SIAM J. OPTIM.
Vol. 11, No. 4, pp. 870–888

A SCALED GAUSS–NEWTON PRIMAL-DUAL SEARCH DIRECTION FOR SEMIDEFINITE OPTIMIZATION*

E. DE KLERK†, J. PENG†, C. ROOS†, AND T. TERLAKY‡

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. As opposed to 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. [*The Gauss–Newton Direction in Semidefinite Programming*, Research report CORR 98-16, University of Waterloo, Waterloo, Canada, 1998] 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 and computational evaluation of the resulting primal-dual algorithm.

Key words. semidefinite optimization, primal-dual search directions, interior point algorithms

AMS subject classification. 65K

PII. S1052623499352632

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 ways to obtain primal-dual search directions. Different directions arise when the perturbed optimality conditions are linearized and subsequently symmetrized (see section 1); a quite comprehensive survey of the search directions obtained this way may be found in [11]. 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). 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. We further show that the new direction is closely related to the well-known (primal) H..K..M and dual H..K..M directions (see the definitions in section 1); the primal part of the new direction coincides with the dual part of the (primal) H..K..M direction, and the dual part of the new direction is simply the primal

*Received by the editors February 22, 1999; accepted for publication (in revised form) August 14, 2000; published electronically March 15, 2001.

http://www.siam.org/journals/siopt/11-4/35263.html

†Faculty of Information Technology and Systems, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands (E.deKlerk@ITS.TUDelft.nl, j.peng@ITS.TUDelft.nl, c.roos@ITS.TUDelft.nl).

‡Department of Computing and Software, McMaster University, Hamilton, ON, Canada (Terlaky@cas.mcmaster.ca). This research was done while the author was working at Delft University of Technology.

part of the dual H..K..M direction. Finally, we present some numerical experiments with the new direction.

Preliminaries. We consider the SDO problem in the standard form. Thus the primal problem (P) is given by

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

and its dual problem (D) is

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

where C and the A_i's are symmetric $n \times n$ matrices, $b, y \in \mathbb{R}^m$, and $X \succeq 0$ means that 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

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

If these conditions are perturbed to

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

for some $\mu > 0$ 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 center 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

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

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 $X + \Delta X \succeq 0$, $S + \Delta S \succeq 0$

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

TABLE 1
Choices for the scaling matrix P.

P	Reference	Abbreviation
$\left[X^{\frac{1}{2}}\left(X^{\frac{1}{2}}SX^{\frac{1}{2}}\right)^{-\frac{1}{2}}X^{\frac{1}{2}}\right]^{-\frac{1}{2}}$	Nesterov and Todd [9]	NT
$X^{-\frac{1}{2}}$	Monteiro [7], Kojima, Shindoh, and Hara [5];	DH..K..M
$S^{\frac{1}{2}}$	Monteiro [7], Helmberg et al. [3], Kojima, Shindoh, and Hara [5];	PH..K..M
I	Alizadeh, Haeberly, and Overton [1]	AHO

and

$$(X + \Delta X)(S + \Delta S) = \mu I, \tag{1}$$

$$\mathrm{Tr}\,(A_i \Delta X) = 0, \quad i = 1, \ldots, m, \tag{2}$$

$$\sum_{i=1}^{m} \Delta y_i A_i + \Delta S = 0, \tag{3}$$

$$(\Delta X)^T = \Delta X, \quad (\Delta S)^T = \Delta S. \tag{4}$$

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, (1) is nonlinear, and primal-dual methods differ with regard to how it is linearized. Care must be taken to ensure that the resulting linear system is not overdetermined. Zhang [14] suggested discarding the symmetry requirements (4) and replacing the nonlinear equation by

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

where H_P is the linear transformation given by

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

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 solvable (for the AHO direction ($P = I$) solvability is only guaranteed if (X, S) lies in a certain neighborhood of the central path), and the solution matrices $\Delta X$and ΔS are symmetric.

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 (s.t.)

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

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 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 above-mentioned 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 $\hat{D}_S$:

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

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:

$$\left(V + \hat{D}_X \right) \left(V + \hat{D}_S \right) = \mu I, \tag{6}$$

$$\mathrm{Tr} \left(\tilde{A}_i \hat{D}_X \right) = 0, \quad i = 1, \ldots, m, \tag{7}$$

$$\sum_{i=1}^{m} \Delta y_i \tilde{A}_i + \hat{D}_S = 0, \tag{8}$$

$$\left(\hat{D}_X \right)^T = \hat{D}_X. \tag{9}$$

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. We now omit the nonlinear term $\hat{D}_X \hat{D}_S$ from the objective function of the least squares problem. 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:

$$\langle A, B\rangle := \mathrm{Tr}\left(V^{-1}AV^{-1}B^T\right) \quad \forall A \in \mathbb{R}^{n\times n},\ B \in \mathbb{R}^{n\times n}.$$

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

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

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

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

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

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

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

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

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, $f\left(D_X, D_S\right) = \frac{1}{2}\|R\|^2$ is the residual of the least squares problem (LQ). Note that the derivatives of f with respect to D_X and D_S are, respectively, given by

$$\nabla_{D_X} f(D_X, D_S) = U^{-\frac{1}{2}}RU^{\frac{1}{2}}, \quad \nabla_{D_S} f(D_X, D_S) = U^{\frac{1}{2}}RU^{-\frac{1}{2}}. \tag{11}$$

Optimality conditions for the least squares problem. We can formulate the optimality conditions for the least squares problem (LQ) by forming the Lagrangian:

$$L(D_X, D_S, \Delta\tilde{y}, \lambda, M_1, M_2) := f(D_X, D_S) - \sum_{i=1}^m \lambda_i \mathrm{Tr}\left(\tilde{A}_i D_X\right) + \mathrm{Tr}\left(\left(D_X - D_X^T\right)M_1\right)$$
$$+\mathrm{Tr}\left(M_2\left(D_S + \sum_{i=1}^m \Delta\tilde{y}_i\tilde{A}_i\right)\right),$$

where $\lambda \in \mathbb{R}^m$, $M_1 \in \mathbb{R}^{n\times n}$, and $M_2 \in \mathbb{R}^{n\times n}$ are Lagrange multipliers. Using the expressions in (11), one can easily rewrite the optimality condition

$$\nabla L(D_X, D_S, \Delta\tilde{y}, \lambda, M_1, M_2) = 0$$

as

$$U^{-\frac{1}{2}} R U^{\frac{1}{2}} = \sum_{i=1}^{m} \lambda_i \tilde{A}_i + M, \tag{12}$$

$$\mathrm{Tr}\left(\tilde{A}_i U^{\frac{1}{2}} R U^{-\frac{1}{2}}\right) = 0, \quad i = 1, \dots, m, \tag{13}$$

$$\mathrm{Tr}\left(\tilde{A}_i D_X\right) = 0, \quad i = 1, \dots, m, \tag{14}$$

$$D_X - D_X^T = 0, \tag{15}$$

$$D_S = -\sum_{i=1}^{m} \Delta\tilde{y}_i \tilde{A}_i, \tag{16}$$

where $M = M_1^T - M_1$ is a skew-symmetric matrix.

Existence and uniqueness of the new direction. We now state an existence and uniqueness result for the new search direction.

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

This result can be proved by using the optimality conditions of (LQ). We omit such a proof here, since the theorem will follow from results in section 5, where we will explore the relation between the new direction and directions from literature.

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) if the current iterate is close enough to the central path. The residual can be bounded from above in terms of the proximity to the target point μI, where the proximity is measured by

$$\delta(X, S, \mu) := \frac{1}{2}\left\|U - U^{-1}\right\|. \tag{17}$$

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 $D_V \equiv D_X + D_S$ is chosen 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

$$(18)\; \|R\|^2 \le 4\delta^2 + 2\mathrm{Tr}\left((U^{-1} - U)D_V\right) + \|D_V\|^2 + \left\| \frac{1}{2}\left(U^{-\frac{1}{2}} Q_V U^{\frac{1}{2}} - U^{\frac{1}{2}} Q_V U^{-\frac{1}{2}}\right)\right\|^2,$$

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 (18)) we use the following lemma.

LEMMA 3.1. *Suppose that the $n \times n$ matrix A is symmetric positive definite and $\xi(A) = \mathrm{Tr}\left(A^2\right) - 2n + \mathrm{Tr}\left(A^{-2}\right)$. 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

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

The above relation means that

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

where the second inequality can easily be verified by calculus, and the third inequality follows by noting that

$$\xi(A^2) = \sum_{i=1}^{n}\left(a_i^4 + \frac{1}{a_i^4} - 2\right). \qquad \square$$

The lemma implies that

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

where we have used $\|D_V\| = \|Q_V\|$ and $\xi(U^{-1}) = 4\delta^2$. Substituting the bound for $\|R_{skew}\|$ into (18) yields

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

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

$$D_V^{NT} = -\frac{1}{1 + \frac{1}{2}\delta^2}\left(U^{-1} - U\right), \tag{20}$$

which happens to be a damped step along the Nesterov–Todd direction (see, e.g., de Klerk [2]). Substituting (20) into (19) yields

$$\begin{aligned}
\|R\|^2 &\leq 4\delta^2 + 2\mathrm{Tr}\left(D_V^{NT}(U - U^{-1})\right) + \left(1 + \frac{1}{2}\delta^2\right)\|D_V^{NT}\|^2 \\
&= 4\delta^2 - \frac{2}{1 + \frac{1}{2}\delta^2}(4\delta^2) + \frac{1}{1 + \frac{1}{2}\delta^2}(4\delta^2) \\
&= 4\delta^2\left(\frac{\delta^2}{2 + \delta^2}\right). \qquad (21)
\end{aligned}$$

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

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

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

LEMMA 3.2. *Let δ be defined by* (17). *One has*

$$\frac{2\delta}{\sqrt{1 + \frac{1}{2}\delta^2}} \leq \|R_U\| \leq 2\delta. \tag{22}$$

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

$$\mathrm{Tr}\left(R^T R_U\right) = 0,$$

by noting that (12), (15), and (14) imply

$$\mathrm{Tr}\left(R^T U^{-\frac{1}{2}} D_X U^{\frac{1}{2}}\right) = 0$$

and (13) and (16) imply

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

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

$$4\delta^2 \equiv \left\|U^{-1} - U\right\|^2 = \|R\|^2 + \|R_U\|^2 \leq 4\delta^2\left(\frac{\delta^2}{2 + \delta^2}\right) + \|R_U\|^2, \tag{23}$$

where the inequality follows from (21). The equations in (23) together with the nonnegativity of $\|R\|$ imply

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

and the inequality in (23) implies

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

Thus we have shown that

$$2\delta \geq \|R_U\| \geq \frac{2\delta}{\sqrt{1+\frac{1}{2}\delta^2}}. \qquad \square \tag{24}$$

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 we will subsequently perform a complexity analysis of the algorithm.

GENERIC PRIMAL-DUAL PATH FOLLOWING ALGORITHM.

Input
  A strictly feasible starting pair (X^0, S^0), satisfying $\delta\left(X^0, S^0, \mu^0\right) \leq \tau$.
Parameters
  A centering parameter $\tau > 0$;
  An accuracy parameter $\epsilon > 0$;
  An updating parameter $\theta < 1$;
  An initial centering parameter $\mu^0 > 0$.
$X := X^0$; $S := S^0$;
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 $\Phi(X, S, \mu) - \Phi(X + \alpha\Delta X, S + \alpha\Delta S, \mu)$ is sufficiently large;
    (A suitable default choice for α is given by (26).)
    $X := X + \alpha\Delta X$, $S := S + \alpha\Delta S$;
  end
end

Recall that

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

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

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

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

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

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

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

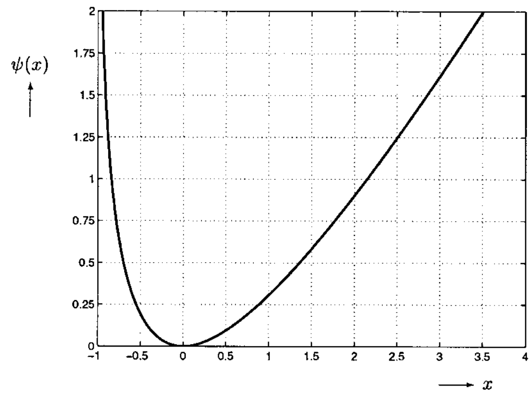

FIG. 1. *The graph of* ψ.

proof see, e.g., Jiang [4] and Roos, Terlaky, and Vial [10, Lemma II.69] for the linear optimization case.)

THEOREM 4.1. *Let* (X, S) *be a strictly feasible pair and let* (D_X, D_S) *be any feasible solution to problem (LQ); define* $D_V := D_X + D_S$. *Then*

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

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

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

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

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

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

Proof. Using the definition of R in (10) and $\mathrm{Tr}\,\left(R^T R_U\right) = 0$ one has

$$-\mathrm{Tr}\,(UD_V) + \mathrm{Tr}\,\left(U^{-1}D_V\right) = \mathrm{Tr}\,\left(R_U^T\left(U^{-1} - U\right)\right) = \mathrm{Tr}\,\left(R_U^T\left(R_U - R\right)\right) = \|R_U\|^2. \tag{25}$$

The required result now follows from Theorem 4.1. □

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*

$$h \leq \rho(\delta)\|R_U\|,$$

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

Proof. By definition,

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

where the last inequality is a result by Jiang [4]. (See Roos, Terlaky, and Vial [10, Lemma II.60] for the analogous result in the linear optimization case.) ∎

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

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

for

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

Proof. From Corollary 4.1 we have

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

The right-hand side of the inequality is maximized by

$$\bar{\alpha} = \frac{1}{h} - \frac{1}{\|R_U\|^2 + h}. \tag{26}$$

This maximizer yields the bound

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

which, by Lemma 4.1, implies

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

Finally we use Lemma 3.2 to complete the proof. ∎

Now we show that δ is bounded in terms of the barrier function Φ, and vice versa. To this end, we use the following lemma which was proved for linear optimization by Roos, Terlaky, and Vial [10, Lemma II.67]. 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.

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

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

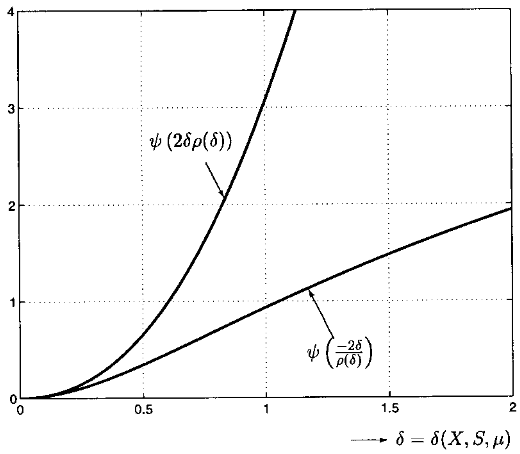

FIG. 2. *Bounds for* $\Phi(X, S, \mu)$.

We assume that 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 [10])

$$\delta(X, S, \mu^+) \leq \frac{2\delta + \theta\sqrt{n}}{2\sqrt{1-\theta}} \leq \frac{2\tau + \sqrt{n}\theta}{2\sqrt{1-\theta}} < 0.58.$$

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

$$\Phi(X, S, \mu^+) \leq \psi(2\delta\rho(\delta)) \leq 0.910.$$

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

$$\delta \leq 0.98.$$

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

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

by Lemma 3.2. Lemma 4.2 shows that the reduction of the barrier function is at least

$$\psi\left(\frac{\|R_U\|}{\rho(\delta)}\right) \geq 0.062. \tag{27}$$

In order to guarantee that $\delta(X, S, \mu^+) \leq \frac{1}{2}$ at the end of the inner iteration phase, one must reduce the value of Φ to below 0.344 (see Figure 2). The bound in (27) implies that, after at most

$$\lceil (0.910 - 0.344)/0.062 \rceil = 10 \tag{28}$$

inner iterations, we have computed a pair (X, S) such that $\delta(X, S, \mu^+) \leq \frac{1}{2}$. 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*

$$\left\lceil 100\sqrt{n}\log\frac{2.5n\mu^0}{\epsilon}\right\rceil .$$

Proof. It can easily be shown that after

$$\left\lceil \frac{1}{\theta}\log\frac{n\mu^0}{\epsilon}\right\rceil \tag{29}$$

barrier parameter updates (outer iterations) one has $n\mu \le \epsilon$ (cf. Lemma II.17 in [10]).

At the end of the inner iterations with respect to μ one has computed a pair (X, S) such that $\delta(X, S, \mu) \le \frac{1}{2}$. Using the definition of δ, it is trivial to show that this implies

$$\mathrm{Tr}\,(XS) \le 2.5n\mu,$$

and consequently $\mathrm{Tr}\,(XS) \le 2.5\epsilon$.

Replacing ϵ by $\epsilon/2.5$ and multiplying the number of outer iterations in (29) by the bound (28) yields the theorem. ∎

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

5. Relation to other search directions. In this section we show that the scaled Gauss–Newton (SGN) direction introduced in this paper is closely related to the primal and dual H..K..M directions (see Table 1). In particular, the ΔX part of the SGN direction is simply the ΔX part of the dual H..K..M direction, while the ΔS part of the SGN direction is the same as the ΔS part of the primal H..K..M direction. Note that this relationship implies Theorem 2.1.

The key in proving this is to decompose problem (LQ) into two independent subproblems. To this end, recall that for all feasible D_X and D_S, it holds that $\mathrm{Tr}\,(D_X D_S) = 0$. Using this fact, we can rewrite the objective of problem (LQ) as

$$\left\|U + U^{\frac{1}{2}} D_S U^{-\frac{1}{2}} - U^{-1}\right\|^2 + \left\|U + U^{-\frac{1}{2}} D_X U^{\frac{1}{2}} - U^{-1}\right\|^2 - \left\|U - U^{-1}\right\|^2 . \tag{30}$$

Omitting the last (constant) term in the last expression, we can separate problem (LQ) into two subproblems,

$$\text{(SGN1)} \quad \min_{D_X} \left\|U + U^{-\frac{1}{2}} D_X U^{\frac{1}{2}} - U^{-1}\right\|^2, \qquad \mathrm{Tr}\left(\tilde{A}_i D_X\right) = 0, \quad D_X = D_X^T;$$

and

$$\text{(SGN2)} \quad \min_{D_S} \left\|U + U^{\frac{1}{2}} D_S U^{-\frac{1}{2}} - U^{-1}\right\|^2, \qquad D_S = -\sum_{i=1}^{m} \Delta y_i \tilde{A}_i.$$

To compute the SGN direction, one can solve the two independent subproblems (SGN1) and (SGN2). Now let us recall the definition of the primal H..K..M direction. As observed by Monteiro (see Lemma 2.1 in [7] and Kojima, Shindoh, and Hara [5]), the primal H..K..M direction is the unique solution of the following linear system:

$$\begin{aligned} XS + X\Delta S + (\Delta X + W)S = \mu I, \\ \mathrm{Tr}(A_i \Delta X) = 0; \quad i = 1, \ldots, m, \\ \sum_{i=1}^{m} \Delta y_i A_i + \Delta S = 0, \quad W + W^T = 0, \quad \Delta X = \Delta X^T. \end{aligned}$$

Premultiplying the first equation in the above system by $D^{-1/2}$ and postmultiplying by $D^{1/2}$, and then dividing by μ, we can rewrite the above system in the scaled space as

$$U^2 + UD_S + (D_X + \tilde{W})U = I,$$

$$\mathrm{Tr}\left(\tilde{A}_i D_X\right) = 0, \quad i = 1, \ldots, m, \tag{31}$$

$$\sum_{i=1}^{m} \Delta\tilde{y}_i \tilde{A}_i + D_S = 0, \quad \tilde{W} + \tilde{W}^T = 0, \quad D_X = D_X^T, \tag{32}$$

where $\tilde{W} = \frac{1}{\mu} D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$ is skew symmetric and $\Delta\tilde{y}_i = \frac{1}{\mu}\Delta y_i$ as before. Again by pre- and postmultiplying the first equation by $U^{-1/2}$ we obtain

$$U - U^{-1} + U^{\frac{1}{2}} D_S U^{-\frac{1}{2}} + U^{-\frac{1}{2}}(D_X + \tilde{W})U^{\frac{1}{2}} = 0. \tag{33}$$

Now we state our main result in this section.

PROPOSITION 5.1. *Suppose that ΔS^* is the solution of the primal H..K..M direction. Then $D_S^* = \frac{1}{\mu} D^{\frac{1}{2}} \Delta S^* D^{\frac{1}{2}}$ is the unique solution of the problem* (SGN2).

Proof. The KKT system for problem (SGN2) can easily be written in the form (33), (31), and (32). ∎

We can approach the solution of problem (SGN1) in exactly the same way, by observing that the dual H..K..M direction is the unique solution of the following problem (see [5]):

$$\begin{array}{rcll} XS + X(\Delta S + W) + \Delta XS &=& \mu I, & \\ \mathrm{Tr}(A_i \Delta X) &=& 0, & i = 1, \ldots, m, \\ \sum_{i=1}^{m} \Delta y_i A_i + \Delta S &=& 0, & \\ \Delta X &=& \Delta X^T, & W + W^T = 0. \end{array} \tag{34}$$

In the same way as before, one can now prove the following.

PROPOSITION 5.2. *Suppose that ΔX^* is the solution of the dual H..K..M directions. Then*

$$D_X^* = \frac{1}{\mu} D^{-\frac{1}{2}} \Delta X^* D^{-\frac{1}{2}}$$

is the unique solution of the problem (SGN1).

TABLE 2
Average number of iterations for the SDTP3 algorithm IPF using various search directions.

Test set	AHO	PH..K..M	NT	GT	SGN	DH..K..M
1: Random SDP $n = m = 50$	16.2	19.3	17.7	16.6	17.9	16.8
2: Norm min. problem $n = 100, m = 26$	18.9	21.3	20.4	19.9	21.1	19.5
3: Cheby. approx. in $\mathbb{R}^{n\times n}$ $n = 100, m = 26$	16.6	19.6	17.6	16.9	19.0	17.0
4: Max-cut $n = m = 50$	15.4	17.7	16.1	16.0	18.4	15.7
5: ETP $n = 100, m = 50$	29.5	34.0	31.2	30.7	32.1	30.4
6: Lovász θ function $n = 30, m = 220$	20.2	23.1	21.5	20.8	22.7	21.8
7: Log. Cheby. prob. $n = 300, m = 51$	21.5	22.7	22.6	21.1	23.8	21.2
8: Cheby. approx. on $\mathcal{C}$	16.1	16.6	16.3	16.1	19.0	16.1

REMARK 5.1. *The relation between the SGN direction and the H..K..M directions implies that the SGN direction shares the same scale-invariance properties as the H..K..M directions; see, e.g.,* [11] *for the definition of scale-invariance.*

In the appendix to this paper we show how the SGN direction can be computed via the solution of the primal and dual H..K..M directions. In particular, we show there that the computational complexity of the SGN direction is upper bounded by

$$2mn^3 + m^2n^2 + \frac{2}{3}m^3 + O\left(n^3 + mn^2 + m^2n\right)$$

flops.[2] In comparison, one has the bound

$$\frac{2}{3}mn^3 + \frac{1}{2}m^2n^2 + \frac{1}{3}m^3 + O\left(n^3 + mn^2 + m^2n\right)$$

flops for the NT direction, and

$$3\frac{2}{3}mn^3 + m^2n^2 + \frac{2}{3}m^3 + O\left(n^3 + mn^2 + m^2n\right)$$

flops for the AHO direction [8].

6. Numerical results. We have implemented two algorithms based on the SGN direction and the dual H..K..M direction by changing the main subroutine of SDPT3 (SDP.m in version 1.3) slightly to admit these two additional search directions. The algorithm we tested is the infeasible path following algorithm without second-order corrector (Algorithm IPF in [13]). Tables 2, 3, and 4 show the performance of this algorithm for various search directions. The test problems are taken from [13], and each test set consists of ten random instances generated by the subroutines in SDPT3. The convergence criterion was to reduce the initial duality gap by a factor of 10^{10}.

The tables show that the algorithm based on the SGN and the dual H..K..M directions are comparable to other search directions with respect to the number of iterations. As for the required CPU time, the SGN direction requires slightly less

[2]We follow the convention in, e.g., [8] that one flop is any floating point operation, i.e., addition and multiplication of two floating point numbers both constitute one flop.

TABLE 3
Average running time of the algorithms.

Test set	AHO	PH..K..M	NT	GT	SGN	DH..K..M
1	32.4	19.4	19.6	23.4	26.8	21.6
2	78.3	42.0	46.3	58.4	72.2	60.6
3	65.0	37.1	38.1	47.3	61.1	50.2
4	20.3	10.7	9.3	13.2	19.7	13.3
5	49.0	26.8	23.5	32.3	38.2	28.3
6	111.0	48.6	40.1	59.1	107.6	68.9
7	45.5	22.7	26.9	34.2	39.5	28.1
8	28.3	14.6	18.1	22.5	29.4	20.7

TABLE 4
Average absolute value of the logarithm of the duality gap at termination, i.e., $|\log_{10} \mathrm{Tr}(XS)|$ *where* (X, S) *are the final iterates.*

Test set	AHO	PH..K..M	NT	GT	SGN	DH..K..M
1	9.4	7.8	7.0	9.1	7.1	6.8
2	12.4	9.5	8.5	12.1	9.1	8.8
3	13.4	10.8	9.4	13.2	9.8	9.6
4	10.9	8.8	7.8	10.6	8.2	7.8
5	7.8	7.0	6.6	8.6	6.4	6.8
6	11.7	9.9	9.3	10.7	9.5	9.2
7	10.8	10.9	10.9	10.9	10.8	10.7
8	13.1	10.4	10.5	13.1	10.9	10.4

than the AHO direction and the dual H..K..M direction less than GT direction. As for the accuracy, both methods are comparable to the primal H..K..M and NT directions. Overall, the performance of the SGN method is somewhat disappointing. In particular, the method does not require fewer iterations than the related primal or dual H..K..M directions in general, even though it is more expensive to compute.

Note, however, that we used the default setting for all parameters in the SDPT3 algorithm IPF; it is reasonable to expect that the iteration count of the algorithms based on the SGN and dual H..K..M directions can be improved by implementing a different line search strategy. Also, the test problems used here are of moderate size. These computational results are therefore of a preliminary nature.

7. Conclusions.

We have presented a primal-dual SGN 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. In particular, the usual $O(\sqrt{n})$ iteration complexity was derived in this paper for the standard small update (short step) primal-dual path following algorithm. The complexity for methods using larger updates remains a topic for future research.

The new direction is closely related to the primal and dual H..K..M directions—it uses the ΔX part of the dual H..K..M direction and the ΔS part of the primal H..K..M direction. As a by-product, we have shown how the dual H..K..M direction can be computed at a cost of at most

$$2mn^3 + \frac{1}{2}m^2n^2 + \frac{1}{3}m^3 + O\left(n^3 + mn^2 + m^2n\right)$$

flops, and the SGN direction can subsequently be computed at a total cost of at most

$$2mn^3 + m^2n^2 + \frac{2}{3}m^3 + O\left(n^3 + mn^2 + m^2n\right)$$

flops.

A preliminary numerical evaluation of the performance of the SGN search direction is somewhat disappointing. The implementation was done using the infeasible path following algorithm in the Matlab code SDPT3. Since we used the default parameter settings for the step lengths and barrier parameter updates in SDTP3, we hope that these results can be improved by finding more suitable (dynamic) parameter settings for the new direction. This is a subject for future research.

Appendix. Computation of the SGN direction. In this appendix, we consider how to compute the SGN direction by first computing the dual H..K..M direction. To this end, we rewrite the linear system (34) (which yields the dual H..K..M direction) by using the Cholesky decompositions $X = L_X^T L_X$ and $S = L_S^T L_S$ as follows:

$$\begin{aligned} L_X L_S^T + L_X(\Delta S + W)L_S^{-1} + L_X^{-T}\Delta X L_S^T &= \mu L_X^{-T} L_S^{-1}, \\ \mathrm{Tr}(A_i \Delta X) &= 0, \quad i = 1, \dots, m, \\ \sum_{i=1}^m \Delta y_i A_i + \Delta S &= 0, \\ \Delta X = \Delta X^T, \quad W + W^T &= 0. \end{aligned}$$

We wish to solve problem SGN1 which is equivalent to solving

$$\min_{\Delta X} \left\| L_X L_S^T + L_X^{-T}\Delta X L_S^T - \mu L_X^{-T} L_S^{-1} \right\|^2 \tag{35}$$

subject to

$$\mathrm{Tr}(A_i \Delta X) = 0, \quad i = 1, \dots, m, \quad \Delta X = \Delta X^T.$$

We now perform a singular value decomposition of $L_X L_S^T$ or an eigenvalue decomposition of $L_X S L_X^T$ to obtain

$$Q^T L_X S L_X^T Q = \Lambda,$$

where Λ is a positive definite diagonal matrix and Q an orthonormal matrix. By defining

$$\Delta \bar{X} = Q^T L_X^{-T} \Delta X L_X^{-1} Q, \quad \bar{A}_i = Q^T L_X A_i L_X^T Q, \quad i = 1, \dots, m,$$

we can rewrite problem (35) as

$$\left\{ \begin{array}{l} \min_{\Delta \bar{X}} \left\| \Lambda^{1/2} + \Delta \bar{X} \Lambda^{1/2} - \mu \Lambda^{-1/2} \right\|^2, \\ \mathrm{Tr}\left(\bar{A}_i \Delta \bar{X}\right) = 0, \quad i = 1, \dots, m, \quad \Delta \bar{X} = \Delta \bar{X}^T, \end{array} \right. \tag{36}$$

which is the same as

$$\left\{ \begin{array}{l} \min_{\Delta \bar{X}} \frac{1}{2} \left\| \Lambda^{1/2} + \Delta \bar{X} \Lambda^{1/2} - \mu \Lambda^{-1/2} \right\|^2 + \frac{1}{2} \left\| \Lambda^{1/2} + \Lambda^{1/2} \Delta \bar{X} - \mu \Lambda^{-1/2} \right\|^2, \\ \mathrm{Tr}\left(\bar{A}_i \Delta \bar{X}\right) = 0, \quad i = 1, \dots, m, \quad \Delta \bar{X} = \Delta \bar{X}^T. \end{array} \right. \tag{37}$$

since $\|A\| = \left\|A^T\right\|$.

In what follows, we will use this notation:

- $\mathbf{svec}\,(X) := \left(X_{11},\ \sqrt{2}X_{12},\ \ldots\ \sqrt{2}X_{1n},\ X_{22},\ \sqrt{2}X_{23},\ \ldots,\ X_{nn}\right)^T \quad \forall\ X = X^T$;
- The *symmetric Kronecker product* $G \otimes_s K$ of $G, K \in \mathbb{R}^{n\times n}$ is implicitly defined by

$$(G \otimes_s K)\,\mathbf{svec}\,(H) := \frac{1}{2}\mathbf{svec}\left(KHG^T + GHK^T\right) \quad \forall\ H = H^T.$$

Now let

$$G^T = (\mathbf{svec}\left(\bar{A}_1\right), \ldots, \mathbf{svec}\left(\bar{A}_m\right)), \quad d_x = \mathbf{svec}\left(\Delta\bar{X}\right). \tag{38}$$

The KKT system of (37) takes the form

$$\begin{cases} \mathcal{E}d_x + G^T v & = & -\mathbf{svec}\,(\Lambda - \mu I_n)\,, \\ Gd_x & = & 0, \end{cases} \tag{39}$$

where $\mathcal{E} = \Lambda \otimes_s I_n$ and v is a variable vector in the suitable space. Premultiplying the first equation in (39) by $G\mathcal{E}^{-1}$, we obtain a linear system in $\mathbb{R}^m$ such that

$$G\mathcal{E}^{-1}G^T v = -G\mathcal{E}^{-1}\mathbf{svec}\,(\Lambda - \mu I_n)\,. \tag{40}$$

Note that $\mathcal{E} = \Lambda \otimes_s I_n$ is a diagonal matrix (see, e.g., the appendix in [12]).

To compute the dual H..K..M direction, we therefore need only solve the system (40) first and then compute $\Delta X, \Delta S$ subsequently. In particular, Δy of the dual H..K..M direction is immediately available from the solution of (40).

PROPOSITION 7.1. *Suppose that $\Delta X, \Delta S$ are solutions of the dual H..K..M. direction and that*

$$\Delta S = -\sum_{i=1}^{m} \Delta y_i A_i.$$

Then $\Delta y = -v$ where v is the solution of the problem (40).

The proof of this proposition is straightforward and therefore omitted.

We can summarize the sequence of steps for the computation of the dual H..K..M direction as follows:

1. Compute G by computing $\bar{A}_i$ for $i = 1, \ldots, m$. Since all $\bar{A}_i$ are symmetric, the computation of all $\bar{A}_i$ requires at most $2mn^3 + O\left(n^3\right)$ flops (see Lemma A.10 in [8]);
2. compute $\mathcal{E}^{-1}G^T$ at a cost of $O(mn^2)$ flops;
3. form the Schur matrix $G\mathcal{E}^{-1}G^T$ ($\frac{1}{2}m^2n^2$ flops);
4. solve the linear system (40) ($\frac{1}{3}m^3$ flops).

Hence the total computation complexity for the dual H..K..M direction is upper bounded by $2mn^3 + \frac{1}{2}m^2n^2 + \frac{1}{3}m^3 + O\left(n^3 + n^2m + nm^2\right)$ flops.

Now recall that the SGN direction uses the ΔS part of the primal H..K..M direction. It is easy to show that the Schur matrix for the primal H..K..M direction has entries $\mathrm{Tr}\left(\bar{A}_i\Lambda^{-1}\bar{A}_j\right)$ $(i, j = 1, \ldots, m)$. We can therefore utilize the fact that G had already been computed when we formed the primal H..K..M Schur matrix. In other words, we only have to perform the analogous steps to steps 3 and 4 above. Hence, the total computational complexity for the SGN direction is $2mn^3 + m^2n^2 + \frac{2}{3}m^3 + O\left(n^3 + n^2m + nm^2\right)$ flops.

Acknowledgments. The authors would like to thank Prof. M. Todd, Prof. M. Kojima, Prof. H. Wolkowicz, and Dr. K. C. Toh for their generous help in the preparation of this work.

REFERENCES

[1] F. Alizadeh, J.-P.A. Haeberly, and M.L. Overton, *Primal-dual interior-point methods for semidefinite programming: Convergence rates, stability and numerical results*, SIAM J. Optim., 8 (1998), pp. 746–768.

[2] E. de Klerk, *Interior Point Methods for Semidefinite Programming*, Ph.D 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 J. Optim., 6 (1996), pp. 342–361.

[4] J. Jiang, *A long step primal–dual path following method for semidefinite programming*, Oper. Res. Lett., 23 (1998), pp. 53–62.

[5] M. Kojima, S. Shindoh, and S. Hara, *Interior point methods for the monotone semidefinite linear complementarity problem in symmetric matrices*, SIAM J. Optim., 7 (1997), pp. 86–125.

[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 J. Optim., 7 (1997), pp. 663–678.

[8] R.D.C. Monteiro and P.R. Zanjácomo, *Implementation of primal-dual methods for semidefinite programming based on Monteiro and Tsuchiya directions and their variants*, Optim. Methods Softw., 11 (1999), pp. 91–140.

[9] Yu. Nesterov and M.J. Todd, *Self–scaled barriers and interior–point methods for convex programming*, Math. Oper. Res., 22 (1997), pp. 1–42.

[10] C. Roos, T. Terlaky, and J.-Ph. Vial, *Theory and Algorithms for Linear Optimization: An Interior Point Approach*, John Wiley & Sons, New York, 1997.

[11] M.J. Todd, *A study of search directions in primal-dual interior-point methods for semidefinite programming*, Optim. Methods Softw., 11 (1999), pp. 1–46.

[12] M.J. Todd, K.C. Toh, and R.H. Tütüncü, *On the Nesterov–Todd direction in semidefinite programming*, SIAM J. Optim., 8 (1998), pp. 769–796.

[13] K.C. Toh, M.J. Todd, and R.H. Tütüncü, *SDPT*3 *— a Matlab software package for semidefinite programming*, Optim. Methods Softw., 11 (1999), pp. 545–581.

[14] Y. Zhang, *On extending some primal-dual interior-point algorithms from linear programming to semidefinite programming*, SIAM J. Optim., 8 (1998), pp. 365–386.