Symmetric primal-dual path following algorithms for semidefinite programming

Jos F. Sturm *
Shuzhong Zhang †

Printed on November 13, 1995

Abstract

In this paper a symmetric primal-dual transformation for positive semidefinite programming is proposed. For standard SDP problems, after this symmetric transformation the primal variables and the dual slacks become identical. In the context of linear programming, existence of such a primal-dual transformation is a well known fact. Based on this symmetric primal-dual transformation we derive Newton search directions for primal-dual path-following algorithms for semidefinite programming. In particular, we generalize: (1) the short step path following algorithm, (2) the predictor-corrector algorithm and (3) the largest step algorithm to semidefinite programming. It is shown that these algorithms require at most $\mathcal{O}(\sqrt{n} \mid \log \epsilon \mid)$ main iterations for computing an ϵ-optimal solution.

Key words: Semidefinite programming; Primal-dual transformation; Primal-dual interior point method.

1. Introduction

The semidefinite programming problem is a generalization of linear programming and has various applications in, among others, system and control theory [2] and combinatorial optimization [1]. A very good overview of the applications is provided by Vandenberghe and Boyd [14]. So far, a significant number of reports has been devoted to generalizing the interior point method to semidefinite programming. The first results were obtained for barrier and potential reduction methods, see e.g. Nesterov and Nemirovsky [11], Vandenberghe and Boyd [13] and Nesterov and Todd [12].

Recently, Kojima, Shindoh and Hara [7], Monteiro [10] and Y. Zhang [15] presented primal-dual interior point algorithms for semidefinite programming (SDP) (or for the complementarity version of the problem), that are generalized from similar algorithms designed for linear programming (LP) (or the linear complementarity problem (LCP)). Their search directions are obtained from a modified Newton equation for approximating a point on the central path. In this paper, we first concentrate on generalizing the v-space concept [6] of linear programming towards semidefinite programming. A Newton equation then follows naturally.

The v-space concept is based on the symmetric duality and the existence of a primal-dual

* Tinbergen Institute Rotterdam, The Netherlands, sturm@opres.few.eur.nl.
† Erasmus University Rotterdam, The Netherlands, zhang@opres.few.eur.nl.

scaling for linear programming. By symmetric duality we mean that the roles of the primal variable x and the dual slack z are interchangeable. In case of the standard LP, there exists for any interior feasible solution pair (x, z) a positive diagonal transformation matrix D, viz. the primal-dual scaling, such that $D^{-1}x = Dz = v$. Moreover, this matrix D and the vector v are uniquely determined by x and z. Conversely, given a positive vector v we will uniquely find feasible solutions x and z such that there is a positive diagonal matrix D satisfying $D^{-1}x = Dz = v$. In this paper we show that this nice property of linear programming can be inherited, to some extent, by semidefinite programming. In view of this symmetric primal-dual transformation, we generalize in this paper three primal-dual interior point algorithms from LP to SDP: the short step primal-dual path following algorithm [5, 9], the predictor-corrector algorithm of Mizuno, Todd and Ye [8] and the largest step algorithm of Gonzaga [4]. These generalized algorithms all possess an iteration bound of $\mathcal{O}(\sqrt{n} \mid \log \epsilon \mid)$, where ϵ is the required precision.

The organization of the paper is as follows. The basic ideas leading to our generic primal-dual method are described in Section 2. In particular, the notions of *symmetric primal-dual transformation* and *primal-dual central path* are introduced. We will then derive the Newton direction for approximating a point on the central path in Section 3. Some technical lemmas bounding the deviation of the iterates from the central path are provided in Section 4. A generic symmetric primal-dual path following method is then proposed in Section 5, and we derive an $\mathcal{O}(\sqrt{n} \log \frac{1}{\epsilon})$ bound on the number of main iterations for three algorithms belonging to this generic class. We conclude the paper in Section 6.

Notation and terminology. The space of symmetric $n \times n$ matrices is denoted by $\mathcal{S}$. Its orthoplement in $\Re^{n\times n}$, viz. the space of skew symmetric $n \times n$ matrices, is denoted by $\mathcal{S}^{\perp}$. Given X and Y in $\Re^{n\times n}$, the standard inner product is defined by

$$X \bullet Y = \mathrm{tr} X^{\mathrm{T}} Y.$$

The Euclidean norm and its associated operator norm, viz. the spectral norm, are both denoted by $\|\cdot\|$. The Frobenius norm of a matrix $X \in \Re^{n\times n}$ is $\|X\|_F = \sqrt{X \bullet X}$. If $X \in \mathcal{S}$ is positive (semi-) definite, we write $X \succ 0$ ($X \succeq 0$). The cone of positive semi-definite matrices is denoted by $\mathcal{S}_+$ and the cone of positive definite matrices is $\mathcal{S}_{++}$. The order n identity matrix is denoted by I.

2. Semidefinite programming and primal-dual transformations

A semidefinite programming (SDP) problem is given as

$$\begin{array}{lll} (P) & \min & C \bullet X \\ & \text{s.t.} & A_i \bullet X = b_i \text{ for } i = 1, 2, \ldots, m \\ & & X \succeq 0 \end{array}$$

where $C \in \mathcal{S}$, $A_1, A_2, \ldots, A_m \in \mathcal{S}$ and $b \in \Re^m$. The decision variable is $X \in \mathcal{S}$.
It is well known that

$$X \succeq 0$$

if and only if

$$\bar{X} := L^{-1} X L^{-\mathrm{T}} \succeq 0$$

for some invertible $L \in \Re^{n \times n}$. The feasible region of (P) can therefore be written in terms of the transformed variable $\bar{X}$ as follows:

$$\mathcal{F}_P(L) := \{\bar{X} \mid (L^{\mathrm{T}} A_i L) \bullet \bar{X} = b_i, i = 1, 2, \ldots, m, \bar{X} \succeq 0\}.$$

Obviously, X is a solution of (P) if and only if $\bar{X}$ is a solution of the SDP

$$\min\{(L^{\mathrm{T}} C L) \bullet \bar{X} \mid \bar{X} \in \mathcal{F}_P(L)\}.$$

Associated with (P) is a dual SDP problem

$$\begin{array}{lll} (D) & \max & b^{\mathrm{T}} y \\ & \text{s.t.} & Z = C - \sum_{i=1}^m y_i A_i \\ & & Z \succeq 0 \end{array}$$

where y is known as the multiplier and Z is the dual slack variable. Transforming the matrices C and $A_1, A_2, \ldots, A_m$ as discussed before yields the transformed dual feasible region

$$\mathcal{F}_D(L) := \{(y, \bar{Z}) \mid \bar{Z} = L^{\mathrm{T}} C L - \sum_{i=1}^m y_i L^{\mathrm{T}} A_i L, \bar{Z} \succeq 0\}.$$

It follows that a dual pair (y, Z) is a solution of (D) if and only if $(y, \bar{Z})$, where $\bar{Z} := L^{\mathrm{T}} Z L$, is a solution of the transformed SDP

$$\max\{b^{\mathrm{T}} y \mid (y, \bar{Z}) \in \mathcal{F}_D(L)\}.$$

Let

$$\mathcal{F}_P^0(L) := \mathcal{F}_P(L) \cap \mathcal{S}_{++}$$

and

$$\mathcal{F}_D^0(L) := \mathcal{F}_D(L) \cap (\Re^m \times \mathcal{S}_{++}).$$

We assume that $\mathcal{F}_P^0(L) \neq \emptyset$ and $\mathcal{F}_D^0(L) \neq \emptyset$. It is well known that under this assumption we have

$$\min\{C \bullet X \mid X \in \mathcal{F}_P(I)\} = \max\{b^{\mathrm{T}} y \mid (y, S) \in \mathcal{F}_D(I)\},$$

i.e. strong duality holds, see e.g. [11]. It is also known, and in fact easy to check, that for given $X \in \mathcal{F}_P(I)$ and $(y,Z) \in \mathcal{F}_D(I)$ there holds

$$C \bullet X - b^{\mathrm{T}} y = X \bullet Z \geq 0.$$

We remark that the duality gap $X \bullet Z$ is invariant under the transformation L, i.e.

$$\bar{X} \bullet \bar{Z} = \mathrm{tr}\bar{X}\bar{Z} = \mathrm{tr}L^{-1}XL^{-\mathrm{T}}L^{\mathrm{T}}ZL = X \bullet Z. \tag{1}$$

Now consider $X \in \mathcal{F}_P^0(I)$ and $(y,Z) \in \mathcal{F}_D^0(I)$. From a primal point of view, an interesting tranformation is $L = X^{1/2}$, yielding

$$\bar{X} = L^{-1}XL^{-\mathrm{T}} = I.$$

Hence,

$$\bar{X} + \{\Delta\bar{X} \in \mathcal{S} \mid \left\|\Delta\bar{X}\right\|_F < 1\} \subset \mathcal{S}_{++}. \tag{2}$$

This means that we can take a large step in the transformed primal space without leaving the positive definite cone. The transformed dual solution $\bar{Z} = X^{1/2}ZX^{1/2}$ however, may be close to the boundary of $\mathcal{S}_{++}$. Based on this transformation, it is therefore not clear whether the duality gap can be reduced substantially. One would rather have a transformation that is neither preoccupied with the primal problem, nor with the dual problem. Such a transformation will be called a symmetric primal-dual transformation.
A *symmetric primal-dual transformation* is by definition a transformation L such that

$$L^{-1}XL^{-\mathrm{T}} = L^{\mathrm{T}}ZL.$$

Lemma 2.1. *Let $X \in \mathcal{S}_{++}$ and $Z \in \mathcal{S}_{++}$. Then*

$$L^{-1}XL^{-\mathrm{T}} = L^{\mathrm{T}}ZL$$

if and only if

$$LL^{\mathrm{T}} = Z^{-1/2}(Z^{1/2}XZ^{1/2})^{1/2}Z^{-1/2}.$$

Proof:
We have $L^{-1}XL^{-\mathrm{T}} = L^{\mathrm{T}}ZL$ if and only if

$$X = LL^{\mathrm{T}}ZLL^{\mathrm{T}} = Z^{-1/2}(Z^{1/2}LL^{\mathrm{T}}Z^{1/2})^2Z^{-1/2}$$

or equivalently,

$$(Z^{1/2}LL^{\mathrm{T}}Z^{1/2})^2 = Z^{1/2}XZ^{1/2}$$

from which it follows that

$$LL^{\mathrm{T}} = Z^{-1/2}(Z^{1/2}XZ^{1/2})^{1/2}Z^{-1/2}.$$

□

Based on Lemma 2.1, we let

$$D(X,Z) := Z^{-1/2}(Z^{1/2}XZ^{1/2})^{1/2}Z^{-1/2}$$

for $X, Z \in \mathcal{S}_{++}$, and we let L be a symmetric primal-dual transformation, i.e. it fulfills $LL^{\mathrm{T}} = D(X,Z)$. Applying this transformation L, we obtain a transformed solution V,

$$V = \bar{X} = \bar{Z}, \tag{3}$$

where $\bar{X} = L^{-1}XL^{-\mathrm{T}}$ and $\bar{Z} = L^{\mathrm{T}}ZL$. Remark that

$$\{V\} = \mathcal{F}_P(L) \cap \mathcal{F}_D(L).$$

From (1), we have

$$X \bullet Z = \|V\|_F^2 .$$

Unlike the primal transformation $LL^{\mathrm{T}} = X$, the symmetric transformation $LL^{\mathrm{T}} = D(X,Z)$ does not discriminate between the primal and the dual. On the other hand, the advantage of the primal transformation, viz. that it maps to the central solution I (see (2)), is also lost in general. However, a property that is similar to (2) does hold with the symmetric transformation $LL^{\mathrm{T}} = D(X,Z)$ if the triple (X,y,Z) is on the *central path.*

Definition 2.1. *The primal-dual central path is the set*

$$\mathcal{C} := \{(X,y,Z) \in \mathcal{F}_P(I) \times \mathcal{F}_D(I) \mid XZ = \mu I, \mu = \frac{trXZ}{n}\}.$$

We remark that the central path is invariant under transformations L, viz.

$$XZ = \mu I$$

if and only if

$$\bar{X}\bar{Z} = L^{-1}XZL = \mu I.$$

In particular, we have for the symmetric transformation $LL^{\mathrm{T}} = D(X,Z)$ that if $(X,y,Z) \in \mathcal{C}$ then

$$V = L^{-1}XL^{-\mathrm{T}} = L^{\mathrm{T}}ZL = \sqrt{\mu}I.$$

This implies that, similar to (2),

$$V + \{\Delta V \in \mathcal{S} \mid \|\Delta V\|_F < \sqrt{\mu}\} \subset \mathcal{S}_{++}$$

if $(X,y,Z) \in \mathcal{C}$. Hence, a large step can be taken in the transformed primal space as well as in the transformed dual space if the current iterate is on the central path. We introduce $\delta(V)$ as a measure of distance to the central path as follows:

$$\delta(V) := \left\| I - \frac{1}{\mu}V^2 \right\|_F ,$$

where

$$\mu := \frac{I \bullet V^2}{n}$$

or equivalently,

$$\mu = X \bullet Z/n.$$

As

$$\mu^2 \delta(V)^2 = \left\|V^2\right\|_F^2 - n\mu^2 = \mathrm{tr}(XZ)^2 - n\mu^2, \tag{4}$$

it follows that the δ measure is not influenced by the way in which we factorize the matrix $D(X,Z)$. To be more precise, if $\hat{L}\hat{L}^{\mathrm{T}} = LL^{\mathrm{T}} = D(X,Z)$, then

$$\delta(V) = \delta(L^{\mathrm{T}}ZL) = \delta(\hat{L}^{\mathrm{T}}Z\hat{L}).$$

Of course, the transformed solution V itself does depend on the way of factorization, i.e.

$$L^{\mathrm{T}}ZL \neq \hat{L}^{\mathrm{T}}Z\hat{L}$$

in general. This is in contrast with linear programming, where there is a one to one correspondence between untransformed primal-dual pairs (X,Z) and symmetrically transformed solutions V, cf. Kojima et al. [6].
Based on the $\delta(V)$ measure, we define a neighborhood of the central path as follows:

$$\mathcal{N}(\beta) := \{V \in \mathcal{S}_{++} \mid \delta(V) \leq \beta\},$$

where $0 < \beta < 1$ is a given constant. If $V \in \mathcal{N}(\beta)$, then obviously

$$V^2 \succeq (1-\beta)\mu I. \tag{5}$$

3. The Newton direction

We will now describe how a direction $(\Delta X, \Delta y, \Delta Z)$ can be determined to improve the current solution (X,y,Z). We let $\mathcal{A}(L)$ denote the span of the matrices $L^{\mathrm{T}}A_iL$, $i = 1,2,\ldots,m$, i.e.

$$\mathcal{A}(L) := \{D_Z \in \mathcal{S} \mid D_Z = -\sum_{i=1}^{m} \Delta y_i L^{\mathrm{T}} A_i L \text{ for some } \Delta y \in \Re^m\}.$$

Its orthoplement in $\mathcal{S}$ is $\mathcal{A}^{\perp}(L)$,

$$\mathcal{A}^{\perp}(L) := \{\Delta X \in \mathcal{S} \mid (L^{\mathrm{T}}A_iL) \bullet \Delta X = 0 \text{ for } i = 1,2,\ldots,m\}.$$

In order to satisfy the equality constraints of (P) and (D), a search direction $(\Delta X, \Delta Z)$ has to satisfy

$$\Delta X \in \mathcal{A}^{\perp}(I) \text{ and } \Delta Z \in \mathcal{A}(I). \tag{6}$$

Letting

$$D_X := L^{-1}\Delta X L^{-\mathrm{T}} \text{ and } D_Z := L^{\mathrm{T}}\Delta Z L \tag{7}$$

denote the search direction in the transformed space, we can rewrite (6) as follows:

$$D_X \in \mathcal{A}^{\perp}(L) \text{ and } D_Z \in \mathcal{A}(L).$$

The maximal feasible step length is

$$t^* := \max\{t \mid V + tD_X \succeq 0 \text{ and } V + tD_Z \succeq 0\}.$$

Let a search direction $(\Delta X, \Delta y, \Delta Z)$ be given. We consider the primal solution $X(t)$ and the dual solution $Z(t)$ along the search direction for a step of length t, i.e.

$$X(t) := X + t\Delta X$$

and

$$Z(t) := Z + t\Delta Z.$$

Given a step length $t \in [0, t^*)$, we let $L(t)$ denote the symmetric primal-dual transformation for which $L^{-1}L(t) \in \mathcal{S}$, i.e.

$$L(t) := L(L^{-1}D(X(t), Z(t))L^{-\mathrm{T}})^{1/2}.$$

It is elementary to verify that indeed $L(t)L(t)^{\mathrm{T}} = D(X(t), Z(t))$. We let $V(t)$ denote the symmetrically transformed solution,

$$V(t) = L(t)^{-1}X(t)L(t)^{-\mathrm{T}} = L(t)^{\mathrm{T}}Z(t)L(t).$$

Letting $\bar{L}(t) := L^{-1}L(t)$, it follows that

$$\begin{aligned} V(t)^2 &= L(t)^{-1}X(t)Z(t)L(t) \\ &= \bar{L}(t)^{-1}(V + tD_X)(V + tD_Z)\bar{L}(t). \end{aligned} \tag{8}$$

Remark that because $\bar{L}(0) = I$, there holds

$$\frac{1}{2}\bar{L}(t)^{-1}V^2\bar{L}(t) + \frac{1}{2}\bar{L}(t)V^2\bar{L}(t)^{-1} = V^2 + o(t). \tag{9}$$

Using the symmetricity of $V(t)$ and $\bar{L}(t)$, we obtain from (8) and (9) that

$$\begin{aligned} V(t)^2 &= \frac{1}{2}\bar{L}(t)^{-1}(V + tD_X)(V + tD_Z)\bar{L}(t) \\ &\quad + \frac{1}{2}\bar{L}(t)(V + tD_Z)(V + tD_X)\bar{L}(t)^{-1} \\ &= V^2 + \frac{1}{2}t(D_XV + VD_Z) + \frac{1}{2}t(D_ZV + VD_X) + o(t) \\ &= V^2 + \frac{1}{2}t(D_X + D_Z)V + \frac{1}{2}tV(D_X + D_Z) + o(t). \end{aligned} \tag{10}$$

Now we can compute the Newton direction towards a solution on the central path $\mathcal{C}$ with a duality gap of $n\gamma\mu$ for some $\gamma \in [0,1]$. Namely, we propose to linearize the nonlinear system

$$V(1)^2 = \gamma\mu I$$

$$D_X \in \mathcal{A}^{\perp}(L), D_Z \in \mathcal{A}(L).$$

Using (10), this yields

$$\frac{1}{2}(D_X + D_Z)V + \frac{1}{2}V(D_X + D_Z) = \gamma\mu I - V^2, \tag{11}$$

$$D_X \in \mathcal{A}^{\perp}(L), D_Z \in \mathcal{A}(L). \tag{12}$$

It is well known from Lyapunov theory [2, 13] that the Sylvester equation (11) has a unique symmetric solution $D_X + D_Z$. Here, the unique solution is

$$D_X + D_Z = \gamma\mu V^{-1} - V. \tag{13}$$

From (12), the Newton direction (D_X, D_Z) now follows as an orthogonal decomposition of $D_X + D_Z$, and therefore,

$$\|D_X\|_F^2 + \|D_Z\|_F^2 = \|D_X + D_Z\|_F^2 . \tag{14}$$

From $V \perp \mu V^{-1} - V$, we obtain

$$\|D_X + D_Z\|_F^2 = \left\|\gamma(\mu V^{-1} - V) - (1-\gamma)V\right\|_F^2 = \gamma^2 \left\|\mu V^{-1} - V\right\|_F^2 + (1-\gamma)^2 n\mu. \tag{15}$$

For the duality gap, we have

$$\|V(t)\|_F^2 = \operatorname{tr}(V + tD_X)(V + tD_Z),$$

so that using $D_X \perp D_Z$ and (13) it follows that

$$\|V(t)\|_F^2 = \|V\|_F^2 + tV \bullet (D_X + D_Z) = (1 - t + \gamma t) \|V\|_F^2 ,$$

i.e.

$$\mu(t) := \|V(t)\|_F^2 /n = (1 - t + \gamma t)\mu. \tag{16}$$

The Newton equations (11)-(12) can be stated in terms of the untransformed variables, using (3) and (7) as follows. Pre-multiplying (13) with L and post-multiplying with L^{T} yields, using (7),

$$\Delta X + D(X,Z)\Delta Z D(X,Z) = \gamma\mu L V^{-1} L^{\mathrm{T}} - L V L^{\mathrm{T}}.$$

Applying (3) to the above relation, we obtain

$$\Delta X + D(X,Z)\Delta Z D(X,Z) = \gamma\mu Z^{-1} - X. \tag{17}$$

We remark that in order to compute the untransformed direction $(\Delta X, \Delta Z)$, the factor L is not needed. In fact, L and V are only used in the analysis and not in the algorithms.

4. Technical results

In this section, we derive some technical lemmas that will be used in proving our main results. We define

$$U(t) := \frac{1}{2}(V + tD_X)(V + tD_Z) - \frac{1}{2}(V + tD_Z)(V + tD_X)$$

and

$$W(t) := \frac{1}{2}(V + tD_X)(V + tD_Z) + \frac{1}{2}(V + tD_Z)(V + tD_X).$$

Notice that $U(t)$ is skew-symmetric, whereas $W(t)$ is symmetric. This implies that

$$U(t) \perp W(t). \tag{18}$$

Lemma 4.1. *Suppose that $\delta(V) < 1$ and $0 \le t < t^*$. There holds*

$$\delta(V(t))^2 = \left\| \frac{1}{\mu(t)} W(t) - I \right\|_F^2 - \frac{\|U(t)\|_F^2}{\mu(t)^2}.$$

Proof:
As in (4), we remark that

$$\mu(t)^2 \delta(V(t))^2 = \left\| V(t)^2 \right\|_F^2 - n\mu(t)^2, \tag{19}$$

where

$$\left\| V(t)^2 \right\|_F^2 = \operatorname{tr}((V + tD_X)(V + tD_Z))^2 = \operatorname{tr}(W(t) + U(t))^2.$$

Now using the skew-symmetricity of $U(t)$ and using (18),

$$\left\| V(t)^2 \right\|_F^2 = (W(t) - U(t)) \bullet (W(t) + U(t)) = \|W(t)\|_F^2 - \|U(t)\|_F^2. \tag{20}$$

As

$$I \bullet W(t) = n\mu(t),$$

it now follows together with (19) and (20) that

$$\begin{aligned} \delta(V(t))^2 &= \frac{\|W(t)\|_F^2 - n\mu(t)^2}{\mu(t)^2} - \frac{\|U(t)\|_F^2}{\mu(t)^2} \\ &= \left\| \frac{1}{\mu(t)} W(t) - I \right\|_F^2 - \frac{\|U(t)\|_F^2}{\mu(t)^2}. \end{aligned}$$

$\square$

Letting

$$G := \frac{1}{2\mu} D_X D_Z + \frac{1}{2\mu} D_Z D_X,$$

it follows from (11) that

$$\begin{aligned} W(t) &= V^2 + \frac{t}{2} V(D_X + D_Z) + \frac{t}{2}(D_X + D_Z)V + t^2 \mu G \\ &= (1-t)V^2 + t\gamma\mu I + t^2 \mu G. \end{aligned}$$

Combining this relation with (16) and Lemma 4.1, it follows that

$$\begin{aligned} \mu(t)^2 \delta(V(t))^2 &= \left\| (1-t)(V^2 - \mu I) + t^2 \mu G \right\|_F^2 - \| U(t) \|_F^2 \\ &\leq \left\| (1-t)(V^2 - \mu I) + t^2 \mu G \right\|_F^2 . \end{aligned}$$

Applying the triangle inequality to the above relation, we obtain

$$\mu(t)\delta(V(t)) \leq (1-t)\mu\delta(V) + t^2 \left\| \mu G \right\|_F . \tag{21}$$

Lemma 4.2. *There holds*

$$\|G\|_F \leq \frac{1}{2\mu} \|D_X + D_Z\|_F^2 .$$

Proof:
Using the triangle inequality and the geometric-arithmetic mean inequality respectively, we have

$$\begin{aligned} \|D_X D_Z + D_Z D_X\|_F &\leq \|D_X D_Z\|_F + \|D_Z D_X\|_F \\ &\leq 2 \|D_X\|_F \|D_Z\|_F \\ &\leq \|D_X\|_F^2 + \|D_Z\|_F^2 , \end{aligned}$$

which together with (14) implies that

$$\|\mu G\|_F = \frac{1}{2} \|D_X D_Z + D_Z D_X\|_F \leq \frac{1}{2} \|D_X + D_Z\|_F^2 .$$

□

Lemma 4.3. *There holds*

$$\frac{1}{\mu} \|D_X + D_Z\|_F^2 \leq \frac{\gamma^2 \delta(V)^2}{1 - \delta(V)} + n(1-\gamma)^2.$$

Proof:
From (15) we know that

$$\|D_X + D_Z\|_F^2 = \gamma^2 \left\| \mu V^{-1} - V \right\|_F^2 + (1-\gamma)^2 n\mu.$$

However,

$$\left\|\mu V^{-1} - V\right\|_F^2 \leq \left\|\mu V^{-1}\right\|^2 \left\| I - \frac{1}{\mu} V^2 \right\|_F^2 = \delta(V)^2 \left\|\mu V^{-1}\right\|^2$$

As $V^2 \succeq (1-\delta(V))\mu I$, it follows that $V^{-2} \preceq \frac{1}{(1-\delta(V))\mu} I$ and therefore,

$$\left\|\mu V^{-1}\right\|^2 \leq \mu/(1-\delta(V)).$$

Combining the above relations, the lemma follows.

□

To summarize the above lemmas, we have

Lemma 4.4. *Suppose $\delta(V) < 1$. For $0 \leq t < t^*$ there holds*

$$(1 - t + \gamma t)\delta(V(t)) \leq (1-t)\delta(V) + \frac{t^2}{2}\left(\frac{\gamma^2 \delta(V)^2}{1-\delta(V)} + n(1-\gamma)^2\right).$$

In the sequel of this section, we investigate how we can choose the parameter γ to guarantee feasibility of the full Newton step. Our basic observation is that if

$$\delta(V(t)) < 1 \text{ for } 0 \leq t \leq 1$$

then, by continuity, $V(1) \succ 0$, implying that the full Newton step is feasible.
Suppose $V \in \mathcal{N}(\beta)$. From Lemma 4.4, it follows for $\gamma \in (0,1]$ that

$$(1 - t + \gamma t)\delta(V(t)) \leq (1-t)\beta + \frac{1}{2}(\gamma t)^2\left(\frac{\beta^2}{1-\beta} + n\left(\frac{1-\gamma}{\gamma}\right)^2\right) \tag{22}$$

yielding the following lemma.

Lemma 4.5. *Let $\gamma \in (0,1]$ and $\beta \in (0,1)$. If $V \in \mathcal{N}(\beta)$ and*

$$\frac{\beta^2}{1-\beta} + n\left(\frac{1-\gamma}{\gamma}\right)^2 \leq 1 \tag{23}$$

then $t^* > 1$ and $V(1) \succ 0$.

Proof:
Let $t \in [0, t^*)$. Based on (22) and (23) we obtain

$$(1 - t + \gamma t)\delta(V(t)) \leq (1-t)\beta + \frac{1}{2}(\gamma t)^2$$

so that

$$\delta(V(t)) < 1 \text{ for } 0 \leq t \leq 1.$$

By continuity this implies that

$$X(t) \succ 0 \text{ and } Z(t) \succ 0 \text{ for } 0 \leq t \leq 1.$$

The lemma is proved.

□

5. Three path-following algorithms

In [3], Gonzaga provides a clear survey of interior point methods for linear programming and linear complementarity problems. He discussed three algorithms, viz. (1) the short step algorithm, (2) the predictor-corrector algorithm and (3) the largest step algorithm. In this section, we generalize these algorithms to semidefinite programming.
We first present a generic path-following algorithm with unspecified parameters $\gamma^{(0)}, \gamma^{(1)}, \ldots$, $t^{(0)}, t^{(1)}, \ldots$ and β. Specific choices for these parameters will then lead to the three above mentioned algorithms.

Algorithm 1.
Data: $\epsilon > 0$, $(X^{(0)}, Z^{(0)})$ *with* $V^{(0)} \in \mathcal{N}(\beta)$.

Step 0 *Set* $k = 0$.

Step 1 *If* $X^{(k)} \bullet Z^{(k)} < \epsilon$ *then stop.*

Step 2 *Choose* $\gamma^{(k)} \in [0,1]$ *and solve* $(\Delta X^{(k)}, \Delta Z^{(k)})$ *from*

$$\Delta X^{(k)} + D(X^{(k)}, Z^{(k)}) \Delta Z^{(k)} D(X^{(k)}, Z^{(k)}) = \gamma^{(k)} \mu^{(k)} (Z^{(k)})^{-1} - X^{(k)},$$

$$\Delta X^{(k)} \in \mathcal{A}^{\perp}(I) \text{ and } \Delta Z^{(k)} \in \mathcal{A}(I),$$

with $\mu^{(k)} = X^{(k)} \bullet Z^{(k)}/n$.

Step 3 *Choose* $t^{(k)}$ *and let* $X^{(k+1)} = X^{(k)} + t^{(k)} \Delta X^{(k)}$, $Z^{(k+1)} = Z^{(k)} + t^{(k)} \Delta Z^{(k)}$.

Step 4 *Set* $k = k+1$ *and return to Step 1.*

We will choose the parameters such that $V^{(k)} \in \mathcal{N}(\beta)$ for all k.

5.1. The short step algorithm

In the generic algorithm, the choice

$$\beta = \frac{1}{2},$$

$$\gamma^{(k)} = \frac{1}{1 + 1/\sqrt{2n}} \text{ for } k = 0, 1, 2, \ldots$$

and

$$t^{(k)} = 1 \text{ for } k = 0, 1, 2, \ldots$$

leads to the so-called short step algorithm. This type of algorithm was studied for linear programming by Monteiro and Adler [9] and Kojima, Mizuno and Yoshise [5]. We let

$$\gamma := \frac{1}{1 + 1/\sqrt{2n}}$$

so that

$$\gamma^{(k)} = \gamma \text{ for } k = 0, 1, 2, \ldots .$$

Based on the technical results of the previous section, we obtain the following result for the short step algorithm.

Lemma 5.1. *For the short step algorithm, we have*

$$V^{(k)} \in \mathcal{N}(\beta)$$

for $k = 0, 1, \ldots$.

Proof:

Let $k \in \{0, 1, 2, \ldots\}$. For the given choice of parameters, we have

$$\frac{\beta^2}{1-\beta} + n(\frac{1-\gamma}{\gamma})^2 = \frac{1}{2} + \frac{1}{2} = 1.$$

Hence, we have from Lemma 4.5 that the maximum step length t^* is greater than 1. Using (22) with $t = 1$ it thus follows that

$$\gamma\delta(V^{(k+1)}) \leq \frac{\gamma^2}{2}$$

which implies

$$\delta(V^{(k+1)}) < \frac{1}{2}.$$

As $V^{(0)} \in \mathcal{N}(\beta)$ by hypothesis, the lemma follows by induction.

□

We are now in a position to prove the polynomiality of the algorithm for obtaining an ϵ-optimal solution.

Theorem 5.1. *The short step algorithm computes an* ϵ*-optimal solution in*

$$\mathcal{O}(\sqrt{n} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$$

iterations.

Proof:
We know from Lemma 5.1 that all iterates of the short step algorithm are contained in $\mathcal{N}(\beta)$. Therefore, the algorithm is well defined. For given $k \in \{0, 1, \ldots\}$ we know from (16) that

$$X^{(k+1)} \bullet Z^{(k+1)} = n\mu^{(k)}(1) = \gamma X^{(k)} \bullet Z^{(k)}.$$

Taking the logarithm on both sides and using the definiton of γ, we arrive at the relation

$$\begin{aligned}
\log X^{(k+1)} \bullet Z^{(k+1)} &= \log \gamma + \log X^{(k)} \bullet Z^{(k)} \\
&= (k+1) \log \gamma + \log X^{(0)} \bullet Z^{(0)} \\
&= \log X^{(0)} \bullet Z^{(0)} - (k+1) \log(1 + \frac{1}{\sqrt{2n}}) \\
&\leq \log X^{(0)} \bullet Z^{(0)} - \frac{k+1}{2\sqrt{2n}}.
\end{aligned}$$

The theorem is proved.

□

5.2. The predictor-corrector algorithm

The short step path following algorithm is of little practical value, because it has a fixed rate of convergence, viz. $\gamma = \frac{1}{1+1/\sqrt{2n}}$. In contrast to the short step algorithm, one has to determine the rate of convergence dynamically in order to perform better than the worst case behavior. The predictor-corrector algorithm of Mizuno, Todd and Ye [8] is such an adaptive step algorithm. The following choice of parameters leads to the predictor-corrector algorithm:

Let $\beta = \frac{1}{2}$. For $k = 0, 1, 2, \ldots$ let
(even iterations: corrector)

$$\gamma^{(2k)} = 1, t^{(2k)} = 1$$

(odd iterations: predictor)

$$\gamma^{(2k+1)} = 0,$$

and $t^{(2k+1)}$ is maximal with respect to

$$V^{(2k+1)}(t) \in \mathcal{N}(\beta) \text{ for } 0 \leq t \leq t^{(2k+1)}.$$

At the start of a corrector iteration, we have $V \in \mathcal{N}(\frac{1}{2})$. From (22) we have

$$\delta(V(1)) \leq \frac{1}{2}\frac{\beta^2}{1-\beta} = \frac{1}{4}.$$

Predictor iterations therefore always start with $V \in \mathcal{N}(\frac{1}{4})$. Using Lemma 4.4, it follows with $\gamma = 0$ that

$$\delta(V(t)) \leq \frac{1}{4} + \frac{nt^2}{2(1-t)}.$$

This means that if

$$0 \leq t \leq \frac{2}{1+\sqrt{1+8n}}$$

then

$$\delta(V(t)) \leq \frac{1}{2},$$

implying that the step lengths in the predictor iterations are never shorter than $\frac{2}{1+\sqrt{1+8n}}$. Similar to Theorem 5.1, this yields:

Theorem 5.2. *The predictor-corrector algorithm computes an ϵ-optimal solution in*

$$\mathcal{O}(\sqrt{n} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$$

iterations.

5.3. The largest step algorithm

The practical performance of the predictor-corrector algorithm can be enhanced by combining the predictor and corrector steps in a single iteration. This idea leads to the largest step algorithm of Gonzaga [4]. In each iteration of the largest step algorithm, we determine the smallest $\gamma \in [0,1]$ such that $V(t) \in \mathcal{N}(\beta)$ for $0 \leq t \leq 1$. In particular, we compute a centering direction $(\Delta X^c, \Delta Z^c)$ from

$$\Delta X^c + D(X,Z)\Delta Z^c D(X,Z) = \mu Z^{-1} - X$$

$$\Delta X^c \in \mathcal{A}^{\perp}(I) \text{ and } \Delta Z^c \in \mathcal{A}(I)$$

and the so-called affine scaling direction $(\Delta X^a, \Delta Z^a)$ from

$$\Delta X^a + D(X,Z)\Delta Z^a D(X,Z) = -X$$

$$\Delta X^a \in \mathcal{A}^{\perp}(I) \text{ and } \Delta Z^a \in \mathcal{A}(I).$$

We can now rewrite (17),(6) as

$$\Delta X_\gamma = \gamma \Delta X^c + (1-\gamma)\Delta X^a \text{ and } \Delta Z_\gamma = \gamma \Delta Z^c + (1-\gamma)\Delta Z^a,$$

where we added a subscript γ to stress the dependence of ΔX and ΔZ on γ.

In the largest step algorithm we set $\gamma = \gamma^*$ where γ^* is obtained by computing

$$\gamma^* := \min\{\gamma' \mid \delta(V_\gamma(1)) \le \beta \text{ for } \gamma' \le \gamma \le 1\},$$

which amounts to solving a quartic equation, as can be seen from Lemma 4.1. We choose $\beta = \frac{1}{2}$. Our analysis of the short step algorithm implies $\gamma^* < \frac{1}{1+1/\sqrt{2n}}$, and this gives the following result:

Theorem 5.3. *The largest step algorithm computes an ϵ-optimal solution in*

$$\mathcal{O}(\sqrt{n} \log \frac{X^{(0)} \bullet Z^{(0)}}{\epsilon})$$

iterations.

6. Concluding remarks

We have made an attempt to generalize the v-space approach for LP and LCP of Kojima et al. [6] to semidefinite programming. Unlike the LCP case, it appeared that there is no one to one correspondence between a primal-dual pair (X, Z) and a symmetrically transformed solution V. However, there does exist a symmetric primal-dual transformation for semidefinite programming, and application of this transformation leads to efficient and elegant algorithms.

References

[1] Alizadeh, F., "Interior point methods in semidefinite programming with applications to combinatorial optimization," *SIAM Journal on Optimization* 5 (1995) 13-51.

[2] Boyd, S., El Ghaoui, L., Feron, E. and Balakrishnan, V., "Linear matrix inequalities in system and control theory," Studies in Applied Mathematics 15 (SIAM, Philadelphia, PA, 1994).

[3] Gonzaga, C.C., "Interior path following algorithms," in J.R. Birge and K.G. Murty, eds., *Mathematical Programming, State of the Art 1994* (University of Michigan, Ann Arbor, 1994) 93-101.

[4] Gonzaga, C.C., "The largest step path following algorithm for monotone linear complementarity problems," Technical Report Nr. 94-07, Technical University of Delft, The Netherlands, 1994.

[5] Kojima, M., Mizuno, S. and Yoshise, A., "A polynomial algorithm for a class of linear complementarity problems," *Mathematical Programming* 44 (1989) 1-26.

[6] Kojima, M., Megiddo, N., Noma, T. and Yoshise, A., *A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems*, Springer-Verlag, Berlin, 1991.

[7] Kojima, M., Shindoh, S. and Hara, S., "Interior-point methods for the monotone linear complementarity problem in symmetric matrices," Research Reports on Information Sciences, B-282, Dept. of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, 1994.

[8] Mizuno, S., Todd, M.J. and Ye, Y., "On Adaptive- Step Primal-Dual Interior-Point Algorithms for Linear Programming," *Mathematics of Operations Research* 18 (1993) 4, 964-981.

[9] Monteiro, R.D.C., Adler, I., "Interior path following primal-dual algorithms. Part 1: linear programming," *Mathematical Programming* 44 (1989) 27-41.

[10] Monteiro, R.D.C., "Primal-dual path following algorithms for semidefinite programming," Technical Report, School of Industrial and Systems Engineering, Georgia Tech, Atlanta, Georgia, U.S.A., 1995.

[11] Nesterov, Y. and Nemirovsky, A., "Interior point polynomial methods in convex programming," Studies in Applied Mathematics 13 (SIAM, Philadelphia, PA, 1994).

[12] Nesterov, Y. and Todd, M.J., "Primal-dual interior-point methods for self-scaled cones," Technical Report 1125, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York, 1995.

[13] Vandenberghe, L. and Boyd, S., "A primal-dual potential reduction method for problems involving matrix inequalities," *Mathematical Programming* 69 (1995) 205-236.

[14] Vandenberghe, L. and Boyd, S., "Semidefinite Programming," submitted to SIAM Review.

[15] Zhang, Y., 'On extending primal-dual interior-point algorithms from linear programming to semidefinite programming," Technical Report, Dept. of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, Maryland, U.S.A., 1995.