A scaled Gauss-Newton Primal–Dual Search Direction for Semidefinite Optimization

E. de Klerk, J. Peng, C. Roos, T. Terlaky

Faculty of Information Technology and Systems

Department Technical Mathematics and Informatics

Delft University of Technology

P.O. Box 5031, 2600 GA Delft, The Netherlands.

e–mail: [e.deklerk,j.peng,c.roos,t.terlaky]@twi.tudelft.nl

February 10, 1999

Abstract

Interior point methods for semidefinite optimization (SDO) have recently been studied intensively, due to their polynomial complexity and practical efficiency. Most of these methods are extensions of linear optimization (LO) algorithms. Unlike in the LO case, there are several different ways of constructing primal-dual search directions in SDO. The usual scheme is to apply linearization in conjunction with symmetrization to the perturbed optimality conditions of the SDO problem. Symmetrization is necessary since the linearized system is overdetermined. A way of avoiding symmetrization is to find a least squares solution of the overdetermined system. Such a 'Gauss Newton' direction was investigated by Kruk *et al.* [6], without giving any complexity analysis. In this paper we present a similar direction where a local norm is used in the least squares formulation, and we give a polynomial complexity analysis of the resulting primal-dual algorithm.

Keywords: Semidefinite optimization, primal-dual search directions, interior point algorithms

1 Introduction

Interior point methods for semidefinite optimization (SDO) became a popular research area when it became clear that the algorithms for linear optimization (LO) can often be extended to the more general SDO case. Following the trend in LO, primal–dual algorithms soon enjoyed the most attention. Unlike the LO-case, however, there are many possibilities to obtain primal–dual search directions. Different directions arise when the perturbed optimality conditions are linearized and subsequently symmetrized (see Section 1.1); a quite comprehensive survey of the search directions obtained this way may be found in [10]. The need for symmetrization arises from the fact that the system of linearized perturbed optimality conditions is overdetermined.

A recent idea by Kruk *et al.* [6] was to avoid symmetrization by solving a least squares problem by the Gauss-Newton method (see Section 1.1). The authors obtained a numerically robust search direction in this way, but did not give convergence proofs for their search direction. The work in our paper was inspired by their approach: here we show that, by using scaling and a different (local) norm in the definition of the least squares problem, a direction is obtained which allows a polynomial time convergence analysis.

1.1 Preliminaries

We consider the semidefinite optimization problem in the standard form. Thus the primal problem (P) is given by:

$$(P) \quad p^* = \inf \{ \operatorname{Tr} CX \;:\; \operatorname{Tr}(A_i X) = b_i \, (1 \le i \le m), \, X \succeq 0 \}$$

and its dual problem (D) is:

$$(D) \quad d^* = \sup \left\{ b^T y \;:\; \sum_{i=1}^m y_i A_i + S = C, \, S \succeq 0 \right\},$$

where C and the A_i's are symmetric $n \times n$ matrices, $b, y \in \mathbb{R}^m$ and '$X \succeq 0$' means X is symmetric positive semidefinite. The matrices A_i are further assumed to be linearly independent. We will assume that a strictly feasible pair $(X \succ 0, S \succ 0)$ exists. This ensures the existence of an optimal primal-dual pair (X^*, S^*) with zero duality gap $(\operatorname{Tr}(X^* S^*) = 0)$.

The optimality conditions for the pair of problems are

$$\begin{aligned} \operatorname{Tr}(A_i X) &= b_i, \; i = 1, \ldots, m, \; X \succeq 0 \\ \sum_{i=1}^m y_i A_i + S &= C, \qquad\qquad\qquad S \succeq 0 \\ XS &= 0. \end{aligned}$$

If these conditions are perturbed to

$$\begin{aligned} \operatorname{Tr}(A_i X) &= b_i, \quad i = 1, \ldots, m, \; X \succeq 0 \\ \sum_{i=1}^m y_i A_i + S &= C, \qquad\qquad\qquad S \succeq 0 \\ XS &= \mu I, \end{aligned}$$

for some $\mu > 0$ and where I denotes the identity matrix, then a unique solution of the perturbed system exists. This solution is denoted by $\{X(\mu), y(\mu), S(\mu)\}$. This solution may be seen as a parameterized curve in the Cartesian product of the primal and dual feasible regions[1], called the *central path*, which converges to the analytic centre of the optimal primal–dual set as $\mu \to 0$. The existence and uniqueness of the central path follow from the fact that $\{X(\mu), y(\mu), S(\mu)\}$ corresponds to the unique minimum of the strictly convex primal–dual barrier function

$$\Phi(X, S, \mu) = \frac{1}{\mu} \operatorname{Tr}(XS) - \ln \det(XS) - n + n \ln(\mu)$$

[1]This Cartesian product of the primal and dual feasible sets will be called the *primal-dual feasible region.*

defined on the primal–dual feasible region. Because of the two different associations, the parameter μ is called either the *barrier parameter*, or the *centering* parameter.

Primal-dual interior point methods solve the system of perturbed optimality conditions approximately, followed by a reduction in μ. Ideally, the goal is to obtain primal and dual steps ΔX and ΔS, respectively, which satisfy $X + \Delta X \succeq 0$, $S + \Delta S \succeq 0$ and

$$
\begin{aligned}
(X + \Delta X)(S + \Delta S) &= \mu I && (1)\\
\operatorname{Tr}(A_i \Delta X) &= 0, \quad i = 1, \ldots, m && (2)\\
\sum_{i=1}^{m} \Delta y_i A_i + \Delta S &= 0 && (3)\\
(\Delta X)^T = \Delta X, &\quad (\Delta S)^T = \Delta S. && (4)
\end{aligned}
$$

Note that the requirement $\Delta S^T = \Delta S$ in (4) is redundant, due to the fact that the matrices A_i in (3) are symmetric. Furthermore, equation (1) is nonlinear, and primal-dual methods differ with regard to how it is linearized. Moreover, care must be taken to ensure that the resulting linear system is not overdetermined. Zhang [11] suggested to discard the symmetry requirements (4) and to replace the nonlinear equation by

$$H_P(XS + \Delta XS + X\Delta S - \mu I) = 0,$$

where H_P is the linear transformation given by

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

for any matrix M, and where the *scaling matrix* P determines the symmetrization strategy. Some popular choices for P are listed in Table 1. The resulting linear systems are now

P	Reference
$\left[X^{\frac{1}{2}}\left(X^{\frac{1}{2}}SX^{\frac{1}{2}}\right)^{-\frac{1}{2}}X^{\frac{1}{2}}\right]^{\frac{1}{2}}$	Nesterov and Todd (NT) [8];
$X^{-\frac{1}{2}}$	Monteiro [7], Kojima *et al.* [5];
$S^{\frac{1}{2}}$	Monteiro [7], Helmberg *et al.* [3], Kojima *et al.* [5];
I	Alizadeh, Haeberley and Overton (AHO) [1];

Table 1: Choices for the scaling matrix P.

solvable (for the AHO direction ($P = I$) solvability is only guaranteed if (X, S) lie in a certain neighbourhood of the central path). The choice $P = X^{-\frac{1}{2}}$ in Table 1 does not automatically lead to a symmetric solution ΔX, and the solution is therefore replaced by its symmetric part. The other three choices of scaling matrices in Table 1 yield symmetric ΔX.

In the recent paper by Kruk *et al.* [6], the symmetrization operator H_P is not used, and the following least squares problem is solved instead:

$$\min \|XS + \Delta XS + X\Delta S - \mu I\|^2 \tag{5}$$

subject to

$$
\begin{aligned}
\operatorname{Tr}(A_i \Delta X) &= 0, \quad i = 1, \ldots, m \\
\sum_{i=1}^{m} \Delta y_i A_i + \Delta S &= 0 \\
\Delta X &= \Delta X^T,
\end{aligned}
$$

where the norm is the Frobenius norm. Note that the symmetry of ΔX is forced. The authors proved (among other things) the following about the resulting *Gauss-Newton* (GN) direction:

- its existence and uniqueness;
- it is always a descent direction for the least squares objective (5);
- it is scale invariant;
- it reduces to the familiar primal-dual direction in the special case of linear optimization;
- it coincides with all the other primal-dual directions from Table 1 if the least squares residual in (5) is zero at optimality.

The new direction we propose can be introduced in a similar way as the GN direction – as will be shown in the next section – and it shares all the abovementioned features of the GN direction. Moreover, it allows a polynomial convergence analysis in the usual primal–dual algorithmic framework, as will become clear in Section 4.

2 The new search direction

Using the well-known NT-scaling (see Table 1), we now reformulate the system (1) – (3). Defining

$$
D = S^{-\frac{1}{2}} \left(S^{\frac{1}{2}} X S^{\frac{1}{2}} \right)^{\frac{1}{2}} S^{-\frac{1}{2}} = X^{\frac{1}{2}} \left(X^{\frac{1}{2}} S X^{\frac{1}{2}} \right)^{-\frac{1}{2}} X^{\frac{1}{2}},
$$

one has $D^{-1}X = SD$. Using this, we introduce

$$
V := D^{-\frac{1}{2}} X D^{-\frac{1}{2}} = D^{\frac{1}{2}} S D^{\frac{1}{2}}.
$$

The matrices D and V are symmetric positive definite. We also introduce the scaled search directions $\hat{D}_X$ and $\hat{D}_S$:

$$
\hat{D}_X := D^{-\frac{1}{2}} \Delta X D^{-\frac{1}{2}}, \quad \hat{D}_S := D^{\frac{1}{2}} \hat{\Delta} S D^{\frac{1}{2}}.
$$

Finally, scaling the data matrices A_i to

$$
\tilde{A}_i := D^{\frac{1}{2}} A_i D^{\frac{1}{2}}, \quad 1 \le i \le m,
$$

the system (1)–(4) can be reformulated as follows:

$$
\begin{aligned}
\left(V + \hat{D}_X\right)\left(V + \hat{D}_S\right) &= \mu I &\quad (6) \\
\operatorname{Tr}\left(\tilde{A}_i \hat{D}_X\right) &= 0, \quad i = 1, \ldots, m &\quad (7) \\
\sum_{i=1}^{m} \Delta y_i \tilde{A}_i + \hat{D}_S &= 0 &\quad (8) \\
\left(\hat{D}_X\right)^T &= \hat{D}_X. &\quad (9)
\end{aligned}
$$

The equation (6) can be rewritten as

$$V^2 + V\hat{D}_S + \hat{D}_X V + \hat{D}_X \hat{D}_S - \mu I = 0.$$

Thus the desired scaled displacements are the (unique) solutions of the least squares problem

$$\min \left\| V^2 + V\hat{D}_S + \hat{D}_X V + \hat{D}_X \hat{D}_S - \mu I \right\|^2,$$

subject to the constraints (7)–(9), and the optimal value of this problem is zero. The above problem is computationally intractable due to the nonlinear term $\hat{D}_X \hat{D}_S$. Therefore we omit this term. This omission makes it important to specify which norm is used, since the optimal solution to our new least squares problem will depend on the norm. The norm which we choose is the norm induced by the inner product:

$$\langle A, B \rangle := \text{Tr}\left(V^{-1} A V^{-1} B\right), \quad \forall A = A^T,\ B = B^T.$$

This can also be viewed as the local norm induced by the Hessian of the self-concordant barrier

$$f(V) = -\ln \det(V),$$

since the Hessian of f evaluated at V is the linear operator:

$$\nabla^2 f(V) : H \mapsto V^{-1} H V^{-1}.$$

Thus we obtain the least squares problem

$$\min \left\| V^{-\frac{1}{2}} \left(V^2 + V\hat{D}_S + \hat{D}_X V - \mu I \right) V^{-\frac{1}{2}} \right\|^2,$$

subject to the constraints (7)–(9), and where the norm now indicates the Frobenius norm. For convenience, we also introduce the notation

$$U := \frac{1}{\sqrt{\mu}} V, \quad D_X := \frac{1}{\sqrt{\mu}} \hat{D}_X, \quad D_S := \frac{1}{\sqrt{\mu}} \hat{D}_S.$$

Using this notation, we can reformulate the above least squares problem as follows

$$(LQ) \quad \begin{cases} \min f(D_X, D_S) & := \left\| U + U^{\frac{1}{2}} D_S U^{-\frac{1}{2}} + U^{-\frac{1}{2}} D_X U^{\frac{1}{2}} - U^{-1} \right\|^2 \\ \text{s.t. } \text{Tr}\left(\tilde{A}_i D_X\right) & = 0, \quad i = 1 \ldots, m \\ {D_X}^T & = D_X \\ D_S & = -\sum_{i=1}^m \Delta y_i \tilde{A}_i. \end{cases}$$

In what follows we will frequently use the notation:

$$R := U + U^{\frac{1}{2}} D_S U^{-\frac{1}{2}} + U^{-\frac{1}{2}} D_X U^{\frac{1}{2}} - U^{-1}. \tag{10}$$

In other words, $\|R\|^2$ is the residual of the least squares problem (LQ).

Theorem 2.1 (Existence and uniqueness of the new direction) *The problem (LQ) determines the displacements D_X, Δy and D_S uniquely. Furthermore, one has $D_X = 0$ and $\Delta y = 0$ (whence $D_S = 0$), if and only if $U = I$ or, equivalently, $XS = \mu I$.*

Proof: We first show that (LQ) has a unique solution. Note that if we cast the elements of D_X and D_S as variables, then (LQ) is a feasible convex quadratic optimization problem (CQP) on a polyhedral set in the space $\mathbb{R}^{2n^2+m}$ and with an a priori lower bound on the optimal value (zero). The existence of the optimal solution for such a problem is well known. Hence, we need only consider the uniqueness of the solution. The system of linear optimality conditions of (LQ) has a unique solution if and only if the corresponding homogeneous system has only the zero solution. This homogeneous linear system is simply the system of optimality conditions of the CQP which is obtained by discarding the terms U and $-U^{-1}$ in the objective of (LQ). We will call this problem the *homogeneous* (LQ). In other words, (LQ) will have a unique solution if and only if the homogeneous (LQ) only has the zero solution. The homogeneous (LQ) has the objective function

$$\|R\|^2 = \left\| U^{\frac{1}{2}} D_S U^{-\frac{1}{2}} + U^{-\frac{1}{2}} D_X U^{\frac{1}{2}} \right\|^2.$$

Since $D_X = D_S = 0$ is a trivially an optimal solution of the homogeneous (LQ), every optimal solution of the homogeneous (LQ) must satisfy $\|R\| = 0$, which means

$$U^{\frac{1}{2}} D_S U^{-\frac{1}{2}} + U^{-\frac{1}{2}} D_X U^{\frac{1}{2}} = 0. \tag{11}$$

Pre-multiplying the right side of the above equality by $U^{\frac{1}{2}} D_X U^{-\frac{1}{2}}$ we obtain

$$U^{\frac{1}{2}} D_X D_S U^{-\frac{1}{2}} + U^{\frac{1}{2}} D_X U^{-1} D_X U^{\frac{1}{2}} = 0.$$

It follows that

$$\mathrm{Tr}\left(U^{\frac{1}{2}} D_X D_S U^{-\frac{1}{2}} + U^{\frac{1}{2}} D_X U^{-1} D_X U^{\frac{1}{2}} \right) = \mathrm{Tr}\left(U^{\frac{1}{2}} D_X U^{-1} D_X U^{\frac{1}{2}} \right) = 0.$$

Since U^{-1} is symmetric positive definite, the above relation means $U^{\frac{1}{2}} D_X U^{-\frac{1}{2}} = 0$ and hence $D_X = 0$. Substituting into (11), we get $D_S = 0$. It follows that $D_X = D_S = 0$ is the unique solution of the homogeneous (LQ). This proves the first statement of the theorem.

Now consider the case when the unique solution of (LQ) is the zero solution. Then we have

$$R = U - U^{-1},$$

whence

$$U^{-\frac{1}{2}} R U^{\frac{1}{2}} = U^{\frac{1}{2}} R U^{-\frac{1}{2}} = U - U^{-1}.$$

Hence R and $U^{-\frac{1}{2}} R U^{\frac{1}{2}}$ are symmetric in this case. Note that we can write the optimality conditions for (LQ) as follows:

$$\begin{aligned} U^{-\frac{1}{2}} R U^{\frac{1}{2}} &= \sum_{j=1}^{m} \lambda_j \tilde{A}_j + \sum_{k<\ell} \beta_{k\ell} E_{k\ell} \\ \mathrm{Tr}\left(\tilde{A}_i U^{\frac{1}{2}} R U^{-\frac{1}{2}} \right) &= 0, \quad i = 1, \ldots, m \end{aligned}$$

together with the constraints in (LQ). Here $E_{k\ell}$ denotes the $n \times n$ matrix that has a 1 in the (k, ℓ)-position, a -1 in the (ℓ, k)-position, and zeros elsewhere. The coefficients λ_j and $\beta_{k\ell}$

are arbitrary real numbers. Since the matrices $E_{k\ell}$ form a linearly independent basis of the space of all skew-symmetric matrices we can rewrite the optimality conditions as

$$
\begin{aligned}
U^{-\frac{1}{2}}RU^{\frac{1}{2}} &= \sum_{j=1}^{m} \lambda_j \tilde{A}_j + M && (12)\\
\mathrm{Tr}\left(\tilde{A}_i U^{\frac{1}{2}} R U^{-\frac{1}{2}}\right) &= 0, \quad i = 1, \ldots, m && (13)\\
\mathrm{Tr}\left(\tilde{A}_i D_X\right) &= 0, \quad i = 1, \ldots, m &&\\
D_X - D_X^T &= 0 && (14)\\
D_S &= -\sum_{i=1}^{m} \Delta y_i \tilde{A}_i, &&
\end{aligned}
$$

where M is a skew-symmetric matrix. Multiplying (12) from the right by $U^{\frac{1}{2}}RU^{-\frac{1}{2}}$ and using (13), we obtain

$$
\mathrm{Tr}\left(U^{-\frac{1}{2}}RU^{\frac{1}{2}}\left(U^{\frac{1}{2}}RU^{-\frac{1}{2}}\right)\right) = \mathrm{Tr}\left(U^{-\frac{1}{2}}RU^{\frac{1}{2}}U^{-\frac{1}{2}}RU^{\frac{1}{2}}\right) = \mathrm{Tr}\left(R^2\right) = \|R\|^2 = 0.
$$

Therefore, $R = 0$, and thus we obtain $U^2 = I$ $(XS = \mu I)$. Conversely, if $U = I$ then $R = D_X + D_S$ (see (10)) and the unique optimal solution of (LQ) is $D_X = D_S = 0$. This completes the proof. □

Remark 2.1 *If the optimal solution of (LQ) corresponds to $R = 0$, then the new direction coincides with all the search directions discussed in Section 1.1. This happens for example if $XS = \bar{\mu} I$ for some $\bar{\mu} > 0$, or if all the data matrices A_i and C are diagonal (in which case the SDO problem reduces to an LO problem).*

Remark 2.2 *Substitution of (12) into (13) leads to*

$$
\mathrm{Tr}\left(\tilde{A}_i U \left(\sum_{j=1}^{m} \lambda_j \tilde{A}_j + M\right) U^{-1}\right) = 0, \quad i = 1, \ldots, m,
$$

or

$$
\sum_{j=1}^{m} \lambda_j \mathrm{Tr}\left(\tilde{A}_i U \tilde{A}_j U^{-1}\right) = -\mathrm{Tr}\left(\tilde{A}_i U M U^{-1}\right), \quad i = 1, \ldots, m.
$$

The matrix with (i,j)-entry $\mathrm{Tr}\left(\tilde{A}_i U \tilde{A}_j U^{-1}\right)$ is positive definite, and hence nonsingular (see e.g. Lemma A.2.2 in De Klerk [2])). Hence, if M is known, then the above system determines λ uniquely. In particular, if $M = 0$ then also $\lambda = 0$.

3 Estimating the least squares residual

In the analysis of the new search direction it is essential to show that the residual of the least squares problem, $\|R\|$, is 'small enough' at the optimal solution of (LQ). The residual can be bounded from above in terms of the proximity to the target point μI, where the proximity is measured by:

$$\delta(X,S,\mu) := \frac{1}{2}\left\|U - U^{-1}\right\|. \tag{15}$$

Note that $\delta(X,S,\mu) = 0$ if and only if $XS = \mu I$. In what follows, we will use the notation $\delta := \delta(X,S,\mu)$ if no confusion is possible.

Let us define $D_V := D_X + D_S$ and $Q_V := D_X - D_S$. Note that $\|D_V\| = \|Q_V\|$. We can now decompose $R := U^{-1} - U + U^{\frac{1}{2}}D_S U^{-\frac{1}{2}} + U^{-\frac{1}{2}}D_X U^{\frac{1}{2}}$ into a symmetric and skew-symmetric component, say

$$R := R_{sym} + R_{skew},$$

where

$$R_{sym} = U^{-1} - U + \frac{1}{2}\left(U^{\frac{1}{2}}D_V U^{-\frac{1}{2}} + U^{-\frac{1}{2}}D_V U^{\frac{1}{2}}\right)$$

and

$$R_{skew} = \frac{1}{2}\left(U^{-\frac{1}{2}}Q_V U^{\frac{1}{2}} - U^{\frac{1}{2}}Q_V U^{-\frac{1}{2}}\right).$$

By construction, one has

$$\|R\|^2 = \|R_{sym}\|^2 + \|R_{skew}\|^2.$$

The new direction chooses $D_V \equiv D_X + D_S$ such that $\|R\|$ is minimized. In order to get an upper bound on the value $\|R\|$ for the new direction, we can consider the value of $\|R\|$ for a class of search directions where $U^{\frac{1}{2}}D_V U^{-\frac{1}{2}} = D_V$. In this way we obtain the bound:

$$\|R\|^2 \le 4\delta^2 + 2\mathrm{Tr}\left((U^{-1} - U)D_V\right) + \|D_V\|^2 + \left\|\frac{1}{2}\left(U^{-\frac{1}{2}}Q_V U^{\frac{1}{2}} - U^{\frac{1}{2}}Q_V U^{-\frac{1}{2}}\right)\right\|^2, \tag{16}$$

where we have used $\left\|U - U^{-1}\right\|^2 = 4\delta^2$. In order to get an upper bound on $\|R_{skew}\|^2$ (the last term in (16)) we use the following lemma.

Lemma 3.1 *Suppose that the matrix A is symmetric positive definite and $\xi(A) = \mathrm{Tr}\,(A^2) - 2n + \mathrm{Tr}\,(A^{-2})$. Then for any symmetric matrix $\bar{A}$, one has*

$$\|A\bar{A}A^{-1} - A^{-1}\bar{A}A\|^2 \le \frac{\xi(A^2)}{2}\|\bar{A}\|^2.$$

Proof: Since A is symmetric positive definite, we can assume in general that A is a diagonal matrix with $a_i > 0$ on the ith diagonal position, by taking an orthogonal transformation if necessary. Denoting $\hat{A} = A\bar{A}A^{-1} - A^{-1}\bar{A}A$, one has

$$\hat{A}_{ii} = 0, \quad \hat{A}_{ij} = \left(\frac{a_i}{a_j} - \frac{a_j}{a_i}\right)\bar{A}_{ij}.$$

The above relation means that

$$\begin{aligned}
\|\hat{A}\|^2 &\le \max_{i,j}\left(\frac{a_i^2}{a_j^2} - 2 + \frac{a_j^2}{a_i^2}\right)\|\bar{A}\|^2 \\
&\le \frac{1}{2}\max_{i,j}\left(a_i^4 + a_j^4 - 4 + \frac{1}{a_i^4} + \frac{1}{a_j^4}\right)\|\bar{A}\|^2 \\
&\le \frac{\xi(A^2)}{2}\|\bar{A}\|^2,
\end{aligned}$$

which gives the statement of the lemma. □

The lemma implies that

$$\begin{aligned}\|R_{skew}\|^2 &\equiv \left\|\frac{1}{2}\left(U^{-\frac{1}{2}}Q_V U^{\frac{1}{2}} - U^{\frac{1}{2}}Q_V U^{-\frac{1}{2}}\right)\right\|^2 \\ &\leq \frac{1}{8}\xi\left(U^{-1}\right)\|Q_V\|^2 \\ &= \frac{1}{2}\delta^2\|D_V\|^2,\end{aligned}$$

where we have used $\|D_V\| = \|Q_V\|$ and $\xi(U^{-1}) = 4\delta^2$. Substituting the bound for $\|R_{skew}\|$ into (16) yields

$$\|R\|^2 \leq 4\delta^2 + 2\mathrm{Tr}\left(D_V(U^{-1} - U)\right) + (1 + \frac{1}{2}\delta^2)\|D_V\|^2. \tag{17}$$

The right hand side is a convex quadratic function of D_V and is minimized by

$$D_V^{NT} = -\frac{1}{1 + \frac{1}{2}\delta^2}\left(U^{-1} - U\right), \tag{18}$$

which happens to be a damped step along the Nesterov-Todd direction (see e.g. De Klerk [2]). Substituting (18) into (17) yields

$$\begin{aligned}\|R\|^2 &\leq 4\delta^2 + 2\mathrm{Tr}\left(D_V^{NT}(U - U^{-1})\right) + (1 + \frac{1}{2}\delta^2)\|D_V^{NT}\|^2 \\ &= 4\delta^2 - \frac{2}{1 + \frac{1}{2}\delta^2}(4\delta^2) + \frac{1}{1 + \frac{1}{2}\delta^2}(4\delta^2) \\ &= 4\delta^2\left(\frac{\delta^2}{2 + \delta^2}\right).\end{aligned} \tag{19}$$

Let us now suppose that D_X, D_S are the solutions of (LQ), and denote

$$D_U := U^{\frac{1}{2}}D_S U^{-\frac{1}{2}} + U^{-\frac{1}{2}}D_X U^{\frac{1}{2}}.$$

Our main result in this section can be stated as follows.

Lemma 3.2 *Let δ be defined by (15). One has*

$$\frac{2\delta}{\sqrt{1 + \frac{1}{2}\delta^2}} \leq \|D_U\| \leq 2\delta. \tag{20}$$

Proof: ¿From the optimality conditions of (LQ) we immediately derive that

$$\mathrm{Tr}\left(RD_U^T\right) = 0.$$

Since $R = U^{-1} - U + D_U$ and R and D_U are orthogonal, we have

$$4\delta^2 \equiv \left\|U^{-1} - U\right\|^2 = \|R\|^2 + \|D_U\|^2 \leq 4\delta^2\left(\frac{\delta^2}{2 + \delta^2}\right) + \|D_U\|^2, \tag{21}$$

where the inequality follows from (19). The equations in (21) together with the nonnegativity of $\|R\|$ imply

$$4\delta^2 \equiv \left\|U^{-1} - U\right\|^2 \geq \|D_U\|^2 ,$$

and the inequality in (21) implies

$$\|D_U\|^2 \geq 4\delta^2 - 4\delta^2 \left(\frac{\delta^2}{2+\delta^2}\right) = 4\delta^2 \left(\frac{2}{2+\delta^2}\right).$$

Thus we have shown that

$$2\delta \geq \|D_U\| \geq \frac{2\delta}{\sqrt{1+\frac{1}{2}\delta^2}}. \tag{22}$$

□

4 Complexity analysis of a primal-dual method

In the present section, we will first propose a primal-dual path following method based on the new search direction, and subsequently perform a complexity analysis of the algorithm.

Generic primal-dual path following method

Parameters
  A centering parameter $\tau > 0$;
  An accuracy parameter $\epsilon > 0$;
  An updating parameter $\theta < 1$;
  A constant $\bar{c} > 0$;
  An initial centering parameter $\mu^0 > 0$.
Input
  A strictly feasible starting pair (X^0, S^0),
  satisfying $\delta\left(X^0, S^0, \mu^0\right) \leq \tau$.

begin
  $X := X^0$; $S := S^0$;
  while $\mathrm{Tr}(XS) > \epsilon$ **do**
    if $\delta(X, S, \mu) \leq \tau$ **do** (*outer iteration*)
      $\mu := (1-\theta)\mu$;
    else if $\delta(X, S, \mu) > \tau$ **do** (*inner iteration*)
      Compute $\Delta X, \Delta S$ by solving (LQ);
      Find α such that $\Phi(X, S, \mu) - \Phi(X + \alpha\Delta X, S + \alpha\Delta S, \mu) \geq \bar{c}$;
      (A suitable default choice for α is given by (23).)
      $X := X + \alpha\Delta X$, $S := S + \alpha\Delta S$;

    end
  end
end

Recall that

$$\Phi(X, S, \mu) = \frac{\operatorname{Tr}(XS)}{\mu} - n - \log\det(XS) + n\log\mu.$$

In the update of the iterate, we require that the step length α is chosen such that the potential function $\Phi(X, S, \mu)$ decreases sufficiently. Lemma 4.2 will give a default value for α.

It is easy to verify that the potential function can also be rewritten as

$$f(U) = \Phi(X, S, \mu) = \operatorname{Tr}\left(U^2\right) - n - \ln\det\left(U^2\right).$$

Assuming that D_X, D_S are solutions of (LQ), we want to estimate the decreasing value of the potential function, given by:

$$\begin{aligned}\Delta\Phi(\alpha) &= f(U) - (\operatorname{Tr}((U + \alpha D_X)(U + \alpha D_S)) - n - \ln\det(U + \alpha D_X)(U + \alpha D_S)) \\ &= -\alpha\operatorname{Tr}(UD_S + D_X U) + \ln\det(I + U^{-\frac{1}{2}} D_X U^{-\frac{1}{2}})(U + \alpha U^{-\frac{1}{2}} D_S U^{-\frac{1}{2}}),\end{aligned}$$

where we have used the orthogonality of D_X and D_S. Now we have the following general bound on the reduction $\Delta\Phi(\alpha)$ which holds for any search direction. (For a proof see e.g. Jiang [4] and Roos *et al.* [9] for the linear case.)

Theorem 4.1 *Let (X, S) be a strictly feasible pair and let (D_X, D_S) be any search direction defined at (X, S). Then:*

$$\Delta\Phi(\alpha) \geq -\alpha\operatorname{Tr}(UD_V) + \alpha\operatorname{Tr}\left(U^{-1}D_V\right) - \psi(-\alpha h)$$

where $\psi(t) := t - \ln(1 + t)$ (see Figure 1), and

$$h^2 = \operatorname{Tr}\left(U^{-1}D_X U^{-1}D_X + U^{-1}D_S U^{-1}D_S\right).$$

Moreover any value of α satisfying $\alpha \leq \frac{1}{h}$ is a feasible step length.

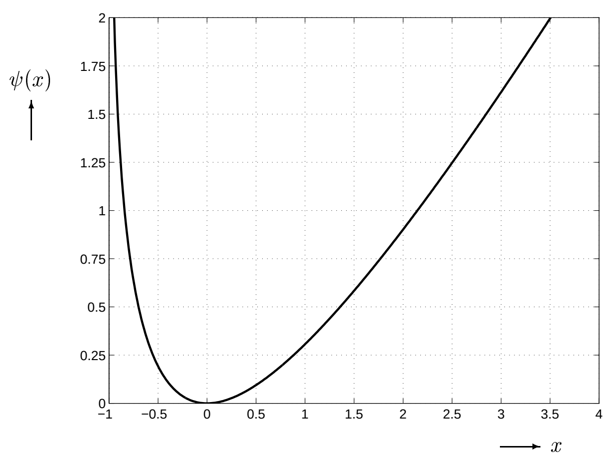

Figure 1: The graph of ψ.

Corollary 4.1 *Let (D_X, D_S) denote the optimal solution of (LQ) (the new search direction). One has:*

$$\Delta\Phi(\alpha) \geq \alpha\|D_U\|^2 - \psi(-\alpha h).$$

Proof: Follows from Theorem 4.1 by using the definition of R in (10) to show that

$$-\mathrm{Tr}\,(UD_V) + \mathrm{Tr}\,\left(U^{-1}D_V\right) = \frac{1}{2}\left(4\delta^2 - \|R\|^2 + \|D_U\|^2\right)$$

and using that

$$\|R\|^2 + \|D_U\|^2 = 4\delta^2$$

holds for the new direction (see (21)). □

All that remains is to give an upper bound for the term $-\psi(-\alpha h)$. This can be done by using the following lemma.

Lemma 4.1 *Let (D_X, D_S) denote the optimal solution of (LQ). One has*

$$h \leq \rho(\delta)\|D_U\|$$

where $\rho(\delta) := \delta + \sqrt{1+\delta^2}$.

Proof: By definition:

$$\begin{aligned} h^2 &= \mathrm{Tr}\,\left(U^{-1}D_XU^{-1}D_X + U^{-1}D_SU^{-1}D_S\right) \\ &= \mathrm{Tr}\,\left(U^{-2}\left(UD_XU^{-1}D_X + UD_SU^{-1}D_S\right)\right) \\ &\leq \lambda_{\max}\left(U^{-2}\right)\mathrm{Tr}\,\left(UD_XU^{-1}D_X + UD_SU^{-1}D_S\right) \\ &\leq \rho^2(\delta)\mathrm{Tr}\,\left(UD_XU^{-1}D_X + UD_SU^{-1}D_S\right) \\ &= \rho^2(\delta)\|D_U\|^2 \end{aligned}$$

where the last inequality is a result by Jiang [4]. (See Roos *at al.* [9] for the analogous result in the linear case.) □

Lemma 4.2 *Let (D_X, D_S) denote the optimal solution of (LQ). One has*

$$\Delta\Phi(\bar{\alpha}) \geq \psi\left(\frac{\|D_U\|}{\rho(\delta)}\right) \geq \psi\left(\frac{2\delta}{\rho(\delta)\sqrt{1+\frac{1}{2}\delta^2}}\right),$$

for

$$\bar{\alpha} := \frac{1}{h} - \frac{1}{\|D_U\|^2 + h}.$$

Proof: From Corollary 4.1 we have

$$\begin{aligned} \Delta\Phi(\alpha) &\geq \alpha\|D_U\|^2 - \psi(-\alpha h) \\ &\equiv \alpha\|D_U\|^2 + \alpha h + \ln(1-\alpha h). \end{aligned}$$

The right hand side of the inequality is maximized by

$$\bar{\alpha} = \frac{1}{h} - \frac{1}{\|D_U\|^2 + h}. \tag{23}$$

This maximizer yields the bound

$$\Delta\Phi(\bar{\alpha}) \geq \psi\left(\frac{\|D_U\|^2}{h}\right),$$

which, by Lemma 4.1, implies:

$$\Delta\Phi(\bar{\alpha}) \geq \psi\left(\frac{\|D_U\|}{\rho(\delta)}\right).$$

Finally we use Lemma 3.2 to complete the proof. □

Now we show that δ is bounded in terms of the potential function Φ, and vice versa. To this end, we use the following lemma which was proved for linear optimization by Roos *et al.* [9]. The extension of the proof to the SDO case is mechanical and is therefore omitted.

Lemma 4.3 *Let* $\delta := \delta(X, S; \mu)$ *and* $\rho(\delta) := \delta + \sqrt{1+\delta^2}$. *Then*

$$\psi\left(\frac{-2\delta}{\rho(\delta)}\right) \leq \Phi(X, S, \mu) \leq \psi\left(2\delta\rho(\delta)\right).$$

The statement of the Lemma is illustrated in Figure 2.

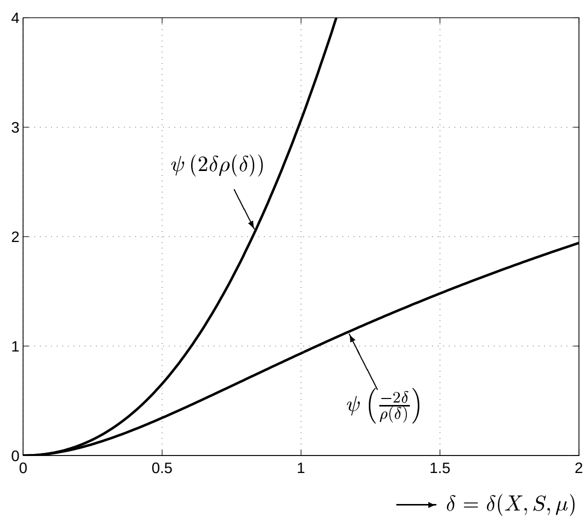

Figure 2: Bounds for $\Phi(X, S, \mu)$.

4.1 Small update methods

We are now in a position to perform the complexity analysis for a small update version of the algorithm. To fix our ideas, we choose the parameters

$$\tau = \frac{1}{2}, \quad \theta = \frac{1}{10\sqrt{n}}.$$

We assume at the current iterates (X, S), the proximity measure satisfies $\delta(X, S, \mu) \leq \tau = \frac{1}{2}$. In this situation, we perform the update $\mu^+ = (1 - \theta)\mu$ (outer iteration). Analogously to the linear optimization case, one has (see Lemma IV. 36 in [9]):

$$\delta(X, S, \mu^+) \leq \frac{2\delta + \theta\sqrt{n}}{2\sqrt{1-\theta}} \leq \frac{2\tau + \sqrt{n}\theta}{2\sqrt{1-\theta}} = \frac{1 + \frac{1}{10}}{2\sqrt{1-\theta}} < 0.571.$$

This also means (by Lemma 4.3) that at the beginning of the inner iterative procedure, one has

$$\Phi(X, S, \mu^+) \leq \psi(2\delta\rho(\delta)) \leq 0.878.$$

This bound implies that the proximity $\delta(X, S, \mu^+)$ is also bounded from above by a constant, by Lemma 4.3 (see Figure 2):

$$\delta \leq 0.57.$$

At each inner iteration one has $\delta \geq \frac{1}{2}$, which implies

$$\|D_U\| \geq \frac{2\delta}{\sqrt{1 + \frac{1}{2}\delta^2}} = \frac{1}{\sqrt{1.125}} \approx 0.9428.$$

Lemma 4.2 shows that the reduction of the potential function is at least

$$\psi\left(\frac{\|D_U\|}{\rho(\delta)}\right) \geq 0.13. \tag{24}$$

In order to guarantee that $\delta(X, S, \mu) \leq \frac{1}{2}$ one must reduce the value of Φ below 0.344 (see Figure 2). The bound in (24) implies that, after at most

$$\lceil (0.878 - 0.344)/0.13 \rceil = 5 \tag{25}$$

inner iterations, the proximity measure again satisfies $\delta(X, S, \mu^+) \leq \frac{1}{2}$. Hence we have the following complexity bound for the algorithm.

Theorem 4.2 *If $\tau = \frac{1}{2}$ and $\theta = \frac{1}{10\sqrt{n}}$, the total number of iterations required by the primal-dual path following algorithm is no more than*

$$\left\lceil 50\sqrt{n} \log \frac{n\mu^0}{\epsilon} \right\rceil.$$

Proof: It can easily be shown that the number of barrier parameter updates (outer iterations) is given by (cf. Lemma II.17, in [9])

$$\left\lceil \frac{1}{\theta} \log \frac{n\mu^0}{\epsilon} \right\rceil.$$

Multiplication of this number by the bound (25) yields the theorem. □

Remark 4.1 *We have only analysed one special small update algorithm, but one can easily derive similar results for any fixed* $\tau > 0$ *and* θ *of the order* $O\left(\frac{1}{\sqrt{n}}\right)$.

4.2 Large update methods

Finally, we consider large update methods based on our search direction. We therefore choose the parameter $\tau > 0$, and $\theta \in (0, 1)$ independently of n. By (22) it follows that

$$\|D_U\| \geq \frac{2\delta}{\sqrt{1 + \frac{1}{2}\delta^2}} \geq 1.$$

After updating the parameter μ, we get, as before,

$$\delta(X, S, \mu^+) \leq \frac{2\tau + \sqrt{n}\theta}{2\sqrt{1-\theta}},$$

which implies

$$\delta^2(X, S, \mu^+) \leq \frac{(2\tau + \sqrt{n}\theta)^2}{4(1-\theta)} \leq O\left(\frac{n\theta^2}{1-\theta}\right).$$

It follows from Lemma 4.3 that

$$\Phi(X, S, \mu^+) \leq \psi(2\delta\rho(\delta)) \leq O\left(\delta^2\right).$$

At each inner iteration, the decrease of the potential function is at least

$$\psi\left(\frac{\|D_U\|}{\rho(\delta)}\right) \geq \Omega\left(\frac{1}{\delta^2}\right),$$

since

$$\frac{\|D_U\|}{\rho(\delta)} \geq \Omega\left(\frac{1}{\delta}\right),$$

and since $\psi(t) \geq t^2$ for sufficiently small t. This means after at most

$$O\left(\delta^4\right) \leq O\left(\left(\frac{n\theta^2}{(1-\theta)}\right)^2\right)$$

inner iterations, we have 'recentered'. Hence the complexity of the large update method follows.

Theorem 4.3 *The total number of iterations required by the large-update primal-dual path following algorithm is not more than*

$$O\left(\left(\frac{n\theta^2}{(1-\theta)}\right)^2 \frac{1}{\theta} \log \frac{n\mu^0}{\epsilon}\right).$$

Remark 4.2 *The bound in Theorem 4.2 yielded the familiar* $O(\sqrt{n})$ *bound for small update methods where* $\theta \leq O\left(\frac{1}{\sqrt{n}}\right)$*, which is the best known bound for these methods. The bound for large update methods of Theorem 4.3 is worse than the best known bound by a factor* $\frac{n\theta^2}{(1-\theta)}$*, though.*

5 Conclusions

We have presented a primal-dual Gauss-Newton-type direction for semidefinite optimization which allows polynomial worst-case iteration complexity analysis. This analysis was inspired by the Gauss-Newton direction of Kruk *et al.* [6], but the new direction seems much more amenable to complexity analysis, due to the use of scaling and a local norm in the definition of the least squares problem. It remains to be seen if the new direction is as numerically robust as the direction of Kruk *et al.* but the close relation between the directions suggests that this might well be the case.

Some important remaining issues are:

- To find an efficient way of computing the new directions;
- To improve the complexity bound for the new direction in the case of large update methods (see Remark 4.2);
- To investigate possible superlinear convergence properties of the new direction. To this end, it will be necessary to analyse the behaviour of the *affine scaling* direction which is defined by setting $\mu = 0$ in the definition of the least squares problem (LQ).

References

[1] F. Alizadeh, J.-P.A. Haeberley, and M.L. Overton. Primal-dual methods for semidefinite programming: convergence rates, stability and numerical results. *SIAM Journal on Optimization*, 8(3):746–768, 1998.

[2] E. de Klerk. *Interior Point Methods for Semidefinite Programming.* PhD thesis, Delft University of Technology, Delft, The Netherlands, 1997.

[3] 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.

[4] J. Jiang. A long step primal–dual path following method for semidefinite programming. *OR Letters*, 23(1,2):53–62, 1998.

[5] M. Kojima, S. Shindoh, and S. Hara. Interior point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM Journal on Optimization*, 7(1):88–125, 1997.

[6] S. Kruk, M. Muramatsu, F. Rendl, R.J. Vanderbei, and H. Wolkowicz. The Gauss–Newton direction in semidefinite programming. Research report CORR 98-16, University of Waterloo, Dept. Combinatorics and Optimization, Waterloo, Canada, 1998.

[7] R.D.C. Monteiro. Primal-dual path-following algorithms for semidefinite programming. *SIAM Journal on Optimization*, 7(3):663–678, 1997.

[8] Yu. Nesterov and M.J. Todd. Self–scaled barriers and interior–point methods for convex programming. *Mathematics of Operations Research*, 22(1):1–42, 1997.

[9] C. Roos, T. Terlaky, and J.-Ph. Vial. *Theory and Algorithms for Linear Optimization: An interior point approach.* John Wiley & Sons, New York, 1997.

[10] M.J. Todd, K.C. Toh, and R.H. Tütüncü. On the Nesterov-Todd direction in semidefinite programming. *SIAM Journal on Optimization*, 8(3):766–796, 1998.

[11] Y. Zhang. On extending some primal-dual interior point algorithms from linear programming to semidefinite programming. *SIAM Journal on Optimization*, 8(2):365–386, 1998.