Annals of Operations Research 103, 115–133, 2001

A Predictor–Corrector Algorithm for QSDP Combining Dikin-Type and Newton Centering Steps *

JIA-WANG NIE and YA-XIANG YUAN ** yyx@lsec.cc.ac.cn
State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, P.O. Box 2719, Beijing 100080, China

Abstract. Recently, we have extended SDP by adding a quadratic term in the objective function and give a potential reduction algorithm using NT directions. This paper presents a predictor–corrector algorithm using both Dikin-type and Newton centering steps and studies properties of Dikin-type step. In this algorithm, when the condition $\mathcal{K}(XS)$ is less than a given number K_0, we use Dikin-type step. Otherwise, Newton centering step is taken. In both cases, step-length is determined by line search. We show that at least a constant reduction in the potential function is guaranteed. Moreover the algorithm is proved to terminate in $O(\sqrt{n}\log(1/\varepsilon))$ steps. In the end of this paper, we discuss how to compute search direction $(\Delta X, \Delta S)$ using the conjugate gradient method.

Keywords: semi-definite programming, quadratic term, potential function, central path, predictor step, corrector step, Dikin-type step, Newton centering step

AMS subject classification: 90C30, 90C99, 65K05

1. Introduction

In recent years, semi-definite programming (SDP) has attracted much attention from researchers. Many interesting and important results on SDP have been obtained. SDP has the following standard form:

$$\min \langle C, X\rangle \tag{1.1}$$
$$\text{s.t.} \quad \langle A_i, X\rangle = b_i, \quad i = 1, \dots, m, \tag{1.2}$$
$$X \succeq 0, \tag{1.3}$$

where $C,\ A_i \in \mathcal{S}\Re^{n\times n}$, $\mathcal{S}\Re^{n\times n}$ being the set of all $n \times n$ real symmetric matrices. The notation $X \succeq 0$ means that $X \in \mathcal{S}\Re^{n\times n}$ and X is positive semidefinite. $\langle A, B\rangle$ denotes the inner product in the space $\Re^{n\times n}$ namely

$$\langle A, B\rangle = \mathrm{Tr}\big(A^{\mathrm{T}}B\big), \quad \forall A, B \in \Re^{n\times n}. \tag{1.4}$$

Most works on SDP are about interior point methods, for example, see [1–3,15,19,20]. The SDP problem (1.1)–(1.3) has a similar form to the linear programming, it is also

* This work is partially supported by Chinese NNSF grants 19731010.
** Corresponding author.

called as SDPLP by [16]. Recently, there are researches on extensions of the SDP, most of them are about semi-definite linear complementarity problems (SDPLCP). For more details about SDPLCP, see [8,13] and [16].

Recently, Nie and Yuan[12] give another extension of SDP problems. Its extension is obtained by adding a quadratic term in $\Re^{n\times n}$. It is known that the general quadratic term in $\Re^{n\times n}$ can be expressed in the standard form $\frac{1}{2}x^{\mathrm{T}}Hx$, where $x \in \Re^n$ and $H \in \Re^{n\times n}$. However, in the matrix space $\Re^{n\times n}$, the general quadratic term in $\Re^{n\times n}$ cannot be written as $\frac{1}{2}x^{\mathrm{T}}Hx$. For example, term $\langle X^{\mathrm{T}}, X\rangle = \mathrm{Tr}(\mathrm{X}^2)$ do not have the above expression. Fortunately, any quadratic term in $\Re^{n\times n}$ can be written as the product of a linear operator $\varphi(X)$ and variable X, that is

$$Q(X) = \frac{1}{2}\big\langle \varphi(X), X\big\rangle. \tag{1.5}$$

Then what is the general expression for $\varphi(X)$? The following theorem answers this question.

Theorem 1. The linear operator $\varphi(X) : \Re^{n\times n} \to \Re^{n\times n}$ has the following standard form

$$\varphi(X) = \sum_{i=1}^{l} H_i^1 X H_i^2, \tag{1.6}$$

where H_i^1 and H_i^2 are matrices in $\Re^{n\times n}$ and l is an integer not greater than n^2.

Proof. Define vec as the linear operator from $\Re^{n\times n}$ to $\Re^{n^2}$ which satisfies that

$$\mathrm{vec}(X) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \tag{1.7}$$

where $X = [x_1, x_2, \ldots, x_n]$. The inverse of vec is denoted by mvec. Because vec, mvec and $\varphi(X)$ are linear, the operator $\mathrm{vec}(\varphi(\mathrm{mvec}(x))) : \Re^{n^2} \to \Re^{n^2}$ is also linear. From basic linear algebra results, we know that there exists a matrix $A \in \Re^{n^2\times n^2}$ such that

$$\mathrm{vec}\big(\varphi\big(\mathrm{mvec}(x)\big)\big) = Ax. \tag{1.8}$$

Assume that $A = (A_{ij})_{n\times n}$ where A_{ij} is a matrix of order n. For any $X = [x_1, x_2, \ldots, x_n] \in \Re^{n\times n}$, we have

$$\begin{aligned}
\varphi(X) &= \mathrm{mvec}\big(\mathrm{vec}\big(\varphi\big(\mathrm{mvec}\big(\mathrm{vec}(X)\big)\big)\big)\big) = \mathrm{mvec}\big(A\mathrm{vec}(x)\big) \\
&= \mathrm{mvec}\begin{pmatrix} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n \\ \vdots \\ A_{n1}x_1 + A_{n2}x_2 + \cdots + A_{nn}x_n \end{pmatrix} \\
&= [A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n, \ldots, A_{n1}x_1 + A_{n2}x_2 + \cdots + A_{nn}x_n]
\end{aligned}$$

$$= \sum_{j=1}^{n} [A_{1j}x_j, A_{2j}x_j, \ldots, A_{nj}x_j] = \sum_{j=1}^{n}\sum_{i=1}^{n} [\ldots, A_{ij}x_j, \ldots]$$
$$= \sum_{j=1}^{n}\sum_{i=1}^{n} A_{ij} X E_{ij} = \sum_{i,j=1}^{n} A_{ij} X E_{ij},$$

where matrix $E_{ij}'s$ only nonzero entry is (i, j)th entry. This proves that expression (1.6) is general. □

Throughout this paper, we make the following assumption.

Assumption 2. Linear operator $\varphi(X)$ is in $\mathcal{S}\Re^{n\times n}$ and is symmetric and positive semidefinite, that is

$$\varphi(X) = \varphi(X)^{\mathrm{T}}, \quad X \in \Re^{n\times n}, \tag{1.9}$$
$$\langle \varphi(X), X \rangle \geqslant 0, \quad X \in \Re^{n\times n}. \tag{1.10}$$

Obviously, linear operator $\varphi(X) = \sum_{i=1}^{l} H_i^{\mathrm{T}} X H_i$ satisfies assumption 2. However, as an anonymous referee pointed out, one interesting question is that whether the linear operators satisfying assumption 2 can always be expressed as $\sum_{i=1}^{l} H_i^{\mathrm{T}} X H_i$. Unfortunately, we find a negative answer. A counterexample is as follows. The linear operator $\varphi(X) = XE_{11} + E_{11}X$ obviously satisfies assumption 2, but it can not be expressed as $\sum_{i=1}^{l} H_i^{\mathrm{T}} X H_i$. Otherwise, by setting X to be the identity matrix, we would have $2E_{11} = \sum_{i=1}^{l} H_i^{\mathrm{T}} H_i$. Assume that $H_i = [h_i, h_2^i, \ldots, h_n^i]$. Comparing the diagonal of both sides of matrix equation $2E_{11} = \sum_{i=1}^{l} H_i^{\mathrm{T}} H_i$, one can see that $2 = \sum_{i=1}^{l} \|h_i\|^2$ and $0 = \sum_{i=1}^{l} \|h_j^i\|^2$, $j = 2, \ldots, n$. This show that $h_j^i = 0$ for $2 \leqslant j \leqslant n$ and $1 \leqslant i \leqslant l$. Thus, the last $n-1$ columns of $\sum_{i=1}^{l} H_i^{\mathrm{T}} X H_i$ are zero. Because

$$\varphi(X) = XE_{11} + E_{11}X = \begin{pmatrix} 2X_{11} & X_{12} & \ldots & X_{1n} \\ X_{21} & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & 0 & \ldots & 0 \end{pmatrix}, \tag{1.11}$$

we see that $\varphi(X) = \sum_{i=1}^{l} H_i^{\mathrm{T}} X H_i$ can not be hold unless $X_{12} = \cdots = X_{1n} = 0$. This example indicates that not all linear operators satisfying assumption 2 may not be expressed by the form $\sum_{i=1}^{l} H_i^{\mathrm{T}} X H_i$.

Now we consider the following extended problem:

$$\text{(QSDP)} \qquad \min\ q(x) = \langle C, X \rangle + \frac{1}{2}\langle \varphi(X), X \rangle \tag{1.12}$$
$$\text{s.t.}\quad \langle A_i, X \rangle = b_i, \quad i = 1, \ldots, m, \tag{1.13}$$
$$X \succeq 0, \tag{1.14}$$

where $\varphi(X)$ has the form (1.6) and satisfies assumption 2. Because of assumption 2, problem (QSDP) can be reformulated as a monotone semi-definite LCP. The dual problem of (QSDP) is

$$\text{(QSDD)} \qquad \max\ d(X, y) = b^{\mathrm{T}} y - \frac{1}{2}\langle \varphi(X), X \rangle \tag{1.15}$$

$$\text{s.t.} \quad \sum_{i=1}^{m} y_i A_i + S = C + \varphi(X), \tag{1.16}$$

$$X, S \succeq 0. \tag{1.17}$$

For simplicity, we denote the primal feasible region and dual feasible region by

$$\mathcal{F}_p = \left\{ X \in \mathcal{S}\Re^{n\times n}\colon\ \langle A_i, X\rangle = b_i,\ i = 1, \ldots, m,\ X \succeq 0 \right\} \tag{1.18}$$

and

$$\mathcal{F}_d = \left\{ (X, y, S) \in \mathcal{S}\Re^{n\times n} \times \Re^m \times \mathcal{S}\Re^{n\times n}\colon\ C + \varphi(X) = \sum_{i=1}^{m} y_i A_i + S,\ X, S \succeq 0 \right\}, \tag{1.19}$$

respectively. In order to guarantee the existence of initial strict interior points, we make the following assumption:

Assumption 3 (Slater regularity condition). There exist $X \succ 0$, $S \succ 0$ and $y \in \Re^m$ such that $X \in \mathcal{F}_p$ and $(X, y, S) \in \mathcal{F}_d$.

Nie and Yuan [12] give the following optimality condition.

Theorem 4. If $A_1, \ldots, A_m$ are linearly independent, and under the assumptions 2 and 3, (1.12)–(1.14) has solutions. Furthermore, φ is given and satisfies assumption (2), $X^\star \in \mathcal{F}_p$ is an solution to the problem (1.12)–(1.14) if and only if there exist $y^\star \in \Re^m$ and $S^\star \in \mathcal{S}\Re^{n\times n}$ such that $S^\star \succeq 0$,

$$\left(X^\star, y^\star, S^\star\right) \in \mathcal{F}_d; \tag{1.20}$$

and

$$X^\star S^\star = 0. \tag{1.21}$$

For any pair $(X, y, S) \in \mathcal{F}_d$, if $X \in \mathcal{F}_p$ it follows that

$$\begin{aligned} q(X) - d(y, X) &= \langle C, X\rangle + \frac{1}{2}\langle \varphi(X), X\rangle - b^{\mathrm{T}} y + \frac{1}{2}\langle \varphi(X), X\rangle \\ &= \left\langle \varphi(X) + C - \sum_{i=1}^{m} y_i A_i, X \right\rangle = \langle S, X\rangle \geqslant 0. \end{aligned} \tag{1.22}$$

Therefore, if we can find a pair $(X^\star, y^\star, S^\star) \in \mathcal{F}_d$ such that $X^\star \in \mathcal{F}_p$ and $\langle X^\star, S^\star\rangle = 0$, problems (QSDP) and (QSDD) are solved simultaneously. To this goal, we introduce the primal–dual potential function:

$$\Psi(X, S) = (n+\rho)\log\langle X, S\rangle - \log\det(X) - \log\det(S), \tag{1.23}$$

where $\rho > 0$ is a parameter. Using inequality (see [17])

$$n\log\langle X, S\rangle - \log\det(XS) \geqslant n\log n, \tag{1.24}$$

we know that $\lim_{k\to\infty}\langle X_k, S_k\rangle = 0$ if a sequence $\{(X_k, S_k)\}$ satisfies that $\Psi(X_k, S_k) \to -\infty$. The main aim of this paper is to study how to generate such a sequence.

The paper is organized as follows. In the next section, we introduce the central path and two measures for the deviation from the central path. In section 3, Dikin-type step is introduced and its property in (QSDP) is studied. Moreover, the reduction in the potential function is estimated. In section 4, we discuss how to reduce the potential function if the pair (X, S) is far away from the central path. In section 5, our algorithm is presented and its polynomial convergence is proved. In section 6, we discuss how to use the conjugate gradient method to compute search directions.

2. The central path and its deviation measure

Most algorithms for SDP and its extended problems use interior point techniques and the central path is often involved. In (QSDP), a pair (X, S) is called in the central path if $X \in \mathcal{F}_p$, $(X, y, S) \in \mathcal{F}_d$ and the following equation holds:

$$XS = \mu I, \tag{2.1}$$

where $\mu > 0$ is a parameter. But in practical implements, equation (2.1) is hardly satisfied. As there are many good properties of the central path, we want the generated sequence $\{(X_k, S_k)\}$ is close to the central path. A "distance" is needed to measure how far a pair is from the central path. Many measures for the deviation from the central path have been given. Two of these measures are discussed in the following.

Because X and S are positive definite, equation (2.1) holds if and only if $\lambda_{\max}(XS) = \lambda_{\min}(XS)$, where $\lambda_{\max}(XS)$ and $\lambda_{\min}(XS)$ denote the largest and smallest eigenvalues of XS respectively. Thus, a measure for the closeness to the central path can be quantified by the following:

$$\mathcal{K}(XS) = \frac{\lambda_{\max}(XS)}{\lambda_{\min}(XS)}. \tag{2.2}$$

The larger $\mathcal{K}(XS)$ is, the farther (X, S) is away from the central path; the smaller $\mathcal{K}(XS)$ is, the closer is (X, S) to the central path. For any $X \succ 0$ and $S \succ 0$, $\mathcal{K}(XS) \geqslant 1$. So when $\mathcal{K}(XS)$ is close to 1, (X, S) is close to the central path. However, $\mathcal{K}(XS)$ is difficult to calculate. Therefore simpler measures are needed.

First, we introduce a transformation. For any given $X_k \succ 0$ and $S_k \succ 0$, let W_k be the symmetric positive definite matrix which satisfies the following relation

$$W_k^{-1/2} X_k W_k^{-1/2} = W_k^{1/2} S_k W_k^{1/2} = V_k. \tag{2.3}$$

The above transformation was introduced by Nesterov and Todd [10]. The matrix W_k is called the scaling point in [10,11] and the symmetric primal–dual transformation in Sturm and Zhang [14]. It has the following expression

$$W_k = X_k^{1/2}\big(X_k^{1/2} S_k X_k^{1/2}\big)^{-1/2} X_k^{1/2}. \tag{2.4}$$

One can easily see that $W_k^{-1/2} X_k S_k W_k^{1/2} = V_k^2$ and $\mathcal{K}(XS) = \mathcal{K}(V_k^2)$. Equation (2.1) is equivalent to

$$V_k^2 = \mu I. \tag{2.5}$$

Thus, another measure can be given as follows:

$$\delta(X_k, S_k) = \left\| \sqrt{\mu}\, V_k^{-1} - \frac{1}{\sqrt{\mu}} V_k \right\|_F, \tag{2.6}$$

where $\mu = \mathrm{Tr}(X_k S_k)/n$. This measure is introduced for SDP by Jiang [6], and it becomes to the measure given by [5] for the special case of linear programming. We can see that $\delta(X_k, S_k) = 0$ if and only if (X_k, S_k) is in the central path. The smaller $\delta(X_k, S_k)$ is, the closer (X_k, s_k) to the central path. Then a question is what is the relationship between the two measures (2.2) and (2.6). The following result from Jiang [6] offers an answer to the question.

Lemma 5. Let X_k, S_k, V_k be the matrices satisfying equation (2.3), we have the following inequality:

$$\sqrt{\max\left\{\frac{\mu}{\lambda_{\min}(X_k S_k)}, \frac{\lambda_{\max}(X_k S_k)}{\mu}\right\}} \leqslant \frac{1}{2}\Big(\delta(X_k, S_k) + \sqrt{4 + \delta^2(X_k, S_k)}\Big). \tag{2.7}$$

Using this lemma and $\mathcal{K}(V_k) = \mathcal{K}((1/\sqrt{\mu})V_k)$, we can see that

$$\sqrt{\mathcal{K}(X_k S_k)} = \mathcal{K}(V_k) \leqslant \left[\frac{1}{2}\Big(\delta(X_k, S_k) + \sqrt{4 + \delta^2(X_k, S_k)}\Big)\right]^2. \tag{2.8}$$

Roughly speaking, if $\delta(X_k, S_k)$ is small, $\mathcal{K}(X_k S_k)$ is also small. A larger $\mathcal{K}(X_k S_k)$ implies a larger $\delta(X_k, S_k)$.

3. Dikin-type step and its properties

In this section, we begin to discuss how to generate a sequence $\{(X_k, S_k)\}$ such that $\lim_{k\to\infty} \Psi(X_k, S_k) = -\infty$. Firstly, we assume that we already have a strict interior point pair (X_k, S_k) such that $X_k \in \mathcal{F}_p$ and $(X_k, y_k, S_k) \in \mathcal{F}_d$. For simplicity, we begin our discussion with the following transformation:

$$\begin{cases} \mathrm{d}X = W_k^{-1/2} \Delta X W_k^{-1/2}, \\ \mathrm{d}S = W_k^{1/2} \Delta S W_k^{1/2}, \\ A_i' = W_k^{1/2} A_i W_k^{1/2}, \\ R' = W_k^{1/2} R W_k^{1/2}, \\ H_i^{1'} = W_k^{1/2} H_i^1 W_k^{1/2}, \\ H_i^{2'} = W_k^{1/2} H_i^2 W_k^{1/2}, \\ \bar{\varphi}(\mathrm{d}X) = W_k^{1/2} \varphi(\Delta X) W_k^{1/2}. \end{cases} \tag{3.1}$$

The direction $(\mathrm{d}X, \mathrm{d}y, \mathrm{d}S)$ is called a Dikin-type step if it is generated by the following linear system:

$$\bar{\varphi}(\mathrm{d}X) - \sum_{i=1}^{m} (\mathrm{d}y)_i A_i' - \mathrm{d}S = 0, \tag{3.2}$$

$$\langle A_i', \mathrm{d}X \rangle = 0, \quad i = 1, \dots, m, \tag{3.3}$$

$$\mathrm{d}X + \mathrm{d}S = -\frac{V_k^3}{\|V_k^2\|_F}. \tag{3.4}$$

Properties of Dikin-type step in SDP are discussed in detail by De Klerk [7]. In this section, we study its properties in (QSDP). In SDP, $\mathrm{d}X$ and $\mathrm{d}S$ are orthogonal, which does not hold any more in (QSDP). However, from (3.3) and (3.4), it is obvious that

$$\langle \mathrm{d}X, \mathrm{d}S \rangle = \langle \bar{\varphi}(\mathrm{d}X), \mathrm{d}X \rangle \geqslant 0, \tag{3.5}$$

since $\bar{\varphi}$ is positive semi-definite.

Once a direction $(\Delta X, \Delta y, \Delta S)$ is computed, how to select step-length α_k such that $X_k + \alpha_k \Delta X \succ 0$ and $S_k + \alpha_k \Delta S \succ 0$? An obvious sufficient condition is that

$$\max\{ \|\alpha_k X_k^{-1/2} \Delta X X_k^{-1/2}\|_F, \|\alpha_k S_k^{-1/2} \Delta S S_k^{-1/2}\|_F \} < 1,$$

that is,

$$\alpha_k < \min\{ \|X_k^{-1/2} \Delta X X_k^{-1/2}\|_F^{-1}, \|S_k^{-1/2} \Delta S S_k^{-1/2}\|_F^{-1} \}.$$

But it is hard to get an exact upper bound for this estimation. Nemirovskii and Gahinet [9] gives another estimation:

$$\varepsilon_{NG} = \|X_k^{-1/2} \Delta X X_k^{-1/2}\|_F^2 + \|S_k^{-1/2} \Delta S S_k^{-1/2}\|_F^2, \tag{3.6}$$

which can be written as

$$\varepsilon_{NG} = \left\| V_k^{-1/2}\,\mathrm{d}X\,V_k^{-1/2} \right\|_F^2 + \left\| V_k^{-1/2}\,\mathrm{d}S\,V_k^{-1/2} \right\|_F^2. \tag{3.7}$$

Furthermore, a often used Dikin-type estimation is

$$\varepsilon_D = \left\| V_k^{-1/2}(\mathrm{d}X + \mathrm{d}S)V_k^{-1/2} \right\|_F^2. \tag{3.8}$$

Now we can see that if $\alpha_k\sqrt{\varepsilon_{NG}} < 1$, then $X_k + \alpha_k\Delta X$ and $S_k + \alpha_k\Delta S$ are strictly feasible. The relationship between (3.7) and (3.8) in SDP is studied by De Klerk [7]. We generalize this relationship to (QSDP) as follows.

Lemma 6. Let $(\mathrm{d}X, \mathrm{d}S)$ be generated by system (3.2)–(3.4), we have

$$\frac{1}{2\mathcal{K}(X_kS_k)}\varepsilon_D \leqslant \varepsilon_{NG} \leqslant \mathcal{K}(X_kS_k)\varepsilon_D. \tag{3.9}$$

Proof. From (3.7), one can see that

$$\begin{aligned}
\varepsilon_{NG} &= \mathrm{Tr}\big(V_k^{-1}\,\mathrm{d}X\,V_k^{-1}\,\mathrm{d}X\big) + \mathrm{Tr}\big(V_k^{-1}\,\mathrm{d}S\,V_k^{-1}\,\mathrm{d}S\big)\\
&\leqslant \frac{1}{\lambda_{\min}(V_k^2)}\big[\mathrm{Tr}\big((\mathrm{d}X)^2 + (\mathrm{d}S)^2\big)\big] \leqslant \frac{1}{\lambda_{\min}(V_k^2)}\mathrm{Tr}\big((\mathrm{d}X + \mathrm{d}S)^2\big)\\
&\leqslant \frac{\lambda_{\max}(V_k^2)}{\lambda_{\min}(V_k^2)}\left\| V_k^{-1/2}(\mathrm{d}X + \mathrm{d}S)V_k^{-1/2} \right\|_F^2 = \mathcal{K}(X_kS_k)\varepsilon_D,
\end{aligned} \tag{3.10}$$

where the third line depends on (3.5). Similarly,

$$\begin{aligned}
\varepsilon_D &= \mathrm{Tr}\big(V_k^{-1}(\mathrm{d}X + \mathrm{d}S)V_k^{-1}(\mathrm{d}X + \mathrm{d}S)\big)\\
&\leqslant \frac{1}{\lambda_{\min}(V_k^2)}\mathrm{Tr}\big((\mathrm{d}X)^2 + (\mathrm{d}S)^2 + \mathrm{d}X\,\mathrm{d}S + \mathrm{d}S\,\mathrm{d}X\big)\\
&\leqslant \frac{2}{\lambda_{min}(V_k^2)}\mathrm{Tr}\big((\mathrm{d}X)^2 + (\mathrm{d}S)^2\big) \leqslant 2\mathcal{K}(X_kS_k)\varepsilon_{NG},
\end{aligned} \tag{3.11}$$

which completes the proof. □

Our goal is to make $\Psi(X_k, S_k)$ converge to negative infinity. Therefore we need to study the reduction in the potential function.

Lemma 7. Let $X_{k+1} = X_k + \alpha\Delta X$, $S_{k+1} = S_k + \alpha\Delta S$. If $X_{k+1} \succ 0$ and $S_{k+1} \succ 0$, then the following holds:

$$\begin{aligned}
\Delta\Psi_k &= \Psi(X_k, S_k) - \Psi(X_{k+1}, S_{k+1})\\
&\geqslant -(n+\rho)\left[\alpha\frac{\mathrm{Tr}(V_k)}{\mathrm{Tr}(V_k^2)} + \alpha^2\frac{\mathrm{Tr}(\mathrm{d}X\,\mathrm{d}S)}{\mathrm{Tr}(V_k^2)}\right] + \alpha\,\mathrm{Tr}\big(V_k^{-1}\mathrm{d}V\big) - h\big(\alpha\sqrt{\varepsilon_{NG}}\big),
\end{aligned} \tag{3.12}$$

where $\mathrm{d}V = \mathrm{d}X + \mathrm{d}S$ and function $h(t)$ is defined by

$$h(t) = \frac{t^2}{2(1-t)}. \tag{3.13}$$

Proof. First, consider the reduction in $\log\langle X, S\rangle$:

$$\begin{aligned}
&\log\langle X_{k+1}, S_{k+1}\rangle - \log\langle X_k, S_k\rangle \\
&\quad = \log\left(\frac{\mathrm{Tr}(X_k S_k) + \alpha \mathrm{Tr}(X_k \Delta S + \Delta X S_k) + \alpha^2 \mathrm{Tr}(\Delta X \Delta S)}{\mathrm{Tr}(X_k S_k)}\right) \\
&\quad = \log\left(1 + \alpha \frac{\mathrm{Tr}(V_k\, \mathrm{d}V)}{\mathrm{Tr}(V_k^2)} + \alpha^2 \frac{\mathrm{Tr}(\mathrm{d}X\, \mathrm{d}S)}{\mathrm{Tr}(V_k^2)}\right),
\end{aligned} \tag{3.14}$$

where we use the inequality

$$\log(1+t) \leqslant t, \quad \forall t > -1. \tag{3.15}$$

In the following proof, we need a more exact inequality (see [17])

$$\mathrm{Tr}(Z) \geqslant \log\det(I + Z) \geqslant \mathrm{Tr}(Z) - \frac{\|Z\|_F^2}{2(1 - \|Z\|_2)} \quad \text{if } \|Z\|_2 < 1, \tag{3.16}$$

where $Z \in \mathcal{S}\mathfrak{R}^{n\times n}$. Using this inequality, we have

$$\begin{aligned}
&\log\det(X_{k+1}) + \log\det(S_{k+1}) - \log\det(X_k) - \log\det(S_k) \\
&\quad = \log\det\big(I + \alpha V_k^{-1/2}\, \mathrm{d}X\, V_k^{-1/2}\big) + \log\det\big(I + \alpha V_k^{-1/2}\, \mathrm{d}S\, V_k^{-1/2}\big) \\
&\quad \geqslant \alpha \mathrm{Tr}\big(V_k^{-1}\, \mathrm{d}V\big) - \frac{\alpha^2 \|V_k^{-1/2}\, \mathrm{d}X\, V_k^{-1/2}\|_F^2}{2(1 - \alpha \|V_k^{-1/2}\, \mathrm{d}X\, V_k^{-1/2}\|_2)} \\
&\quad\quad - \frac{\alpha^2 \|V_k^{-1/2}\, \mathrm{d}S\, V_k^{-1/2}\|_F^2}{2(1 - \alpha \|V_k^{-1/2}\, \mathrm{d}S\, V_k^{-1/2}\|_2)}.
\end{aligned} \tag{3.17}$$

Because $\|V_k^{-1/2}\, \mathrm{d}S\, V_k^{-1/2}\|_2 \leqslant \sqrt{\varepsilon_{NG}}$ and $\|V_k^{-1/2}\, \mathrm{d}X\, V_k^{-1/2}\|_2 \leqslant \sqrt{\varepsilon_{NG}}$, (3.17) yields

$$\log\det(X_{k+1} S_{k+1}) - \log\det(X_k S_k) \tag{3.18}$$

$$\geqslant \alpha \mathrm{Tr}\big(V_k^{-1}\, \mathrm{d}V\big) - h\big(\alpha\sqrt{\varepsilon_{NG}}\big). \tag{3.19}$$

Combing inequalities (3.14) and (3.19), one can see that the lemma is true. □

If $(\mathrm{d}X, \mathrm{d}S)$ satisfies (3.2)–(3.4), the reduction in the potential function can be obtained as follows. From (3.4), it follows that

$$\begin{aligned}
\Delta\Psi_k &\geqslant -\alpha(n+\rho)\left[-\frac{\mathrm{Tr}(V_k^4)}{\mathrm{Tr}(V_k^2)\|V_k^2\|_F} + \alpha^2 \frac{\mathrm{Tr}(\mathrm{d}X\, \mathrm{d}S)}{\mathrm{Tr}(V_k^2)}\right] - \alpha \frac{\mathrm{Tr}(V_k^2)}{\|V_k^2\|_F} - h\big(\alpha\sqrt{\varepsilon_{NG}}\big) \\
&= \alpha(n+\rho)\frac{\|V_k^2\|_F}{\mathrm{Tr}(V_k^2)} - \alpha \frac{\mathrm{Tr}(V_k^2)}{\|V_k^2\|_F} - \alpha^2(n+\rho)\frac{\mathrm{Tr}(\mathrm{d}X\, \mathrm{d}S)}{\mathrm{Tr}(V_k^2)} - h\big(\alpha\sqrt{\varepsilon_{NG}}\big).
\end{aligned} \tag{3.20}$$

From the relation that

$$\mathrm{Tr}\big(V_k^2\big) = \langle V_k^2, I\rangle \leqslant \|V_k^2\|_F \|I\|_F = \sqrt{n}\|V_k^2\|_F$$

and $\mathrm{Tr}(\mathrm{d}X\,\mathrm{d}S) \leqslant \frac{1}{2}\|\mathrm{d}X + \mathrm{d}S\|_F^2$, it follows from inequality (3.20) that

$$\begin{aligned}\Delta\Psi_k &\geqslant \alpha\frac{\rho}{\sqrt{n}} - \frac{\alpha^2(n+\rho)}{2}\frac{\mathrm{Tr}(V_k^6)}{\mathrm{Tr}(V_k^2)\mathrm{Tr}(V_k^4)} - h\big(\alpha\sqrt{\varepsilon_{NG}}\big)\\ &= \alpha\frac{\rho}{\sqrt{n}} - \frac{\alpha^2(n+\rho)}{2}\frac{\sum_{i=1}^n \lambda_i^3}{(\sum_{i=1}^n \lambda_i)(\sum_{i=1}^n \lambda_i^2)} - h\big(\alpha\sqrt{\varepsilon_{NG}}\big). \qquad (3.21)\end{aligned}$$

From (3.9), we have that

$$\varepsilon_{NG} \leqslant \mathcal{K}\big(V_k^2\big)\varepsilon_D = \mathcal{K}\big(V_k^2\big)\big\|V_k^{-1/2}\,\mathrm{d}V\,V_k^{-1/2}\big\|_F^2 = \mathcal{K}\big(V_k^2\big)\frac{\mathrm{Tr}(V_k^4)}{\|V_k^2\|_F^2} = \mathcal{K}\big(V_k^2\big). \qquad (3.22)$$

If we let $\alpha \leqslant \eta/\mathcal{K}(V_k^2)$, where $0 < \eta < 1$, inequalities (3.22) and (3.21) yield

$$\begin{aligned}\Delta\Psi_k &\geqslant \alpha\frac{\rho}{\sqrt{n}} - \frac{\alpha^2}{2}\left(1+\frac{\rho}{n}\right)\mathcal{K}\big(V_k^2\big) - \frac{\alpha^2\varepsilon_{NG}}{2(1-\eta)}\\ &\geqslant \alpha\frac{\rho}{\sqrt{n}} - \frac{1}{2}\alpha^2\left[\frac{1}{1-\eta} + \left(1+\frac{\rho}{n}\right)\right]\mathcal{K}\big(V_k^2\big) = f(\alpha). \qquad (3.23)\end{aligned}$$

Now we consider the following line search:

$$\max\ f(\alpha) \qquad (3.24)$$

$$\text{s.t.}\quad \alpha \leqslant \frac{\eta}{\mathcal{K}(V_k^2)}. \qquad (3.25)$$

Let $\alpha^\star$ be the optimal solution. Simple calculations show that

$$f(\alpha^\star) = \min\left\{\frac{\rho^2}{2\mathcal{K}(V_k^2)[1/(1-\eta) + (1+\rho/n)]n}, \frac{\rho\eta}{2\sqrt{n}\mathcal{K}(V_k^2)}\right\}. \qquad (3.26)$$

Assume that $\mathcal{K}(V_k^2) \leqslant K_0$, where K_0 is a pre-specified constant. If we let $\rho = \sqrt{nK_0}$, we can show that

$$\Delta\Psi_k \geqslant \min\left\{\frac{1}{2[1/(1-\eta) + (1+\rho/n)]}, \frac{\eta}{2}\right\} = \xi_1 > 0. \qquad (3.27)$$

Thus, if $\widehat{\alpha}$ is obtained by the following line search:

$$\max\ \Psi(X_k + \alpha\Delta X, S_k + \alpha\Delta S), \qquad (3.28)$$

$$\text{s.t.}\quad \alpha \leqslant \frac{\eta}{\mathcal{K}(V_k^2)}, \qquad (3.29)$$

it follows that $\Delta\Psi_k \geqslant \xi_1$. Because $\alpha \leqslant \eta/\mathcal{K}(V_k^2)$ ensures that $X_k + \alpha\Delta X \succ 0$ and $S_k + \alpha\Delta S \succ 0$, we have the following lemma.

Theorem 8. If search direction $(\Delta X, \Delta y, \Delta S)$ uses Dikin-type step, that is $(\Delta X, \Delta y, \Delta S)$ is defined by linear system (3.2)–(3.4), $\mathcal{K}(V_k^2) \leqslant K_0$ and parameter $\rho = \sqrt{nK_0}$, then the potential function $\Psi(X, S)$ can be reduced by at least a constant.

4. The Newton centering step

In the previous section, it is shown that the potential function can be reduced by at least a constant if $\mathcal{K}(V_k^2) \leqslant K_0$. Now we consider the case when this inequality does not hold. In this case, the current point pair (X_k, S_k) is "far" away from the central path. So we expect to update this point in a direction towards to the central path. A Newton step may be a good choice. Direct calculations show that the Newton step for $V_k V_k = \mu I$ is

$$\frac{1}{2}(V_k \,\mathrm{d}V + \mathrm{d}V\, V_k) = \mu I - V_k^2. \tag{4.1}$$

This is a kind of Sylvester equation. Its unique symmetric solution is given by

$$\mathrm{d}V = \mathrm{d}X + \mathrm{d}S = \mu V_k^{-1} - V_k, \tag{4.2}$$

which is equivalent to

$$W_k^{-1} \Delta X W_k^{-1} + \Delta S = \mu X^{-1} - S. \tag{4.3}$$

Now we consider the the reduction in the potential function. From lemma 7, we know that

$$\begin{aligned}
\Delta\Psi_k &\geqslant -(n+\rho)\left[\alpha \frac{\mathrm{Tr}(V_k(\mu V_k^{-1} - V_k))}{\mathrm{Tr}(V_k^2)} + \frac{\alpha^2 \|\mathrm{d}V\|_F^2}{2\mathrm{Tr}(V_k^2)}\right] \\
&\quad + \alpha \,\mathrm{Tr}\big(V_k^{-1}(\mu V_k^{-1} - V_k)\big) - h\big(\alpha\sqrt{\varepsilon_{NG}}\big) \\
&= -\alpha^2 \frac{(n+\rho)\mu}{2\mathrm{Tr}(V_k^2)} \left\| \sqrt{\mu} V_k^{-1} - \frac{1}{\mu} V_k \right\|_F^2 + \alpha \mathrm{Tr}\left(\left(\sqrt{\mu}\, V_k^{-1} - \frac{1}{\mu} V_k\right)^2\right) - h\big(\alpha\sqrt{\varepsilon_{NG}}\big) \\
&= \alpha \delta^2(X_k, S_k) - \frac{1}{2}\alpha^2\left(1 + \frac{\rho}{n}\right)\delta^2(X_k, S_k) - h\big(\alpha\sqrt{\varepsilon_{NG}}\big),
\end{aligned} \tag{4.4}$$

where $\mu = \mathrm{Tr}(\mathrm{V}_k^2)/n$ is used. Similar to the previous section, if we require $\alpha\sqrt{\varepsilon_{NG}} < \eta$, which implies X_{k+1} and S_{k+1} are strictly feasible, (4.4) gives that

$$\begin{aligned}
\Delta\Psi_k &\geqslant \alpha \delta^2(X_k, S_k) - \frac{\alpha^2}{2}\left(1 + \frac{\rho}{n}\right)\delta^2(X_k, S_k) - \frac{\alpha^2 \varepsilon_{NG}}{2(1-\eta)} \\
&= \alpha \delta^2(X_k, S_k) - \frac{\alpha^2}{2}\left[\left(1 + \frac{\rho}{n}\right)\delta^2(X_k, S_k) + \frac{\varepsilon_{NG}}{1-\eta}\right] = g(\alpha).
\end{aligned} \tag{4.5}$$

Let $\theta_k = (1 + \rho/n)\delta^2(X_k, S_k) + \varepsilon_{NG}/(1 - \eta)$, and let the step-length $\alpha^\star$ be the solution of the following problem:

$$\max \; g(\alpha) = \alpha\delta^2(X_k, S_k) - \frac{\alpha^2}{2}\theta_k \tag{4.6}$$

$$\text{s.t.} \quad \alpha \leqslant \frac{\eta}{\sqrt{\varepsilon_{NG}}}. \tag{4.7}$$

If $\delta^2(X_k, S_k)/\theta_k \leqslant \eta/\sqrt{\varepsilon_{NG}}$, we can see that the $\alpha^\star = \delta^2(X_k, S_k)/\theta_k$. Thus,

$$g(\alpha^\star) = \frac{\delta^4(X_k, S_k)}{\theta_k} - \frac{1}{2}\frac{\delta^4(X_k, S_k)}{\theta_k^2}\theta_k = \frac{\delta^4(X_k, S_k)}{2\theta_k}. \tag{4.8}$$

If $\delta^2(X_k, S_k)/\theta_k > \eta/\sqrt{\varepsilon_{NG}}$, it can be seen that $\alpha^\star = \eta/\sqrt{\varepsilon_{NG}}$, which gives that

$$g(\alpha^\star) = \frac{\eta}{\sqrt{\varepsilon_{NG}}}\delta^2(X_k, S_k) - \frac{1}{2}\frac{\eta^2}{\varepsilon_{NG}}\theta_k > \frac{\eta\delta^2(X_k, S_k)}{2\sqrt{\varepsilon_{NG}}}. \tag{4.9}$$

Using (4.8) and (4.9), we can see that

$$\Delta\Psi_k \geqslant \frac{1}{2}\min\left\{\frac{\delta^4(X_k, S_k)}{\varepsilon_{NG}/(1-\eta) + (1+\rho/n)\delta^2(X_k, S_k)}, \frac{\eta\delta^2(X_k, S_k)}{\sqrt{\varepsilon_{NG}}}\right\}. \tag{4.10}$$

Theorem 9. Let constant $K_0 = (7 + 3\sqrt{5})/2$ and parameter $\rho = \sqrt{nK_0}$. If search direction $(\Delta X, \Delta y, \Delta S)$ uses Newton centering step, that is $(\Delta X, \Delta y, \Delta S)$ satisfies linear system (3.2), (3.3) and (4.2), and $\mathcal{K}(V_k^2) \geqslant K_0$, then the potential function $\Psi(X, S)$ can also be reduced by at least a constant.

Proof. Noting $K_0 = (7 + 3\sqrt{5})/2$, from (2.8) we know that $\mathcal{K}(V_k^2) \geqslant K_0$ implies $\delta(X_k, S_k) \geqslant 1$ and therefore $\delta^2(X_k, S_k) \leqslant \delta^4(X_k, S_k)$. In this case, (4.10) gives that

$$\Delta\Psi_k \geqslant \frac{1}{2}\min\left\{\frac{\tau_k^2}{1/(1-\eta) + (1+\sqrt{K_0/n})\tau_k^2}, \eta\tau_k\right\}, \tag{4.11}$$

where $\tau_k = \delta^2(X_k, S_k)/\sqrt{\varepsilon_{NG}}$. Jiang [6] shows that

$$\tau_k = \frac{\delta^2(X_k, S_k)}{\sqrt{\varepsilon_{NG}}} \geqslant \frac{2\delta(X_k, S_k)}{\delta(X_k, S_k) + \sqrt{4 + \delta^2(X_k, S_k)}} \geqslant \frac{2}{1+\sqrt{5}} = \tau_0. \tag{4.12}$$

The above two inequalities give that

$$\Delta\Psi_k \geqslant \min\left\{\frac{\tau_0^2}{1/(1-\eta) + (1+\sqrt{K_0/n})\tau_0^2}, \eta\tau_0\right\} = \xi_2. \tag{4.13}$$

This shows that the reduction in the potential function can be at least a fixed constant when $\mathcal{K}(V_k^2) \geqslant K_0$. □

5. The predictor–corrector algorithm

Let $K_0 = (7+3\sqrt{5})/2$. We use the primal–dual Dikin-type step if $\mathcal{K}(V_k^2) \leqslant K_0$, otherwise we use the Newton centering step. Thus, from theorems 8 and 9 in the previous two sections, we can see that at every iteration the the potential function is reduced by at least a constant:

$$\Delta\Psi_k \geqslant \min\{\xi_1, \xi_2\} = \xi > 0. \tag{5.1}$$

The following is our predictor–corrector algorithm that use both Dikin-type step and Newton centering step.

Algorithm 10. Given a strict interior point pair $(X_0, y_0, S_0) \in \mathcal{F}_d$ and $X_0 \in \mathcal{F}_p$. Given a small $\varepsilon > 0$ and constant $0 < \eta < 1$. $k = 0$.

Step 1. Let $K_0 = (7+3\sqrt{5})/2$ and $\rho = \sqrt{K_0 n}$.

Step 2. If $\mathrm{Tr}(X_k S_k) \leqslant \varepsilon$ then stop.

Step 3. If $\mathcal{K}(V_k^2) \leqslant K_0$, compute $(\Delta X, \Delta S)$ from system (3.2)–(3.4); otherwise compute $(\Delta X, \Delta S)$ from system (3.2), (3.3) and (4.2).

Step 4. Compute line search interval $[a_1, a_2]$.

$$a_1 = \max\left\{-\frac{\eta}{\lambda}\colon\ \lambda > 0,\ \lambda = \lambda_i\big(X_k^{-1/2}\Delta X X_k^{-1/2}\big) \text{ or } \lambda_i\big(S_k^{-1/2}\Delta S S_k^{-1/2}\big)\right\},$$
$$a_2 = \min\left\{-\frac{\eta}{\lambda}\colon\ \lambda < 0,\ \lambda = \lambda_i\big(X_k^{-1/2}\Delta X X_k^{-1/2}\big) \text{ or } \lambda_i\big(S_k^{-1/2}\Delta S S_k^{-1/2}\big)\right\}.$$

Step 5. Compute step-length by line search:

$$\alpha^\star = \operatorname*{argmin}_{a_1 \leqslant \alpha \leqslant a_2} \Psi(X_k + \alpha\Delta X, S_k + \alpha\Delta S).$$

Step 6. Let $(X_{k+1}, y_{k+1}, S_{k+1}) = (X_k, y_k, S_k) + \alpha^\star(\Delta X, \Delta y, \Delta S)$.

Step 7. $k := k+1$, and go to step 2.

Step 3 of the above algorithm computes search direction $(\Delta X, \Delta y, \Delta S)$. It will be discussed in the next section. The performance of algorithm 10 can be seen in the following theorem.

Theorem 11. Algorithm 10 will terminate in $\mathrm{O}(\sqrt{n}\log(1/\varepsilon))$ iterations.

Proof. From (5.1) we know that

$$\Psi(X_k, S_k) - \Psi(X_0, S_0) \leqslant -k\xi. \tag{5.2}$$

It follows from the definition of $\Psi(X_k, S_k)$ and inequality (1.23) that

$$\rho \log \mathrm{Tr}(X_k S_k) \leqslant \Psi(X_0, S_0) - k\xi, \tag{5.3}$$

which yields that

$$\mathrm{Tr}(X_k S_k) \leqslant \exp\left\{\frac{\Psi(X_0, S_0) - k\xi}{\rho}\right\}. \tag{5.4}$$

Simple calculations show that $\mathrm{Tr}(X_k S_k) \leqslant \varepsilon$ if

$$k > \left\lceil \frac{\sqrt{K_0 n}\log(1/\varepsilon) + \Psi(X_0, S_0)}{\xi} \right\rceil.$$

Therefore the theorem is true. □

6. Search direction calculation

This section considers the calculation of the search direction in the algorithm given in the previous section. The search direction, as we can see, is determined by (3.2)–(3.4) or by (3.2), (3.2) and (4.2). These equations are relations about the variables $\mathrm{d}X$ and $\mathrm{d}S$. Using the transformation (3.1), these equations can be rewritten as follows:

$$\varphi(\Delta X) - \sum_{i=1}^{m} (\Delta y)_i A_i - \Delta S = 0, \tag{6.1}$$

$$\langle A_i, \Delta X\rangle = 0, \quad i = 1, \ldots, m, \tag{6.2}$$

$$W_k^{-1} \Delta X W_k^{-1} + \Delta S = -B_k, \tag{6.3}$$

where B_k equals $SXS/\|XS\|_F$ if a Dikin-type is used, or $-\mu X^{-1} + S^{-1}$ if the Newton centering step is used, depending on the value of $\mathcal{K}(X_k S_k)$.

Therefore, all we need to do is to solve the linear system (6.1), (6.2) and (6.3). (6.1) and (6.3) give the following equation:

$$\varphi(\Delta X) + W_k^{-1} \Delta X W_k^{-1} - \sum_{i=1}^{m} (\Delta y)_i A_i = -B_k. \tag{6.4}$$

$(\Delta X, \Delta y)$ can be computed from (6.2) and (6.4). Then, ΔS could be obtained by (6.3). System (6.2) and (6.4) can be solved by any standard linear system solver (see Golub and Van Loan [4]). As the system involves a n^2 by n^2 matrix, its dimension is extremely high. We use the conjugate gradient method to solve this system. One of the good properties of the conjugate gradient method is that it can exploit sparsity.

For simplicity, define the linear operator $\Phi : \Re^{n\times n} \to \Re^{n\times n}$ as follows:

$$\Phi(\Delta X) = \varphi(\Delta X) + W_k^{-1} \Delta X W_k^{-1}, \quad \forall\, \Delta X \in \Re^{n\times n}. \tag{6.5}$$

Thus we can see that $(\Delta X, \Delta y)$ can be obtained by solving linear system

$$\Phi(\Delta X) - \sum_{i=1}^{m} (\Delta y)_i A_i = -B_k, \tag{6.6}$$

$$\langle A_i, \Delta X\rangle = 0. \tag{6.7}$$

Assumption 2 and (6.5) imply that the operator Φ is positive definite. Furthermore,

$$\big\langle \Phi(\Delta X), \Delta X\big\rangle \geqslant \lambda_{\min}\big(W^{-2}\big)\|\Delta X\|_F^2, \quad \forall\, \Delta X \in \Re^{n\times n}. \tag{6.8}$$

One can easily show that system (6.6), (6.7) is equivalent to the following subproblem

$$\min\ \langle B_k, \Delta X\rangle + \frac{1}{2}\big\langle \Phi(\Delta X), \Delta X\big\rangle \tag{6.9}$$

$$\text{s.t.}\quad \langle A_i, \Delta X\rangle = 0, \quad i = 1, \ldots, m. \tag{6.10}$$

Δy is the corresponding Lagrange multiplier. (6.9), (6.10) is essentially minimizing a strict convex function over a subspace in $\Re^{n\times n}$. By observing this, we can use the generalized conjugate gradient method to compute ΔX.

Define a projective operator $P : S\Re^{n\times n} \to S\Re^{n\times n}$ as follows.

$$P(Z) = \sum_{i=1}^{m} \lambda_i(Z) A_i, \tag{6.11}$$

where coefficients $\lambda_i(Z)$ are chosen so that

$$\big\langle A_i, Z - P(Z)\big\rangle = 0, \quad i = 1, \ldots, m. \tag{6.12}$$

From the definition, one can easily see that

$$\big\langle P(\Delta X), \Delta X\big\rangle = 0, \tag{6.13}$$

$$\big\langle (I-P)(\Delta X), \Delta X\big\rangle = \big\|(I-P)(\Delta X)\big\|^2, \tag{6.14}$$

where ΔX is a feasible point for subproblem (6.9), namely (6.10) holds.

At the beginning of computation, find a feasible point ΔX^0, namely ΔX^0 is symmetric and satisfies (6.12). We generate ΔX^1 in the following way. Let $G^0 = B_k + \Phi(\Delta X^0)$, and $D^0 = -(I-P)G^0$. Step-length α^0 is determined by minimizing the objective function of subproblem (6.9) in the line $\Delta X^0 + \alpha D^0$, i.e.

$$\alpha^0 = -\frac{\langle B_k + \Phi(\Delta X^0), D^0\rangle}{\langle \Phi(D^0), D^0\rangle}. \tag{6.15}$$

Let $\Delta X^1 = \Delta X^0 + \alpha^0 D^0$ and $G^1 = B_k + \Phi(\Delta X^1)$. Since α^0 is an exact line search step-length, it is obvious that:

$$\big\langle (I-P)G^1, D^0\big\rangle = \big\langle G^1, D^0\big\rangle = 0. \tag{6.16}$$

We generate D^1 as follows:

$$D^1 = -(I-P)G^1 + \beta^0 D^0, \tag{6.17}$$

where β^0 is chosen so that D^0 and D^1 are Φ-*conjugate*, i.e., $\langle \Phi(D^0), D^1\rangle = 0$. Simple calculations show that

$$\beta^0 = \frac{\langle \Phi(D^0), (I-P)G^1\rangle}{\langle \Phi(D^0), D^0\rangle}. \tag{6.18}$$

We generate ΔX^2 as $\Delta X^2 = \Delta X^1 + \alpha^1 D^1$, where α^1 is defined by

$$\alpha^1 = -\frac{\langle B_k + \Phi(\Delta X^1), D^1\rangle}{\langle \Phi(D^1), D^1\rangle}. \tag{6.19}$$

Because ΔX^0, D^0 and ΔX^0 are symmetric, ΔX^1 is also symmetric. For any $i \geqslant 1$, if ΔX^{i+1}, D^i and α^i are known, ΔX^{i+2}, D^{i+1} and α^{i+1} are generated as follows:

$$D^{i+1} = -(I-P)G^{i+1} + \beta^i D^i + \sum_{j=0}^{i-1} \beta_j^{(i)} D^j, \tag{6.20}$$

$$\alpha^{i+1} = -\frac{\langle B_k + \Phi(\Delta X^{i+1}), D^{i+1}\rangle}{\langle \Phi(D^{i+1}), D^{i+1}\rangle}, \tag{6.21}$$

$$\Delta X^{i+2} = \Delta X^{i+1} + \alpha^{i+1} D^{i+1}, \tag{6.22}$$

where coefficients β^i and $\beta_j^{(i)}$ are chosen so that so that $D^0, D^1, \ldots, D^{i+1}$ are mutually Φ-conjugate and

$$G^{i+1} = B_k + \Phi\big(\Delta X^{i+1}\big). \tag{6.23}$$

Note that matrices ΔX^{i+1}, G^{i+1} and D^{i+1} are all symmetric, which guarantees the iterate ΔX^{i+2} to be symmetric.

Theorem 12. Sequence $\{G^i\}$ and D^i generated by (6.20) have the following property:

$$\big\langle (I-P)G^{i+1}, D^j\big\rangle = 0, \quad j = 0, 1, \ldots, i. \tag{6.24}$$

Proof. We prove the theorem by induction. When $i = 0$, (6.24) follows from (6.16). Assume that (6.24) holds for $i-1$, i.e.

$$\big\langle (I-P)G^{i}, D^j\big\rangle = 0, \quad j = 0, 1, \ldots, i-1. \tag{6.25}$$

Since $\Delta X^{i+1} = \Delta X^i + \alpha^i D^i$ is defined by exact line search, it can be seen that

$$\big\langle (I-P)G^{i+1}, D^i\big\rangle = 0. \tag{6.26}$$

For any $j \leqslant i-1$, if follows from (6.23), (6.13) and the mutually conjugacy of $D^0, \ldots, D^i$ that

$$\begin{aligned}\big\langle (I-P)G^{i+1}, D^j\big\rangle &= \big\langle (I-P)G^i + \alpha^i (I-P)\Phi\big(D^i\big), D^j\big\rangle \\ &= \big\langle (I-P)G^i, D^j\big\rangle + \alpha^i\big\langle (I-P)\Phi\big(D^i\big), D^j\big\rangle \\ &= \alpha^i\big\{\big\langle \Phi\big(D^i\big), D^j\big\rangle - \big\langle \Phi\big(D^i\big), D^j\big\rangle\big\} = 0. \end{aligned} \tag{6.27}$$

Therefore the theorem is true. □

It follows from (6.20) that $(I-P)G^i$ is a combination of $D^0, \ldots, D^i$. Thus, Theorem 12 tells us that

$$\big\langle (I-P)G^{i+1}, (I-P)G^j\big\rangle = 0, \quad 0 \leqslant j \leqslant i, \tag{6.28}$$

which implies

$$\big\langle (I-P)G^{i+1}, \Phi\big(D^j\big)\big\rangle = \big\langle (I-P)G^{i+1}, (I-P)\big(G^{j+1}-G^j\big)\big\rangle = 0,\quad 0 \leqslant j \leqslant i-1. \tag{6.29}$$

Thus, from (6.20) and (6.29), it can be easily shown that

$$\beta_j^{(i)} = 0, \quad 0 \leqslant j \leqslant i-1, \tag{6.30}$$

$$\beta_i = \frac{\langle (I-P)G^{i+1}, \Phi(D^i)\rangle}{\langle \Phi(D^i, D^i\rangle}. \tag{6.31}$$

Relation (6.28) implies $(I-P)G^0, \ldots, (I-P)G^i$ are orthogonal, and therefore linearly independent. Consequently, there exists an integer q such that

$$(I-P)G^{q+1} = 0, \tag{6.32}$$

which is equivalent to

$$B_k + \Phi\big(\Delta X^{q+1}\big) = \sum_{i=1}^{m} \lambda_i A_i, \tag{6.33}$$

$$\big\langle \Delta X^{q+1}, A_i\big\rangle = 0, \quad i = 1, \ldots, m. \tag{6.34}$$

This is a necessary and sufficient optimality condition for subproblem (6.9), (6.10). Let $\Delta X = \Delta X^{q+1}$, $\Delta y = \lambda$ and $\Delta S = \varphi(\Delta X) - \sum_{i=1}^{m} \Delta y_i A_i$. Thus, the solution for (6.1)–(6.3) is obtained, which is the search direction we want to compute.

In our algorithm, we also need to compute the scaling matrix $W_k^{-1} = S_k^{1/2}(S_k^{1/2} X_k S_k^{1/2})^{-1/2} S_k^{1/2}$. Here we use the technique introduced by Todd, Toh and Tütüncü [15]. Let the Cholesky factorizations of the matrices X_k and S_k be

$$S = L_1 {L_1}^{\mathrm{T}}, \quad X = L_2 {L_2}^{\mathrm{T}}. \tag{6.35}$$

We compute the SVD of $L_2^{\mathrm{T}} L_1$

$${L_2}^{\mathrm{T}} L_1 = U \Lambda V^{\mathrm{T}} \tag{6.36}$$

and define $J = L_1^{-1} S^{1/2}$, which is an orthogonal matrix. From the fact that

$$S_k^{1/2} X_k S_k^{1/2} = J^{\mathrm{T}}\big({L_1}^{\mathrm{T}} L_2\big)\big({L_2}^{\mathrm{T}} L_1\big) J = \big(J^{\mathrm{T}} V\big)\Lambda^2\big(V^{\mathrm{T}} J\big), \tag{6.37}$$

it follows

$$\big(S_k^{1/2} X_k S_k^{1/2}\big)^{-1/2} = \big(J^{\mathrm{T}} V\big)\Lambda^{-1}\big(V^{\mathrm{T}} J\big). \tag{6.38}$$

Thus, $W_k^{-1} = L_1 V \Lambda^{-1} V^{\mathrm{T}} {L_1}^{\mathrm{T}} = F F^{\mathrm{T}}$, where $F = L_1 V \Lambda^{-1/2}$.

Now we can give an algorithm for solving system (6.1)–(6.3).

Algorithm 13.

Step 1. Compute $S = L_1 {L_1}^{\mathrm{T}}$, $X = L_2 {L_2}^{\mathrm{T}}$, and ${L_2}^{\mathrm{T}} L_1 = U \Lambda V^{\mathrm{T}}$.

Step 2. $F = L_1 V \Lambda^{-1/2}$, $W_k^{-1} = L_1 V \Lambda V^{\mathrm{T}} L_1{}^{\mathrm{T}} = F F^{\mathrm{T}}$.

Step 3. If $\mathcal{K}(X_k S_k) \leqslant K_0$ let $B_k = S_k X_k S_k / \|S_k X_k\|_F$ otherwise $B_k = -\mu X^{-1} + S$, where $\mu = \mathrm{Tr}(X_k S_k)/n$.

Step 4. Let $\Delta X^0 = 0$. If $\|(I - P)G^0\| = 0$ then stop; $D^0 = -(I - P)G^0$, $i = 0$.

Step 5. Compute the step-length:

$$\alpha_i = -\frac{\langle B_k + \Phi(\Delta X^i), D^i \rangle}{\langle \Phi(D^i), D^i \rangle}, \qquad \Delta X^{i+1} = \Delta X^i + \alpha_i D^i.$$

Step 6. If $\|(I - P)G^{i+1}\| = 0$ then stop;

$$\beta_i = \frac{\langle (I - P)G^{i+1}, \Phi(D^i) \rangle}{\langle \Phi(D^i), D^i \rangle}, \qquad D^{i+1} = -(I - P)G^{i+1} + \beta_i D^i.$$

Step 7. $i = i + 1$. Go to step 5.

From properties of the conjugate gradient method, one can see that the following theorem holds.

Theorem 14. Algorithm 13 terminates within at most $\frac{1}{2}n(n+1)$ steps. Furthermore, for any $1 \leqslant i \leqslant \frac{1}{2}n(n+1)$, we have

$$\left\langle (I - P)G^i, D^i \right\rangle = -\left\| (I - P)G^i \right\|^2, \tag{6.39}$$

$$\left\langle \Phi\left(D^i\right), D^j \right\rangle = 0, \tag{6.40}$$

$$\left\langle (I - P)G^i, D^j \right\rangle = 0, \tag{6.41}$$

$$\left\langle (I - P)G^i, (I - P)G^j \right\rangle = 0, \tag{6.42}$$

where $j = 0, 1, \ldots, i - 1$.

Acknowledgment

The authors would like to thank two anonymous referees for their valuable comments.

References

[1] F. Alizadeh, Interior point methods in semidefinite programming with application to combinatorial optimization, SIAM J. Opt. 5 (1995) 13–51.

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

[3] K. Fujisawa, M. Kojima and K. Nakata, Exploiting sparsity in primal–dual interior-point methods for semidefinite programming, Math. Programming 79 (1997) 235–253.

[4] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 3rd ed. (Johns Hopkins University Press, Baltimore, 1996).

[5] B. Jansen, C. Roos, T. Terlaky and J.P. Vial, Primal–dual algorithms for linear programming based on the logarithmic barrier method, J. Optim. Theory Appl. 83 (1994) 1–26.

[6] J. Jiang, A long step primal–dual path following method for semidefinite programming, Operations Research Letters 21 (1998) 53–61.

[7] E. De Klerk, Interior point methods for SDP, Ph.D. thesis, Delft University of Technology, Delft, The Netherlands (1997).

[8] M. Kojima, S. Shindoh and S. Hara, Interior-point methods for the monotone semidefinite linear complementarity problems, SIAM J. Opt. 7 (1997) 86–125.

[9] A. Nemirovskii and P. Gahinet, The projective method for solving linear matrix inequalities, Math. Programming, Series B 77 (1997) 163–190.

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

[11] Y. Nesterov and M.J. Todd, Primal–dual interior-point methods for self-scaled cones, SIAM J. Opt. 8 (1998) 32–364.

[12] J.W. Nie and Y. Yuan, A potential reduction algorithm for an extended SDP problem, Science in China 43 (2000) 35–46.

[13] H.D. Qi and X. Chen, On stationary points of merit functions for semi-definite complementarity problems, Report, ICMSEC, Chinese Academy of Sciences (1997).

[14] J.F. Sturm and S. Zhang, Symmetric primal–dual path following algorithms for semidefinite programming. Applied Numerical Mathematics 29 (1999) 301–315.

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

[16] P. Tseng, Search directions and convergence analysis of some infeasible path-following methods for the monotone semi-definite LCP, Optimization Methods and Software 9 (1998) 245–268.

[17] Y. Ye, *Interior Point Algorithm Theory and Analysis* (Wiley, New York, 1998).

[18] Y. Yuan, *Numerical Methods for Nonlinear Programming* (Shanghai Sci. Tech. Publishers, Shanghai, 1993) (in Chinese).

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

[20] Q. Zhao, S.E. Karische, F. Rendl and H. Wolkowicz, Semidefinite programming relaxations for the quadratic assignment problem, CORR Report 95-27, Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada (1996).