

POLYNOMIAL PRIMAL–DUAL CONE AFFINE SCALING FOR SEMIDEFINITE PROGRAMMING

Arjan B. Berkelaar¹, Jos F. Sturm² and Shuzhong Zhang³

November 26, 1996

Report 9667/A

Econometric Institute, Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
URL: <http://www.eur.nl/few/ei/>

ABSTRACT

In this paper we generalize the primal–dual cone affine scaling algorithm of Sturm and Zhang to semidefinite programming. We show in this paper that the underlying ideas of the cone affine scaling algorithm can be naturally applied to semidefinite programming, resulting in a new algorithm. Compared to other primal–dual affine scaling algorithms for semidefinite programming (see, De Klerk, Roos and Terlaky [3]), our algorithm enjoys the lowest computational complexity.

AMS 1991 subject classification: 90C25,90C30.

Key words. Semidefinite Programming, Affine Scaling, Primal–Dual Interior Point Methods.

¹ E-mail: berkelaa@few.eur.nl.

² E-mail: sturm@few.eur.nl.

³ E-mail: zhang@few.eur.nl.

1 Introduction

An important reason for the current interest in semidefinite programming (SDP) is the fact that SDP problems can efficiently be solved by interior point methods, see [1, 8, 3, 15, 16, 18, 20, 21, 23, 24, 28] among others. However, a remarkable but disappointing result was recently obtained by Muramatsu [19], who gave an example of a SDP problem that satisfies all usual regularity conditions, but nonetheless both the short step and the long step variants of the primal affine scaling algorithm converge to a non-optimal point. For linear programming (LP), the primal affine scaling algorithm is one of the more popular interior point methods, since it is both simple and efficient. Although proposed by Dikin [4, 5] as early as in 1967, the affine scaling algorithm only received the proper attention when Barnes [2] and Vanderbei et al.[27] rediscovered it as a natural simplification of Karmarkar’s algorithm [13, 14], by replacing Karmarkar’s projective transformations by affine transformations.

By Karmarkar’s projective scaling transformation, the original linear objective function becomes a fractional linear function. The search direction used in Karmarkar’s method (in the transformed space) is obtained by optimizing only the numerator of the transformed fractional linear function (thus a simplification) over an inscribed sphere of the solution space. This search direction is in general not a descent direction for the original linear objective, and hence a potential function [13] is used. In contrast to this strategy, Padberg [22] derived a search direction by optimizing the *entire fractional objective* over the sphere. Similar algorithms were independently proposed and analyzed by Goldfarb and Xiao [6], Gonzaga [7] and Jan and Fang [9]. In fact, one may obtain the search direction derived in Padberg [22] and Goldfarb and Xiao [6] by optimizing the original linear objective over a conic section, using merely an affine transformation. In this sense, the direction may be called a *cone affine scaling* direction.

Monteiro et al.[17] proposed a variant of the affine scaling algorithm for LP that is symmetric in the duality, henceforth called a *primal–dual* affine scaling algorithm. Other primal–dual affine scaling algorithms were proposed by Jansen et al.[10, 11, 12] and Sturm and Zhang [26]. Although polynomiality of the original method of Dikin is considered unlikely, the primal–dual variants have been shown to have polynomial iteration bounds. The primal–dual affine scaling algorithm of Monteiro, Adler and Resende [17] requires $O(nL^2)$ iterations, whereas the primal–dual Dikin-type affine scaling algorithm of Jansen, Roos and Terlaky [11] and the primal–dual cone affine scaling algorithm of Sturm and Zhang [26] solve linear programs in only $O(nL)$ and $O(\sqrt{n}L)$ main iterations respectively.

Interestingly, De Klerk et al.[3] recently extended both the primal–dual affine scaling algorithm of Monteiro, Adler and Resende [17] and the primal–dual Dikin-affine scaling algorithm of Jansen, Roos and Terlaky [11] to SDP. They derived iteration bounds that generalize the results known for the respective LP counterparts. In this paper we will generalize the primal–dual cone affine scaling algorithm of Sturm and Zhang [26] to SDP. As is the case for the other two primal–dual affine scaling variants, the iteration bound of this algorithm is the same as for LP. Hence, of all affine scaling variants for SDP, the cone affine scaling algorithm has the best worst-case behavior.

This paper is organized as follows. In Section 2, we will discuss the underlying ideas of the cone affine scaling method. We show in Section 3 how the iterates of the cone affine scaling algorithm for semidefinite programming can be computed. The polynomiality of this algorithm will be

established in Section 4.

We will use the following notation. The set \mathcal{S}^n denotes the set of all symmetric matrices in $\mathbb{R}^{n \times n}$. Moreover, we will denote \mathcal{S}_{++}^n (\mathcal{S}_+^n) the set of symmetric positive (semi)definite matrices. The inner product of two matrices X and Y , denoted as $X \bullet Y$, is defined as $\text{tr}(X^T Y)$. The corresponding Frobenius norm $\|X\|_F$ of a matrix X is defined as $\sqrt{X \bullet X}$. The spectral norm of a matrix X is denoted by $\|X\|$. Given $X \in \mathcal{S}^n$, we let $\lambda_{\min}(X)$ denote its smallest eigenvalue. The identity matrix will be denoted as I . The direct sum of two matrices X and Y is denoted by $X \oplus Y$, i.e.

$$X \oplus Y = \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}.$$

2 Cone affine scaling fundamentals

Consider the primal SDP problem (P)

$$(P) \quad \min \{ C \bullet X : A_i \bullet X = b_i, i = 1, \dots, m, X \succeq 0 \},$$

and its dual

$$(D) \quad \max \left\{ b^T y : \sum_{i=1}^m y_i A_i + Z = C, Z \succeq 0 \right\},$$

where $X, Z, C, A_1, \dots, A_m \in \mathcal{S}^n$ and $b, y \in \mathbb{R}^m$. We make the common assumption that positive definite solutions X and Z exist which are feasible for (P) and (D) respectively (primal–dual Slater condition). In addition, we assume that $n \geq 2$.

As is well-known, the primal–dual Slater condition implies that both (P) and (D) have optimal solutions. Moreover, if the triple (X, y, Z) satisfies the feasibility requirements

$$A_i \bullet X = b_i \text{ for } i = 1, 2, \dots, m, \tag{1}$$

$$\sum_{i=1}^m y_i A_i + Z = C, \tag{2}$$

then

$$X \bullet Z = C \bullet X - b^T y.$$

The quantity $X \bullet Z$ is known as the *duality gap*. Therefore, solving the primal–dual pair (P) and (D) is equivalent to minimizing the duality gap:

$$\min \{ X \bullet Z : (X, Z) \in \mathcal{M}, X \oplus Z \in \mathcal{S}_+^{2n} \}, \tag{3}$$

where \mathcal{M} is the linear manifold of pairs $(X, Z) \in \mathcal{S}^n \times \mathcal{S}^n$ that satisfy (1) and (2) for some $y \in \mathbb{R}^m$.

The cone affine scaling algorithm generates a sequence of feasible solution pairs (X^1, Z^1) , (X^2, Z^2) , \dots with

$$(X^{i+1}, Z^{i+1}) = \arg \min \{ X \bullet Z : (X, Z) \in \mathcal{M}, X \oplus Z \in \mathcal{K}_i \}, \tag{4}$$

where for each iteration $i = 0, 1, \dots$, the set $\mathcal{K}_i \subseteq \mathcal{S}_+^{2n}$ is an inscribed convex cone of the semidefinite cone \mathcal{S}_+^{2n} . The cones $\mathcal{K}_0, \mathcal{K}_1, \dots$ will be chosen in such a way that

- the cone program (4) can be solved analytically, and
- the duality gaps $X^1 \bullet Z^1, X^2 \bullet Z^2, \dots$ converge to zero at least linearly.

In particular, we will consider the case where \mathcal{K} is a circular cone (also known as second-order cone). First, we will derive some relations between circular cones and the cone of semidefinite matrices. Then, we will discuss a class of linear transformations that affect the circular cone, but leave the semidefinite cone untouched. In this way, we obtain a class of inscribed cones of the semidefinite cone. The cones $\mathcal{K}_0, \mathcal{K}_1, \dots$ will be chosen from this class.

2.1 The semidefinite cone and circular cones

Consider the circular cone

$$\mathcal{C}_{\text{in}}^n := \{Y \in \mathcal{S}^n : \text{tr}Y \geq \sqrt{n-1} \|Y\|_F\}.$$

The following lemma states that $\mathcal{C}_{\text{in}}^n$ is an inscribed cone of the semidefinite cone \mathcal{S}_+^n .

Lemma 2.1 *There holds*

$$\mathcal{C}_{\text{in}}^n \subseteq \mathcal{S}_+^n.$$

The above result follows immediately from Lemma A.2 in the Appendix of this paper. In fact, $\mathcal{C}_{\text{in}}^n$ is the largest inscribed circular cone of the semidefinite cone \mathcal{S}_+^n . The following lemma characterizes the symmetric matrices that are both on the boundary of \mathcal{S}_+^n and on the boundary of $\mathcal{C}_{\text{in}}^n$.

Lemma 2.2 *Let $Y \in \mathcal{S}^n$, be a symmetric matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. The following two statements are equivalent:*

1. $Y \in \mathcal{C}_{\text{in}}^n$ and $\lambda_n \leq 0$,
2. $\lambda_n = 0$ and $\lambda_1 = \lambda_2 = \dots = \lambda_{n-1} \geq 0$.

Proof: It is straightforward to verify that (2) implies (1). To show that the converse is also true, assume that $Y \in \mathcal{C}_{\text{in}}^n$ and $\lambda_n \leq 0$. Since $\mathcal{C}_{\text{in}}^n \subseteq \mathcal{S}_+^n$, it follows that $\lambda_n = 0$. Let $u \in \mathbb{R}^n$ be defined as

$$u_1 = u_2 = \dots = u_{n-1} = 1, \quad u_n = 0.$$

Then, using the fact that $\lambda_n = 0$,

$$\text{tr}Y = \sum_{i=1}^{n-1} \lambda_i = u^T \lambda, \quad \|Y\|_F = \|\lambda\|.$$

By definition, $Y \in \mathcal{C}_{\text{in}}^n$ implies that

$$u^T \lambda = \text{tr}Y \geq \sqrt{n-1} \|Y\|_F = \|u\| \|\lambda\|,$$

which means, by the inequality of Cauchy-Schwarz, that $u^T \lambda = \|u\| \|\lambda\|$, hence λ is a multiple of u . This completes the proof. \square

The smallest circumscribing circular cone of \mathcal{S}_+^n is given as

$$\mathcal{C}_{\text{out}}^n := \{Y \in \mathcal{S}^n : \text{tr}Y \geq \|Y\|_F\}.$$

Indeed, if $Y \in \mathcal{S}_+^n$ then it has n nonnegative eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, and

$$\text{tr}Y = \|\lambda\|_1 \geq \|\lambda\| = \|Y\|_F.$$

For $n = 2$, we have $\mathcal{C}_{\text{in}}^2 = \mathcal{C}_{\text{out}}^2$, so that in this case the semidefinite cone is circular itself.

An interesting property of the largest inscribed circular cone $\mathcal{C}_{\text{in}}^n$ is that it contains the so-called Dikin-sphere [3]:

$$\{Y \in \mathcal{S}^n : \|Y - I\|_F \leq 1\}.$$

Lemma 2.3 *Let $Y \in \mathcal{S}^n$. If $\|Y - I\|_F \leq 1$ then $Y \in \mathcal{C}_{\text{in}}^n$.*

Proof: Assume that $\|Y - I\|_F \leq 1$. Observe that we have $Y \neq 0$ (since $n \geq 2$) and

$$1 \geq \|Y - I\|_F^2 = \|Y\|_F^2 - 2\text{tr}Y + n.$$

Rearranging terms, we get

$$\text{tr}Y \geq \frac{\|Y\|_F^2 + (n-1)}{2}.$$

However, it follows from the arithmetic-geometric mean inequality that

$$\left(\|Y\|_F + \frac{n-1}{\|Y\|_F} \right) \geq 2\sqrt{n-1},$$

and hence

$$\text{tr}Y \geq \sqrt{n-1} \|Y\|_F.$$

\square

A well known property (Sylvester's law of inertia) of the semidefinite cone is that for any invertible matrix P of order n , we have

$$Y \in \mathcal{S}_+^n \text{ if and only if } PYP^T \in \mathcal{S}_+^n. \quad (5)$$

Now consider the class of linearly transformed circular cones

$$\mathcal{C}^n(\beta, P) := \{Y \in \mathcal{S}^n : \text{tr}PYP^T \geq \sqrt{(1-\beta^2)n} \|PYP^T\|_F\}, \quad (6)$$

with $\beta \in [0, 1]$ and P an invertible matrix. From Lemma 2.1 we know that

$$Y \in \mathcal{C}^n\left(\frac{1}{\sqrt{n}}, P\right) \Rightarrow PYP^T \in \mathcal{C}_{\text{in}}^n \subseteq \mathcal{S}_+^n \Rightarrow Y \in \mathcal{S}_+^n,$$

where the last implication follows from (5). Hence,

$$\mathcal{C}^n(\beta, P) \subseteq \mathcal{C}^n\left(\frac{1}{\sqrt{n}}, P\right) \subseteq \mathcal{S}_+^n,$$

for all $0 \leq \beta \leq 1/\sqrt{n}$ and invertible P . Remark also that

$$\mathcal{C}^n\left(\frac{1}{\sqrt{n}}, Q\right) = \mathcal{C}_{\text{in}}^n, \quad \mathcal{C}^n\left(\sqrt{1 - \frac{1}{n}}, Q\right) = \mathcal{C}_{\text{out}}^n,$$

for any orthogonal matrix Q , because circular cones are invariant under orthogonal transformations.

2.2 The symmetric primal–dual transformation

Consider a pair $(X, Z) \in \mathcal{M}$ such that X and Z are both positive definite. Let Λ_{XZ} denote a positive diagonal matrix whose diagonal entries are the eigenvalues of the matrix XZ . Define

$$V := \Lambda_{XZ}^{1/2}.$$

It is shown in [24] that there exists an invertible matrix L_d such that

$$L_d^{-1} X L_d^{-T} = L_d^T Z L_d = V.$$

We see that the pair (V, V) is feasible for the linearly transformed SDP

$$\min\{\bar{X} \bullet \bar{Z} : (L_d \bar{X} L_d^T, L_d^{-T} \bar{Z} L_d^{-1}) \in \mathcal{M}, (\bar{X} \oplus \bar{Z}) \in \mathcal{S}^{2n}\}.$$

This elucidates that the above transformation is known as the symmetric primal–dual transformation. Due to the invertibility of L_d , the above problem admits a one-to-one correspondence with the untransformed problem (3). Remark that the duality gap of the pair (X, Z) in the original SDP pair is the same as the duality gap of the pair (V, V) for the transformed SDP pair:

$$\text{tr} XZ = \text{tr} \Lambda_{XZ} = \text{tr} V^2.$$

It follows that the optimal solution set will be approached if $V \rightarrow 0$.

2.3 Cone affine scaling algorithm

The cone affine scaling algorithm to be introduced is iterative in nature. Suppose that in the i -th iteration of the cone affine scaling algorithm, we have an iterate $(X^i, Z^i) \in \mathcal{M}$ with X^i and Z^i positive definite. We compute L_d^i such that

$$(L_d^i)^{-1} X^i (L_d^i)^{-T} = (L_d^i)^T Z^i L_d^i,$$

exactly as discussed above. The next iterate is then defined as the solution (X^{i+1}, Z^{i+1}) of the cone program (4), with $\mathcal{K}_i := \mathcal{C}^{2n}(\beta/\sqrt{2}, (L_d^i)^{-1} \oplus (L_d^i)^T)$, for a suitable parameter $\beta \in (0, 1/\sqrt{n})$. We will derive an analytic expression for the solution (X^{i+1}, Z^{i+1}) of (4) in this section. Obviously, this solution does not exist if $\mathcal{K}_i \cap \mathcal{M} = \emptyset$. However, we will show in Section 4 that this will not occur if we choose $\beta = 1/(4\sqrt{n})$.

Cone Affine Scaling Algorithm

Input:

An initial feasible solution (X^0, Z^0) , a parameter $\beta \in (0, 1/\sqrt{n})$ and tolerance $\epsilon > 0$

begin

$i := 0;$

while $X^i \bullet Z^i < \epsilon$ **do**
begin

 Compute L_d^i satisfying

$$(L_d^i)^{-1} X^i (L_d^i)^{-T} = (L_d^i)^T Z^i L_d^i.$$

 Calculate the new iterates

$$(X^{i+1}, Z^{i+1}) := \arg \min \{X \bullet Z : (X, Z) \in \mathcal{M}, (X \oplus Z) \in \mathcal{K}_i\},$$

$$\text{with } \mathcal{K}_i := \mathcal{C}^{2n}(\beta/\sqrt{2}, (L_d^i)^{-1} \oplus (L_d^i)^T),$$

 and let $i := i + 1$.

end

end

3 Search directions

In this section we will prove that the above algorithm can be implemented in an explicit way, i.e. we show that at each iteration the search directions can be computed analytically. Since the algorithm is iterative, we will illustrate this fact by amplifying how one particular iteration should proceed. For notational convenience, let $X \succ 0, Z \succ 0$ and $(X, Z) \in \mathcal{M}$ be the current iterates under consideration. We need to compute a new solution as follows

$$\begin{aligned} X^+ &:= X + 2\Delta X, \\ Z^+ &:= Z + 2\Delta Z, \end{aligned}$$

where ΔX and ΔZ are displacements satisfying the feasibility requirements

$$\begin{aligned} A_i \bullet \Delta X &= 0, \quad i = 1, \dots, m \\ \sum_{i=1}^m \Delta y_i A_i + \Delta Z &= 0 \text{ for some } y \in \mathbb{R}^m, \end{aligned} \tag{7}$$

and the constraint

$$(X^+ \oplus Z^+) \in \mathcal{C}^{2n}(\beta/\sqrt{2}, (L_d)^{-1} \oplus (L_d)^T). \tag{8}$$

The displacements in the transformed space are given by

$$\begin{aligned} D_x &= L_d^{-1} \Delta X L_d^{-T}, \\ D_z &= L_d^T \Delta Z L_d. \end{aligned} \tag{9}$$

Let $D_v := D_x + D_z$ and notice from (7) and (9) that $D_x \perp D_z$. Hence, D_x and D_z form an orthogonal decomposition of D_v . Remark that

$$(L_d^{-1} \oplus L_d^T)(X^+ \oplus Z^+)(L_d^{-T} \oplus L_d) = (V + 2D_x) \oplus (V + 2D_z).$$

Constraint (8) can therefore be rewritten as follows:

$$\text{tr}((V + 2D_x) \oplus (V + 2D_z)) \geq \sqrt{\left(1 - \frac{\beta^2}{2}\right)2n} \|(V + 2D_x) \oplus (V + 2D_z)\|_F. \quad (10)$$

However, using the fact that $D_x \perp D_z$, we have

$$\|D_x\|_F^2 + \|D_z\|_F^2 = \|D_x + D_z\|_F^2 = \|D_v\|_F^2,$$

so that

$$\begin{aligned} \|(V + 2D_x) \oplus (V + 2D_z)\|_F^2 &= \|V + 2D_x\|_F^2 + \|V + 2D_z\|_F^2 \\ &= \|V\|_F^2 + \|V + 2D_v\|_F^2 \\ &= \|V \oplus (V + 2D_v)\|_F^2. \end{aligned} \quad (11)$$

Moreover,

$$\text{tr}((V + 2D_x) \oplus (V + 2D_z)) = \text{tr}(V \oplus (V + 2D_v)). \quad (12)$$

Combining (10)–(12), it follows that (10) is equivalent with

$$\text{tr}(V \oplus (V + 2D_v)) \geq \sqrt{\left(1 - \frac{\beta^2}{2}\right)2n} \|V \oplus (V + 2D_v)\|_F,$$

i.e. $V \oplus (V + 2D_v) \in \mathcal{C}^{2n}(\beta/\sqrt{2}, I)$. In order to solve the cone program (4), we have to minimize the duality gap

$$X^+ \bullet Z^+ = \text{tr}((V + 2D_x)(V + 2D_z)) = \|V\|_F^2 + 2V \bullet D_v.$$

It follows that D_v is the solution of

$$\min\{V \bullet D_v : \text{tr}(V \oplus (V + 2D_v)) \geq \sqrt{\left(1 - \frac{\beta^2}{2}\right)2n} \|V \oplus (V + 2D_v)\|_F\}. \quad (13)$$

Denoting the angle between V and the identity matrix I by ϕ , i.e.

$$\phi = \arccos\left(\frac{\text{tr}V}{\sqrt{n}\|V\|_F}\right),$$

we let

$$\delta := \sin(\phi).$$

In other words

$$\delta = \sqrt{1 - \frac{(\text{tr}V)^2}{n\|V\|_F^2}}. \quad (14)$$

The crux of the cone affine scaling algorithm is that (4), which is equivalent to (13), can be solved analytically as the following lemma shows.

Lemma 3.1 For $\delta < \beta < 1$, the solution of (13) is

$$D_v = \frac{\gamma - 1}{2(1 - \beta^2)\gamma} \frac{\text{tr}V}{n} I - \frac{\gamma + 1}{2\gamma} V \quad (15)$$

where

$$\gamma = \sqrt{\frac{2 - \beta^2 - \delta^2}{\beta^2 - \delta^2}}. \quad (16)$$

Proof: Consider the convex program (13). The Lagrangian is

$$\mathcal{L}_\lambda(D_v) = V \bullet D_v + \lambda \left(\frac{1}{2} \sqrt{(2 - \beta^2)n(\|V\|_F^2 + \|V + 2D_v\|_F^2)} - \text{tr}(V + D_v) \right)$$

with gradient

$$\nabla \mathcal{L}_\lambda(D_v) = V + \eta(V + 2D_v) - \lambda I,$$

where we let

$$\eta := \lambda \sqrt{\frac{n(2 - \beta^2)}{\|V\|_F^2 + \|V + 2D_v\|_F^2}}. \quad (17)$$

The Karush–Kuhn–Tucker optimality conditions are

$$\nabla \mathcal{L}_\lambda(D_v) = V + \eta(V + 2D_v) - \lambda I = 0, \quad (18)$$

$$\lambda \geq 0 \text{ and } \sqrt{\frac{n(1 - \beta^2/2)}{2}(\|V\|_F^2 + \|V + 2D_v\|_F^2)} - \text{tr}(V + D_v) \leq 0, \quad (19)$$

$$\lambda \left(\sqrt{\frac{n(1 - \beta^2/2)}{2}(\|V\|_F^2 + \|V + 2D_v\|_F^2)} - \text{tr}(V + D_v) \right) = 0. \quad (20)$$

Rearranging the terms in (18), we have

$$2\eta(V + D_v) = \lambda I + (\eta - 1)V. \quad (21)$$

From (20) we obtain

$$\begin{aligned} 0 &= 2\eta \left(\sqrt{\frac{n(1 - \beta^2/2)}{2}(\|V\|_F^2 + \|V + 2D_v\|_F^2)} - \text{tr}(V + D_v) \right) \\ &= 2\lambda n(1 - \beta^2/2) - 2\eta \text{tr}(V + D_v) \\ &= 2\lambda n(1 - \beta^2/2) - \lambda n - \eta \text{tr}V + \text{tr}V, \end{aligned}$$

where, in the last equation, we used (21). The above relation implies that

$$\lambda = \frac{(\eta - 1)\text{tr}V}{n(1 - \beta^2)}. \quad (22)$$

However, the quantity η depends on λ . In particular, we have from the definition of η (17) that

$$\begin{aligned} n(2 - \beta^2)\lambda^2 &= \eta^2(\|V\|_F^2 + \|V + 2D_v\|_F^2) \\ &= \eta^2\|V\|_F^2 + \|\lambda I - V\|_F^2, \end{aligned}$$

which can be rewritten as

$$(\eta^2 + 1)\|V\|_F^2 = n(1 - \beta^2)\lambda^2 + 2\lambda\text{tr}V.$$

Substituting λ by (22) in the above relation yields

$$(\eta^2 + 1)\|V\|_F^2 = (\eta - 1)\lambda\text{tr}V + 2\lambda\text{tr}V = (\eta + 1)\lambda\text{tr}V = \frac{(\eta^2 - 1)(\text{tr}V)^2}{n(1 - \beta^2)}.$$

Furthermore, using (14) it follows that

$$(\eta^2 + 1)(1 - \beta^2) = (\eta^2 - 1)(1 - \delta^2).$$

As η is nonnegative, we conclude that

$$\eta = \sqrt{\frac{2 - \beta^2 - \delta^2}{\beta^2 - \delta^2}} = \gamma.$$

Together with (18) and (22), the lemma follows. \square

From the definition of γ , we have

$$\gamma^2 + 1 = 2\frac{1 - \delta^2}{\beta^2 - \delta^2}, \quad \gamma^2 - 1 = 2\frac{1 - \beta^2}{\beta^2 - \delta^2}, \quad (23)$$

so that

$$\frac{\gamma - 1}{2(1 - \beta^2)\gamma} = \frac{\gamma^2 - 1}{2(1 - \beta^2)\gamma(\gamma + 1)} = \frac{1}{(\beta^2 - \delta^2)\gamma(\gamma + 1)},$$

and

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

Applying the above two relations and (14) to (15) it follows that

$$\begin{aligned} D_v &= \frac{\gamma - 1}{2(1 - \beta^2)\gamma} \frac{\text{tr}V}{n} I - \frac{\gamma + 1}{2\gamma} V \\ &= \frac{1 - \delta^2}{(\beta^2 - \delta^2)\gamma(\gamma + 1)} \left(\frac{\|V\|_F^2}{\text{tr}V} I - V \right) - \frac{1}{\gamma + 1} V. \end{aligned} \quad (24)$$

The new value of the duality gap, after taking the cone affine scaling step, is derived in the lemma below. Let us first define Λ_{XZ}^+ as the positive diagonal matrix whose diagonal entries are the eigenvalues of the matrix X^+Z^+ .

Lemma 3.2 *We have*

$$\|V^+\|_F^2 = \frac{\gamma-1}{\gamma+1} \|V\|_F^2,$$

where $V^+ := (\Lambda_{XZ}^+)^{\frac{1}{2}}$.

Proof: Since $D_x \perp D_z$, we have

$$\|V^+\|_F^2 = \text{tr}(V + 2D_x)(V + 2D_z) = V \bullet (V + 2D_v).$$

Now we derive from (24) that

$$V \bullet D_v = -\frac{1}{\gamma+1} \|V\|_F^2.$$

Combining the above two relations, the result follows. \square

4 Polynomiality of the cone affine scaling algorithm

We will show in this section that the cone affine scaling algorithm has a polynomial iteration bound. Observe from Lemma 3.1 that the cone affine scaling step is only defined if $\delta < \beta$. Therefore, it is crucial for the convergence analysis to estimate the next value for δ , viz. the quantity $\delta^+ := \sin(\phi^+)$, where ϕ^+ is the angle between V^+ and the identity matrix.

Lemma 4.1 *For any orthogonal matrix Q , there holds*

$$\delta^+ \leq \frac{1}{2}\delta + \frac{\|V + D_v - QV^+Q^T\|_F}{\|V^+\|_F}.$$

Proof: Since δ^+ is the sine of the angle between V and the identity matrix I , we have

$$\begin{aligned} \delta^+ \|V^+\|_F &= \min_{\alpha} \|\alpha I - V^+\|_F \\ &= \min_{\alpha} \|\alpha I - QV^+Q^T\|_F \\ &\leq \min_{\alpha} \|\alpha I - (V + D_v)\|_F + \|V + D_v - QV^+Q^T\|_F. \end{aligned}$$

However,

$$V + D_v = \frac{\gamma-1}{2(1-\beta^2)\gamma} \frac{\text{tr}V}{n} I + \frac{\gamma-1}{2\gamma} V, \quad (25)$$

so that

$$\min_{\alpha} \|\alpha I - (V + D_v)\|_F = \frac{\gamma-1}{2\gamma} \min_{\alpha} \|\alpha I - V\|_F = \frac{\gamma-1}{2\gamma} \delta \|V\|_F.$$

Using Lemma 3.2, we further have

$$\min_{\alpha} \|\alpha I - (V + D_v)\|_F = \frac{1}{2} \sqrt{\frac{\gamma+1}{\gamma-1} \frac{\gamma-1}{\gamma}} \delta \|V^+\|_F = \frac{1}{2} \sqrt{\frac{\gamma^2-1}{\gamma^2}} \delta \|V^+\|_F \leq \frac{1}{2} \delta \|V^+\|_F.$$

Now, the lemma follows easily from the above derivations. \square

The following result is cited from Sturm and Zhang [25], Corollary 3.1.

Lemma 4.2 *Suppose $V + D_v \succ 0$. Let $\rho := \|D_v\|_F / \lambda_{\min}(V + D_V)$. If $\rho < 2/3$ then there exists an orthogonal matrix Q such that*

$$\|V + D_v - QV^+Q^T\|_F \leq \frac{\rho}{2 - 3\rho} \|D_v\|_F.$$

Based on Lemma 4.1 and Lemma 4.2, a natural way to proceed the estimation of δ^+ is to work out the quantities ρ and $\|D_v\|_F / \|V^+\|_F$. This will be done in the following lemmas.

Lemma 4.3 *There holds*

$$\|D_v\|_F = \sqrt{\frac{\beta^2}{\beta^2 - \delta^2} - \frac{(1 - \beta^2)\delta^2}{(2 - \beta^2 - \delta^2)(\beta^2 - \delta^2)}} \frac{\|V\|_F}{\gamma + 1}.$$

Proof: From (24) and the fact that $((\|V\|_F^2 / \text{tr}V)I - V) \perp V$, we have

$$\|D_v\|_F^2 = \left[\frac{1 - \delta^2}{(\beta^2 - \delta^2)\gamma(\gamma + 1)} \right]^2 \tan^2(\phi) \|V\|_F^2 + \frac{\|V\|_F^2}{(\gamma + 1)^2}.$$

By the definition of δ , we have $\tan^2(\phi) = \delta^2 / (1 - \delta^2)$. Therefore,

$$\begin{aligned} \|D_v\|_F^2 &= \frac{\|V\|_F^2}{(\gamma + 1)^2} \left[1 + \frac{\delta^2(1 - \delta^2)}{(\beta^2 - \delta^2)^2\gamma^2} \right] \\ &\stackrel{(16)}{=} \frac{\|V\|_F^2}{(\gamma + 1)^2} \left[1 + \frac{\delta^2(1 - \delta^2)}{(\beta^2 - \delta^2)(2 - \beta^2 - \delta^2)} \right] \\ &= \frac{\|V\|_F^2}{(\gamma + 1)^2} \left[\frac{\beta^2}{\beta^2 - \delta^2} - \frac{(1 - \beta^2)\delta^2}{(2 - \beta^2 - \delta^2)(\beta^2 - \delta^2)} \right]. \end{aligned}$$

□

Lemma 4.4 *There holds*

$$\frac{\|D_v\|_F}{\|V^+\|_F} = \sqrt{\frac{\beta^2}{2(1 - \beta^2)} - \frac{\delta^2}{2(2 - \beta^2 - \delta^2)}}$$

Proof: Using Lemmas 3.2 and 4.3 and relation (23), it follows that

$$\begin{aligned} \frac{\|D_v\|_F^2}{\|V^+\|_F^2} &= \left[\frac{\beta^2}{\beta^2 - \delta^2} - \frac{(1 - \beta^2)\delta^2}{(2 - \beta^2 - \delta^2)(\beta^2 - \delta^2)} \right] \frac{1}{\gamma^2 - 1} \\ &= \frac{\beta^2}{2(1 - \beta^2)} - \frac{\delta^2}{2(2 - \beta^2 - \delta^2)}. \end{aligned}$$

□

Lemma 4.5 *There holds*

$$\rho = \frac{\|D_v\|_F}{\lambda_{\min}(V + D_v)} \leq \sqrt{2n}\beta.$$

Proof: Let

$$\alpha := \frac{\sqrt{n}}{\|V\|_F} \left(\frac{\text{tr}V}{n} + (1 - \beta^2)\lambda_{\min}(V) \right).$$

We obtain with (25),

$$\lambda_{\min}(V + D_v) = \frac{\gamma - 1}{2(1 - \beta^2)\gamma} \left[\frac{\text{tr}V}{n} + (1 - \beta^2)\lambda_{\min}(V) \right] = \frac{\gamma - 1}{2\sqrt{n}(1 - \beta^2)\gamma} \alpha \|V\|_F.$$

Now using Lemma 4.3, we have

$$\rho = \sqrt{\frac{\beta^2}{\beta^2 - \delta^2} - \frac{(1 - \beta^2)\delta^2}{(2 - \beta^2 - \delta^2)(\beta^2 - \delta^2)}} \frac{2\sqrt{n}\gamma(1 - \beta^2)}{\alpha(\gamma^2 - 1)}$$

Applying (16) and (23),

$$\begin{aligned} \rho &= \frac{1}{\alpha} \sqrt{(2 - \beta^2 - \delta^2)n\beta^2 - (1 - \beta^2)n\delta^2} \\ &= \frac{1}{\alpha} \sqrt{(1 - \beta^2)n\beta^2 + n(\beta^2 - \delta^2)}. \end{aligned}$$

From Lemma A.2 (see the Appendix) we have

$$\begin{aligned} V &\succeq (1 - \sqrt{n-1}\tan(\phi)) \frac{\text{tr}V}{n} I \\ &\succeq (\cos(I, V) - \sqrt{n-1}\sin(\phi)) \frac{\|V\|_F}{\sqrt{n}} I \\ &\succeq (1 - \sqrt{n}\delta) \frac{\|V\|_F}{\sqrt{n}} I. \end{aligned}$$

This yields a lower bound on α

$$\alpha \geq \sqrt{1 - \delta^2} + (1 - \beta^2)(1 - \sqrt{n}\delta).$$

Therefore, we have

$$\begin{aligned} \rho &\leq \frac{\sqrt{(1 - \beta^2)n\beta^2 + n(\beta^2 - \delta^2)}}{\sqrt{1 - \delta^2} + (1 - \beta^2)(1 - \sqrt{n}\delta)} \\ &\leq \frac{\sqrt{(1 - \beta^2)n\beta^2 + n(\beta^2 - \delta^2)}}{\sqrt{1 - \delta^2}}. \end{aligned} \tag{26}$$

Since the above right-hand side of (26) is monotone in δ and $\delta \in (0, 1)$ we obtain

$$\begin{aligned} \rho &\leq \sqrt{2n\beta^2 - n\beta^4} \\ &\leq \sqrt{2n}\beta, \end{aligned}$$

completing the proof. \square

Lemma 4.6 Suppose that $\beta \leq 1/(4\sqrt{n})$. Then, there exists an orthogonal matrix Q such that

$$\frac{\|V + D_v - QV^+Q^T\|_F}{\|V^+\|_F} \leq \frac{1}{2}\beta.$$

Proof: Combining Lemma 4.2 and Lemma 4.4 we have

$$\frac{\|V + D_v - QV^+Q^T\|_F}{\|V^+\|_F} \leq \frac{\rho}{2 - 3\rho} \sqrt{\frac{\beta^2}{2(1 - \beta^2)} - \frac{\delta^2}{2(2 - \beta^2 - \delta^2)}}.$$

Using that $\rho \leq \sqrt{2n}\beta$ and $\beta \leq 1/(4\sqrt{n})$ it follows that

$$\frac{\|V + D_v - QV^+Q^T\|_F}{\|V^+\|_F} \leq \frac{\rho}{2 - 3\rho} \frac{\beta}{\sqrt{2(1 - \beta^2)}} \leq \frac{1}{2}\beta,$$

which yields the proof. \square

Combining the results above, we are now able to prove the following result.

Lemma 4.7 Let $\beta \leq 1/(4\sqrt{n})$. Then, there holds

$$\delta^+ \leq \beta.$$

Proof: From Lemma 4.1 and Lemma 4.6 we conclude that

$$\delta^+ \leq \frac{1}{2}\delta + \frac{1}{2}\beta \leq \beta.$$

\square

We are now in a position to prove polynomiality of the cone affine scaling algorithm.

Theorem 4.8 Suppose X^0 and Z^0 are feasible interior solutions of (P) and (D) respectively. Let ϵ be an accuracy parameter. Moreover, let $\beta = 1/(4\sqrt{n})$, and $\delta^0 = \sin(V^0, I) \leq \beta$. Then the cone affine scaling algorithm yields a pair of primal and dual feasible solutions (X, Z) with $X \bullet Z < \epsilon$ in at most $\mathcal{O}(\sqrt{n} \log(X^0 \bullet Z^0/\epsilon))$ main iterations.

Proof: From Lemma 4.7 we have $\delta^i \leq \beta$ for all i . Now choose $\beta = 1/(4\sqrt{n})$. Hence

$$\begin{aligned} \gamma^i &= \sqrt{\frac{2 - \beta^2 - (\delta^i)^2}{\beta^2 - (\delta^i)^2}} = \sqrt{\frac{2(1 - (\delta^i)^2)}{\beta^2 - (\delta^i)^2} - 1} \\ \gamma^i &\leq \sqrt{\frac{2}{\beta^2} - 1} \leq \sqrt{\frac{2}{\beta^2}} \leq 6\sqrt{n}. \end{aligned}$$

We have from Lemma 3.2

$$X^{i+1} \bullet Z^{i+1} = \left(1 - \frac{2}{\gamma^i + 1}\right) X^i \bullet Z^i \leq \left(1 - \frac{2}{6\sqrt{n} + 1}\right) X^i \bullet Z^i,$$

which implies the theorem. \square

References

- [1] F. Alizadeh, J.A. Heaberly, and M. Overton. Primal–dual interior–point methods for semidefinite programming: convergence rates, stability and numerical results. Technical report, Computer Science Department, New York University, New York, 1996.
- [2] E. R. Barnes. A variation on Karmarkar’s algorithm for solving linear programming problems. *Mathematical Programming*, 36:174–182, 1986.
- [3] E. de Klerk, C. Roos, and T. Terlaky. Polynomial primal-dual affine scaling algorithms in semidefinite programming. Technical Report 96-42, Faculty of Technical Mathematics and Informatics, Delft University of Technology, Delft, The Netherlands, 1996.
- [4] I. I. Dikin. Iterative solution of problems of linear and quadratic programming. *Doklady Akademii Nauk SSSR*, 174:747–748, 1967. Translated in : *Soviet Mathematics Doklady*, 8:674–675, 1967.
- [5] I. I. Dikin. Letter to the editor. *Mathematical Programming*, 41:393–394, 1988.
- [6] D. Goldfarb and D. Xiao. A primal projective interior point method for linear programming. *Mathematical Programming*, 51:17–43, 1991.
- [7] C. C. Gonzaga. Conical projection algorithms for linear programming. *Mathematical Programming*, 43:151–173, 1989.
- [8] C. Helmberg, F. Rendl, R.J. Vanderbei, and H. Wolkowicz. An interior–point method for semidefinite programming. *SIAM Journal on Optimization*, 6:342–361, 1996.
- [9] G. M. Jan and S. C. Fang. A new variant of the primal affine scaling method for linear programs. *Optimization*, 22(5):681–715, 1991.
- [10] B. Jansen. *Interior Point Techniques in Optimization. Complexity, Sensitivity and Algorithms*. PhD thesis, Faculty of Mathematics and Informatics, TU Delft, NL–2628 BL Delft, The Netherlands, December 1995.
- [11] B. Jansen, C. Roos, and T. Terlaky. A polynomial primal–dual Dikin–type algorithm for linear programming. Technical Report 93–36, Faculty of Technical Mathematics and Informatics, TU Delft, NL–2600 GA Delft, The Netherlands, May 1993.
- [12] B. Jansen, C. Roos, T. Terlaky, and Y. Ye. Improved complexity using higher–order correctors for primal–dual Dikin affine scaling. Technical Report 94–75, Faculty of Technical Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands, 1994. (To appear in *Mathematical Programming*).
- [13] N. K. Karmarkar. A new polynomial–time algorithm for linear programming. *Proceedings of the 16th Annual ACM Symposium on Theory of Computing*, pages 302–311, 1984.
- [14] N. K. Karmarkar. A new polynomial–time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

- [15] M. Kojima, S. Shindoh, and S. Hara. Interior–point methods for the monotone linear complementarity problem in symmetric matrices. Technical Report B-282, Dept. of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, 1994. To appear in *SIAM Journal on Optimization*.
- [16] C.J. Lin and R. Saigal. An infeasible start predictor corrector method for semi–definite linear programming. Technical report, Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, USA, 1995.
- [17] R. D. C. Monteiro, I. Adler, and M. G. C. Resende. A polynomial–time primal–dual affine scaling algorithm for linear and convex quadratic programming and its power series extension. *Mathematics of Operations Research*, 15:191–214, 1990.
- [18] R.D.C. Monteiro. Primal–dual path following algorithms for semidefinite programming. Technical report, School of Industrial and Systems Engineering, Georgia Tech, Atlanta, Georgia, U.S.A., 1995. To appear in *SIAM Journal on Optimization*.
- [19] M. Muramatsu. Affine scaling algorithm fails for semidefinite programming. Research Report 16, Department of Mechanical Engineering, Sophia University, USA, 1996.
- [20] Y. Nesterov and M.J. Todd. Self–scaled barriers and interior–point methods for convex programming. Technical Report 1091, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York, 1994. To appear in *Mathematics of Operations Research*.
- [21] Y. Nesterov and M.J. Todd. 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.
- [22] M. W. Padberg. Solution of a nonlinear programming problem arising in the projective method for linear programming. Technical Report, School of Business and Administration, New York University, New York, NY 10003, USA, March 1985.
- [23] F.A. Potra and R. Sheng. A superlinearly convergent primal–dual infeasible–interior–point algorithm for semidefinite programming. Technical Report 78, Department of Mathematics, The University of Iowa, Iowa City, USA, 1995.
- [24] J.F. Sturm and S. Zhang. Symmetric primal–dual path following algorithms for semidefinite programming. Technical Report 9554/A, Econometric Institute, Erasmus University Rotterdam, The Netherlands, 1995.
- [25] J.F. Sturm and S. Zhang. On weighted centers for semidefinite programming. Technical Report 9636/A, Econometric Institute, Erasmus University Rotterdam, The Netherlands, 1996.
- [26] J.F. Sturm and S. Zhang. An $O(\sqrt{n}L)$ iteration bound primal–dual cone affine scaling algorithm for linear programming. *Mathematical Programming*, 72(2):177–194, 1996.
- [27] R. J. Vanderbei, M. S. Meketon, and B. A. Freedman. A modification of Karmarkar’s linear programming algorithm. *Algorithmica*, 1(4):395–407, 1986.

- [28] Y. Zhang. 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.

A Technical Lemmas

Lemma A.1 *Let $Y \in \mathcal{S}^n$. If $\text{tr}Y = 0$, then*

$$\|Y\| \leq \sqrt{\frac{n-1}{n}} \|Y\|_F.$$

Proof: Let us denote the eigenvalues of Y by $\lambda_1, \dots, \lambda_n$, where we assume, without loss of generality, that these eigenvalues are ordered such that

$$|\lambda_1| \leq |\lambda_2| \leq \dots \leq |\lambda_n|.$$

By definition of the Frobenius norm of Y , and using that the trace of a matrix is the sum of its eigenvalues, we have

$$\|Y\|_F^2 = \sum_{i=1}^n \lambda_i^2 = \lambda_n^2 + \sum_{i=1}^{n-1} \lambda_i^2. \quad (27)$$

From $\text{tr}Y = 0$ we have

$$\lambda_n = -\sum_{i=1}^{n-1} \lambda_i,$$

so that

$$\sum_{i=1}^{n-1} \lambda_i^2 - \frac{\lambda_n^2}{n-1} = \sum_{i=1}^{n-1} \left(\lambda_i + \frac{\lambda_n}{n-1}\right)^2 \geq 0. \quad (28)$$

Combining (27) and (28) yields

$$\|Y\|_F^2 \geq \left(1 + \frac{1}{n-1}\right) \lambda_n^2 = \frac{n}{n-1} \|Y\|^2.$$

This completes the proof. \square

Lemma A.2 *Let $Y \in \mathcal{S}^n$ with $\text{tr}Y > 0$. Let*

$$\phi := \arccos\left(\frac{\text{tr}Y}{\sqrt{n} \|Y\|_F}\right)$$

denote the angle between Y and the identity matrix. If $\sqrt{n-1} \tan(\phi) \leq 1$, then

$$\frac{n}{\text{tr}Y} Y \succeq (1 - \sqrt{n-1} \tan(\phi)) I.$$

Proof: For any matrix $A \in \mathcal{S}^n$ we know that $A + \|A\| I \succeq 0$. Applying this property with $A = ((n/\text{tr}Y)Y - I) \in \mathcal{S}^n$ we conclude that

$$\frac{n}{\text{tr}Y} Y = I + \left(\frac{n}{\text{tr}Y} Y - I\right) \succeq \left(1 - \left\|\frac{n}{\text{tr}Y} Y - I\right\|\right) I. \quad (29)$$

Since $\text{tr}((n/\text{tr}Y)Y - I) = 0$, we obtain from Lemma A.1 that

$$\left\|\frac{n}{\text{tr}Y} Y - I\right\| \leq \sqrt{\frac{n-1}{n}} \left\|\frac{n}{\text{tr}Y} Y - I\right\|_F = \sqrt{n-1} \tan(\phi).$$

Together with (29) this implies the lemma. \square