June 18, 1996

ON THE LONG STEP PATH–FOLLOWING METHOD FOR SEMIDEFINITE PROGRAMMING

Jos F. Sturm[1] and Shuzhong Zhang[2]

Report 9638/A, Econometric Institute
Erasmus University Rotterdam
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

ABSTRACT

It has been shown in various recent research reports that the analysis of short step primal-dual path following algorithms for linear programming can be nicely generalized to semidefinite programming. However, the analysis of long step path-following algorithms for semidefinite programming appeared to be less straightforward. For such an algorithm, Monteiro [8] obtained an $O(n^{1.5}\log(1/\epsilon))$ iteration bound for obtaining an ϵ-optimal solution, where n is the order of the semidefinite decision variable. In this paper, we propose to use a different search direction, viz. the so-called V-space direction. It is shown that this modification reduces the iteration complexity to $O(n\log(1/\epsilon))$. Independently, Monteiro and Y. Zhang obtained a similar result using Nesterov-Todd directions.

Key words. semidefinite programming, long step path following, symmetric primal–dual transformation.

[1] Econometric Institute, Erasmus University Rotterdam, The Netherlands, sturm@few.eur.nl.

[2] Econometric Institute, Erasmus University Rotterdam, The Netherlands, zhang@few.eur.nl

1. Introduction

A number of recent reports are dedicated to the generalization of primal-dual path-following algorithms from linear towards semidefinite programming. This research has yielded various ways of generating search directions. Alizadeh, Heaberly and Overton [1] proposed the so-called $XZ+ZX$– and Q–methods. Polynomiality results have been established for neither of these methods, but local quadratic convergence for the $XZ+ZX$–method was shown by Kojima, Shida and Shindoh [5]. Other methods, proposed by Helmberg et al. [2], Nesterov and Todd [10, 11] and Monteiro [8], fit in the general frameworks of Kojima, Shindoh and Hara [4] and Y. Zhang [15], and will henceforth be denoted as KSH-methods. For several KSH-algorithms in which the central path is followed closely, an $O(\sqrt{n}\log(1/\epsilon))$ worst case iteration bound has been obtained, exactly as in the linear programming case. The generalization of the *long step path-following* method of Kojima, Mizuno and Yoshise [3] appeared to be more difficult. A generalization that uses a particular KSH-direction was proposed by Monteiro [8] and subsequently analyzed also by Y. Zhang [15]. Both authors obtained an $O(n^{1.5}\log(1/\epsilon))$ iteration bound, therefore requiring an extra factor $\sqrt{n}$ compared to the well known result for linear programming [3]. Independently of our work, Monteiro and Y. Zhang [9] show that the iteration bound can be reduced to $O(n\log(1/\epsilon))$ if Nesterov-Todd directions [10, 11, 12, 14] are used.

We show in this short paper that we can also get rid of the afore mentioned extra factor $\sqrt{n}$ if we use the so-called *V-space direction* instead of Monteiro's KSH-direction. The V-space direction was recently proposed in our report [13]. To derive our result, we use an essentially different argumentation than Monteiro and Y. Zhang [9]. An important point underlying the analysis of [9] is the fact that the spectrum of a real matrix is contained between the smallest and largest eigenvalue of its symmetric part (See Lemma 5.3 in [9]). Our argumentation however relies on an estimation of the higher order terms of the symmetrically transformed V-solution in terms of the Frobenius norm.

Notation.
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 $\|X\|_F=\sqrt{X\bullet X}$. The space of symmetric $n\times n$ matrices is denoted by $\mathcal{S}$. 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 $\mathcal{S}_{++}$. For $X\in\mathcal{S}_+$ we let $\lambda_{\min}(X)$ denote the smallest eigenvalue of X. The order n identity matrix is denoted by I.

1.1. Problem statement

In this paper, we study the standard primal semidefinite programming problem (SDP)

$$(P)\quad \inf\{C\bullet X \mid X\succeq 0,\ A_i\bullet X=b_i \text{ for } i=1,\ldots,m\}$$

along with its dual

$$(D)\quad \sup\{b^{\mathrm{T}}y \mid Z\succeq 0,\ Z=C-\textstyle\sum_{i=1}^m y_iA_i\}.$$

The primal feasible set is denoted by

$$\mathcal{F}_P := \{X \in \mathcal{S}_+ \mid A_i \bullet X = b_i \text{ for } i = 1, \ldots, m\}$$

and the dual feasible set is

$$\mathcal{F}_D := \{Z \in \mathcal{S}_+ \mid Z = C - \sum_{i=1}^{m} y_i A_i,\ \exists y \in \Re^m\}.$$

We make the standard assumption that both $\mathcal{F}_P$ and $\mathcal{F}_D$ contain positive definite matrices. This assumption guarantees the existence of an optimal solution pair (X, Z) with $X \bullet Z = 0$, and we can replace 'inf' by 'min' and 'sup' by 'max' in problems (P) and (D) respectively.

1.2. Long step path-following

A primal-dual pair of feasible solutions $(X, Z) \in \mathcal{F}_P \times \mathcal{F}_D$ is said to be a *μ-center*, if

$$XZ = \mu I$$

for some $\mu \in \Re_{++}$. The long step primal-dual path-following method for semidefinite programming [8, 15] generates an iterative sequence $(X^{(k)}, Z^{(k)})$ in the *wide neighborhood* of the central path

$$\mathcal{N}(\theta) = \{(X, Z) \in \mathcal{F}_P \times \mathcal{F}_D \mid \lambda_{\min}(XZ) \geq \theta^2 \mu,\ \mu = \frac{X \bullet Z}{n}\}$$

with parameter $\theta \in (0, 1)$. Given an iterate $(X^{(k)}, Z^{(k)}) \in \mathcal{N}(\theta)$, a new iterate

$$X^{(k+1)} = X^{(k)} + 2t_k \Delta X^{(k)}, \quad Z^{(k+1)} = Z^{(k)} + 2t_k \Delta Z^{(k)}$$

is determined for some step length t_k and search direction $(\Delta X^{(k)}, \Delta Z^{(k)})$. The step length t_k is chosen such that $(X^{(k+1)}, Z^{(k+1)}) \in \mathcal{N}(\theta)$. We let $t_k(\theta)$ denote the step length to the boundary of $\mathcal{N}(\theta)$, i.e.

$$t_k(\theta) = \max\{\bar{t} \mid (X^{(k)} + 2t\Delta X^{(k)}, Z^{(k)} + 2t\Delta Z^{(k)}) \in \mathcal{N}(\theta)\, \forall t \in [0, \bar{t}]\}. \tag{1}$$

The long step path-following algorithm is now described as follows:

Algorithm 1 (Long step path-following).
Parameters $\theta \in (0,1)$, $\gamma \in (0,1)$ *and* $\epsilon > 0$. *Initial solution* $(X^{(0)}, Z^{(0)}) \in \mathcal{N}(\theta)$.

Step 0 *Set* $k = 0$.

Step 1 *Stop if* $X^{(k)} \bullet Z^{(k)} \leq \epsilon$.

Step 2 *Compute a Newton direction* $(\Delta X^{(k)}, \Delta Z^{(k)})$ *towards the* $(\gamma^2 \mu^{(k)})$*-center, where*

$$\mu^{(k)} := \frac{X^{(k)} \bullet Z^{(k)}}{n}.$$

Step 3 *Choose a step length* $t_k \geq t_k(\theta)/2$ *such that*

$$(X^{(k)} + 2t_k \Delta X^{(k)}, Z^{(k)} + 2t_k \Delta Z^{(k)}) \in \mathcal{N}(\theta).$$

Step 4 *Let*

$$X^{(k+1)} = X^{(k)} + 2t_k \Delta X^{(k)}, \quad Z^{(k+1)} = Z^{(k)} + 2t_k \Delta Z^{(k)},$$

replace k *by* $k+1$ *and return to Step 1.*

In Step 2 of the above algorithm, different ways of computing a Newton direction are possible. In fact, there is an ongoing debate on the method of generating Newton directions for semidefinite programming. In this paper, we will use the V-space Newton directions as proposed in [13].

1.3. *The V-space*

In this section, we briefly review the V-space approach for semidefinite programming. Consider $X, Z \in \mathcal{S}_{++}$ with $X \in \mathcal{F}_P$ and $Z \in \mathcal{F}_D$. Let Λ_{XZ} denote a positive diagonal matrix whose diagonal entries are the eigenvalues of the matrix XZ. Let Q denote an arbitrary unitary matrix of order n (i.e. $QQ^{\mathrm{T}} = I$), and define

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

It is shown in [12, 13, 14] that there exists an invertible matrix L_d such that

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

We see that V is both feasible for the transformed primal problem

$$\min\{L_d^{\mathrm{T}} C L_d \bullet \bar{X} \mid L_d \bar{X} L_d^{\mathrm{T}} \in \mathcal{F}_P\}$$

and for its dual

$$\max\{b^{\mathrm{T}} y \mid \bar{Z} \succeq 0,\ \bar{Z} = L_d^{\mathrm{T}}(C - \sum_{i=1}^{m} y_i A_i) L_d\}.$$

This explains why the above transformation is known as the *symmetric primal-dual transformation*. Remark that the duality gap of the pair (X, Z) in the original SDP is the same as the duality gap of the pair (V, V) for the transformed SDP. It follows that the optimal solution set will be approached if $V \to 0$.

1.4. Primal-dual search directions

It is easily seen that given a symmetric matrix $D_v \in \mathcal{S}$ and an invertible matrix $L_d \in \Re^{n\times n}$, the relations

$$D_x + D_z = D_v$$

$$L_d^{\mathrm{T}} A_i L_d \bullet D_x = 0 \text{ for } i = 1, 2, \ldots, m$$

$$D_z + \sum_{i=1}^{m} \Delta y_i L_d^{\mathrm{T}} A_i L_d = 0$$

$$\Delta X = L_d D_x L_d^{\mathrm{T}}, \quad \Delta Z = L_d^{-\mathrm{T}} D_z L_d^{-1}$$

uniquely determine the primal search direction ΔX and the dual search direction $(\Delta y, \Delta Z)$, and vice versa. Namely, the transformed search directions D_x amd D_z form an orthogonal decomposition of D_v. In the sequel, "the search direction" will refer to the matrix D_v, which combines the primal and the dual search directions.
Once the search direction D_v is given, we obtain a solution pair

$$X(t) := X + 2t\Delta X, \quad Z(t) := Z + 2t\Delta Z$$

by moving simultaneously along ΔX in the primal space and along ΔZ in the dual-space, with a step length $t \geq 0$. The maximal step length retaining feasibility is

$$t^* := \max\{t \mid X(t) \succeq 0,\, Z(t) \succeq 0\}.$$

Consider a step length $t \in [0, t^*)$ and let $G(t)$ be the positive definite matrix satisfying

$$G(t)^{-1}(V + 2tD_x)G(t)^{-1} = G(t)(V + 2tD_z)G(t).$$

Now define

$$V(t) := G(t)(V + 2tD_z)G(t)$$

and notice that

$$V(t)^2 = G(t)^{-1} L_d^{-1} X(t) Z(t) L_d G(t).$$

It is now easily seen that the eigenvalues of $\mathrm{V(t)}^2$ and the eigenvalues of the matrix $X(t)Z(t)$ coincide.
It is shown in [13] that

$$V(t) = V + tD_v + o(t).$$

This implies that

$$D_v = \gamma\sqrt{\mu} I - V \tag{2}$$

is a Newton direction for obtaining the $(\gamma^2\mu)$-center, which is by definition the solution pair $(\hat{X}, \hat{Z})$ for which all the eigenvalues of the matrix $\hat{X}\hat{Z}$ are equal to the quantity $\gamma^2\mu$. We let $F(t)$ denote the linearization of $V(t)$, i.e.

$$F(t) := V + tD_v. \tag{3}$$

For those step lengths t that result in a positive definite linearization $F(t) \succ 0$, we introduce the quantity

$$\rho(t) := \frac{t \, \|D_v\|_F}{\lambda_{\min}(F(t))}.$$

The following lemma estimates the higher order terms of $V(t)$. For a proof, we refer to Lemma 3.5 in [13].

Lemma 1.1. *Let $t \geq 0$ be such that $F(t) \succ 0$ and $\rho(t) < 2/3$. Then*

$$\|F(t) - V(t)\|_F \leq \frac{\rho(t) t}{2 - 3\rho(t)} \|D_v\|_F .$$

2. Worst-case analysis

For simplicity, we will restrict ourselves to parameter choices with $\gamma \leq 2\theta$ (a similar restriction on the parameters was made in Mizuno, Todd and Ye [7]). Moreover, we will assume $n \geq 4$. Consider a primal-dual pair $(X, Z) \in \mathcal{N}(\theta)$. We let

$$\phi := \frac{\operatorname{tr}(\Lambda_{XZ}^{1/2})}{n}.$$

From the definition of $\mathcal{N}(\theta)$, we obtain

$$\phi \geq \theta \sqrt{\mu}. \tag{4}$$

On the other hand, application of the Cauchy-Schwartz inequality yields

$$\phi \leq \frac{\left\| \Lambda_{XZ}^{1/2} \right\|_F}{\sqrt{n}} = \sqrt{\mu}. \tag{5}$$

Since $D_v = \gamma \sqrt{\mu} I - V$, we have

$$\|D_v\|_F^2 = \gamma^2 n \mu - 2\gamma n \sqrt{\mu} \phi + n\mu \leq n\mu, \tag{6}$$

where we used (4) and the fact that $\gamma \leq 2\theta$.

Lemma 2.1. *Let $t \in [0, 1)$ be such that $F(t) \succ 0$. Then*

$$\rho(t) \leq \frac{\sqrt{n} t}{(1-t)\theta}.$$

Proof:
By definitions (3) and (2), we have

$$F(t) = V + tD_v = (1-t)V + t\gamma\sqrt{\mu} I \succeq (1-t)V$$

so that, using the definition of $\mathcal{N}(\theta)$,

$$\lambda_{\min}(F(t)) \geq (1-t)\lambda_{\min}(V) \geq (1-t)\theta\sqrt{\mu}.$$

Combining the above relation with (6), we obtain

$$\rho(t) = \frac{t \, \|D_v\|_F}{\lambda_{\min}(F(t))} \leq \frac{\sqrt{n} t}{(1-t)\theta}.$$

□

We let

$$\mu(t) := \frac{X(t) \bullet Z(t)}{n},$$

so that $n\mu(t)$ denotes the duality gap after taking a step of length t. An upper bound on $\mu(t)$ is derived below.

Lemma 2.2. *There holds*

$$\mu(t) \leq (1 - 2(1-\gamma)t)\mu.$$

Proof:
First remark that

$$\begin{aligned} n\mu(t) = \mathrm{tr} X(t)Z(t) &= \mathrm{tr} L_d^{-1} X(t)Z(t)L_d \\ &= \mathrm{tr}(V + 2tD_x)(V + 2tD_z) \\ &= V \bullet (V + 2tD_v) + 4t^2 D_x \bullet D_z. \end{aligned}$$

Now using the orthogonality $D_X \perp D_Z$, we have

$$n\mu(t) = V \bullet (V + 2tD_v) = (1-2t)n\mu + 2nt\gamma\phi\sqrt{\mu}.$$

The lemma follows by applying (5) to the above relation.

□

The following lemma states a lower bound on $t(\theta)$, the step length towards the boundary of $\mathcal{N}(\theta)$, see definition (1).

Lemma 2.3. *There holds*

$$t(\theta) \geq \frac{\gamma\theta(1-\theta)}{n}.$$

Proof:
Let

$$\bar{t} := \frac{\gamma\theta(1-\theta)}{n}.$$

For $0 \leq t \leq \bar{t}$, we wish to show that

$$(X(t), Z(t)) \in \mathcal{N}(\theta).$$

Let $0 \leq t \leq \bar{t}$. Using definitions (3) and (2), we write

$$V(t) = F(t) + (V(t) - F(t)) = V + t(\gamma\sqrt{\mu}I - V) + (V(t) - F(t)). \tag{7}$$

As $(X, Z) \in \mathcal{N}(\theta)$, we have $V \succeq \theta\sqrt{\mu}I$. We can therefore estimate (7) as follows:

$$V(t) \succeq (1 - (1-\gamma)t)\theta\sqrt{\mu}I + \gamma t(1-\theta)\sqrt{\mu}I - \|F(t) - V(t)\|\, I.$$

Since the eigenvalues of $V(t)^2$ coincide with the eigenvalues of $X(t)Z(t)$, we can rewrite the above relation as

$$\sqrt{\lambda_{\min}(X(t)Z(t))} \geq (1 - (1-\gamma)t)\theta\sqrt{\mu} + \gamma t(1-\theta)\sqrt{\mu} - \|F(t) - V(t)\|\,. \tag{8}$$

Using Lemma 1.1, Lemma 2.1 and (6) respectively, we have

$$\|F(t)-V(t)\| \le \|F(t)-V(t)\|_F \le \frac{\rho(t)t}{2-3\rho(t)}\,\|D_v\|_F \le \frac{n\sqrt{\mu}t^2}{\theta(1-t)(2-3\rho(t))}. \tag{9}$$

Using Lemma 2.1 once more, we obtain

$$(1-t)(2-3\rho(t)) = 2-2t-3(1-t)\rho(t) \ge 2-2t-3\frac{\sqrt{n}t}{\theta},$$

and using $t \le \bar{t}$,

$$(1-t)(2-3\rho(t)) \ge 2-\frac{2\gamma\theta(1-\theta)}{n}-3\frac{\gamma(1-\theta)}{\sqrt{n}} \ge 1, \tag{10}$$

where the last inequality follows from our assumptions that $\gamma \le 2\theta$ and $n \ge 4$. Combining the estimations (9) and (10), we arrive at

$$\|F(t)-V(t)\| \le \frac{n\sqrt{\mu}t^2}{\theta} \le \gamma t(1-\theta)\sqrt{\mu}. \tag{11}$$

Applying relation (11) to (8) yields

$$\begin{aligned}\sqrt{\lambda_{\min}(X(t)Z(t))} &\ge (1-(1-\gamma)t)\theta\sqrt{\mu}+\gamma t(1-\theta)\sqrt{\mu}-\|V(t)-F(t)\| \\ &\ge (1-(1-\gamma)t)\theta\sqrt{\mu},\end{aligned}$$

which together with Lemma 2.2 implies that

$$\lambda_{\min}(X(t)Z(t)) \ge (1-2(1-\gamma)t+(1-\gamma)^2t^2)\theta^2\mu \ge \theta^2\mu(t),$$

i.e. $(X,Z)\in\mathcal{N}(\theta)$.

□

We immediately obtain from Lemma 2.3 the main result of this short paper:

Theorem 2.1. *Consider the long step path following algorithm with parameters $\gamma \in (0,1)$, $\theta \in (0,1)$, $\gamma \le 2\theta$, and the Newton direction defined by (2). Given a tolerance $\epsilon > 0$, the algorithm terminates with an iterate satisfying $X^{(k)} \bullet Z^{(k)} \le \epsilon$ in*

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

main iterations.

Proof:
From Lemma 2.3 we know for $k=0,1,\ldots$ that

$$t_k(\theta) \ge \frac{\gamma\theta(1-\theta)}{n} = \frac{1}{O(n)}.$$

Since $t_k \ge t_k(\theta)/2$, we also have $1/t_k = O(n)$. Applying Lemma 2.2, we thus obtain

$$X^{(k+1)}\bullet Z^{(k+1)} \le (1-\frac{1}{O(n)})X^{(k)}\bullet Z^{(k)}$$

for all $k=0,1,\ldots$. The theorem follows immediately from the above inequality.

□

References

[1] Alizadeh, F., Heaberley, J.-P.A. and Overton, M.L., "Primal-dual interior point methods for semidefinite programming: convergence rates, stability and numerical results," Report 721, New York University Computer Science Department, New York, NY, 1996.

[2] Helmberg, C., Rendl, F., Vanderbei, R.J. and Wolkowicz, H., "An interior point method for semidefinite programming," Program in Statistics and Operations Research, Princeton University, Princeton, NJ 08544, 1994. To appear in *SIAM Journal on Optimization.*

[3] Kojima, M., Mizuno, S. and Yoshise, A., "A primal-dual interior point algorithm for linear programming," in *Progress in Mathematical Programming: Interior point and related methods* pp. 29-37, (ed. Megiddo, N.), Springer Verlag, New York, 1989.

[4] 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. Revised 1995. To appear in *SIAM Journal on Optimization.*

[5] Kojima, M., Shida, M. and Shindoh, S., "A predictor-corrector interior-point algorithm for the semidefinite linear complementarity problem using the Alizadeh-Heaberley-Overton search direction," Research Reports on Information Sciences, B-311, Dept. of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, 1996.

[6] Kojima, M., Shida, M. and Shindoh, S., "A note on the Nesterov-Todd and the Kojima-Shindoh-Hara search directions in semidefinite programming," Research Reports on Information Sciences, B-313, Dept. of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, 1996.

[7] 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.

[8] 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. To appear in *SIAM Journal on Optimization.*

[9] Monteiro, R.D.C. and Zhang, Y., "A unified analysis for a class of path-following primal-dual interior point algorithms for semidefinite programming," Working paper (first draft), June 1996.

[10] Nesterov, Y. and Todd, M.J., "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.*

[11] 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. To appear in *SIAM Journal on Optimization.*

[12] Sturm, J.F. and Zhang, S., "Symmetric primal-dual path following algorithms for semidefinite programming," Report 9554/A, Econometric Institute, Erasmus University Rotterdam, Rotterdam, The Netherlands, 1995.

[13] Sturm, J.F. and Zhang, S., "On weighted centers for semidefinite programming," Report

9636/A, Econometric Institute, Erasmus University Rotterdam, Rotterdam, The Netherlands, 1996.

[14] Todd, M.J., Toh, K.C. and Tütüncü, R.H., "On the Nesterov-Todd direction in semidefinite programming," Technical Report 1154, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York, 1996.

[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.