product and the **vec** operator comes from the following property:

$$\mathbf{vec}(GHK^T) \;=\; (K \otimes G)\mathbf{vec}H. \tag{4.15}$$

Using (4.15) we now can write the system (4.10-a), (4.10-b) and (4.10-d$'$) in a 3×3 block equation

$$\begin{pmatrix} 0 & \mathcal{A} & 0 \\ \mathcal{A}^T & 0 & \mathcal{I} \\ 0 & E & F \end{pmatrix} \begin{pmatrix} \Delta y \\ \mathbf{vec}(\Delta X) \\ \mathbf{vec}(\Delta S) \end{pmatrix} \;=\; \begin{pmatrix} r_p \\ \mathbf{vec}(R_d) \\ \mathbf{vec}(R_c) \end{pmatrix}, \tag{4.16}$$

where $\mathcal{I}$ is the identity matrix of order n^2,

$$E \;:=\; P \otimes P^{-T}S + P^{-T}S \otimes P, \tag{4.17}$$

$$F \;:=\; PX \otimes P^{-T} + P^{-T} \otimes PX, \tag{4.18}$$

$$r_p \;:=\; b - \mathcal{A}\,\mathbf{vec}(X), \tag{4.19}$$

$$R_d \;:=\; C - S - \sum_{i=1}^{m} y_i A_i, \tag{4.20}$$

$$R_c \;:=\; 2(\sigma\mu I \;-\; H_P(XS)). \tag{4.21}$$

Here, $\mathcal{A}$ is a matrix of dimension $m \times n^2$ defined by

$$\mathcal{A}^T := [\mathbf{vec}(A_1)\; \cdots \;\mathbf{vec}(A_m)]\,. \tag{4.22}$$

The assumption that the A_i's are linearly independent implies that $\mathcal{A}$ has full row rank. It is interesting to note that (4.16) resembles the Newton systems encountered in the interior-point methods for LP. Also, we remark that the system (4.16) can be expressed in a more economical form by using the symmetrized versions of the Kronecker product and the **vec** operator, see [TTT96].

We now list some useful properties of the Kronecker products that we will use in the rest of the chapter. The reader is referred to, e.g., [Gra81] for further details on the Kronecker product and the **vec** operator.

1. For arbitrary matrices G and K, $(G \otimes K)^T = G^T \otimes K^T$.

2. For matrices G and H of the same dimension, and an arbitrary matrix K, $G \otimes K + H \otimes K = (G + H) \otimes K$.

3. Let G be a square matrix. Then, $G \otimes I$, $I \otimes G$, and $G \otimes I + I \otimes G$ are symmetric if and only if G is.

4. Let G and K be square invertible matrices. Then, $(G \otimes K)^{-1} = G^{-1} \otimes K^{-1}$.

5. Let G, K, H, and L be arbitrary matrices provided the dimensions of matrices are such that the following expressions exist. Then, $(G \otimes K)(H \otimes L) = GH \otimes KL$.

6. Let $\Lambda(G)$ denote the set of eigenvalues of the matrix G. If $\Lambda(G) = \{\lambda_i\}$ and $\Lambda(K) = \{\mu_j\}$, then $\Lambda(G \otimes K) = \{\lambda_i \mu_j\}$. If x_i and y_j are the eigenvectors corresponding to the eigenvalues λ_i and μ_j of G and K, then $\mathbf{vec}(y_j x_i^T)$ is the eigenvector corresponding to the eigenvalue $\lambda_i \mu_j$ of $G \otimes K$.

7. Let G and K be square matrices. $\mathbf{vec}(G)^T \mathbf{vec}(K) = G \bullet K$.

Finally, we present a less straightforward property of the Kronecker product. In the statement and proof of this lemma we do not assume that positive definiteness or semidefiniteness imply symmetry.

Lemma 4.3.1 *Let G be a square (not necessarily symmetric) matrix. Then, $G \otimes I$ and $I \otimes G$ are positive definite (semidefinite) if and only if G is positive definite (semidefinite).*

Proof:

If G is symmetric then the assertion of the lemma follows from Properties 3 and 6 above using the following argument: If G is positive definite (semidefinite), all the eigenvalues of G, I and therefore $G \otimes I$ and $I \otimes G$ are positive (nonnegative). Furthermore, $G \otimes I$ and $I \otimes G$ are also symmetric, therefore they are both positive definite (semidefinite). On the other hand, if $G \otimes I$ and $I \otimes G$ are both positive definite (semidefinite), then their eigenvalues must be all positive (nonnegative) since they are both symmetric. But then, eigenvalues of the symmetric matrix G must also be positive (nonnegative) by Property 5 which indicates that G is positive definite (semidefinite).

Let us now assume that G is not symmetric. For an arbitrary vector x of appropriate dimension we have that

$$x^T (G \otimes I) x = x^T (G \otimes I)^T x = x^T (G^T \otimes I) x = x^T (\frac{G + G^T}{2} \otimes I) x.$$

But the symmetric matrix $\frac{G + G^T}{2}$ is positive definite (semidefinite) if and only if G is and therefore the previous argument indicates that $G \otimes I$ and $I \otimes G$ are positive definite (semidefinite) if and only if G is. $\square$

4.4 Uniqueness of symmetric search directions

In this section we analyze the left-hand-side matrix of the system (4.16) to understand the conditions under which this system has a unique solution with symmetric ΔX and ΔS. In their recent paper [SSK96], Shida, Shindoh, and Kojima established the existence and uniqueness of the solutions of (4.16) for some choices of the matrix P, as well as some other systems that may not be expressible in the form (4.16). When P is chosen to be $X^{-1/2}$, they show that the system (4.16) has a unique solution for all positive definite X and S; however when $P = I$, the condition that $XS + SX$ is positive semidefinite is used to ensure the uniqueness of the solution. Our findings generalize their result to arbitrary invertible matrices P. We first establish the invertibility of the matrices E and F.

Lemma 4.4.1 *Let P be invertible and let X and S be positive definite. Then the matrices $E := P \otimes P^{-T}S + P^{-T}S \otimes P$ and $F := PX \otimes P^{-T} + P^{-T} \otimes PX$ in (4.17) and (4.18) are invertible.*

Proof:

We only prove that E is invertible; the proof for F can be carried out similarly. The proof uses the properties of the Kronecker product that are listed in Section 4.3.

To show that E is invertible, we show that the condition $Ex = 0$ implies that $x = 0$. Let $r = (P \otimes P)x$. Noting that $E = (P^{-T}SP^{-1} \otimes I + I \otimes P^{-T}SP^{-1})(P \otimes P)$, we have $(P^{-T}SP^{-1} \otimes I + I \otimes P^{-T}SP^{-1})r = 0$. Letting $R = \mathbf{mat}(r)$, this can be rewritten as

$$(P^{-T}SP^{-1})R + R(P^{-T}SP^{-1}) \quad = \quad 0. \tag{4.23}$$

Since $P^{-T}SP^{-1}$ is positive definite (because S is), $P^{-T}SP^{-1}$ and $-P^{-T}SP^{-1}$ have no eigenvalues in common. This implies that $R = 0$ is the only solution to the Lyapunov equation (4.23) (see [Gra81, p. 39]). Now $R = 0$ implies that $x = 0$, because $P \otimes P$ is invertible. $\square$

Before we present the main theorem of this section we introduce the following notation:

$$\mathcal{U} \quad := \quad \{\mathbf{vec}\,(U) : U \in S\mathbb{R}^{n\times n}\}. \tag{4.24}$$

Theorem 4.4.1 *The system of equations (4.16) has a unique solution* $(\Delta y, \Delta X, \Delta S)$ $\in \mathbb{R}^m \times S\mathbb{R}^{n\times n} \times S\mathbb{R}^{n\times n}$ *if* $E^{-1}F$ *is positive definite (not necessarily symmetric) on the set* $\mathcal{U}$*. In particular, this condition holds when* $H_P(XS)$ *is positive semidefinite.*

Proof:

For this proof, positive definiteness does not imply symmetry; if matrices are symmetric, we will explicitly say so.

For the first part of the theorem it suffices to show that the 3×3 block system (with an $(m + 2n^2)$-dimensional square matrix)

$$\begin{pmatrix} 0 & \mathcal{A} & 0 \\ \mathcal{A}^T & 0 & \mathcal{I} \\ 0 & E & F \end{pmatrix} \begin{pmatrix} \Delta y \\ \mathbf{vec}(\Delta X) \\ \mathbf{vec}(\Delta S) \end{pmatrix} \quad = \quad 0 \tag{4.25}$$

has only the trivial solution.

Since E is invertible by Lemma 4.4.1, we can use block Gaussian elimination to reduce (4.25) to a Schur complement equation:

$$(\mathcal{A}E^{-1}F\mathcal{A}^T)\,\Delta y \quad = \quad 0. \tag{4.26}$$

Now we claim that $\mathcal{A}E^{-1}F\mathcal{A}^T$ is positive definite. To see this, consider an arbitrary m-dimensional vector $v \neq 0$. Note that $\mathcal{A}^T v \in \mathcal{U}$ since the A_i's are in $S\mathbb{R}^{n\times n}$ and that $\mathcal{A}^T v \neq 0$ since $\mathcal{A}$ has full row rank. Now, since $E^{-1}F$ is assumed to be positive definite on $\mathcal{U}$ it follows that $v^T\mathcal{A}E^{-1}F\mathcal{A}^T v > 0$ so that $\mathcal{A}E^{-1}F\mathcal{A}^T$ is positive definite and thus $\Delta y = 0$ is the only solution to (4.26).

Now the second block equation in (4.25) gives $\Delta S = -\mathbf{mat}(\mathcal{A}^T\,\Delta y) = 0$, and the third block equation gives $\Delta X = -E^{-1}F\,\Delta S = 0$. This shows that the solution to (4.16) exists and is unique.

Now, we will show that the condition that $H_P(XS)$ is positive semidefinite implies that $E^{-1}F$ is positive definite (not only on $\mathcal{U}$ but on the whole space of n^2-dimensional vectors). Consider an arbitrary nonzero vector $g \in \Re^{n^2}$. We shall prove that $g^TE^{-1}Fg > 0$. Let $k := E^{-T}g$ and $K := \mathbf{mat}(k)$, so that k is also nonzero.

Then we have

$$\begin{aligned}
g^TE^{-1}Fg &= k^T\,FE^T\,k \\
&= k^T\,(PX\otimes P^{-T} + P^{-T}\otimes PX)(P^T\otimes SP^{-1} + SP^{-1}\otimes P^T)\,k \\
&= k^T\,(PXP^T\otimes P^{-T}SP^{-1} + P^{-T}SP^{-1}\otimes PXP^T)\,k \\
&\quad +k^T\,(PXSP^{-1}\otimes I + I\otimes PXSP^{-1})\,k \\
&> k^T\,(PXSP^{-1}\otimes I + I\otimes PXSP^{-1})\,k \\
&\geq 0.
\end{aligned}$$

The strict inequality above holds because $PXP^T\otimes P^{-T}SP^{-1}$ and $P^{-T}SP^{-1}\otimes PXP^T$ are symmetric positive definite, a consequence of the fact that PXP^T and $P^{-T}SP^{-1}$ are symmetric positive definite (see Section 4.3). The last inequality follows from the assumption that $H_P(XS)$ is positive semidefinite, which indicates that $PXSP^{-1}$ and $P^{-T}SXP^T$ are positive semidefinite. But then, using Lemma 4.3.1 we see that $PXSP^{-1} \otimes I$ and $I \otimes PXSP^{-1}$ are both positive semidefinite and that the last inequality holds. This completes the proof. $\square$

We note that, if $P = X^{-1/2}$ or, more generally, $P = (X^{1/2}SX^{1/2})^{\alpha}X^{-1/2}$ for some α, then $H_P(XS) = X^{1/2}SX^{1/2}$, which is positive definite whenever X and S are positive definite. Similarly, when $P = S^{1/2}$, or $P = (S^{1/2}XS^{1/2})^{\alpha}S^{1/2}$, we have $H_P(XS) = S^{1/2}XS^{1/2}$, which again is positive definite whenever X and S are. So, as in [SSK96], we see that the system (4.16) always has a unique solution with symmetric ΔX and ΔS for Monteiro's choices of P. When $P = I$ (Alizadeh et al.'s choice of P), the sufficient condition given in Theorem 4.4.1 (i.e. that $H_P(XS)$ is positive semidefinite) coincides with the condition in [SSK96], but there are examples where the coefficient matrix in (4.16) is singular. One such example is given by

$$\mathcal{A} = [-1 \quad 2 \quad 0], \quad X = \begin{bmatrix} 1 & \sqrt{2} \\ \sqrt{2} & 3 \end{bmatrix}, \quad S = \begin{bmatrix} 1 & 0 \\ 0 & 11 \end{bmatrix}.$$

Indeed, in this case the Schur complement matrix in (4.26) is the 1×1 zero matrix.

The following proposition describes a class of matrices P that satisfy the sufficient condition of Theorem 4.4.1.

Proposition 4.4.1 *Let X and S be symmetric and positive definite matrices and let P be an invertible matrix. Let E and F be as in Lemma 4.4.1. Then the following are equivalent:*

(a) PXP^T and $P^{-T}SP^{-1}$ commute;

(b) $PXSP^{-1}$ *is symmetric;*

(c) FE^T *is symmetric; and*

(d) $E^{-1}F$ *is symmetric.*

Condition (a) implies that $H_P(XS)$ *is positive definite. Consequently, any of the conditions above imply that* $E^{-1}F$ *and* FE^T *are both positive definite.*

Proof:

The equivalence of (a) and (b) is immediate: PXP^T and $P^{-T}SP^{-1}$ commute if and only if

$$PXSP^{-1} = (PXP^T)(P^{-T}SP^{-1}) = (P^{-T}SP^{-1})(PXP^T) = P^{-T}SXP^T. \quad (4.27)$$

On the other hand,

$$\begin{aligned} FE^T &= (PX \otimes P^{-T} + P^{-T} \otimes PX)(P^T \otimes SP^{-1} + SP^{-1} \otimes P^T) \\ &= (PXP^T \otimes P^{-T}SP^{-1} + P^{-T}SP^{-1} \otimes PXP^T) \\ &\quad + (PXSP^{-1} \otimes I + I \otimes PXSP^{-1}) \end{aligned}$$

The first two terms in the last equality above are symmetric, since PXP^T and $P^{-T}SP^{-1}$ are. Therefore FE^T is symmetric if and only if the sum of the last two terms are. But this holds if and only if $PXSP^{-1}$ is symmetric. This shows the equivalence of (b) and (c). Finally, the equivalence of (c) and (d) is obvious: $E^{-1}F = E^{-1}(FE^T)E^{-T}$ is symmetric if and only if FE^T is.

Now suppose (a) holds (thus (b) holds). Then

$$H_P(XS) = PXSP^{-1} = (PXP^T)(P^{-T}SP^{-1}),$$

which is the product of two commuting positive definite matrices. This implies that $H_P(XS)$ is positive definite. Consequently, by the proof of Theorem 4.4.1, $E^{-1}F$ and hence FE^T are positive definite. $\square$

The first of the equivalent conditions of Proposition 4.4.1 holds for the two Monteiro choices for P (one of the two matrices is the identity), and also for $P = W^{-1/2}$ with $W = X^{1/2}(X^{1/2}SX^{1/2})^{-1/2}X^{1/2}$, since

$$W^{-1/2}XW^{-1/2} = W^{1/2}SW^{1/2}; \tag{4.28}$$

we will see more of this choice in the next section.

4.5 The Nesterov-Todd direction as a Newton direction

In this section, we describe the system of equations that are used in the primal-dual interior-point methods developed by Nesterov and Todd for a class of problems including SDP and show that the system defining their direction is actually equivalent to (4.16) for some choices of P.

In their two recent papers [NT94,NT95] Nesterov and Todd developed efficient primal-dual algorithms for convex programming problems expressed in a conic form when the cone and its associates barrier are *self-scaled.* In the case of SDP, the search direction they use solves the following system of equations for some σ:

$$\begin{array}{rcrcl} & A_i \bullet \Delta X & & = & 0, \quad i = 1, \ldots, m, \\ \sum_{i=1}^m \Delta y_i A_i & & + \;\; \Delta S & = & 0, \\ & W^{-1}\Delta X W^{-1} & + \;\; \Delta S & = & \sigma\mu X^{-1} - S, \end{array} \tag{4.29}$$

where $W := X^{1/2}(X^{1/2}SX^{1/2})^{-1/2}X^{1/2}$ and $\mu := X \bullet S/n$. The unique solution to

(4.29) is an affine-scaling direction for $\sigma = 0$, and a centering direction for $\sigma = 1$ — see [NT95].

Let us describe how the above system arises. A self-scaled barrier function for the cone of positive semidefinite matrices of order n is

$$F(X) := -\ln \det(X). \tag{4.30}$$

We will not give the definition of self-scaled barriers here, but note that a key consequence is that, for every X in the interior of the primal cone K and every S in the interior of the corresponding dual cone

$$K^* := \{S : \langle S, X\rangle \geq 0 \, \text{for all} \, X \in K\},$$

there is a unique *scaling* point W in the interior of K with

$$F''(W)X = S. \tag{4.31}$$

Here, $F'(X) = X^{-1}$ and $F''(W)$ is the operator taking X to $W^{-1}XW^{-1}$. Therefore, equation (4.31) becomes

$$W^{-1}XW^{-1} = S, \tag{4.32}$$

which leads to

$$W = X^{1/2}(X^{1/2}SX^{1/2})^{-1/2}X^{1/2}$$

after some algebra.

We have to mention that the Nesterov-Todd approach provides general methods, convergence proofs, and complexity estimates applying to any convex programming problem in conic form, as long as the cone is self-scaled and has a self-scaled barrier. Such cones include the nonnegative orthant and the second-order cone in addition to the cone of positive semidefinite matrices and their analysis extend to classes of problems defined on such cones. In contrast, the other methods considered in the previous sections such as the methods using AHO or H..K..M directions are specific to SDP, and require new convergence proofs.

Let us denote by (4.29′) the system (4.29) where the two zeroes on the right-hand side are replaced by r_p and R_d respectively. We now will show that the system (4.29′) defining the Nesterov-Todd (NT) direction is equivalent to (4.16) for special choices of the matrix P. We note that, independently of our work, Sturm and Zhang [SZ95] recently derived the Nesterov-Todd direction as a Newton direction for path-following algorithms using the symmetric transformation given in (4.28).

We first note that if an invertible matrix P satisfies $P^T P = W^{-1}$, then

$$P^T P X S P^{-1} P^{-T} = W^{-1} X S W = S W W^{-1} X = S X. \tag{4.33}$$

Above we used the fact that

$$W^{-1} X = S W, \tag{4.34}$$

which follows from (4.32). Now, (4.33) indicates that

$$P X S P^{-1} \quad = \quad P^{-T} S X P^T.$$

Now using Proposition 4.4.1, we conclude that $H_P(XS)$ is positive definite.

There are several matrices P that satisfy $P^T P = W^{-1}$; for example $P = W^{-1/2}$ works. Indeed, any P matrix of the form $Q\,W^{-1/2}$ for some orthogonal matrix Q satisfies $P^T P = W^{-1}$. Now, we are ready to show that the NT direction is a Newton direction.

Theorem 4.5.1 *Let P be a invertible matrix such that $P^T P = W^{-1}$. Then the solutions to (4.16) and (4.29′) are identical.*

Proof:

It suffices to show the equivalence of the last equation in (4.29) and (4.11). We observe that (4.11) can be rewritten as

$$W^{-1}(X\Delta S + \Delta X S)W + \Delta S X + S\Delta X = 2\sigma\mu I - 2SX \tag{4.35}$$

by premultiplying it with P^T, postmultiplying it with P^{-T}, and using $P^T P = W^{-1}$. On the other hand, the last equation of (4.29) is equivalent to the following two equations:

$$\begin{aligned} W^{-1}\Delta X W^{-1} X + \Delta S X &= \sigma\mu I - SX, & (4.36)\\ S\Delta X + SW\Delta S W &= \sigma\mu SWX^{-1}W - SWSW. & (4.37) \end{aligned}$$

For the first equality above, we postmultiplied the last equation of (4.29) by X, and for the second, we premultiplied it with SW, and postmultiplied it with W. Now adding up these two equations, observing that $SWX^{-1}W = I$, $WSW = X$, and using (4.34), we get (4.35). This proves that the solution to (4.29′) will also satisfy (4.16). However, (4.16) has a unique solution as indicated by the remarks above and Theorem 4.4.1, and therefore these two systems are equivalent. □

4.6 A Mehrotra predictor-corrector modification of the NT direction

Some of the most successful implementations of interior-point methods use Mehrotra-type predictor-corrector approaches [Meh92]. This approach can be seen as a power-series variant of primal-dual algorithms. Every iteration consists of two parts: an affine scaling step where the duality gap is decreased considerably (predictor), and then a centering step that takes the iterate within a small neighborhood of the central path (corrector). This approach was introduced by Mizuno, Todd, and Ye [MTY93]. The novelty of Mehrotra's variant comes from the way he incorporates the information from the predictor step to get a computationally cheaper corrector step. The original description of the Nesterov-Todd directions in (4.29) does not lead to a Mehrotra-type corrector direction (neither does the approach of Sturm and Zhang [SZ95]). Our results in the previous section helps us to fill this gap and describe corrector directions for the NT approach.

To derive the Mehrotra predictor-corrector direction for the NT approach, we suppose that we have approximations δX and δS to the search directions (coming from the predictor step), and substitute them on the right-hand side of (4.10-c$'$). As Theorem 4.5.1 indicates, a choice of P satisfying $P^T P = W^{-1}$ will give us the NT direction. For convenience, we will choose $P = W^{-1/2}$ in the rest of our discussion, but we note that there is a computationally more attractive choice for P, see [TTT96]. The resulting system is:

$$\begin{pmatrix} 0 & \mathcal{A} & 0 \\ \mathcal{A}^T & 0 & \mathcal{I} \\ 0 & E & F \end{pmatrix} \begin{pmatrix} \Delta y \\ \mathbf{vec}(\Delta X) \\ \mathbf{vec}(\Delta S) \end{pmatrix} = \begin{pmatrix} r_p \\ \mathbf{vec}(R_d) \\ \mathbf{vec}(R_s) \end{pmatrix}, \tag{4.38}$$

where E, F, r_p, and R_d are defined in equations (4.17), (4.18), (4.19), and (4.20), and

$$R_s \;=\; 2(\sigma\mu I - H_{W^{-1/2}}(XS + \delta X\,\delta S)) \tag{4.39}$$

(subscript s for second-order). Then,

$$R_s = R_c + R_q,$$

where R_c is as in (4.21) and R_q is the "quadratic correction" $-2\,H_{W^{-1/2}}(\delta X\,\delta S)$.

The system (4.38) involves $m + 2n^2$ linear equations in as many variables. A standard method to solve this large system efficiently is to use block Gaussian elimination to reduce it to a Schur complement equation involving only Δy:

$$\begin{aligned}(\mathcal{A}E^{-1}F\mathcal{A}^T)\,\Delta y &= r_p + \mathcal{A}E^{-1}F\,\mathbf{vec}(R_d) \\ &\quad - \mathcal{A}E^{-1}\,\mathbf{vec}(R_s). \qquad (4.40)\end{aligned}$$

In our case, E^{-1} is not easy to obtain, but we only need $E^{-1}F$ and $E^{-1}\,\mathbf{svec}(R_s)$. It is easy to check directly, by premultiplying by E, that

$$E^{-1}F = W \otimes W, \qquad E^{-1}\,\mathbf{vec}(R_c) = \mathbf{vec}(\sigma\mu S^{-1} - X). \qquad (4.41)$$

However, computing $E^{-1}\,\mathbf{svec}(R_s)$ is somewhat more involved. In addition to $E^{-1}\,\mathbf{vec}(R_c)$, we need to determine

$$\mathbf{vec}(R_{NT}) \;:=\; E^{-1}\,\mathbf{vec}(R_q) = -2E^{-1}\mathbf{vec}(H_{W^{-1/2}}(\delta X\,\delta S)).$$

In other words $E\mathbf{vec}(R_{NT}) = -2\mathbf{vec}(H_{W^{-1/2}}(\delta X\,\delta S))$. Letting $\tilde{R}_{NT} = W^{-1/2}\,R_{NT}\,W^{-1/2}$, we get the following Lyapunov equation:

$$\begin{aligned}(W^{1/2}SW^{1/2})\tilde{R}_{NT} + \tilde{R}_{NT}(W^{1/2}SW^{1/2}) &= -W^{-1/2}(\delta X\,\delta S)W^{1/2} \qquad (4.42) \\ &\quad -W^{1/2}(\delta S\,\delta X)W^{-1/2}.\end{aligned}$$

This system can be solved efficiently by calculating the eigenvalue decomposition of $(W^{1/2}SW^{1/2})$ as in [AHO96,TTT96]. Since the details of the solution to equations like (4.42) is given in these references, we will omit them here.

Once $\tilde{R}_{NT}$ is calculated, it is straightforward to get R_{NT} and $E^{-1}\mathbf{vec}(R_s)$, and then one can solve the Schur complement system (4.40) either by Cholesky factorization of the left-hand-side matrix, or by QR factorization treating the system as a linear least squares problem. This is how we obtain Δy. Then, ΔS and ΔX can be obtained from the following equations:

$$\Delta S \quad = \quad R_d - \mathbf{mat}(\mathcal{A}^T \, \Delta y), \tag{4.43}$$

which follows from the second block equation in (4.38) and

$$\Delta X \quad = \quad \mathbf{mat}[E^{-1} \, \mathbf{svec}(R_s) - E^{-1} F \, \mathbf{vec}(\Delta S)] \tag{4.44}$$

which follows from the third block equation.

We conclude by noting that computational results indicating the good behavior of the Mehrotra-type predictor-corrector implementation of the NT approach are given in [TTT96].