A FULL NT-STEP INFEASIBLE INTERIOR-POINT ALGORITHM FOR SEMIDEFINITE OPTIMIZATION BASED ON A SELF-REGULAR PROXIMITY

B. KHEIRFAM

Abstract

We introduce a full NT-step infeasible interior-point algorithm for semidefinite optimization based on a self-regular function to provide the feasibility step and to measure proximity to the central path. The result of polynomial complexity coincides with the best known iteration bound for infeasible interior-point methods.

2010 *Mathematics subject classification*: primary 90C51; secondary 90C22.

Keywords and phrases: semidefinite optimization, infeasible interior-point method, polynomial complexity, self-regular.

1. Introduction

Semidefinite optimization (SDO) problems are convex optimization problems over the intersection of an affine set and the cone of positive semidefinite matrices. SDO has wide application in continuous and combinatorial optimization [1, 4, 20]. It has become a popular research area in the past decade, after it became clear that the algorithm for linear optimization (LO) can often be extended to the more general SDO case. Several interior-point methods (IPMs) designed for LO have been successfully extended to SDO [8, 16, 19, 21]. An important contribution to this field was made by Nesterov and Todd [13, 14], who showed that the primal-dual algorithm maintains its theoretical efficiency when the nonnegativity constraints in LO are replaced by a convex cone, as long as the cone is homogeneous and self-dual, or, in the terminology of Nesterov and Todd, as long as the cone is self-scaled. Self-scaled cones are cones that have a self-scaled barrier; the nonnegative orthant and the cone of positive semidefinite matrices are special cases. Recently, Peng et al. [16, 17] studied primal-dual feasible IPMs for SDO by using self-regular functions, and showed that the complexity bounds are $O(\sqrt{n} \log(n/\epsilon))$ and $O(\sqrt{n} \log n \log(n/\epsilon))$ for small-update

Department of Mathematics, Azarbaijan University of Tarbiat Moallem, Tabriz, Iran;
e-mail: b.kheirfam@azaruniv.edu.

methods and large-update methods, respectively. Wang et al. [19] presented primal-dual feasible IPMs for SDO by using a kernel function.

Infeasible IPMs (IIPMs) start with an arbitrary positive point, and feasibility is reached as optimality is approached. The choice of the starting point in IIPMs is crucial to their performance. Lustig [9] was the first to present IIMPs for LO. Zhang [21] was the first to present a primal-dual IIPM for SDO with polynomial iteration complexity. Recently, Mansouri and Roos [12] proposed a new IIPM to solve the SDO problem. It is different from the classical IIPMs, in that it uses only full steps. Our motivation for the use of full Nesterov and Todd (NT) steps is that we use another definition for the feasibility step, which is induced by a proximity function, and show that the complexity result coincides with the best known complexity of IIPMs.

2. Preliminaries

We consider the SDO problem in standard form,

$$\min\{C \bullet X : A_i \bullet X = b_i, i = 1, 2, \ldots, m, X \geq 0\}, \tag{P}$$

and its dual problem,

$$\max\left\{b^T y : \sum_{i=1}^{m} y_i A_i + S = C, S \geq 0\right\}, \tag{D}$$

where each $A_i \in \mathbf{S}^n$, $b \in \mathbf{R}^m$ and $C \in \mathbf{S}^n$. Here $\mathbf{S}^n$ denotes the set of all symmetric $n \times n$ matrices, and $X \geq 0$ ($X > 0$) means that X is symmetric and positive semidefinite (symmetric and positive definite). Without loss of generality, we assume that the matrices A_i are linearly independent.

As usual for IIPMs, it was assumed by Mansouri and Roos [12] that one knows a positive scalar ξ such that $X^* + S^* \leq \xi I$ for some optimal solution (X^*, y^*, S^*) of (P) and (D), and that the initial iterates are $(X^0, y^0, S^0) = \xi(I, 0, I)$, where I denotes the identity matrix. Using $\mu^0 = \xi^2$, the total number of iterations for the algorithm of Mansouri and Roos [12] is bounded above by

$$20n \log \frac{\max\{n\xi^2, \|r_b^0\|_F, \|R_c^0\|_F\}}{\epsilon}, \tag{2.1}$$

where r_b^0 and R_c^0 are the initial residual matrices:

$$(r_b^0)_i = b_i - A_i \bullet X^0, \quad i = 1, 2, \ldots, m, \quad R_c^0 = C - \sum_{i=1}^{m} y_i^0 A_i - S^0.$$

Up to a constant factor, the iteration bound (2.1) was first obtained by Kojima et al. [5] and Potra and Sheng [18]. It is still the best known iteration bound for IIPMs.

To describe our aim, we recall the main ideas underlying the algorithm of Mansouri and Roos [12]. For any ν with $0 < \nu \leq 1$, we consider the perturbed problem

$$\min\{(C - \nu R_c^0) \bullet X : A_i \bullet X = b_i - \nu(r_b^0)_i, i = 1, 2, \ldots, m, X \geq 0\}, \tag{P_ν}$$

and its dual problem,

$$\max\left\{(b - \nu r_b^0)^T y : \sum_{i=1}^{m} y_i A_i + S = C - \nu R_c^0, S \geq 0\right\}. \tag{D_ν}$$

Note that if $\nu = 1$ then $X = X^0$ yields a strictly feasible solution of (P_ν) and $(y, S) = (y^0, S^0)$ is a strictly feasible solution of (D_ν). We conclude that if $\nu = 1$ then (P_ν) and (D_ν) are strictly feasible, which means that both perturbed problems (P_ν) and (D_ν) satisfy the well-known interior-point condition (IPC). More generally, one has the following result [12, Lemma 4.1].

LEMMA 2.1. *If the original problems* (P) *and* (D) *are feasible then, for each ν satisfying $0 < \nu \leq 1$, the perturbed problems* (P_ν) *and* (D_ν) *are strictly feasible.*

Assuming that (P) and (D) are feasible, it follows from Lemma 2.1 that the problems (P_ν) and (D_ν) satisfy the IPC for each $0 < \nu \leq 1$. Then their central paths exist, meaning that the system

$$b_i - A_i \bullet X = \nu(r_b^0)_i, \quad i = 1, 2, \ldots, m, \quad X \geq 0, \tag{2.2}$$

$$C - \sum_{i=1}^{m} y_i A_i - S = \nu R_c^0, \quad S \geq 0, \tag{2.3}$$

$$XS = \mu I$$

has a unique solution for every $\mu > 0$. If $0 < \nu \leq 1$ and $\mu = \nu \xi^2$, we denote this unique solution in the sequel by $(X(\nu), y(\nu), S(\nu))$. These are the μ-centers of the perturbed problems (P_ν) and (D_ν). In this notation, if we take $\nu = 1$, then

$$(X(1), y(1), S(1)) = (X^0, y^0, S^0) = \xi(I, 0, I).$$

We measure the proximity of iterates (X, y, S) to the μ-centers of the perturbed problems (P_ν) and (D_ν) by the quantity

$$\delta(X, S; \mu) := \delta(V) := \tfrac{1}{2}\|V^{-1} - V\|_F,$$

where

$$V := \frac{1}{\sqrt{\mu}} D^{-1} X D^{-1} = \frac{1}{\sqrt{\mu}} DSD$$

and

$$D^2 = X^{1/2}\{X^{1/2} S X^{1/2}\}^{-1/2} X^{1/2} = S^{-1/2}\{S^{1/2} XS^{1/2}\}^{1/2} S^{-1/2}.$$

Initially, we have $\delta(X, S; \mu) = 0$. In the sequel, we assume that at the start of each iteration, $\delta(X, S; \mu)$ is smaller than or equal to a threshold value $\tau > 0$. This certainly holds at the start of the first iteration.

We now describe one main iteration of the algorithm given by Mansouri and Roos [12]. The algorithm begins with an infeasible interior point (X, y, S) such that (X, y, S) is feasible for the perturbed problems (P_ν) and (D_ν), $\mathbf{Tr}(XS) = n\mu$

and $\delta(X, S; \mu) \leq \tau$, where $\mu = \nu\xi^2$. Each main iteration consists of one so-called feasibility step, a μ-update, and a few centering steps. First we find a new point (X^f, y^f, S^f) which is feasible for the perturbed problems with ν replaced by $\nu^+ := (1 - \theta)\nu$. Then μ is decreased to $\mu^+ := (1 - \theta)\mu$. Generally, there is no guarantee that $\delta(X^f, S^f; \mu^+) \leq \tau$. So a limited number of centering steps are applied to produce a new point (X^+, y^+, S^+) such that $\mathbf{Tr}(X^+ S^+) = n\mu^+$ and $\delta(X^+, S^+; \mu^+) < \tau$, where $\mu^+ = \nu^+\xi^2$. This process is repeated until the norms of the residuals and $\mathbf{Tr}(X^+ S^+)$ are less than some prescribed accuracy parameter ϵ.

We now describe the search directions used in the feasibility and centering steps. For the feasibility step, Mansouri and Roos [12] used search directions $\Delta^f X$, $\Delta^f y$ and $\Delta^f S$ defined by the system

$$A_i \bullet \Delta^f X = \theta\nu(r_b^0)_i, \quad i = 1, 2, \ldots, m,$$

$$\sum_{i=1}^m \Delta^f y_i A_i + \Delta^f S = \theta\nu(R_c^0),$$

$$\Delta^f X + P\Delta^f S P^T = \mu S^{-1} - X,$$

where

$$P := X^{1/2}\{X^{1/2} S X^{1/2}\}^{-1/2} X^{1/2} = S^{-1/2}\{S^{1/2} X S^{1/2}\}^{1/2} S^{-1/2},$$

which is symmetric and nonsingular.

If (X, y, S) is feasible for the perturbed problems (P_ν) and (D_ν) then after the feasibility step the iterates satisfy the affine equations in (2.2) and (2.3), with $\nu = \nu^+$. Assuming that $\delta(X, S; \mu) \leq \tau$ holds before the step, and by taking θ small enough, it can be guaranteed that after the step, the iterates

$$X^f = X + \Delta^f X, \quad y^f = y + \Delta^f y, \quad S^f = S + \Delta^f S \tag{2.4}$$

are positive definite and $\delta(X^f, S^f; \mu^+) \leq 1/\sqrt[4]{2}$, where $\mu^+ = (1 - \theta)\mu$.

In the centering steps, starting at the iterates $(X, y, S) = (X^f, y^f, S^f)$ and targeting at the μ-centers, the search directions ΔX, Δy and ΔS are the usual primal-dual Newton directions, defined by

$$A_i \bullet \Delta X = 0, \quad i = 1, 2, \ldots, m,$$

$$\sum_{i=1}^m \Delta y_i A_i + \Delta S = 0, \tag{2.5}$$

$$\Delta X + P\Delta S P^T = \mu S^{-1} - X.$$

Denoting the iterates after a centering step by X^+, y^+ and S^+, we recall the following result of Peng et al. [15].

LEMMA 2.2. *If $\delta := \delta(X, S; \mu) \leq 1$ then the primal-dual NT-step is feasible, that is, X^+ and S^+ are feasible. Moreover, if $\delta \leq 1/\sqrt[4]{2}$ then $\delta(X^+, S^+; \mu) \leq \delta^2$.*

Define

$$D_X^f := \frac{1}{\sqrt{\mu}} D^{-1} \Delta^f X D^{-1}, \quad D_S^f := \frac{1}{\sqrt{\mu}} D \Delta^f S D, \tag{2.6}$$

$$(V^f)^2 := \frac{1}{\mu} D^{-1} X^f S^f D,$$

where $D = P^{1/2}$ with P defined as above. The system defining the search directions $\Delta^f X$, $\Delta^f y$ and $\Delta^f S$ can be expressed in terms of the scaled search directions D_X^f and D_S^f as follows:

$$DA_i D \bullet D_X^f = \frac{1}{\sqrt{\mu}} \theta \nu (r_b^0)_i, \quad i = 1, \dots, m,$$

$$\sum_{i=1}^{m} \frac{\Delta^f y_i}{\sqrt{\mu}} DA_i D + D_S^f = \frac{1}{\sqrt{\mu}} \theta \nu DR_c^0 D, \tag{2.7}$$

$$D_X^f + D_S^f = V^{-1} - V.$$

It is surprising that the right-hand side in the third equation of (2.7), $V^{-1} - V$, equals the negative gradient of the classical logarithmic barrier function

$$\Psi(V) := \sum_{i=1}^{n} \psi(\lambda_i(V)), \quad \lambda_i(V) = \frac{1}{\mu} \lambda_i(D^{-1} X S D),$$

whose kernel function is

$$\psi(t) = \frac{t^2 - 1}{2} - \log t.$$

The main contribution of this paper is a modification of the feasibility step. We present a slightly different algorithm, obtained by changing the definition of the feasibility step via replacing the third equation of (2.7) by $D_X^f + D_S^f = -\nabla \Phi(V)$, where the kernel function of $\Phi(V)$ is

$$\phi(t) = \frac{1}{2} \left(t - \frac{1}{t} \right)^2.$$

Therefore, the system of the new feasibility step becomes

$$DA_i D \bullet D_X^f = \frac{1}{\sqrt{\mu}} \theta \nu (r_b^0)_i, \quad i = 1, \dots, m,$$

$$\sum_{i=1}^{m} \frac{\Delta^f y_i}{\sqrt{\mu}} DA_i D + D_S^f = \frac{1}{\sqrt{\mu}} \theta \nu DR_c^0 D, \tag{2.8}$$

$$D_X^f + D_S^f = -\nabla \Phi(V).$$

Since $\phi'(t) = t - 1/t^3$, the third equation in (2.8) can be written as

$$D_X^f + D_S^f = V^{-3} - V. \tag{2.9}$$

Note that $\|\nabla \Phi(V)\|_F = 0$ if and only if $V = I$, that is, if and only if $XS = \mu I$. Thus $\|\nabla \Phi(V)\|_F$ is also a suitable proximity. This norm-based proximity is used to define

the central neighborhood. We now give a more formal description of Algorithm 1 below.

Algorithm 1: Primal-dual infeasible IPM

Input: accuracy parameter $\epsilon > 0$;
            barrier update parameter θ, $0 < \theta < 1$;
            threshold parameter τ, $0 < \tau \le 1/\sqrt[4]{2}$;
            $X^0 = \xi I$, $S^0 = \xi I$, $y^0 = 0$, $\nu = 1$;
begin
    $X := X^0$, $S := S^0$, $y := y^0$, $\mu := \mu^0$;
    while $\max(\mathbf{Tr}(XS), \|r_b\|, \|R_c\|) \ge \epsilon$ **do**
    begin
        feasibility step: solve (2.8) and use (2.6) for
        $\Delta^f X, \Delta^f y, \Delta^f S$; set $(X, y, S) := (X, y, S) + (\Delta^f X, \Delta^f y, \Delta^f S)$;
        μ-update: $\mu := (1 - \theta)\mu$;
        centering steps: solve (2.5) for $\Delta X, \Delta y, \Delta S$;
        while $\frac{1}{2}\|\nabla\Phi(V)\|_F^2 \ge \tau$ **do**
        begin
            $(X, y, S) := (X, y, S) + (\Delta X, \Delta y, \Delta S)$;
        end
    end
end

---

3. Technical results

We now give some lemmas which are used in the analysis later. Recall that we denote by $\mathbf{S}^n$ the set of all symmetric $n \times n$ matrices. Here, for any $A \in \mathbf{S}^n$, we let $\lambda_i(A)$, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the ith eigenvalue, smallest eigenvalue and largest eigenvalue of A, respectively. Moreover, we let $\mathbf{S}^n_{++}$ denote the set of all symmetric positive definite $n \times n$ matrices.

LEMMA 3.1 [2, Lemma A.1]. *Let $Q \in \mathbf{S}^n_{++}$ and let $M \in \mathbf{R}^{n \times n}$ be skew-symmetric (that is, $M = -M^T$). Then $\det(Q + M) > 0$. Moreover, if the eigenvalues of $Q + M$ are real then*

$$0 < \lambda_{\min}(Q) \le \lambda_{\min}(Q + M) \le \lambda_{\max}(Q + M) \le \lambda_{\max}(Q).$$

LEMMA 3.2 [15]. *Let $Q \in \mathbf{S}^n$ with $Q > 0$, and let $M \in \mathbf{R}^{n \times n}$ be skew-symmetric. Then $\mathbf{Tr}((Q + M)^{-1}) \le \mathbf{Tr}(Q^{-1})$.*

LEMMA 3.3 [10]. *Let $A, B \in \mathbf{S}^n$ with $A + B > 0$ and $\lambda_{\min}(A) + \lambda_{\min}(B) > 0$. Then*

$$\sum_{i=1}^n \lambda_i(A + B) = \sum_{i=1}^n (\lambda_i(A) + \lambda_i(B)), \quad \sum_{i=1}^n \frac{1}{\lambda_i(A + B)} = \sum_{i=1}^n \frac{1}{\lambda_i(A) + \lambda_i(B)}.$$

LEMMA 3.4 [11, Lemma A.1]. *For $i = 1, 2, \ldots, n$, let $f_i : \mathbb{R}_+ \to \mathbb{R}$ be a convex function. Then, for any nonzero vector $z \in \mathbb{R}_+^n$,*

$$\sum_{i=1}^{n} f_i(z_i) \leq \frac{1}{e^T z} \sum_{j=1}^{n} z_j \Big(f_j(e^T z) + \sum_{i \neq j} f_i(0) \Big).$$

LEMMA 3.5 [7, Lemma 3]. *Let $f(x) = f_1(x) + f_2(x)$, with $f_1(x)$ and $f_2(x)$ both strictly monotonically increasing in a given interval. Let x_1 and x_2 denote the roots of $f_1(x) = 0$ and $f_2(x) = 0$, respectively. Then the root x^* of $f(x) = 0$ satisfies $x^* \geq \min\{x_1, x_2\}$.*

LEMMA 3.6 [2, Lemma 6.1]. *Suppose that $X, S > 0$ and, for $0 \leq \alpha \leq 1$, $X(\alpha) = X + \alpha \Delta X$ and $S(\alpha) = S + \alpha \Delta S$. If*

$$\det(X(\alpha)S(\alpha)) > 0 \quad \text{for all } 0 \leq \alpha \leq \bar{\alpha},$$

then $X(\bar{\alpha}) > 0$ and $S(\bar{\alpha}) > 0$.

Let X^+ and S^+ denote the iterates after a centering step, that is,

$$X^+ = X + \Delta X, \quad S^+ = S + \Delta S,$$

where ΔX and ΔS are the centering steps. The next lemma focuses on the effect of the feasible search direction induced by the self-regular proximity function.

LEMMA 3.7. *If $\Phi(V) := \Phi(X, S; \mu) \leq 2$ then the primal-dual NT-step is feasible, that is, X^+ and S^+ are feasible and $\mathbf{Tr}(X^+ S^+) = n\mu$. Moreover, if $\Phi(V) \leq \sqrt{2}$ then $\Phi(X^+, S^+; \mu) \leq (\Phi(V)/\sqrt{2})^2$.*

PROOF. By the relation $\Phi(V)/2 = \delta(V)^2$, if $\Phi(V) := \Phi(X, S; \mu) \leq 2$ then $\delta \leq 1$, and by Lemma 2.2, X^+ and S^+ are feasible. If $\Phi(V) \leq \sqrt{2}$ then

$$\tfrac{1}{2}\Phi(X^+, S^+; \mu) = \delta(X^+, S^+; \mu)^2 \leq \delta(V)^4 = \tfrac{1}{4}\Phi(V)^2,$$

where the inequality follows from Lemma 2.2. □

The following lemma shows the effect on the proximity measure if V is replaced by $\tilde{V} := \sqrt{1 - \theta} V$.

LEMMA 3.8. *Let (X, S) be a positive definite primal-dual pair and $\mu > 0$ such that $\mathbf{Tr}(XS) = n\mu$. Moreover, let $\Phi(V) = \Phi(X, S; \mu)$ and $\tilde{V} := \sqrt{1 - \theta} V$. Then*

$$\Phi(\tilde{V}) = \frac{1}{1 - \theta} \Phi(V) + \frac{n\theta^2}{2(1 - \theta)}.$$

PROOF. From the definition of the kernel function,

$$\Phi(\tilde{V}) = \frac{1}{2} \|\tilde{V} - \tilde{V}^{-1}\|_F^2 = \frac{1}{2} \left\| \sqrt{1 - \theta} V - \frac{1}{\sqrt{1 - \theta}} V^{-1} \right\|_F^2$$

$$
\begin{aligned}
&= \frac{1}{2} \left\| \left(\sqrt{1-\theta}\, V - \frac{1}{\sqrt{1-\theta}} V \right) + \frac{1}{\sqrt{1-\theta}} (V - V^{-1}) \right\|_F^2 \\
&= \frac{1}{2} \left\| \frac{1}{\sqrt{1-\theta}} (V - V^{-1}) - \frac{\theta}{\sqrt{1-\theta}} V \right\|_F^2 \\
&= \frac{1}{2(1-\theta)} \|V - V^{-1}\|_F^2 + \frac{\theta^2}{2(1-\theta)} \|V\|_F^2 - \frac{\theta}{1-\theta} \mathbf{Tr}(V^T(V - V^{-1})) \\
&= \frac{1}{1-\theta} \Phi(V) + \frac{n\theta^2}{2(1-\theta)},
\end{aligned}
$$

due to $\mu V^2 \sim XS$, $\mathbf{Tr}(I) = n$ and

$$
\|V\|_F^2 = \mathbf{Tr}(V^T V) = \sum_{i=1}^{n} \lambda_i(V^2) = \sum_{i=1}^{n} \lambda_i\left(\frac{XS}{\mu} \right) = \frac{\mathbf{Tr}(X^T S)}{\mu} = n. \qquad \square
$$

The following lemma states the relation between the proximity function $\Phi(V)$ and the norm-based proximity $\|\nabla\Phi(V)\|_F$.

LEMMA 3.9 [6, Lemma 6]. $\frac{1}{2}\|\nabla\Phi(V)\|_F^2 \geq \Phi(V)$.

PROOF. Since

$$
\tfrac{1}{2}(t^{-3} - t)^2 - \tfrac{1}{2}(t - t^{-1})^2 = \tfrac{1}{2}(t^{-2} - 1)^2(t^{-2} + 2) \geq 0 \quad \text{for all } t > 0,
$$

the lemma follows. $\qquad \square$

4. Analysis of the feasibility step

Let X, y and S denote the iterates at the start of an iteration, with $\mathbf{Tr}(XS) = n\mu$ and $\frac{1}{2}\|\Phi(V)\|_F^2 \leq \tau$. Recall that at the start of the first iteration these certainly hold, because $\mathbf{Tr}(X^0 S^0) = n\mu^0$ and $\|\Phi(V)\|_F = 0$.

4.1. Effect of the feasibility step According to Lemma 3.7, we need to show that $\Phi(X^f, S^f; \mu^+) \leq \sqrt{2}$ after the feasibility step, that is, that the new iterates are within the region where the NT process targeting at the μ^+-centers of (P_{ν^+}) and (D_{ν^+}) is quadratically convergent. Using (2.4) and (2.6),

$$
\begin{aligned}
X^f &= X + \Delta^f X = \sqrt{\mu}\, D(V + D_X^f) D, \\
S^f &= S + \Delta^f S = \sqrt{\mu}\, D^{-1}(V + D_S^f) D^{-1}.
\end{aligned}
$$

Therefore,

$$
X^f S^f = \mu D(V + D_X^f)(V + D_S^f) D^{-1}.
$$

This shows that

$$X^f S^f \sim \mu(V + D_X^f)(V + D_S^f).$$

Defining

$$D_{XS}^f := \tfrac{1}{2}(D_X^f D_S^f + D_S^f D_X^f), \quad M := (D_X^f V - V D_X^f) + \tfrac{1}{2}(D_X^f D_S^f - D_S^f D_X^f),$$

noting that D_{XS}^f is symmetric and M is skew-symmetric, and using (2.9),

$$
\begin{aligned}
X^f S^f &\sim \mu(V^2 + V D_S^f + D_X^f V + D_X^f D_S^f) \\
&= \mu(V^{-2} - V D_X^f + D_X^f V + D_X^f D_S^f) \\
&= \mu(V^{-2} - V D_X^f + D_X^f V + D_X^f D_S^f - \tfrac{1}{2}D_S^f D_X^f + \tfrac{1}{2}D_S^f D_X^f) \\
&= \mu(V^{-2} + \tfrac{1}{2}(D_X^f D_S^f + D_S^f D_X^f) + (D_X^f V - V D_X^f) + \tfrac{1}{2}(D_X^f D_S^f - D_S^f D_X^f)) \\
&= \mu(V^{-2} + D_{XS}^f + M).
\end{aligned}
\tag{4.1}
$$

We want the new iterates to be positive definite. We call the NT-step strictly feasible if $X^f > 0$ and $S^f > 0$. The next lemma gives conditions for strict feasibility of the full NT-step; its proof is similar to the proof of Lemma 5.4 of Mansouri and Roos [12].

LEMMA 4.1. *Let $X > 0$ and $S > 0$. Then the iterates (X^f, y^f, S^f) are strictly feasible if $V^{-2} + D_{XS}^f > 0$.*

PROOF. Introduce a step length α with $0 \le \alpha \le 1$, and define

$$X^\alpha = X + \alpha \Delta^f X, \quad y^\alpha = y + \alpha \Delta^f y, \quad S^\alpha = S + \alpha \Delta^f S.$$

We thus have $X^0 = X$, $X^1 = X^f$, $y^0 = y$, $y^1 = y^f$, $S^0 = S$ and $S^1 = S^f$. Note that $\det(X^0 S^0) = \det(\mu^0 I) = (\mu^0)^n > 0$. Our aim is to show that the determinant of $X^\alpha S^\alpha$ remains positive for all $0 \le \alpha \le 1$. We may write

$$
\begin{aligned}
\frac{X^\alpha S^\alpha}{\mu} &\sim (V + \alpha D_X^f)(V + \alpha D_S^f) = V^2 + \alpha(D_X^f V + V D_S^f) + \alpha^2 D_X^f D_S^f \\
&= V^2 + \alpha(V D_X^f + V D_S^f) + \alpha(D_X^f V - V D_X^f) + \alpha^2 D_X^f D_S^f.
\end{aligned}
$$

Using (2.9),

$$\frac{X^\alpha S^\alpha}{\mu} = V^2 + \alpha(V^{-2} - V^2) + \alpha(D_X^f V - V D_X^f) + \alpha^2(D_X^f D_S^f).$$

By subtracting $\alpha^2(D_X^f V - V D_X^f)$ and adding $(\alpha^2/2)D_S^f D_X^f$ to the right-hand side of the above equality,

$$
\begin{aligned}
\frac{X^\alpha S^\alpha}{\mu} &= V^2 + \alpha(V^{-2} - V^2) + \alpha^2 D_{XS}^f + \frac{\alpha^2}{2}(D_X^f D_S^f - D_S^f D_X^f) \\
&\quad + \alpha^2(D_X^f V - V D_X^f) + \alpha(1 - \alpha)(D_X^f V - V D_X^f) \\
&= \alpha V^{-2} + (1 - \alpha)V^2 + \alpha^2 D_{XS}^f + \alpha^2 M + \alpha(1 - \alpha)(D_X^f V - V D_X^f).
\end{aligned}
$$

The matrix $(1 - \alpha)(D_X^f V - V D_X^f) + \alpha M$ is skew-symmetric for $0 \le \alpha \le 1$. Lemma 3.1 therefore implies that the determinant of $X^\alpha S^\alpha$ will be positive if the symmetric matrix

$$\alpha V^{-2} + (1 - \alpha) V^2 + \alpha^2 D_{XS}^f$$

is positive definite. The latter is true for $0 \le \alpha \le 1$, because

$$\alpha V^{-2} + (1 - \alpha) V^2 + \alpha^2 D_{XS}^f = (1 - \alpha)(V^2 + \alpha V^{-2}) + \alpha^2 (V^{-2} + D_{XS}^f) > 0.$$

So $X^\alpha S^\alpha$ has positive determinant for $0 \le \alpha \le 1$. Therefore, by Lemma 3.6, X^1 and S^1 are positive definite. This completes the proof. $\square$

LEMMA 4.2. *The following inequalities hold:*

$$\frac{1}{\rho(\Phi(V))} \le \lambda_i(V) \le \rho(\Phi(V)), \quad i = 1, 2, \ldots, n, \tag{4.2}$$

where

$$\rho(\Phi(V)) := \frac{\sqrt{\Phi(V)} + \sqrt{\Phi(V) + 2}}{\sqrt{2}}. \tag{4.3}$$

PROOF. By definition of $\Phi(V)$,

$$2\Phi(V) = \|V - V^{-1}\|_F^2 = \mathbf{Tr}((V - V^{-1})^T (V - V^{-1}))$$

$$= \mathbf{Tr}(V^2 - 2I + V^{-2}) = \sum_{i=1}^n \left(\lambda_i(V)^2 - 2 + \frac{1}{\lambda_i(V)^2} \right) = \sum_{i=1}^n \left(\lambda_i(V) - \frac{1}{\lambda_i(V)} \right)^2.$$

This implies that

$$\left(\lambda_i(V) - \frac{1}{\lambda_i(V)} \right)^2 \le 2\Phi(V).$$

It follows that

$$\frac{\sqrt{2}}{\sqrt{\Phi(V)} + \sqrt{\Phi(V) + 2}} \le \lambda_i(V) \le \frac{\sqrt{\Phi(V)} + \sqrt{\Phi(V) + 2}}{\sqrt{2}}. \qquad \square$$

In the sequel, we denote

$$w := w(V) := \tfrac{1}{2} \sqrt{\|D_X^f\|_F^2 + \|D_S^f\|_F^2}. \tag{4.4}$$

This implies that $\|D_X^f\|_F \le 2w$ and $\|D_S^f\|_F \le 2w$. We recall two properties of the Frobenius matrix norm [3]. For any real symmetric $n \times n$ matrix A,

$$\|A\|_F^2 = \mathbf{Tr}(A^2) = \sum_{i=1}^n \lambda_i(A^2) = \sum_{i=1}^n \lambda_i(A)^2.$$

This norm is also submultiplicative, that is, for any two square matrices A and B,

$$\|AB\|_F \le \|A\|_F \|B\|_F.$$

Using these properties and (4.4),

$$\|D^f_{XS}\|_F \le \|D^f_X\|_F \|D^f_S\|_F \le \frac{1}{2}(\|D^f_X\|^2_F + \|D^f_S\|^2_F) = 2w^2,$$

$$2w_i^2 := |\lambda_i(D^f_{XS})| \le \|D^f_{XS}\|_F \le 2w^2, \quad i = 1, \dots, n. \tag{4.5}$$

LEMMA 4.3. *Let* $1 - 2\rho(\Phi(V))^2 w^2 > 0$. *Then* $\lambda_{\min}(V^{-2}) + \lambda_{\min}(D^f_{XS}) > 0$.

PROOF. By Lemma 4.2,

$$\frac{1}{\rho(\Phi(V))} \le \lambda_{\max}(V) \le \rho(\Phi(V)).$$

This implies that

$$\frac{1}{\rho(\Phi(V))^2} \le \lambda_{\min}(V^{-2}) \le \rho(\Phi(V))^2. \tag{4.6}$$

From (4.5),

$$- 2w^2 \le \lambda_{\min}(D^f_{XS}) \le 2w^2. \tag{4.7}$$

Using (4.6) and (4.7),

$$0 < 1 - 2\rho(\Phi(V))^2 w^2 = \rho(\Phi(V))^2 \Big(\frac{1}{\rho(\Phi(V))^2} - 2w^2 \Big)$$

$$\le \rho(\Phi(V))^2 (\lambda_{\min}(V^{-2}) + \lambda_{\min}(D^f_{XS})).$$

The result follows. □

LEMMA 4.4. *If* $V^{-2} + D^f_{XS} > 0$ *and* $1 - 2\rho(\Phi(V))^2 w^2 > 0$ *then*

$$2\Phi(V^f) \le \frac{2\Phi(V)}{1 - \theta} + \frac{n\theta^2}{1 - \theta} + \frac{2w^2}{1 - \theta} + \frac{2(1 - \theta)w^2\rho(\Phi(V))^4}{1 - 2\rho(\Phi(V))^2 w^2}.$$

PROOF. Using (4.1), after division of both sides by $\mu^+ = (1 - \theta)\mu$ we obtain

$$(V^f)^2 \sim \frac{\mu(V^{-2} + D^f_{XS} + M)}{\mu^+} = \frac{V^{-2} + D^f_{XS} + M}{1 - \theta}.$$

Hence

$$2\Phi(V^f) = \|V^f - (V^f)^{-1}\|^2_F = \mathbf{Tr}((V^f)^2) + \mathbf{Tr}((V^f)^{-2}) - 2n$$

$$\le \mathbf{Tr}\Big(\frac{V^{-2} + D^f_{XS}}{1 - \theta} \Big) + \mathbf{Tr}((1 - \theta)(V^{-2} + D^f_{XS})^{-1}) - 2n$$

$$= \sum_{i=1}^n \frac{\lambda_i(V)^{-2} + \lambda_i(D^f_{XS})}{1 - \theta} + \sum_{i=1}^n \frac{1 - \theta}{\lambda_i(V)^{-2} + \lambda_i(D^f_{XS})} - 2n$$

$$\le \sum_{i=1}^n \Big(\frac{\lambda_i(V)^{-2} + 2w_i^2}{1 - \theta} + \frac{1 - \theta}{\lambda_i(V)^{-2} - 2w_i^2} - 2 \Big).$$

Here for the first inequality we use Lemma 3.2 and the fact that $\mathbf{Tr}(A + M) = \mathbf{Tr}(A)$ for skew-symmetric M, and for the third equality we use Lemma 3.3.

Define

$$f_i(2w_i^2) := \frac{\lambda_i(V)^{-2} + 2w_i^2}{1 - \theta} + \frac{1 - \theta}{\lambda_i(V)^{-2} - 2w_i^2} - 2, \quad i = 1, 2, \ldots, n.$$

Using (4.2), (4.5) and the hypothesis $1 - 2\rho(\Phi(V))^2 w^2 > 0$,

$$\lambda_i(V)^{-2} - 2w_i^2 > 0.$$

This implies that $f_i(2w_i^2)$ is convex in $2w_i^2$. Therefore, by Lemma 3.4,

$$2\Phi(V^f) \leq \sum_{i=1}^{n} f(2w_i^2) \leq \frac{1}{2w^2} \sum_{j=1}^{n} 2w_j^2 \Big(f_j(2w^2) + \sum_{i \neq j} f_i(0) \Big)$$

$$= \frac{1}{2w^2} \sum_{j=1}^{n} \Big[2w_j^2 \Big(\Big(\frac{\lambda_j(V)^{-2} + 2w^2}{1 - \theta} + \frac{1 - \theta}{\lambda_j(V)^{-2} - 2w^2} - 2 \Big)$$

$$+ \sum_{i \neq j} \Big(\frac{\lambda_i(V)^{-2}}{1 - \theta} + \frac{1 - \theta}{\lambda_i(V)^{-2}} - 2 \Big) \Big) \Big].$$

Using Lemma 3.8,

$$\sum_{i \neq j} \Big(\frac{\lambda_i(V)^{-2}}{1 - \theta} + \frac{1 - \theta}{\lambda_i(V)^{-2}} - 2 \Big)$$

$$= \sum_{i=1}^{n} \Big(\frac{\lambda_i(V)^{-2}}{1 - \theta} + \frac{1 - \theta}{\lambda_i(V)^{-2}} - 2 \Big) - \Big(\frac{\lambda_j(V)^{-2}}{1 - \theta} + \frac{1 - \theta}{\lambda_j(V)^{-2}} - 2 \Big)$$

$$= \frac{2\Phi(V)}{1 - \theta} + \frac{n\theta^2}{1 - \theta} - \Big(\frac{\lambda_j(V)^{-2}}{1 - \theta} + \frac{1 - \theta}{\lambda_j(V)^{-2}} - 2 \Big).$$

This gives the following upper bound for $2\Phi(V^f)$:

$$2\Phi(V^f) \leq \frac{1}{2w^2} \sum_{j=1}^{n} \Big[2w_j^2 \Big(\Big(\frac{\lambda_j(V)^{-2} + 2w^2}{1 - \theta} + \frac{1 - \theta}{\lambda_j(V)^{-2} - 2w^2} - 2 \Big)$$

$$+ \frac{2\Phi(V)}{1 - \theta} + \frac{n\theta^2}{1 - \theta} - \Big(\frac{\lambda_j(V)^{-2}}{1 - \theta} + \frac{1 - \theta}{\lambda_j(V)^{-2}} - 2 \Big) \Big) \Big]$$

$$= \frac{2\Phi(V)}{1 - \theta} + \frac{n\theta^2}{1 - \theta} + \frac{2w^2}{1 - \theta} + \frac{1}{2w^2} \sum_{j=1}^{n} 2w_j^2 \Big(\frac{2(1 - \theta)w^2}{\lambda_j(V)^{-2}(\lambda_j(V)^{-2} - 2w^2)} \Big)$$

$$\leq \frac{2\Phi(V)}{1 - \theta} + \frac{n\theta^2}{1 - \theta} + \frac{2w^2}{1 - \theta} + \frac{2(1 - \theta)w^2}{1/\rho(\Phi(V))^2(1/\rho(\Phi(V))^2 - 2w^2)}$$

$$= \frac{2\Phi(V)}{1 - \theta} + \frac{n\theta^2}{1 - \theta} + \frac{2w^2}{1 - \theta} + \frac{2(1 - \theta)w^2\rho(\Phi(V))^4}{1 - 2\rho(\Phi(V))^2 w^2}.$$

The proof is complete. $\square$

We conclude this section by presenting a value that we do not allow w to exceed. Needing $\Phi(V^f) \leq \sqrt{2}$, it follows from Lemma 4.4 that it is sufficient to have

$$\frac{2\Phi(V)}{1-\theta} + \frac{n\theta^2}{1-\theta} + \frac{2w^2}{1-\theta} + \frac{2(1-\theta)w^2\rho(\Phi(V))^4}{1-2\rho(\Phi(V))^2w^2} \leq 2\sqrt{2}.$$

Now,

$$f(t) = \frac{2\Phi(V)}{1-\theta} + \frac{n\theta^2}{1-\theta} + \frac{2t}{1-\theta} + \frac{2(1-\theta)\rho(\Phi(V))^4 t}{1-2\rho(\Phi(V))^2 t} - 2\sqrt{2},$$

$$f_1(t) = \frac{2\Phi(V)}{1-\theta} + \frac{n\theta^2}{1-\theta} + \frac{2t}{1-\theta} - \frac{3\sqrt{2}}{2},$$

$$f_2(t) = \frac{2(1-\theta)\rho(\Phi(V))^4 t}{1-2\rho(\Phi(V))^2 t} - \frac{\sqrt{2}}{2}.$$

Note that $f(t) = f_1(t) + f_2(t)$, and that both $f_1(t)$ and $f_2(t)$ are monotonically increasing in t. By Lemma 3.5, the root t^* of $f(t) = 0$ satisfies $t^* \geq \min\{t_1^*, t_2^*\}$, where t_1^* and t_2^* are the roots of $f_1(t) = 0$ and $f_2(t) = 0$, respectively.

Since $f_1(t_1^*) = 0$,

$$t_1^* = \left(\frac{3\sqrt{2}}{2} - \frac{2\Phi(V)}{1-\theta} - \frac{n\theta^2}{1-\theta}\right)\frac{1-\theta}{2},$$

and from $f_2(t_2^*) = 0$,

$$t_2^* = \frac{\sqrt{2}}{4(1-\theta)\rho(\Phi(V))^4 + 2\sqrt{2}\rho(\Phi(V))^2}.$$

At this stage, we choose

$$\tau = \frac{1}{16}, \quad \theta = \frac{\alpha}{4\sqrt{n}}, \quad \alpha \leq 1. \tag{4.8}$$

Then, for $n \geq 1$ and

$$\Phi(V) \leq \tfrac{1}{2}\|\nabla\Phi(V)\|_F^2 \leq \tau,$$

it is easily verified that

$$t_1^* \geq \frac{3}{8}\left(\frac{3\sqrt{2}}{2} - \frac{8\tau}{3}\right) - \frac{1}{32} \geq \frac{11}{16}, \quad t_2^* \geq \frac{14}{125}. \tag{4.9}$$

Using (4.9) and the assumption that $1 - 2\rho(\Phi(V))^2 w^2 > 0$, it is easily verified that if $w^2 \leq \min\{11/16, 14/125, 4/11\}$, that is,

$$w \leq \sqrt{\frac{14}{125}}, \tag{4.10}$$

then $\Phi(V^f) \leq \sqrt{2}$.

4.2. Upper bound for $\|D_X^f\|_F^2 + \|D_S^f\|_F^2$

In this subsection, we obtain an upper bound for $\|D_X^f\|_F^2 + \|D_S^f\|_F^2$, which enables us to find a default value for θ.

Consider the system

$$DA_iD \bullet D_X^f = \frac{1}{\sqrt{\mu}}\theta\nu(r_b^0)_i, \quad i = 1, \ldots, m,$$

$$\sum_{i=1}^m \frac{\Delta y_i}{\sqrt{\mu}}DA_iD + D_S^f = \frac{1}{\sqrt{\mu}}\theta\nu DR_c^0D, \tag{4.11}$$

$$D_X^f + D_S^f = V^{-3} - V,$$

which defines the search directions $\Delta^f X$, $\Delta^f y$ and $\Delta^f S$ in terms of the scaled search directions D_X^f and D_S^f. For the moment, let us define

$$(r_b)_i = \theta\nu(r_b^0)_i, \quad i = 1, 2, \ldots, m, \quad R_c = \theta\nu R_c^0, \quad R := V^{-3} - V = -\nabla\Phi(V).$$

With $\xi_i = -\Delta^f y_i$, by eliminating D_S^f from (4.11) we obtain

$$DA_iD \bullet D_X^f = \frac{1}{\sqrt{\mu}}(r_b)_i, \quad i = 1, 2, \ldots, m,$$

$$\sum_{i=1}^m \frac{\xi_i}{\sqrt{\mu}}DA_iD + D_X^f = R - \frac{1}{\sqrt{\mu}}DR_cD. \tag{4.12}$$

Recall the definition of the Kronecker product $M \otimes N$: if $M \in R^{k \times \ell}$ and $N \in R^{r \times s}$ then $M \otimes N \in R^{kr \times \ell s}$ is a $k \times \ell$ block matrix whose (i, j)th block is $m_{ij}N$. The following properties of the Kronecker product are used later [10, pp. 19–20]. Here, for any $M \in R^{k \times \ell}$, **vec**(M) denotes the $k\ell$-vector obtained by stacking the columns of M:

- $(M \otimes N)(K \otimes L) = MK \otimes NL$;
- $(M \otimes N)^{-1} = M^{-1} \otimes N^{-1}$;
- $(M \otimes N)\mathbf{vec}(K) = \mathbf{vec}(NKM^T)$;
- $(M \otimes N)^T = M^T \otimes N^T$;
- if $M > 0$ and $N > 0$ then $M \otimes N > 0$.

Using these properties and the definition of the inner product of two matrices, the system (4.12) can be rewritten as follows:

$$\mathbf{vec}(A_i)^T(D \otimes D)\mathbf{vec}(D_X^f) = \frac{1}{\sqrt{\mu}}(r_b)_i, \quad i = 1, 2, \ldots, m,$$

$$\sum_{i=1}^m \frac{\xi_i}{\sqrt{\mu}}(D \otimes D)\mathbf{vec}(A_i) + \mathbf{vec}(D_X^f) = \mathbf{vec}(R) - \frac{1}{\sqrt{\mu}}(D \otimes D)\mathbf{vec}(R_c). \tag{4.13}$$

Let $\mathcal{A}^T = [\mathbf{vec}(A_1), \ldots, \mathbf{vec}(A_m)]$ and $\xi = (\xi_1, \xi_2, \ldots, \xi_m)$. We may rewrite (4.13) as

$$\mathcal{A}(D \otimes D)\mathbf{vec}(D_X^f) = \frac{1}{\sqrt{\mu}}r_b,$$

$$(D \otimes D)\mathcal{A}^T\frac{\xi}{\sqrt{\mu}} + \mathbf{vec}(D_X^f) = \mathbf{vec}(R) - \frac{1}{\sqrt{\mu}}(D \otimes D)\mathbf{vec}(R_c). \tag{4.14}$$

By multiplying the second equation of (4.14) on the left by $\mathcal{A}(D \otimes D)$ and using the first equation of (4.14), we obtain

$$\mathcal{A}(D \otimes D)^2 \mathcal{A}^T \frac{\xi}{\sqrt{\mu}} + \frac{1}{\sqrt{\mu}} r_b = \mathcal{A}(D \otimes D)\Big(\mathbf{vec}(R) - \frac{1}{\sqrt{\mu}}(D \otimes D)\mathbf{vec}(R_c)\Big).$$

Therefore,

$$\xi = \sqrt{\mu}(\mathcal{A}(D \otimes D)^2 \mathcal{A}^T)^{-1}\Big[\mathcal{A}(D \otimes D)\Big(\mathbf{vec}(R) - \frac{1}{\sqrt{\mu}}(D \otimes D)\mathbf{vec}(R_c)\Big) - \frac{1}{\sqrt{\mu}} r_b\Big].$$

Substitution into (4.14) gives

$$\mathbf{vec}(D_X^f) = [I - (D \otimes D)\mathcal{A}^T (\mathcal{A}(D \otimes D)^2 \mathcal{A}^T)^{-1}\mathcal{A}(D \otimes D)]$$
$$\times \Big(\mathbf{vec}(R) - \frac{1}{\sqrt{\mu}}(D \otimes D)\mathbf{vec}(R_c)\Big)$$
$$+ \frac{1}{\sqrt{\mu}}(D \otimes D)\mathcal{A}^T (\mathcal{A}(D \otimes D)^2 \mathcal{A}^T)^{-1} r_b.$$

Let $(\bar{X}, \bar{y}, \bar{S})$ be such that

$$\mathcal{A}\mathbf{vec}(\bar{X}) = b, \quad \mathcal{A}^T \bar{y} + \mathbf{vec}(\bar{S}) = \mathbf{vec}(C).$$

Then we may write

$$r_b = \theta \nu r_b^0 = \theta \nu (b - \mathcal{A}\mathbf{vec}(X^0)) = \theta \nu \mathcal{A}(\mathbf{vec}(\bar{X}) - \mathbf{vec}(X^0))$$

and

$$\mathbf{vec}(R_c) = \theta \nu \mathbf{vec}(R_c^0) = \theta \nu (\mathbf{vec}(C) - \mathcal{A}^T y^0 - \mathbf{vec}(S^0))$$
$$= \theta \nu (\mathcal{A}^T (\bar{y} - y^0) + \mathbf{vec}(\bar{S}) - \mathbf{vec}(S^0)).$$

To simplify the notation, we set

$$\tilde{P} = (D \otimes D)\mathcal{A}^T (\mathcal{A}(D \otimes D)^2 \mathcal{A}^T)^{-1}\mathcal{A}(D \otimes D).$$

Then

$$\mathbf{vec}(D_X^f) = [I - \tilde{P}]\Big[\mathbf{vec}(R) - \frac{\theta \nu}{\sqrt{\mu}}(D \otimes D)(\mathcal{A}^T (\bar{y} - y^0) + \mathbf{vec}(\bar{S}) - \mathbf{vec}(S^0))\Big]$$
$$+ \frac{\theta \nu}{\sqrt{\mu}}\tilde{P}(D \otimes D)^{-1}(\mathbf{vec}(\bar{X}) - \mathbf{vec}(X^0)).$$

Since

$$[I - \tilde{P}](D \otimes D)\mathcal{A}^T (\bar{y} - y^0) = 0,$$

it follows that

$$\mathbf{vec}(D_X^f) = [I - \tilde{P}]\Big[\mathbf{vec}(R) - \frac{\theta \nu}{\sqrt{\mu}}(D \otimes D)(\mathbf{vec}(\bar{S}) - \mathbf{vec}(S^0))\Big]$$
$$+ \frac{\theta \nu}{\sqrt{\mu}}\tilde{P}(D \otimes D)^{-1}(\mathbf{vec}(\bar{X}) - \mathbf{vec}(X^0)).$$

To simplify the above expression, denote

$$U^S = \frac{\theta \nu}{\sqrt{\mu}}[I - \tilde{P}](D \otimes D)(\mathbf{vec}(\bar{S}) - \mathbf{vec}(S^0)),$$

$$U^X = \frac{\theta \nu}{\sqrt{\mu}}\tilde{P}(D \otimes D)^{-1}(\mathbf{vec}(\bar{X}) - \mathbf{vec}(X^0)),$$

$$R_1 = [I - \tilde{P}]\mathbf{vec}(R), \quad R_2 = \tilde{P}\mathbf{vec}(R).$$

Then $\mathbf{vec}(D_X^f) = [R_1 - U^S] + U^X$. For $\mathbf{vec}(D_S^f)$, using (4.11) we obtain

$$\begin{aligned}
\mathbf{vec}(D_S^f) &= \mathbf{vec}(R) - \mathbf{vec}(D_X^f) = \mathbf{vec}(R) - [R_1 - U^S] - U^X \\
&= \tilde{P}\mathbf{vec}(R) + U^S - U^X = R_2 + U^S - U^X.
\end{aligned}$$

Therefore, by the orthogonality and $\|\mathbf{vec}(W)\|_F = \|W\|_F$,

$$\begin{aligned}
\|D_X^f\|_F^2 &= \|\mathbf{vec}(D_X^f)\|_F^2 = \|R_1 - U^S + U^X\|_F^2 = \|R_1 - U^S\|_F^2 + \|U^X\|_F^2 \\
&= \|R_1\|_F^2 + \|U^S\|_F^2 - 2\mathbf{Tr}(R_1^T U^S) + \|U^X\|_F^2 \\
&\leq \|R_1\|_F^2 + \|U^S\|_F^2 + \|U^X\|_F^2 + \|R_1\|_F^2 + \|U^S\|_F^2 \\
&= 2\|R_1\|_F^2 + 2\|U^S\|_F^2 + \|U^X\|_F^2,
\end{aligned} \tag{4.15}$$

where the inequality follows from the fact that $-2\mathbf{Tr}(R_1^T U^S) \leq \|R_1\|_F^2 + \|U^S\|_F^2$. For $\|D_S^f\|_F$,

$$\begin{aligned}
\|D_S^f\|_F^2 &= \|R_2 + U^S - U^X\|_F^2 = \|R_2 - U^X\|_F^2 + \|U^S\|_F^2 \\
&= \|R_2\|_F^2 + \|U^X\|_F^2 - 2\mathbf{Tr}(R_2^T U^X) + \|U^S\|_F^2 \\
&\leq 2\|R_2\|_F^2 + 2\|U^X\|_F^2 + \|U^S\|_F^2.
\end{aligned} \tag{4.16}$$

Due to the definitions of U^X and U^S,

$$\begin{aligned}
\|U^S\|_F^2 + \|U^X\|_F^2 &= \frac{\theta^2 \nu^2}{\mu}(\|(D \otimes D)(\mathbf{vec}(\bar{S}) - \mathbf{vec}(S^0))\|_F^2 \\
&\quad + \|(D \otimes D)^{-1}(\mathbf{vec}(\bar{X}) - \mathbf{vec}(X^0))\|_F^2) \\
&= \frac{\theta^2 \nu^2}{\mu}(\|\mathbf{vec}(D(\bar{S} - S^0)D)\|_F^2 + \|\mathbf{vec}(D^{-1}(\bar{X} - X^0)D^{-1})\|_F^2) \\
&= \frac{\theta^2 \nu^2}{\mu}(\|D(\bar{S} - S^0)D\|_F^2 + \|D^{-1}(\bar{X} - X^0)D^{-1}\|_F^2).
\end{aligned} \tag{4.17}$$

To proceed, we must specify our initial iterate (X^0, y^0, S^0). We assume that ξ is such that

$$\|X^* + S^*\|_\infty \leq \xi \tag{4.18}$$

for some optimal solution (X^*, y^*, S^*) of (P) and (D); as usual, we start the algorithm with

$$(X^0, y^0, S^0) = (\xi I, 0, \xi I), \quad \mu^0 = \xi^2. \tag{4.19}$$

We are still free to choose $\bar{X}$ and $\bar{S}$, subject to the constraints

$$A_i \bullet \bar{X} = b_i, \quad i = 1, 2, \ldots, n, \quad \sum_{i=1}^{m} \bar{y}_i A_i + \bar{S} = C.$$

Taking $\bar{X} = X^*$ and $\bar{S} = S^*$, the entries of the matrices $\bar{X} - X^0$ and $\bar{S} - S^0$ satisfy

$$0 \le X^0 - X^* \le \xi I, \quad 0 \le S^0 - S^* \le \xi I.$$

In order to obtain upper bounds for $\|D(\bar{S} - S^0)D\|_F$ and $\|D^{-1}(\bar{X} - X^0)D^{-1}\|_F$, we write, using $\nu = \mu/\mu^0$ and $V = D^{-1}XD^{-1}/\sqrt{\mu} = DSD/\sqrt{\mu}$,

$$\|D(\bar{S} - S^0)D\|_F^2 \le \|(\bar{S} - S^0)\|^2 \|D^2\|_F^2 \le \xi^2 \mathbf{Tr}(P^2) \le \xi^2 \frac{\mathbf{Tr}(X^2)}{\mu\lambda_{\min}(V^2)}. \tag{4.20}$$

Similarly,

$$\|D^{-1}(\bar{X} - X^0)D^{-1}\|_F^2 \le \xi^2 \|D^{-2}\|_F^2 = \xi^2 \mathbf{Tr}(P^{-2}) \le \xi^2 \frac{\mathbf{Tr}(S^2)}{\mu\lambda_{\min}(V^2)}. \tag{4.21}$$

Substituting (4.20) and (4.21) into (4.17),

$$\begin{aligned}
\|U^S\|_F^2 + \|U^X\|_F^2 &\le \frac{\theta^2\nu^2}{\mu}\Big(\xi^2 \frac{\mathbf{Tr}(X^2)}{\mu\lambda_{\min}(V^2)} + \xi^2 \frac{\mathbf{Tr}(S^2)}{\mu\lambda_{\min}(V^2)}\Big) \\
&= \frac{\theta^2}{\xi^2\lambda_{\min}(V^2)}(\mathbf{Tr}(X^2) + \mathbf{Tr}(S^2)) \\
&\le \frac{\theta^2}{\xi^2\lambda_{\min}(V)^2}(\mathbf{Tr}(X + S)^2).
\end{aligned} \tag{4.22}$$

LEMMA 4.5 [12, Lemma 5.14]. *Let* (X, y, S) *be feasible for the perturbed problems* (P_ν) *and* (D_ν), *and let* (X^*, y^*, S^*) *and* (X^0, y^0, S^0) *be as defined in* (4.18) *and* (4.19). *Then*

$$\nu\xi\mathbf{Tr}(X + S) \le S \bullet X + \nu n\xi^2. \tag{4.23}$$

LEMMA 4.6. *Using the same notation as in Lemma* 4.5,

$$\mathbf{Tr}(X + S) \le (1 + \rho(\Phi(V))^2)n\xi, \tag{4.24}$$

with $\rho(\Phi(V))$ *defined as in* (4.3).

PROOF. Using (4.23) and $\mu = \nu\xi^2$, $\mathbf{Tr}(X + S) \le \mathbf{Tr}(XS/\mu)\xi + n\xi$. Since $XS \sim \mu V^2 = D^{-1}XSD$, we deduce that

$$\mathbf{Tr}\Big(\frac{XS}{\mu}\Big) = \mathbf{Tr}(V^2) = \sum_{i=1}^{n} \lambda_i(V)^2 \le \sum_{i=1}^{n} \rho(\Phi(V))^2 = n\rho(\Phi(V))^2,$$

where the inequality follows from (4.2).   □

Using (4.15), (4.16), (4.22) and (4.24),

$$\|D_X^f\|_F^2 + \|D_S^f\|_F^2 \le 2\|\nabla\Phi(V)\|_F^2 + 3n^2\theta^2(1 + \rho(\Phi(V))^2)^2\rho(\Phi(V))^2.$$

4.3. Value for θ At this stage, we choose

$$\tau = \tfrac{1}{16}.$$

Since

$$\Phi(V) \le \tfrac{1}{2}\|\nabla\Phi(V)\|_F^2 \le \tau = \tfrac{1}{16}$$

and $\rho(\Phi(V))$ is monotonically increasing in $\Phi(V)$,

$$\|D_X^f\|_F^2 + \|D_S^f\|_F^2 \le \tfrac{4}{16} + 3n^2\theta^2(1 + \rho(\tfrac{1}{16})^2)^2\rho(\tfrac{1}{16})^2 \le \tfrac{1}{4} + 3n^2\theta^2 \times 10.8387.$$

Using $\theta = \alpha/4\sqrt{n}$, the above inequality reads

$$\|D_X^f\|_F^2 + \|D_S^f\|_F^2 \le \frac{1}{4} + \frac{8.129}{4}n\alpha^2. \tag{4.25}$$

From (4.10), we know that $w \le \sqrt{14/125}$ is needed to guarantee that $\Phi(V^f) \le \sqrt{2}$. By (4.25) and (4.4), this holds if $1/4 + 8.129n\alpha^2/4 \le 56/125$. This means that if we take

$$\alpha = \frac{1}{4\sqrt{n}} \tag{4.26}$$

then it is guaranteed that $\Phi(V^f) \le \sqrt{2}$.

4.4. Complexity analysis We have seen that if the iterates satisfy $\tfrac{1}{2}\|\Phi(V)\|_F^2 \le \tau$ at the start of an iteration, with τ as defined in (4.8), then after the feasibility step, with θ as defined in (4.8) and α as defined in (4.26), the iterates satisfy $\Phi(X^f, S^f; \mu^+) \le \sqrt{2}$.

After the feasibility step, we perform some centering steps in order to make the iterates (X^+, y^+, S^+) satisfy $\mathbf{Tr}(X^+S^+) = n\mu^+$ and $\delta(X^+, S^+; \mu^+) \le \tau$, where τ is (much) smaller than $1/\sqrt[4]{2}$. This process is repeated until the maximum of the norms of the residual and $\mathbf{Tr}(X^+S^+)$ is less than ϵ.

The next theorem gives an upper bound for the total number of iterations. This bound coincides with the current best known bound for IIPMs for SDO [12].

THEOREM 4.7. *If* (P) *and* (D) *have optimal solutions* (X^*, y^*, S^*) *such that* $\|X^* + S^*\|_\infty \le \xi$, *then after at most*

$$80n \log \frac{\max\{n\xi^2, \|r_b^0\|_F, \|R_c^0\|_F\}}{\epsilon}$$

iterations the algorithm finds an ϵ-optimal solution of (P) *and* (D).

PROOF. Let k denote the number of centering steps. By Lemma 3.7, after k centering steps, the iterates (X^+, y^+, S^+) are still feasible for (P_{ν^+}) and (D_{ν^+}) and satisfy

$$\Phi(X^+, S^+; \mu^+) \le (1/2)^{2^k-1}(\sqrt{2})^{2^k}.$$

From this inequality, $\Phi(X^+, S^+; \mu^+) \le \tau$ will hold if $(1/2)^{2^k-1}(\sqrt{2})^{2^k} \le \tau$, which implies that $k \ge \log_2 \log_2(4/\tau^2)$. Since $\tau = 1/16$, at most $k = \lceil \log_2 \log_2(4/\tau^2) \rceil = 4$ centering

steps are required to obtain the iterate (X^+, y^+, S^+) satisfying $\Phi(X^+, S^+; \mu^+) \leq \tau$. So each main iteration consists of at most five inner iterations. In each main iteration, both the value of $X \bullet S$ and the norm of the residual are reduced by the factor $1 - \theta$. Hence the total number of main iterations is bounded above by

$$\frac{1}{\theta} \log \frac{\max\{X^0 \bullet S^0, \|r_b^0\|, \|R_c^0\|\}}{\epsilon}.$$

By multiplying the number of inner iterations by the number of main iterations, the total number of iterations is bounded above by

$$\frac{5}{\theta} \log \frac{\max\{X^0 \bullet S^0, \|r_b^0\|_F, \|R_c^0\|_F\}}{\epsilon},$$

where $\theta = 1/16n$, by (4.8) and (4.26), and hence we obtain the upper bound

$$80n \log \frac{\max\{\mathbf{Tr}(X^0 S^0), \|r_b^0\|_F, \|R_c^0\|_F\}}{\epsilon}.$$

This completes the proof.   □

Acknowledgements

References

[1] F. Alizadeh, "Interior point methods in semidefinite programming with applications to combinatorial optimization", *SIAM J. Optim.* **5** (1995) 13–51; doi:10.1137/0805002.

[2] E. de Klerk, *Aspects of semidefinite programming. Applied optimization*, 1st edn (Kluwer Academic, Dordrecht, 2002).

[3] R. A. Horn and C. H. R. Johnson, *Topics in matrix analysis* (Cambridge University Press, Cambridge, 1991).

[4] E. S. Karisch, F. Rendl and J. Clausen, "Solving graph bisection problems with semidefinite programming", *INFORMS J. Comput.* **12** (2000) 177–191; doi:10.1287/ijoc.12.3.177.12637.

[5] M. Kojima, M. Shida and S. Hara, "Local convergence of predictor–corrector infeasible-interior-point algorithms for SDPs and SDLCPs", *Math. Program.* **80** (1998) 129–160; doi:10.1007/BF01581723.

[6] Z. Y. Liu and Y. Chen, "A full-Newton step infeasible interior-point algorithm for linear programming based on a self-regular proximity", *J. Appl. Math. Informatics* **29** (2011) 119–133.

[7] Z. Y. Liu and W. Y. Sun, "An infeasible interior-point algorithm with full-Newton step for linear optimization", *Numer. Algorithms* **46** (2007) 173–188; doi:10.1007/s11075-007-9135-x.

[8] Z. Q. Luo, J. Sturm and S. Z. Zhang, "Superlinear convergence of a symmetric primal-dual path following algorithm for semidefinite programming", *SIAM J. Optim.* **8** (1998) 59–81; doi:10.1137/S1052623496299187.

[9] I. J. Lustig, "Feasible issues in a primal-dual interior-point method for linear programming", *Math. Program.* **49** (1991) 145–162; doi:10.1007/BF01588785.

[10] H. Lütkepohl, *Handbook of matrices* (Wiley, Chichester, 1996).

[11] H. Mansouri and C. Roos, "Simplified $O(nL)$ infeasible interior-point algorithm for linear optimization using full-Newton steps", *Optim. Methods Soft.* **22** (2007) 519–530; doi:10.1080/10556780600816692.

[12]   H. Mansouri and C. Roos, "A new full-Newton step $O(n)$ infeasible interior-point algorithm for semidefinite optimizationg", *Numer. Algorithms* **52** (2009) 225–255; doi:10.1007/s11075-009-9270-7.

[13]   Yu. E. Nesterov and M. J. Todd, "Self-scaled barriers and interior-point methods for convex programming", *Math. Oper. Res.* **22** (1997) 1–42; doi:10.1287/moor.22.1.1.

[14]   Yu. E. Nesterov and M. J. Todd, "Primal-dual interior-point methods for self-scaled cones", *SIAM J. Optim.* **8** (1998) 324–364; doi:10.1137/S1052623495290209.

[15]   J. Peng, C. Roos and T. Terlaky, "New complexity analysis of the primal-dual method for semidefinite optimization based on the NT-direction", *J. Optim. Theory Appl.* **109** (2001) 327–343.

[16]   J. Peng, C. Roos and T. Terlaky, "Self-regular functions and new search directions for linear and semidefinite optimization", *Math. Program.* **93** (2002) 129–171; doi:10.1007/s101070200296.

[17]   J. Peng, C. Roos and T. Terlaky, *Self-regularity. A new paradigm for primal-dual interior-point algorithms* (Princeton University Press, Princeton, NJ, 2002).

[18]   F. A. Potra and R. Sheng, "A superlinear convergent primal-dual infeasible-interior-point algorithm for semidefinite programming", *SIAM J. Optim.* **8** (1998) 1007–1028; doi:10.1137/S1052623495294955.

[19]   G. Q. Wang, Y. Q. Bai and C. Roos, "Primal-dual interior-point algorithms for semidefinite optimization based on a simple kernel function", *J. Math. Model. Algorithms* **4** (2005) 409–433; doi:10.1007/s10852–005–3561–3.

[20]   H. Wolkowicz, R. Saigal and L. Vandenberghe, *Handbook of semidefinite programming, theory, algorithm, and applications* (Kluwer Academic Publishers, Dordrecht, 2000).

[21]   Y. Zhang, "On extending some primal-dual interior-point algorithms from linear programming to semidefinite programming", *SIAM J. Optim.* **8** (1998) 365–386; doi:10.1137/S1052623495296115.